IUTAM Bookseries

Hiroshi Yabuno · Walter Lacarbonara · Balakumar Balachandran · Alexander Fidlin · Giuseppe Rega · Masaharu Kuroda · Shinichi Maruyama *Editors*

IUTAM Proceedings of the IUTAM Symposium on Nonlinear Dynamics for Design of Mechanical Systems Across Different Length/Time Scales

Springer

IUTAM Bookseries **43**

The IUTAM Bookseries publishes the refereed proceedings of symposia organised by the International Union of Theoretical and Applied Mechanics (IUTAM).

Every two years the IUTAM General Assembly decides on the list of IUTAM Symposia. The Assembly calls upon the advice of the Symposia panels. Proposals for Symposia are made through the Assembly members, the Adhering Organizations, and the Affiliated Organizations, and are submitted online when a call is launched on the IUTAM website.

The IUTAM Symposia are reserved to invited participants. Those wishing to participate in an IUTAM Symposium are therefore advised to contact the Chairman of the Scientific Committee in due time in advance of the meeting. From 1996 to 2010, Kluwer Academic Publishers, now Springer, was the preferred publisher of the refereed proceedings of the IUTAM Symposia. Proceedings have also been published as special issues of appropriate journals. From 2018, this bookseries is again recommended by IUTAM for publication of Symposia proceedings.

Indexed in Ei Compendex and Scopus.

Hiroshi Yabuno · Walter Lacarbonara ·
Balakumar Balachandran · Alexander Fidlin ·
Giuseppe Rega · Masaharu Kuroda ·
Shinichi Maruyama
Editors

Proceedings of the IUTAM Symposium on Nonlinear Dynamics for Design of Mechanical Systems Across Different Length/Time Scales

Springer

Editors
Hiroshi Yabuno
University of Tsukuba
Tsukuba Science City, Ibaraki, Japan

Balakumar Balachandran
Department of Mechanical Engineering
University of Maryland
College Park, MD, USA

Giuseppe Rega
DISG (Department of Structural
and Geotechnical Engineering)
Sapienza Università di Roma
Rome, Italy

Shinichi Maruyama
Gunma University
Kiryu, Japan

Walter Lacarbonara
DISG (Department of Structural
and Geotechnical Engineering)
Sapienza Università di Roma
Rome, Italy

Alexander Fidlin
Institute of Technical Mechanics
Karlsruhe Institute of Technology
Karlsruhe, Baden-Württemberg, Germany

Masaharu Kuroda
University of Hyogo
Himeji, Japan

ISSN 1875-3507 ISSN 1875-3493 (electronic)
IUTAM Bookseries
ISBN 978-3-031-72793-1 ISBN 978-3-031-72794-8 (eBook)
https://doi.org/10.1007/978-3-031-72794-8

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Switzerland AG 2025

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Preface

This book contains papers presented at International Union of Theoretical and Applied Mechanics (IUTAM) Symposium on Nonlinear dynamics for design of mechanical systems across different length/time scales held in Tsukuba Science City.

This volume of IUTAM bookseries includes many papers of the symposium after undergoing review. The papers are categorized into nine areas: calculus and computation, fluid mechanics and fluid-structure interactions, fundamental mechanics, composite materials and soft matter, nonlinear dynamics and control, nonlinear dynamics in practical systems, sports dynamics and biomechanics, stochastic dynamics, and synchronization. The book editors are as follows: Balakumar Balachandran, Alexander Fidlin, Masaharu Kuroda, Walter Lacarbonara, Shinichi Maruyama, Giuseppe Rega, and Hiroshi Yabuno.

Nonlinear phenomena arise in a variety of mechanical and physical systems. Also, the experimental analyses for nonlinear dynamics, together with theoretical analyses, confirm not only the reliability of the theoretical results but also deliver meaningful hints on realizing high-performance machines, which exhibit features of high-accuracy, high-speed, high-flexibility, high-reliability, and so on. In fact, the positive utilization of resonant vibrations and their nonlinear characteristics can be found in various devices such as auto-parametric vibration absorbers, gyroscopic dampers, ultrasonic vibrational machine tools, atomic force microscopes, ultrasensitive mass sensors, in the fields across diverse length scales, i.e., from macroscopic applications to micro/nanoscopic ones. This symposium, while shedding light onto state-of-the-art knowledge and advancements, was eminently expected to pave the way towards future directions and perspectives in the field. Share of the knowledge on theoretical and experimental approaches to nonlinear dynamics by extensive discussions and exchange of ideas in the symposium may help to devise novel comprehensive methods to understand nonlinear phenomena much more deeply and lead to paradigm shifts in mechanical system design towards next-generation design by positive exploitation of nonlinear phenomena.

The keynote lectures of the symposium were on some relevant paradigmatic topics as follows:

- “Constructive Utilization of Nonlinear Dynamics in Micro-scale Systems” by Professor Hanna Cho (Ohio State University).
- “Koopman as a gate to Hill: a quest for a stability criterion” by Professor Dr. ir. habil. Remco Leine (University of Stuttgart) and M.Sc. Fabia Bayer.
- “Forecasting Critical Transitions Using Data-Driven Methods” by Professor Bogdan I. Epureanu (University of Michigan).
- “Melnikov’s methods and nonintegrability of forced nonlinear oscillators” by Professor Kazuyuki Yagasaki (Kyoto University).

As most presentations were related to more than one topic, regular presentations were divided into sessions

- Day one: Duffing oscillator, Nonlinear dynamical systems in practical systems.
- Day two: Stochastic dynamics, Nonlinear fluid mechanics and fluid-solid interactions, Sports dynamics, Synchronization.
- Day three: Large deformation and soft robotics, Analysis and reduction of multi-degree-of-freedom nonlinear systems, Design of resonators and sensors, Multi-stability.
- Day four: Vibration control and energy harvesting, Metamaterials and soft matter.
- Day five: Fractional calculus, Energy transfer, Wave and propagation in practical systems.

The presentations in detail are shown below.

Symposium Schedule

Jul. 31st, 2023

12:30 – 13:05 **Registration**

13:05 – 13:15 **Opening address**

13:15 – 13:40 **Tutorial**
Exploiting global nonlinear dynamics for analysis, control and design of mechanical systems and structures across different length/time scales [online]
Giuseppe Rega

13:40 – 14:30 **Mon-f1: Duffing (Chair: Alexander Fidlin, Karlsruhe Institute of Technology)**
Bursting oscillations across different time scales: models and quantification of their responses
Ivana Kovacic
An asymptotic solution for near peak behavior of the Duffing equation [online]
Stefano Lenci

14:30 – 15:20 **Keynote 1: Constructive Utilization of Nonlinear Dynamics in Micro-scale Systems (Chair: Walter Lacarbonara, Sapienza University of Rome)**
Associate Professor Hanna Cho
Micro/Nano Multiphysical Dynamics Laboratory
Department of Mechanical and Aerospace Engineering
College of Engineering
The Ohio State University
https://mae.osu.edu/mnmdl

15:20 – 15:40 **Coffee break**

15:40 – 17:45 **Mon-f2: Nonlinear dynamics in practical systems (Chair: Tamás Kalmár-Nagy, Budapest University of Technology and Economics)**
Asymptotic Approximation of the Maximal Lyapunov Exponent of Moore-Greitzer PDE model with small Multiplicative Noise close to Stall Bifurcation
Yiming Meng, N. Sri Namachchivaya, Nicolas Perkowski
Detect and evaluate undesired nonlinearities in engineering structural dynamics
Mayuko Nishio, Sifan Wang
Intermodal Targeted Energy Transfer
Oleg Gendelman, Majdi Gzal, Alexander Vakakis
Wave propagation in metamaterial honeycombs with embedded nonlinear membrane resonators
Yichang Shen, Walter Lacarbonara
Non-linear characteristics of a two-DOF shaft-system coupled by a universal joint with clearance
Junaid Ali, Anil K. Bajaj, Gregory Shaver

18:00 – 20:00 **Welcome reception**

Aug. 1st, 2023

8:45 – 9:05 **Registration**

9:05 – 9:55 **Keynote 2: Koopman as a gate to Hill: a quest for a stability criterion (Chair: Bala Balachandran, University of Maryland)**
Prof. Dr. ir. habil. Remco Leine, M.Sc. Fabia Bayer
Institute for Nonlinear Mechanics, University of Stuttgart
www.inm.uni-stuttgart.de

9:55 – 10:45 **Tue-f1: Stochastic dynamics (Chair: Oleg V. Gendelman, Technion-Israel Institute of Technology)**
Stability and bifurcation of a noisy nonlinear auto-parametric vibration absorber
Peter H. Baxendale, N. Sri Namachchivaya, Momoiyioluwa Oluyemi
Nonlinear stochastic global dynamics of an imperfect, electrically actuated microbeam through an operator approach
Kaio C. B. Benedetti, Paulo B. Goncalves, Stefano Lenci, Giuseppe Rega

10:45 – 11:05 **Coffee break**

11:05 – 12:45 **Tue-f2: Nonlinear fluid mechanics and fluid-structure interactions (Chair: Anil Bajaj, Purdue University)**
A novel pair of non-dimensional measures to predict the order-to-chaos wake transition in the flow-field of a flapping foil

Dipanjan Majumdar, Chandan Bose, Sunetra Sarkar
Coupled Bluff Body Energy Harvesters: Vortex-Influenced Dynamics
Khawar Zamman Wani, Manoj Pandey, Balakumar Balachandran
Dynamical Modeling and Analysis of Fluid-Structure Interaction Phenomena and Bifurcation Occurring in the Inducer of a Turbopump for Rocket Engines
Tsuyoshi Inoue, Hayate Okawa, Hironori Horiguchi
Nonlinear phenomena in perturbed Rayleigh-Benard convection
Hiroaki Yoshimura, Masahito Watanabe

12:45 – 13:40 **Lunch**

13:40 – 14:30 **Tue-f3: Sports dynamics (Chair: Masaharu Kuroda, University of Hyogo)**
Head speed generating mechanisms during turning phases of athletic hammer throwing
Sekiya Koike, Tempei Tominaga, Keigo Byun Ohyama, Alexander P. Willmott
Nonlinear Finite Element Analysis of the Impact Between a Tennis Racket and a Tennis Ball
Toki Shimada, Yudai Washida, Akihiro Matsuda

14:30 – 15:30 **Shotgun session 1 (Chair: Triet Nguyen-Van, University of Tsukuba)**

[2-01 Viscoelasticity]

PS01 **Gripping force of pneumatically actuated flexible arms**
Kazuya Sakai, Tomohiko G. Sano
PS02 **Reaction-diffusion model for the tape-peeling trace by deformed adhesives**
Keisuke Taga, Hiroya Nakao, Yoshihiro Yamazaki
PS04 **Theory and computation of weakly nonlinear ultrasound propagation in a viscoelastic bubbly liquid**
Takeru Hasegawa, Tetsuya Kanagawa

[2-02 Materials]

PS05 **Modeling of Porous Hyperelastic Materials Containing Compressible Voids**
Xiao Li, Akihiro Matsuda
PS06 **Shrinkage and shape memory of gelated cellulose nanopaper in the drying process**
Shohei Moriwaki, Itsuo Hanasaki

[2-03 Soft actuators]

PS07 J**umping mechanism of elastic shells on rigid or granular substrates**
Takara Abe, Shuhei Shimizu, Genya Ishigami, Tomohiko G. Sano
PS08 **Experimental analysis on jumping mechanism using snap-through buckling of an elastic strip**
Shuhei Shimizu, Takara Abe, Tomohiko G. Sano, Genya Ishigami

PS09 **Modelling of series inflatable actuators for sequential activation control**
Kiyohiro Araki, Diego Paez-Granados, Modar Hassan, Kenji Suzuki

[2-04 Biomechanical systems]

PS10 **Nonlinear material modeling of human skeletal muscle with muscle contraction**
Tomotaka Hamajima, Akihiro Matsuda

PS11 **An Instrumented Pole for Measuring the Individual Hand Forces in Pole Vaulting**
Rinri Uematsu, Kiyoshi Hirose, Osamu Takeda, Alexander P. Willmott, Sekiya Koike

PS12 **Optimization-based Kinetic Synthesis Method to Produce Nonlinear Response Property of Prosthetic Legs**
Sergio Alberto Galindo Leon, Diego Paez-Granados, Modar Hassan, Kenji Suzuki

PS13 **Mechanical Characterization of Multi-Degree-of-Freedom Elastic Neck Exoskeleton for Persons with Dropped Head Syndrome**
Santiago Price Torrendell, Hideki Kadone, Modar Hassan, Kenji Suzuki

[2-05 Bubble]

PS14 **Weakly nonlinear ultrasound propagation in liquids containing multiple ultrasound contrast agents with shell in buckled or ruptured states**
Quoc Nam Nguyen, Tetsuya Kanagawa

PS15 **Thermal effects on bubble dynamics under long-term ultrasound irradiation in a gel**
Satoshi Kusakabe, Keita Ando

PS16 **Effect of microbubble coated with anisotropic shell on ultrasound propagation in liquid containing multiple microbubbles**
Ryoki Kawahata, Tetsuya Kanagawa

[2-06 CFRP]

PS17 **Inelastic two-scale analysis of dovetail joints of CFRP fan blades for jet engines**
Eiichiro Mori, Masayoshi Akaza, Chiharu Tsujikawa, Tetsuya Matsuda, Naoki Morita, Nobuhiro Yoshikawa

PS18 **Vibration characteristics of curvilinear CFRP prepared by electrodeposition resin molding manufacturing method**
Md Tansirul Islam, Shinya Honda, Kazuaki Katagiri, Katsuhiko Sasaki, Ryo Takeda, Isamu Saiwaki, Yuto Shimizu

PS19 **Experiments on Nonlinear and Chaotic Vibrations of a Bi-stable Unsymmetrically Laminated CFRP Plate**
Kodai Hirano, Shinichi Maruyama, Ken-ichi Nagai, Takao Yamaguchi, Chihiro Kamio

[2-07 Machine tools]

PS20 **Monitoring system for Laser Welding and Emitting Energy Distribution Model to Explain Nonlinear Relationship**
Ichiro Ogura

PS21 **Feasibility studies of using vibration sensor to monitor tool wear progress during finishing step of hardened steel machining**
Jonny Herwan, German Herrera-Granados, Ichiro Ogura, Yoshiyuki Furukawa, Takashi Misaka, Hitoshi Komoto

PS22 **Experimental study of the nanoscale cutting using the self-excited microcantilever**
Linjun An, Ichiro Ogura, Kiwamu Ashida, Hiroshi Yabuno

[2-08 Synchronization]

PS23 **Phase-Amplitude Reduction of Limit Cycling Networks for Optimal Synchronization**
Petar Mircheski, Hiroya Nakao

PS24 **Design of artificial limit-cycle oscillators with global synchronization properties**
Norihisa Namura, Hiroya Nakao

PS25 **Synchronization in Switching Networks of Nonlinear Systems**
Shuma Kamo, Toshiki Oguchi

PS26 **Rigorous numerics for synchronized solutions of coupled time delay Duffing systems**
Kazuki Takahashi, Akitoshi Takayasu

[2-09 Data-driven dynamics]

PS27 **A case study of applying virtual node approach to multiple oscillators for reservoir computing**
Takeshi Shibuya

PS28 **Automatic Generation of Governing Equations for Mechanical Systems with Piecewise-Linear Systems using Sparse Regression**
Ryosuke Kanki, Akira Saito

PS29 **Analysis of nonlinear acoustic waves using physics-informed neural networks**
Kazuya Yokota, Takahiko Kurahashi, Masajiro Abe

[2-10 Numerical analysis and computation]

PS30 **A GPU-based multi-sphere DE-FE method and its application in the simulations of tire-terrain interaction**
Xiaobing Guo, Shunhua Chen, Mengyan Zang, Naoto Mitsume

PS31 **Parallel Implementation of a rigorous contour integral based eigensolver and its performance evaluation**
Shota Seto, Akitoshi Takayasu

PS32 **Rigorous numerical computations with the Chebyshev interpolation using the Julia language**
Shinsuke Kondo, Akitoshi Takayasu

15:30 – 16:30 **Poster session 1 with Coffee break**

16:30 – 18:00 **Short oral presentations (Chair: Rajesh Chaunsali, Indian Institute of Science & Tomohiko G. Sano, Keio University)**
Two-scale analysis of magnetoactive polymers and bifurcation studies on slender ferromagnetic structures
Vivekanand Dabade, Chinika Dangi, Krishna Chand Avatar
Two-scale damage propagation analysis of CFRP considering randomness of fiber distribution
Yukinobu Shimura, Kazuma Akashi, Tetsuya Matsuda
Rheological measurement of Tetra-PEG deformed by laser-induced microbubble oscillation
JianYi Zhao, Takuya Katashima, Makoto Asai, Keita Ando
Nonlinear dynamics of topological Kagome lattice
K. Prabith, R. Chaunsali
Crack propagation on bilayer spherical shells
Naoki Sekiya, Yuri Akiba, Tomohiko G. Sano
Motion of a variable length elastic system made periodic or quasi-periodic through configurational forces
Francesco Dal Corso, Panagiotis Koutsogiannakis, Diego Misseroni, Davide Bigoni

18:00 – 18:20 **Coffee break**

18:20 – 19:35 **Tue-f4: Synchronization (Chair: Harry Dankowicz, University of Illinois at Urbana-Champaign)**
Date-driven reconstruction of asymptotic phase and amplitude functions of stochastic oscillators using extended dynamic mode decomposition
Shohei Takata, Yuzuru Kato, Hiroya Nakao
Different coherent states for lightly supported coupled pendula
Dawid Dudkowski, Patrycja Jaros, Tomasz Kapitaniak
Exciting of slowly modulated vibrations with unbalances rotor exciters using nonlinear resonance and avoiding self-synchronization
Alexander Fidlin, Tunc Yüzbasioglu

Aug. 2nd, 2023

8:45 – 9:05 **Registration**

9:05 – 9:55 **Keynote 3: Forecasting Critical Transitions Using Data-Driven Methods (Chair: George Haller, ETH Zürich)**
Arthur F. Thurnau Professor Bogdan I. Epureanu
Professor of Mechanical Engineering, and Professor of Electrical Engineering and Computer Science
Director, Automotive Research Center
University of Michigan, Ann Arbor

9:55 – 10:45 **Wed-f1: Large deformation and soft robotics (Chair: JC Ji, University of Technology Sydney)**
Geometrically nonlinear deformation of hyperbolic shells: Experimental and numerical studies
Kosei Kataoka, Tomohiko G. Sano
Soft Robotics powered by Rod Theory
Hiromi Mochiyama

10:45 – 11:05 **Coffee break**

11:05 – 12:45 **Wed-f2: Analysis and reduction of multi-degree-of-freedom nonlinear systems (Chair: Oriel Shoshani, Ben-Gurion University of the Negev)**
Data-Driven Nonlinear Reduced-Order Modeling and Control via Spectral Submanifolds
George Haller
Unified perspectives on model reduction of geometrically nonlinear structures
Tieding Guo, Giuseppe Rega
Frequency Domain Identification of Nonlinear, Multi-degree-of-freedom Oscillators
Thomas Breunung, Balakumar Balachandran
Compact weighted residual formulation for periodic solutions of systems undergoing frictional occurrences
Mathias Legrand, Christophe Pierre

12:45 – 13:40 **Lunch**

13:40 – 14:30 **Wed-f3: Design of resonators and sensors (Chair: Manoj Pandey, Indian Institute of Technology Madras)**
Design of MEMS resonators for phase noise reduction using internal resonance [online]
Steven W Shaw, Oriel Shoshani, Scott Strachan, Daniel Lopez, David Czaplewski
Using Self-Excited Template Dynamics and Root-Finding Algorithms for Sensor Design
Yu Mao, Harry Dankowicz

14:30 – 15:30 **Shotgun session 2 (Chair: Shin Kawai, University of Tsukuba)**

[3-01 Nonlinear resonance of elastic systems]

PS33 (withdraw)
PS34 **Analysis on Nonlinear Vibrations of a Shallow Rectangular Shell-panel of which One Edge Clamped and the Others Simply-Supported**
Tomoaki Tsuji, Shinichi Maruyama, Ken-ichi Nagai, Takao Yamaguchi, Chihiro Kamio

PS35 **Dynamic characteristics of a long-period, cable-stayed bridge infrastructure subjected to seismic waveforms**
Kakeru Sagara, Hiroharu Utsunomiya, Gaku Shoji
PS36 **Real-time additive sound synthesis of plucked strings considering inharm geometrical nonlinearity using the method of multiple scales**
Shoya Saito, Naoto Wakatsuki, Tadashi Ebihara, Yuka Maeda, Koichi Mizutani
PS37 **Non-planar self-excited oscillation of a string due to velocity feedback control**
Xinzhe Xu, Hiroshi Yabuno

[3-02 Vibration control]

PS38 **Active Wave Control of a Mass–Spring–Damper System using Filtered PI-Controller**
Kairi Okamura, Natsuki Kawaguchi, Masaharu Kuroda
PS39 **Oscillator design for energy harvesters with nonlinear snap-through motion**
Takumi Furihata, Akira Saito
PS40 **Experimental Study of Vibration Control Using Autoparametric Vibratoion Absorber**
Yuma Kawasumi, Kenji Yasui, Hiroshi Yabuno
PS41 **Influence of viscous damping in autoparametric vibration absorber on its performance at the anti-resonance point (Experimental study by a simple apparatus)**
Chao Zhang, Hiroshi Yabuno
PS42 **Study on Box-shaped Jig for a Vibration Test(Reinforcement of a Resin Jig by Metal Parts for a Heavy Test Object)**
Ayato Toriya, Shinichi Maruyama, Yukinori Suzuki, Tetsuya Nakamura, Tomoaki Sakata, Hiroshi Aoyagi

[3-03 Fundamental mechanics]

PS43 **Bifurcation phenomena and hyperbolic Lagrangian coherent structures in Rayleigh-Benard convection**
Yuta Haga, Yudai Mohri, Yuta Nakahara, Yutaro Ohtsuka, Masahito Watanabe, Hiroaki Yoshimura
PS44 **The influence of accuracy of initial values on the discrete energy in variational integrator**
Mamoru Gunji, Yusuke Ono, Linyu Peng

[3-04 Fluid-conveying pipes]

PS45 **Experimental study on non-planar motions of a standing pipe conveying fluid**
Buana Mahkota Raja, Kiyotaka Yamashita
PS46 **Analysis of 1:1 internal resonance of a flexible pipe conveying fluid**
Tong Shen, Eisuke Higuchi, Hiroshi Yabuno, Kiyotaka Yamashita

[3-05 Self-excited oscillations]

PS47 **Vibration energy flow behaviour of self-excited dynamic systems**
Jian Yang, Cui Chao

PS48 **Observation of Self-Excited Vibrations of Strings on Bowed String Instrument Using Displacement Measurement Method Based on Light Intensity**
Yuto Hara, Ryodai Izaki, Naoto Wakatsuki, Tadashi Ebihara, Yuka Maeda
PS49 **Selective self-excited oscillations in a two-degree-of-freedom spring-mass-damper system in two modes by using band-pass filter**
Mari Nishijima, Hiroshi Yabuno

[3-06 Sensing devices]

PS50 **Liquid density measurement using a microchannel stainless cantilever**
Takumi Nakamura, Hiroshi Yabuno, Yasuyuki Yamamoto, Sohei Matsumoto
PS51 **Dynamic Behavior of an Acoustic Warning Device Utilizing Impact Vibration**
Hideaki Nishio, Shinichi Maruyama, Kenta Watanabe
PS52 **Effect of Contact Pressure on Reflective Photolethysmogram and Waveform Parameters**
Yuka Maeda, Masaki Sekine, Toshiyo Tamura, Koichi Mizutani

[3-07 Acoustic applications]

PS53 **Modulation and Demodulation Techniques for Indoor Acoustic Communications Using Parametric Loudspeaker**
Tadashi Ebihara, Naoto Wakatsuki, Koichi Mizutani, Riku Fukuda
PS54 **Utilization of biomimetic pulse trains using cetacean vocalizations for acoustic localization in coastal areas**
Hanako Ogasawara, Saki Watanabe, Eri Sato, Takanobu Kuroyama, Kazuyoshi Mori

[3-08 Motion control]

PS55 **Stability and nonlinearity in the flapping flight of butterfly**
Narumi Fujii, Kei Senda
PS56 **Trajectory Control of Manipulator's End-effector using SDRE Method**
Ryosuke Kita, Natsuki Kawaguchi, Masaharu Kuroda

[3-09 Predictive control]

PS57 **Sample-Based Adaptive Monte Carlo Model Predictive Control and Its Experimental Verification for Swing-up and Stabilization of a Pendulum on a Cart**
Koki Tachibana, Hisashi Date
PS58 **Performance evaluation of model predictive control for active suspension systems employing an inerter with road preview**
Wanshu Wang, Takehiko Asai
PS59 **A Study on Damping Characteristics of Accelerator Pedal Operating Leg for Vehicle Driving**
Kazuhiro Tanaka, Takashi Tsuchiya

[3-10 Discrete time systems]

PS60 **Discrete-time state observer for nonlinear continuous-time systems**
Triet Nguyen-Van

PS61 **Matched Pole–Zero Model for Multi-Input and Multi-Output Systems**
Shin Kawai, Keisuke Yagi

PS62 **Interpretation of frequency transfer functions based on discrete-time response**
Yuichi Tokutomi, Triet Nguyen-Van, Shin Kawai

PS63 **Discrete-time state feedback control for continuous-time system transformations**
Tatsuya Oshima, Shin Kawai, Triet Nguyen-Van

PS64 **State estimation and control considering nonlinear characteristics of stepper motor with noise**
Noritaka Kinjo, Shin Kawai, Triet Nguyen-Van

15:30 – 16:30 **Poster session 2 with Coffee break**

16:30 – 18:00 **Short oral presentations (Chair: Valery Pilipchuk, Wayne State University & Li-Qun Chen, Shanghai University)**

Differences among different types of modes in SPS modeling and model reduction by singular perturbation method
Kei Senda, Kazuya Hata

Nonlinear guided control of a human driver via an automated vehicle
Bence Szaksz, Gábor Orosz, Gábor Stépán

Estimating the Curvature of a Two Dimensional Time-Invariant External Potential through Swarm Formation: Theory and Experiments
Yanran Wang, Takashi Hikihara

Bias of the steady-state averaged solutions of a strongly overdamped particle in a cosine potential under harmonic excitation
Attila Genda, Alexander Fidlin, Oleg Gendelman

Inter-modal interactions in a chain of mass-in-mass non-linear oscillators
Jean Flosi, Alireza Ture Savadkoohi, Claude-Henri Lamarque

Rigorous forward numerical integration of general evolutionary partial differential equations based on semigroup theory
Akitoshi Takayasu, Gabriel W. Duchesne, Jean-Philippe Lessard

18:00 – 18:20 **Coffee break**

18:20 – 20:00 **Wed-f4: Multistability (Chair: Ivana Kova Kovačič, University of Novi Sad)**

Bistable Microstructures and their Implementation in Sensors
Slava Krylov, Dan Refaeli, Alex Liberzon

Nonlinear dynamics and energy harvesting from multi-stable shells
Jerzy Warminski, Andrzej Mitura, Lukasz Kloda, Francesco Romeo, Matteo Brunetti

Modelling and Design of Straw-based Truss Metamaterials
Dotan Ilssar, Dennis M. Kochmann

Nonlinear dynamics of ball rolling of radial bearings in an assembly with an unbalanced multi-stage rotor [online]
Katica R. (Stevanović) Hedrih

Aug. 3rd, 2023

8:45 – 9:05 **Registration**

9:05 – 10:20 **Thu-f1: Vibration control and energy harvesting (Chair: Jerzy Warminski, Lublin University of Technology)**
Tailoring the dynamical response of a mechanical resonator via nonlinear coupling to secondary resonators
Oriel Shoshani
Developments of energy flow theory for nonlinear dynamical systems and applications to relativistic dynamics
Jing Tang Xing

10:20 – 10:40 **Coffee break**

10:40 – 11:55 **Thu-f2: Metamaterials and soft matter (Chair: Dotan Ilssar, ETH Zürich)**
Coarse-graining for bridging spatio-temporal scales of soft matter dynamics for engineering science
Itsuo Hanasaki
Energy absorbing devices and metamaterials using properties of soft-walled billiards
Valery Pilipchuk
Numerical and Experimental Study of Spatially Variant Truss-based Metamaterials under Dynamic Loading
Kaoutar Radi, Dennis M. Kochmann

11:55 – 12:50 **Lunch**

12:50 – 16:30 **Science Tour**

19:00 – Banquet

Aug. 4th, 2023

8:45 – 9:05 **Registration**

9:05 – 9:55 **Keynote 4: Melnikov's methods and nonintegrability of forced nonlinear oscillators (Chair: N Sri Namachchivaya, University of Waterloo)**
Prof. Kazuyuki Yagasaki
Graduate School of Informatics, Department of Applied Mathematics and Physics Professor, Kyoto University
https://kdb.iimc.kyoto-u.ac.jp/profile/en.9388421097ed837e.html

9:55 – 10:45 **Fri-f1: Fractional calculus (Chair: Tomaz Kapitaniak, Technical University of Lodz)**
Autotuners and fractional calculus. Applications to the control of nonlinear processes

Cristina I Muresan, Isabela R. Birs, Robin De Keyser
Fractional-order Servo LQR Control for a Magnetic Levitation System
Yuki Moriguchi, Masaharu Kuroda, Natsuki Kawaguchi

10:45 – 11:05 **Coffee break**

11:05 – 12:45 **Fri-f2: Energy transfer (Chair: Jian Yang, University of Nottingham Ningbo China)**
Energy transfer in a tree-structured multi-DoF oscillator
Tamás Kalmár-Nagy, Róbert Rochlitz, Bendegúz Dezső Bak
Chimera States in Elastically Coupled Self-Excited Underactuated Rigid-Body Arrays
Yuval Levi, Gilad Yakir, Oded Gottlieb
Modal analysis for localization phenomena of parametric resonance in pendulum arrays subjected to vertical excitation
Yuji Harata, Takashi Ikeda
Integrated nonlinear design of vibration isolation and energy harvesting
Ze-Qi Lu, Li-Qun Chen

12:45 – 13:40 **Lunch**

13:40 – 15:10 **Short oral presentations (Chair: Paulo Goncalves, Pontifical Catholic University of Rio de Janeiro)**
Vibration Design of Composites with Nonlinear Fiber Shapes Fabricated by Tailored Fiber Placement Machine and Electrodeposition Resin Molding Method
Shinya Honda, Isamu Saiwaki, Kazuaki Katagiri, Katsuhiko Sasaki, Ryo Takeda
Numerical Study of Chatter Control in End Milling Process by Workpiece Excitation
Yutaka Nakano, Satoshi Sekikawa, Hiroki Takahara
Inverse Magic Formula Capable of Tire Performance Requirement Analysis at Early Stage of Vehicle Development [online]
Takao Kobayashi
Linear/Nonlinear system identification based on response probability density function of 1-DOF system which is subjected to white noise excitation
Soichiro Takata, Kaito Araki
Evaluation method for the effectiveness of ultrasonic vibration assisted drilling
Shigeru Aoki
Frequency and amplitude dependence of self-tuning resonance
Tatsuki Tagashira, Toshihiko Sugiura

15:10 – 15:30 **Coffee break**

15:30 – 16:20 **Fri-f3: Wave and propagation in practical systems (Chair: Keita Ando, Keio University)**
Study on Unstable Oscillation of Pantograph under Friction
Yuki Amano, Shigeyuki Kobayashi, Hiroki Mori, Nobuyuki Sowa
Theoretical and numerical analysis on nonlinear propagation of focused ultrasound in bubbly liquids toward cancer treatment by microbubble-enhanced HIFU
Tetsuya Kanagawa, Shunsuke Kagami

16:20 – 16:30 **Closing address**

The scientific committee in this symposium was appointed by the Bureau of IUTAM including the following members: Hiroshi Yabuno (University of Tsukuba, Japan, Chair), Walter Lacarbonara (La Sapienza University of Rome, Italy, Co-Chair), Giuseppe Rega (La Sapienza University of Rome, Italy, IUTAM Representative), Balakumar Balachandran (University of Maryland, United States), Felix Chernousko (Russian Academy of Sciences, Russia), Alexander Fidlin (Karlsruhe Institute of Technology, Germany), Marian Wiercigroch (University of Aberdeen, United Kingdom), Steven Shaw (Florida Institute of Technology, United States), and Masaharu Kuroda (University of Hyogo, Japan). There were regular and short presentations registered participants from 14 different countries according to the following geophysical distribution: Belgium, Canada, China, Germany, Hungary, India, Israel, Italy, Japan, Poland, Serbia, Switzerland, UK, and USA. We thank all presenters and participants for their valuable contributions.

The financial support of IUTAM and Tsukuba Tourism and Convention Association is most gratefully acknowledged. Finally, we would like to express our gratitude to Springer, especially to Dr. Mayra Castro and Mr. Ramasubramaniyan Velu for their support in producing this volume of IUTAM Bookseries.

May 2024 Hiroshi Yabuno

Organization

Chair

Yabuno, Hiroshi	University of Tsukuba

Co-chair

Lacarbonara, Walter	Sapienza University of Rome

Scientific Committees

Balachandran, Balakumar	University of Maryland
Chernousko, Felix	Russian Academy of Sciences
Fidlin, Alexander	Karlsruhe Institute of Technology
Kuroda, Masaharu	University of Hyogo
Rega, Giuseppe	Sapienza University of Rome
Shaw, Steven	Florida Institute of Technology
Wiercigroch, Marian	University of Aberdeen

IUTAM Representative

Rega, Giuseppe	Sapienza University of Rome

Local Organizing Committee

Kawai, Shin	University of Tsukuba
Kuroda, Masaharu	University of Hyogo
Maruyama, Shinichi	Gunma University
Nguyen-Van, Triet	University of Tsukuba
Shibuya, Takeshi	University of Tsukuba
Yamashita, Kiyotaka	Fukui University of Technology

Contents

Tutorial

Exploiting Global Dynamics for Analysis, Control and Design of Mechanical Systems and Structures Across Different Length/time Scales 3
Giuseppe Rega

Keynote

Melnikov's Methods and Nonintegrability of Forced Nonlinear Oscillators 21
Kazuyuki Yagasaki

Fundamental Mechanics

Intermodal Targeted Energy Transfer-Beyond Common Simplifications 43
Oleg V. Gendelman

Bias of the Steady-State Averaged Solutions of a Strongly Overdamped Particle in a Cosine Potential Under Harmonic Excitation 58
Attila Genda, Alexander Fidlin, and Oleg V. Gendelman

Nonlinear Oscillations of a Bistable Composite Shell: Experimental Studies ... 75
Jerzy Warminski, Andrzej Mitura, Lukasz Kloda, Danyil Melnyk, Matteo Brunetti, and Francesco Romeo

The Effect of Accuracy of Initial Velocity Discretisations on Discrete Energy in Variational Integration 88
Mamoru Gunji, Yusuke Ono, and Linyu Peng

A Harmonic-Balance Power Flow Analysis Method for Nonlinear Vibration Systems 99
Jian Yang, Cui Chao, and Yuhao Liu

Stochastic Dynamics

Asymptotic Approximation of the Maximal Lyapunov Exponent of Moore-Greitzer PDE Model with Small Multiplicative Noise Close to Stall Bifurcation 111
Yiming Meng, N. Sri Namachchivaya, and Nicolas Perkowski

Nonlinear Stochastic Global Dynamics of an Imperfect, Electrically Actuated Microcantilever Through an Operator Approach 125
Kaio C. B. Benedetti, Paulo B. Gonçalves, Stefano Lenci, and Giuseppe Rega

Definition and Data-Driven Reconstruction of Asymptotic Phase and Amplitudes of Stochastic Oscillators via Koopman Operator Theory 141
Shohei Takata, Yuzuru Kato, and Hiroya Nakao

Comparison Between Linear and Nonlinear System Identification Based on Response Probability Density Function of 1-DOF System Which is Subjected to White Noise Excitation 154
Soichiro Takata, Hiroharu Matsubara, Naoko Watanabe, Shuya Kubota, and Kaito Araki

Calculus and Computation

Autotuners and Fractional Calculus. Applications to the Control of Nonlinear Processes 165
Cristina Muresan, Isabela Birs, and Robin De Keyser

Fractional-Order Servo Linear Quadratic Regulator Control for a Magnetic Levitation System 184
Yuki Moriguchi, Masaharu Kuroda, and Natsuki Kawaguchi

Rigorous Numerical Computations with the Chebyshev Interpolation Using the Julia Language 199
Shinsuke Kondo and Akitoshi Takayasu

Parallel Implementation of a Verified Contour Integral Based Eigensolver and Its Performance Evaluation Using the Julia Language 210
Shota Seto and Akitoshi Takayasu

Analysis of Nonlinear Acoustic Waves Using Physics-Informed Neural Networks 221
Kazuya Yokota, Takahiko Kurahashi, and Masajiro Abe

Synchronization

Exciting of Slowly Modulated Vibrations with Unbalanced Rotor Exciters Using Nonlinear Resonance 231
Alexander Fidlin and Tunc Yüzbasioglu

Phase-Amplitude Reduction of Limit-Cycling Networks for Optimal Synchronization 245
Petar Mircheski, Jinjie Zhu, and Hiroya Nakao

Design of Artificial Limit-Cycle Oscillators with Global Synchronization Properties 255
Norihisa Namura and Hiroya Nakao

Composite Materials and Soft Matter

Coarse-Graining for Bridging Spatio-Temporal Scales of Soft Matter Dynamics in Engineering Science 265
Itsuo Hanasaki

Two-Scale Tensile Analysis of Dovetail Joints for CFRP Fan Blades Considering Material Nonlinearity of Resins 281
Eiichiro Mori, Tetsuya Matsuda, Chiharu Tsujikawa, Naoki Morita, Gai Kubo, Masahiro Hojo, Toshihiko Hosaka, Shinya Fukushige, and Nobuhiro Yoshikawa

Two-Scale Damage/Strength Analysis of CFRP Considering Randomness of Fiber Distribution 293
Yukinobu Shimura and Tetsuya Matsuda

Pattern Formation in Tape-Peeling Traces of Deformed Adhesives 305
Keisuke Taga, Hiroya Nakao, and Yoshihiro Yamazaki

Modeling of Porous Hyperelastic Materials Containing Compressible Voids ... 312
Xiao Li and Akihiro Matsuda

Sports Dynamics and Biomechanics

Head Speed Generating Mechanisms During Turning Phases of Athletic Hammer Throwing 323
Sekiya Koike, Tempei Tominaga, Keigo B. Ohyama, and Alexander P. Willmott

Nonlinear Material Modeling of Human Skeletal Muscle with Muscle Contraction 335
Tomotaka Hamajima and Akihiro Matsuda

An Instrumented Pole for Measuring the Individual Hand Forces in Pole Vaulting 343
Rinri Uematsu, Kiyoshi Hirose, Osamu Takeda, Alexander P. Willmott, and Sekiya Koike

Fluid Mechanics and Fluid-Structure Interaction

Coupled Bluff Body Energy Harvesters: Vortex Influenced Dynamics 355
Khawar Zamman Wani, Manoj Pandey, and Balakumar Balachandran

Dynamical Modeling and Analysis of Fluid-Structure Interaction Phenomena and Bifurcation Occurring in the Inducer of a Turbopump for Rocket Engines 370
Tsuyoshi Inoue, Hayate Okawa, and Hironori Horiguchi

Experimental Study on Non-planar Motions of a Standing Pipe Conveying Fluid 383
Buana Mahkota Raja and Kiyotaka Yamashita

Nonlinear Dynamics and Control

Numerical Study of Chatter Control in End Milling Process by Workpiece Excitation 397
Yutaka Nakano, Satoru Sekikawa, and Hiroki Takahara

Nonlinear Guidance of a Human Driver via an Automated Vehicle 405
Bence Szaksz, Gábor Orosz, and Gábor Stépán

Performance Evaluation of Model Predictive Control for Active Suspension Systems Employing an Inerter with Road Preview 417
Wanshu Wang and Takehiko Asai

Experimental Study of Vibration Control Using Autoparametric Vibration Absorber 426
Yuma Kawasumi, Kenji Yasui, and Hiroshi Yabuno

Nonlinear Dynamics in Practical Systems

Numerical Bifurcation Analysis of a Two-DOF Shaft-System Coupled by a Universal Joint with Clearance 441
Junaid Ali, Gregory Shaver, and Anil K. Bajaj

Soft Robotics Powered by Rod Theory 457
Hiromi Mochiyama

Dynamic Characteristics of a Long-Period, Cable-Stayed Bridge Infrastructure Subjected to Seismic Waveforms 471
Kakeru Sagara and Gaku Shoji

Feasibility Studies of Using Vibration Sensor to Monitor Tool Wear Progress During Finishing Step of Hardened Steel Machining 486
Jonny Herwan, German Herrera-Granados, Ichiro Ogura, Yoshiyuki Furukawa, Takashi Misaka, and Hitoshi Komoto

Monitoring System for Laser Welding and Emitting Energy Distribution Model to Explain Nonlinear Relationship 496
Ichiro Ogura

Author Index 503

Tutorial

Exploiting Global Dynamics for Analysis, Control and Design of Mechanical Systems and Structures Across Different Length/time Scales

Giuseppe Rega(✉)

Sapienza University, 00197 Rome, Italy
giuseppe.rega@uniroma1.it

Abstract. An overview of some main research problems and challenges on how to exploit global dynamics for the analysis, control and design of mechanical systems and structures at different length and time scales is presented.

Keywords: Global Dynamics · Mechanical Systems and Structures · Different Length/Time Scales · Analysis Control and Design

1 Introduction

The huge development of nonlinear dynamics in mechanics and engineering occurred around the millennium turn has brought enormous and distributed knowledge on a rich variety of methodological and phenomenological issues. After such an intense and long-lasting growth, the related research activity and outcomes are deemed to have attained a kind of plateau, corresponding to a stage of intrinsic full maturity of the area, from which moving forward to its full exploitation in a wide variety of technological areas of interest across different space and time scales. A characteristic that in fact is promptly reflected in the present IUTAM Symposium Proceedings.

This certainly holds also for the sub-area of global dynamics, whose adequate consideration, in combination with local bifurcation/dynamics, may bring nontrivial benefits in the analysis, control and design of mechanical systems and structures, mostly in the presence of strong nonlinearities. This chapter aims at providing an overview of some of the main key open research issues and the challenges that arise from them.

The overview moves from a scheme of the different phases of development of nonlinear dynamics in mechanics in about the last forty years [1], to arrive at the current phase of its marked hybridization with other scientific/technological fields, in which nonlinear dynamics concepts, methods and phenomena can fruitfully contribute to interpreting and governing major issues of application interest [2]. Within such framework, the use of geometric/computational techniques and tools, properly enriched by the predictive capability of analytical methods and the validating role of experimental investigations, may allow to exploit global dynamics, too, for analysing, controlling and designing

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 3–18, 2025.
https://doi.org/10.1007/978-3-031-72794-8_1

reduced order models (ROMs) of systems in macro- to micro/nano-mechanics, even in multi-physics and/or non-deterministic contexts (Fig. 1) [3].

Global analysis of dynamical systems relies on the knowledge and aware use of sophisticated theoretical concepts (e.g., invariant manifolds, attractor-basin portraits, global bifurcations, a huge variety of topological events), on one hand, and of advanced computational techniques for their characterization, construction, and geometric representation, on the other hand. When suitably used in conjunction with local analysis of bifurcation/response features, global analysis not only enhances understanding of systems' dynamics in incremental terms, but also allows to unveil specific important phenomena, along with the need for their proper control.

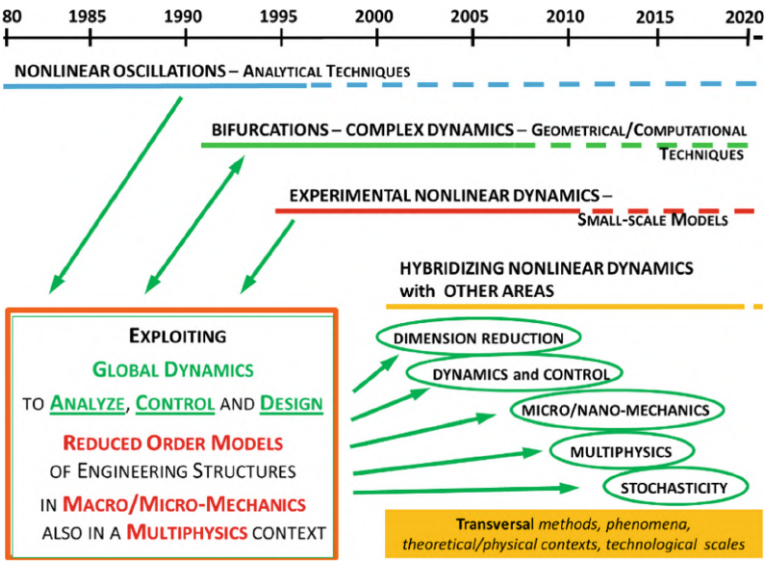


Fig. 1. Exploiting global dynamics in the frame of historical phases of development of nonlinear dynamics in mechanics.

2 Global Dynamics for Analysis and Safety

To highlight the role played by global dynamics in the analysis and safety of systems and structures, in the sequel we dwell on three main and distinct issues. The first refers to an exemplary situation in which local dynamics outcomes must be paralleled by a comprehensive use of global dynamics, in order to get a full understanding of the seemingly odd features of a structure's coupled response in a multi-field and multi-time-scale environment. The second addressed topic summarizes classical topological and computational knowledge in global dynamics, to the goal of outlining a systematic approach to the analysis of global phenomena and the assessment of their possibly unsafe effects in a variety of mechanical systems and structures. The third topic widens the

analysis/safety perspective to the non-deterministic context, dwelling on how classical nonlinear dynamics concepts and tools need to be revisited appropriately to allow for a reliable description and understanding of the effects produced by the uncertainties always occurring in the real world.

2.1 Slow Transient Effects on Fast Steady Response

Within a first mode discretization of mechanical and thermal fields, the unknowns $W(t)$, $T_{R0}(t)$, and $T_{R1}(t)$ of a dimensionless ROM of thermo-mechanically coupled plate

$$\begin{gathered} \ddot{W} + a_{12}\dot{W} + a_{13}W + a_{14}W^3 + a_{15}T_{R1} + a_{16}T_{R0}W + a_{17}cos(t) = 0 \\ \dot{T}_{R0} + a_{22}T_{R0} + a_{23}\alpha_1 T_\infty + a_{24}W\dot{W} = 0 \\ \dot{T}_{R1} + a_{32}T_{R1} + a_{33}\dot{W} = 0 \end{gathered} \tag{1}$$

with harmonic excitation $a_{17}cos(t)$, are the transverse displacement at plate midpoint, and the membrane and bending temperatures, respectively [4].

Starting from a pre-buckling configuration with the sole cross-well periodic solution $P1$, and imposing a constant temperature difference $T_\infty = 100$ between plate and environment, the local bifurcation diagram of transverse displacement versus varying thermal excitation T_∞ highlights the onset of couples of in-well buckled mechanical responses, which add to the persisting unbuckled solution (Fig. 2a). In global dynamics terms, the mechanical cross-section $(w, \dot{w})$ of the four-dimensional (4D) phase space $(w, \dot{w}, T_{R0}, T_{R1})$ at null initial values of thermal variables (i.e., $T_{R0}(0) = T_{R1}(0) = 0.0$) exhibits the sole pre-buckled solution $P1$ (Fig. 2b), both before and after application of thermal excitation T_∞. In contrast, the 2D phase portrait of the corresponding uncoupled mechanical system – i.e., sole Eq. (1a) with $T_{R0} = T_{R0}(T_\infty = 100)$ in the term $a_{16}T_{R0}W$ – highlights the occurrence of both pre-buckling $P1$ (dark gray) and buckled $P1^{III}/P1^{IV}$ (cyan/blue) solutions (Fig. 2c), consistent with the local bifurcation outcomes. This discrepancy ensues from the coexistence of *slow* (thermal) and *fast* (mechanical) dynamics in the coupled system, and is unveiled by comprehensively looking at its global dynamics.

Indeed, in the coupled system, the long transient response of the membrane temperature $T_{R0}(t)$ (Fig. 3b) entails a slow parametric-like contribution of the associated term $a_{16}T_{R0}(t)W(t)$ to the mechanical stiffness $a_{13}W(t)$ in Eq. (1a). Thus, the much faster mechanical response falls back onto the pre-buckling solution $P1$ (stable also upon applying thermal excitation) well before the complete deployment of T_∞ dynamic effects (Fig. 3a). The response evolution in the mechanical cross-sections of 4D phase space at increasing initial values of membrane temperature (Fig. 4a) – which correspond to progressively reducing transients in Fig. 3b (see different colors) – ends up to the same phase portrait as that of the uncoupled mechanical system (Fig. 2c) for the $T_{R0}(0) = 0.9$ initial value assumed as steady thermal excitation in the analysis of the latter.

The complex scenario of transient response is further apparent in the $\left(T_{R0}, \dot{W}\right)$ cross-section of 4D basins. Indeed, Fig. 4b highlights the evolution from a monostable $P1$ response for $T_{R0}(0) < 0.2$, to the appearance/enlargement of the compact part of $P1^{III}/P1^{IV}$ buckled basins in the range $0.2 < T_{R0}(0) < 0.9$ – with the response scenario for $T_{R0}(0) = 0.9$ being the sole one provided by the uncoupled mechanical system –, up to a strong fractalization significantly reducing basins' volume and making them unsafe

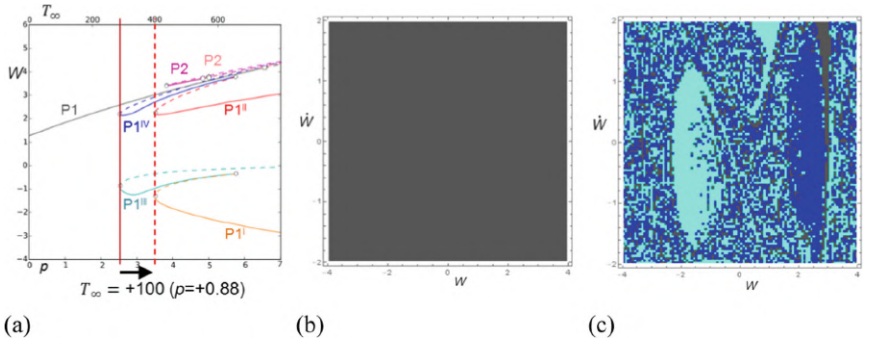


Fig. 2. (a) Local bifurcation diagram of transverse displacement vs. varying thermal excitation T_∞ (or equivalent mid-plane traction p). (b) Mechanical cross-sections $(w, \dot{w})$ of 4D $(w, \dot{w}, T_{R0}, T_{R1})$ phase space at $(T_{R0}(0) = T_{R1}(0) = 0.0)$ before and after application of thermal excitation T_∞ . (c) 2D phase portrait of corresponding uncoupled mechanical system ($T_\infty = 100$).

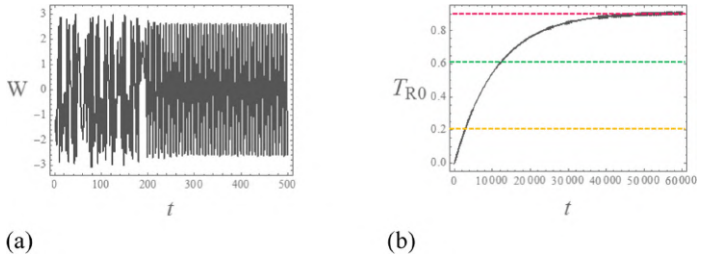


Fig. 3. (a) Fast mechanical response vs. (b) slow transient dynamics of membrane temperature.

for $T_{R0}(0) > 0.9$. In summary, global dynamics fully succeeds in unveiling the strong modification of steady mechanical vibrations occurring in thermo-mechanically coupled systems due to the long transients of the coexisting thermal field, as also observed when considering different sources of thermal excitation [3].

2.2 Global Bifurcations, Basin Erosion, Escape, Dynamic Integrity

In globally (locally) softening, damped, forced mechanical systems with a single (multi-well) potential, the topological scenario in phase space is governed by the invariant manifolds of one (or more) hilltop saddle(s). When varying a control parameter (typically, when increasing the excitation amplitude), the previously detached stable and unstable manifolds of the saddle(s) undergo a *global* (homoclinic or heteroclinic) *bifurcation* where they become tangent to each other, and then develop an infinite number of intersections. Correspondingly, the previously compact basin(s) of in-well bounded (typically, resonant and non-resonant) solutions, with smooth in-between boundaries, are eroded by the fractal tongues of out-of-well (and unbounded, in globally softening systems) solution, with a (though variable) swift decrease of basin(s) compactness and general progression of the system response towards final escape.

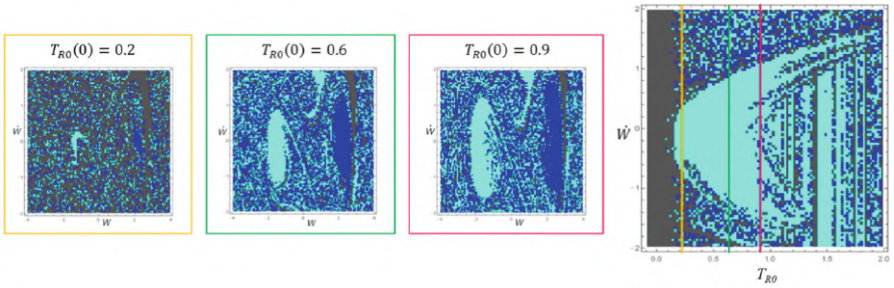


Fig. 4. (a) Mechanical response scenario in cross-sections of 4D phase space at increasing initial values of membrane temperature. (b) Complex scenario of transient response in $\left(T_{R0}, \dot{W}\right)$ cross-section of 4D basins at $T_{R1}(0) = 0$, $W(0) = -1.3$ (core of buckled $P1^{III}$ solution).

Escape is indeed the extreme dynamical condition representing the physical failure for a variety of softening mechanical systems, where it may correspond to, e.g., ship capsizing, rigid block overturning, dynamic buckling of slender structures, dynamic pull-in in micro-electro-mechanical-systems (MEMS) used as sensors, jump-to-contact in non-contact atomic force microscopes (AFMs). Even though there are also conditions of physically desired out-of-well solutions (e.g., in MEMS used as switches), an issue of general interest is how the basin *erosion* phenomenon develops up to escape when varying a control parameter, i.e., with either smooth or sudden decrease of response robustness. Indeed, attractors' safety strongly depends on the topological features of their basins, which must be enough large and compact to guarantee robustness of the corresponding dynamical outcomes with respect to also finite, though small, variations of initial conditions (i.c.) and/or system/excitation parameters. Moreover, in practical applications and experiments, problem-dependent disturbances generally entail further unexpected variations of i.c. and parameters, which may significantly affect the attractor-basin observability and ensuing safety. All of this must be assessed confidently in qualitative, and mostly quantitative, terms.

Dynamic integrity (DI) is a geometric-computational approach [5] to (i) reliably analyze erosion phenomena and their evolution in parameter space, and (ii) safely operate a system with the desired behavior depending on expected disturbances. It marks the transition from the sole locally-oriented perspective in the evaluation of attractors' safety based on Lyapunov stability, to a more comprehensive, i.e. also globally-driven, safety assessment taking the topological features of attractor-basin portraits into account, along with their evolution with a varying parameter. Overall, this corresponds to suitably complementing the local viewpoint to dynamic stability allowed by the classical, and mostly analytical, approach – which of course remains fundamental –, with also a global viewpoint entailing a wider understanding and knowledge of system dynamics, generally achievable only through heavy and systematic numerical simulations.

DI approach to stability is based on the following main concepts/tools (e.g., [6]).

(i) Safe basin, i.e. stating the initial conditions to be considered safe.
(ii) Measures, allowing to quantitatively assess the robustness of a safe basin.

(iii) Profiles/Charts, informing about basins' evolution due to parameters' variations.

Nominal safe basins are represented typically by the entire potential well of a softening system, or by the basin of a single attractor of interest. The most natural measure of a safe basin (GIM, Global Integrity Measure) is given by its area (in 2D), which however includes also unsafe fractal parts. If aiming to suitably eliminating them from the integrity evaluation, indicators of the sole compact part of the nominal safe basin must be adopted, such as the Local Integrity Measure (LIM) or the Integrity Factor (IF), which are radiuses of largest circles entirely belonging to the safe basin, centered at the corresponding attractor or not, respectively. For a softening system undergoing erosion up to final escape, the typical profile of a given integrity measure is shown in Fig. 5a. It summarizes how both Lyapunov and DI approaches allow to characterize the initial stability (safe existence) and final instability (disappearance) of a solution/attractor, whereas the sole DI approach provides information on the features of the basin's erosion process entailing loss of robustness in the mid-part of the profile, up to the inevitable escape. A process which may indeed develop in a sharp (i.e., Dover cliff-like) and very dangerous way, as exemplarily highlighted in Fig. 5b by the 3D chart of DI of a non-contact AFM over the 2D control space of excitation (amplitude and frequency) parameters [7].

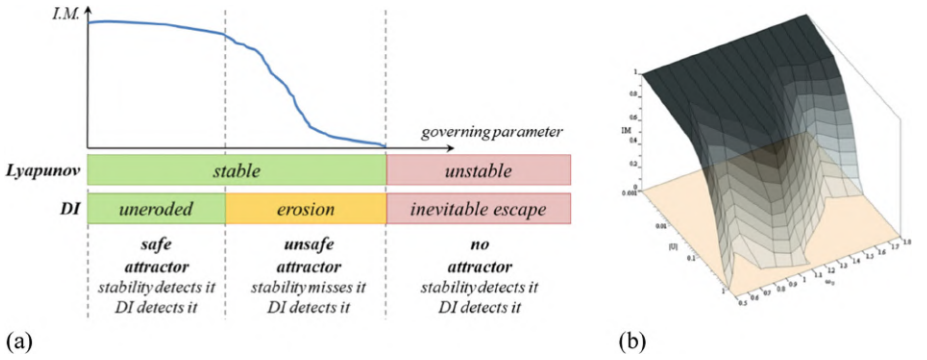


Fig. 5. (a) Schematic of erosion profile of a potential well in a generic softening system; Lyapunov (local) vs. DI (global) safety assessments. (b) Dover-cliff sharp loss of DI for a ROM of non-contact AFM.

2.3 Non-deterministic Global Analysis

Non-determinism is due to different sources, which include uncertainties in a system parameter or i.c., as well as stochastic effects such as those due to a white noise added to a deterministic excitation. All these cases can be dealt with effectively via a unified phase space operator approach, combined with a computationally suitable adaptive discretization in which refinements of attractors and basins boundaries are performed [8]. Perron-Frobenius and Foias operators are used to obtain attractors' distributions in systems with parameter uncertainty and noise, respectively, and Koopman operator to obtain basins' distributions.

For the illustrative case of a dimensionless Helmholtz oscillator with parametric uncertainty $\sigma\lambda$ added to the deterministic mean linear vibration frequency α,

$$\ddot{x} + \delta\dot{x} + (\alpha + \sigma\lambda)x + \beta x^2 = \gamma \sin \omega t \tag{2}$$

(λ being a random variable with a truncated standard normal distribution and σ its scaling factor), the following main effects occur when increasing σ. (i) The Dirac deltas representing non-resonant and resonant attractors of the deterministic system (top panels in Fig. 6) turn into attractors' mean distributions (first color bar in mid/bottom panels of Fig. 6) spreading over one-dimensional manifolds in phase space, with a varying probability. (ii) The responses from i.c. around formerly deterministic (and smooth, at the considered excitation amplitude $\gamma = 0.06$) basins boundary converge in mean sense to the two spread attractors, still with varying probability (second color bar). Basins' uncertain regions spread progressively over the phase space, up to no more regions of the resonant basin surely converging to the corresponding attractor in the mean sense for $\sigma = 0.06$ (right panel), where the probability is less than 0.8 everywhere.

Within a revisited definition of dynamic integrity [8], a weighted normalized basin area can be computed, based on the stochastic basin of attraction, an assumed probability threshold p in between 0 and 1, and an indicator function depending on the basin probability range. In the deterministic limit, the new definition reduces to the GIM. For both non-resonant and resonant attractors, Fig. 7 highlights not only the basin area decrease occurring for increasing values of the parametric uncertainty σ, but also, and even more important, the progressively steeper descent of the integrity profile that occurs when moving from an unsafely assumed null value of probability threshold p (top profile) up to its largest value 1 (bottom profile). Such a decreasing trend of integrity profiles adequately translates a less reliable estimate of the expected uncertainties into a more conservative safety assessment.

3 Global Dynamics for Control and Safety Assessment

To highlight the role played by global dynamics in control and safety of systems and structures, we dwell now on three main issues. First, we discuss how to leverage global concepts/phenomena for control purposes aimed at favorably modifying an overall region of stability. Second, we highlight the role of DI in interpreting/predicting global safety mismatches between theoretical and practical results associated with disturbances/uncertainties always existing in real systems, as well as global safety shortcomings induced by a solely locally-tailored control. Third, we deal with the possible assessment of an enhanced load carrying capacity of systems and structures based on a novel, global dynamics-informed, paradigm for engineering design, also accounting for the effects of uncertainties/perturbations.

3.1 Unified Control of Global Dynamics

Being DI reduction responsible for a dangerous safety loss of attractors/basins, triggered by an underlying global bifurcation, it is natural to investigate whether controlling the

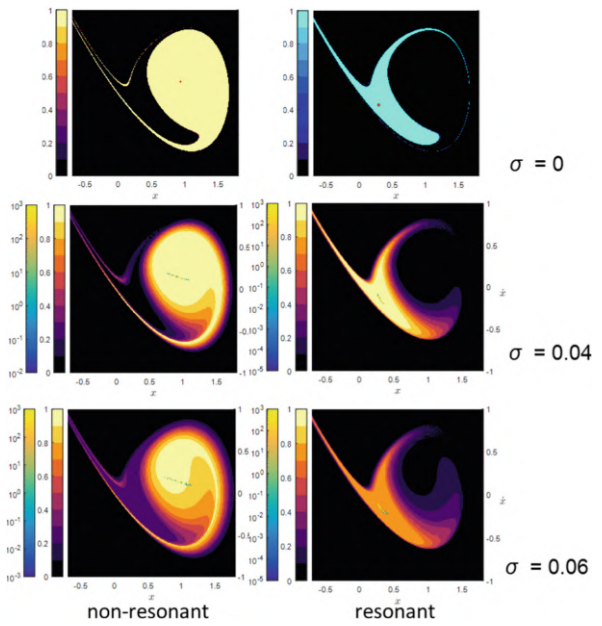


Fig. 6. Helmholtz oscillator with parametric uncertainty. Evolution of non-resonant and resonant attractor-basin portraits with increasing σ, moving from deterministic case (top panels). Mean attractor distributions (first color bar) and mean basins of attraction (second color bar) for non-deterministic cases (mid and bottom panels).

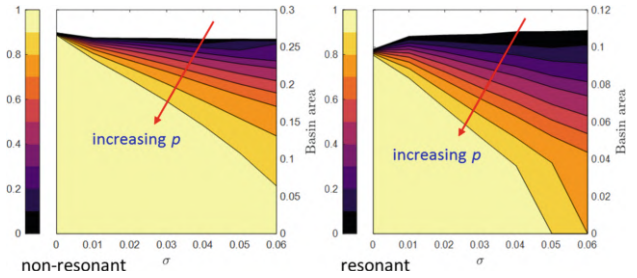


Fig. 7. Helmholtz oscillator with parametric uncertainty. Variation of basins' area as a function of σ, for increasing values (top to bottom profiles) of the assumed probability threshold p.

latter it is possible to affect favorably the former, too. Among various possible ways to avoid/delay the intersection of saddle manifolds in a harmonically forced system, there is the addition of appropriately chosen superharmonic components to the reference harmonic excitation. This entails keeping the period of the overall excitation fixed (as required by the system operating conditions), while changing its shape.

Let $\gamma^h_{1,cr}(\omega)$ be the critical amplitude for global bifurcation with an ω-harmonic excitation. If stable and unstable manifolds of a hilltop saddle intersect with each other for an excitation amplitude $\gamma_1 > \gamma^h_{1,cr}$ (Fig. 8a, upper panel), adding even a single

controlling superharmonic with properly selected amplitude h_2 and phase Ψ_2 may change the minimum of the distance along those manifolds,

$$d(\omega s) = c\left\{1 + \gamma_1/\gamma^h_{1,cr}(\omega)[\cos(\omega s) + h_2\cos(2\omega s + \Psi_2)]\right\} \tag{3}$$

(where c is a constant and s an arc length parameter), from negative (intersecting manifolds) to positive (Fig. 8a, lower panel), overall entailing manifolds' detachment at the γ_1 amplitude value of the resulting periodic excitation.

On this basis, a unified and system-independent control of global dynamics has been formulated [9] in which exploiting a global property (i.e., the intersection of stable and unstable manifolds) allows us to *control homo/heteroclinic bifurcations* triggering the erosion. Within this general framework:

(i) optimal choice of controlling superharmonics allows to maximize the excitation amplitude for occurrence of global bifurcation, with a significant gain $G = \gamma_{1,cr}/\gamma^h_{1,cr}$ (ratio between critical amplitudes for controlled, i.e., harmonic plus superharmonics, and reference/uncontrolled, i.e., sole harmonic, systems) being achieved in the overall control parameter space (see Fig. 8b for the deterministic Helmholtz oscillator).
(ii) localized or distributed (i.e., only one or more bifurcations) control may be implemented in phase space, depending on the governing global mechanism;
(iii) analytical control via the Melnikov function can be developed, if the global mechanism is governed by one (or more) hilltop saddle(s), as in many situations of technical interest.

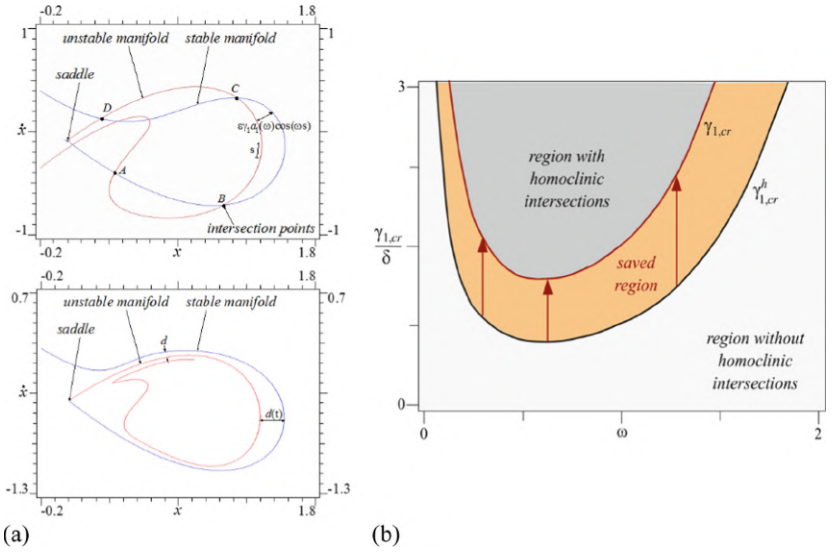


Fig. 8. (a) Manifolds intersection (upper panel) and detachment (lower panel) at a given value of excitation amplitude γ_1 in a harmonically excited Helmholtz oscillator, without or with a single added and properly chosen superharmonic component. (b) Optimal saved region in control space.

Global bifurcation control has been seen to work well in a large variety of mechanical/dynamical conditions, including systems hardening or softening, symmetric or asymmetric, perfect or imperfect, smooth or non-smooth, resonant or non-resonant, single- or multi-well, as well as bifurcations distributed or localized, homoclinic or heteroclinic, and transient or steady, deterministic or stochastic, dynamics, theoretical or experimental investigations. Even more important, global bifurcation control entails non-trivial advantages also in terms of *DI increase*, with the sharp fall down of erosion profiles being meaningfully delayed, i.e., shifted towards higher excitation amplitudes,. This is definitely a circumstance of major relevance from the safety viewpoint [6, 9].

Of course, this does not occur if the global bifurcation triggering erosion involves the manifolds of a saddle(s) different from the hilltop one(s), with the increase of critical amplitude ensuing from the realized separation of correspondingly controlled manifolds not entailing any improvement (i.e., shift towards higher excitation amplitudes) in the erosion profile. In such a case, first an extended geometric-computational investigation aimed at identifying the saddle(s) whose manifolds tangency actually triggers erosion, which strongly depends on systems' topological features, has to be accomplished. Then, a procedure aimed at realizing an optimal numerical control of the intersection of those specific manifolds must be implemented, to also succeed in delaying the sharp fall down of the profiles. A corresponding illustrative case has been dealt with in [10] for a single-mode model of non-contact AFM (for numerical control see also [11]), for which an internal saddle was found to be responsible for the in-well basins' separation via fractal tongues from the escape solution, with ensuing sudden loss of DI.

3.2 Systems' Global Safety

Theoretical vs. Practical Mismatches. Ranges of actual existence and stability of a system response in physical experiments and/or technical applications are sometimes different from, and usually unexpectedly smaller than, those predicted by theoretical/numerical analyses. This is generally associated with the uncertainties inevitably occurring in real situations as regards i.c. and/or system/operating parameters, which are not taken into adequate account in the mathematical models adopted in those investigations. The rationale behind such *mismatches* can be unveiled by a global analysis of system dynamics capable to highlight the loss of robustness that makes those responses practically unfeasible. A relevant, and somehow paradigmatic, situation [12] is summarized in the sequel.

Periodic rotations of a parametrically excited down-hanging pendulum have small basins of attractors with respect to those of coexisting oscillations, which makes it difficult to detect them in practical situations. Indeed, theoretical analysis predicts occurrence and stability of period-1 rotations in the quite large region of the excitation frequency-amplitude parameter plane of Fig. 9a in between the bifurcation loci of lower saddle-node (SN, where they originate) and upper period-doubling (PD, where period-2 rotations are established, thereafter evolving to chaos and successive boundary crisis (BC)). Conversely, in an experimental physical pendulum mounted on a vertically moving platform in a wave flume, period-1 rotations are detected only in the quite narrow intermediate orange strip of Fig. 9a approximately encompassing the blue triangles marking their physically robust occurrence. Such a meaningful mismatch is explained by looking at

the IF profiles of period-1 rotations with increasing excitation amplitude p, at different values of excitation frequency, as per the illustrative case in Fig. 9b obtained for $\omega = 1.30$ [12]. Indeed, within the overall IF profile – increasing up to meaningful values and then decreasing down to final disappearance –, DI is high enough to guarantee low response sensitiveness to finite variations of i.c. (as those inevitably occurring in experiments) only in the orange central strip approximately corresponding to the vertical range of practical occurrence of period-1 rotations in the physical system.

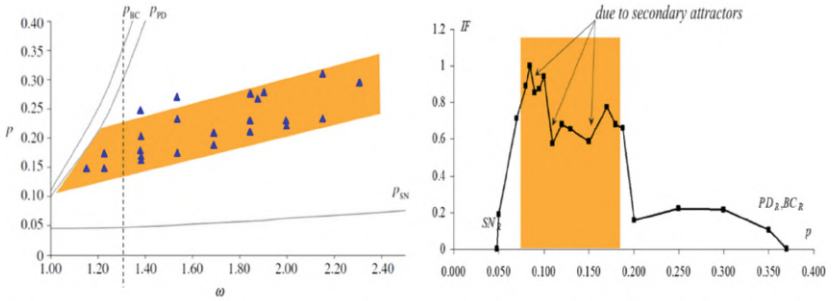


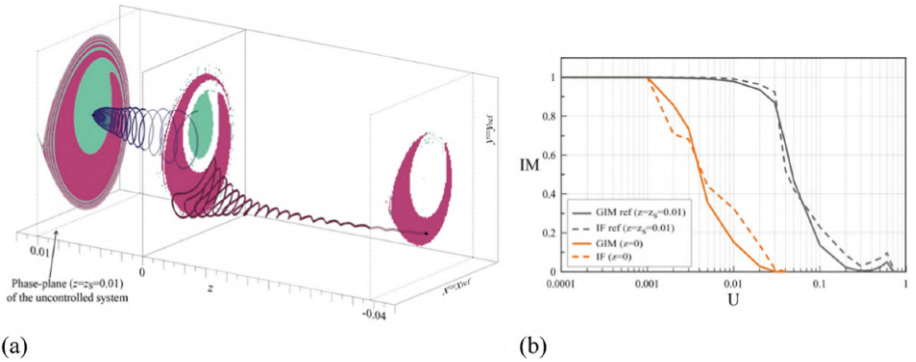
Fig. 9. Parametrically excited pendulum. (a) Theoretical (in between SN and PD bifurcation loci) versus experimental (orange strip, with blue triangles) regions of period-1 rotations. (b) Sample IF profile, with the central orange region of relatively high integrity (and substantially uneroded basin) approximately corresponding to the range of experimentally observed rotations [3].

Global Shortcomings of a Locally-Tailored Control. External feedback controls tailored for stabilizing a reference response of operating interest are successfully implemented, e.g. for micro/nano-systems. However, adding a control variable means enlarging the phase space dimension of the original (uncontrolled) system, and generally modifying the scenario of its local bifurcations. In turn, this may affect the overall features of global dynamic stability of the newly considered (i.e., controlled) system, so that it is of major importance to verify the extent of possibly occurred global changes.

In a single-mode model of softening non-contact AFM, for the excitation amplitude value $U = 0.01$ the 2D phase space of reference (i.e., uncontrolled) system exhibits in-well non-resonant and resonant solutions with smooth basins. In the 3D phase space of the controlled system of Fig. 10a (including the new control variable z and obtained for $x = x_{ref}, y = y_{ref}$), it is represented by the leftmost phase-portrait at $z_s = 0.01$, which is the z value identifying the solutions of the reference system to be stabilized via feedback control, related to an appropriate distance between microcantilever tip and sample surface [13]. However, the control modifies substantially the topological scenario of basins erosion. Indeed, the (intermediate) phase portrait obtained with fixed i.c. of the control variable $z(0) = 0$ highlights penetration of (white) tongues from the escape basin in between resonant and non-resonant basins. This occurs also for the low value of excitation amplitude $U = 0.01$ for which the reference system exhibits smoothly connected basins. 3D trajectories starting from i.c. within the (green) non-resonant basin of the controlled system at $z = 0$ correctly end up to the corresponding attractor on the

leftmost phase-plane (at $z = z_s$) of the reference system. In contrast, trajectories starting from i.c. within the (purple) resonant basin at $z = 0$ fail to reach the same cross-section, and end up to a new resonant attractor on the rightmost phase-plane at $z = -0.037 \neq z_s$, different from the reference one to be stabilized and thus unacceptable.

In terms of DI (Fig. 10b), this means moving from the GIM or IF (black) profiles of the reference system, whose safe basin includes resonant and non-resonant basins (leftmost phase-portrait in Fig. 10a), back to the (orange) profiles of the controlled system, whose safe basin consists of the sole correctly controlled non-resonant basin (intermediate phase-portrait in Fig. 10a). Overall, the locally-tailored control entails a dramatically *detrimental effect* in terms of system stability and global safety [13].



(a) (b)

Fig. 10. (a) 3D phase space $\left(x = x_{ref}\,, y = y_{ref}\right)$ of the controlled system at $U = 0.01$. Trajectories from i.c. within the (green) non-resonant basin at $z = 0$ end up to the corresponding reference attractor on the leftmost phase-plane (at $z = z_s$); trajectories from i.c. within the (purple) resonant basin at $z = 0$ end up to a resonant attractor on the rightmost phase-plane at $z = -0.037 \neq z_s$, different from the reference one and thus unacceptable. (b) Erosion profiles for the safe basin of the reference system (black, corresponding to union of solutions at $z = z_s = 0.01$) and the controlled system (orange, corresponding to the sole non-resonant solution at $z = 0$), evaluated through GIM (solid) and IF (dashed) measures.

3.3 Dynamic Integrity as a Possible Innovative Design Paradigm

In addition to highlighting the valuable *interpretative* role that DI can play in explaining theoretical vs. practical mismatches and global shortcomings of locally-tailored controls, previous analyses provide meaningful hints also on the even more appealing *predictive* role that DI can play for an informed, yet safe, innovative design properly accounting for the system nonlinear resources within a global dynamics framework. The matter is summarized by the schematic safety chart in the excitation parameter plane of Fig. 11, which refers to a generic softening system subjected to harmonic excitation [6], with further addition of some unavoidable source of (inherent or external) randomness.

The top stability threshold (usually resulting from the worst combination of local bifurcation loci above which bounded responses vanish) marks the transition to the

above (pink) region of escape, where the considered system is theoretically unstable. Generally, conventional approaches to safety do not allow operating conditions in which nonlinear phenomena come more or less significantly into play, to this aim introducing some properly selected safety factor (depending on the technological field) which allow to confidently operate a system only in the rather restricted (dark green) region below a corresponding safe threshold. Accordingly, the whole (light green) region in between safety factor bottom threshold and stability top threshold, wherein nonlinear and global dynamic phenomena take place progressively, is precluded from design. However, in the deterministic context, an intermediate DI safe threshold marked by an acceptable residual value of integrity can be suitably, and conservatively, identified, e.g. before the possible steep decrease of the DI profile to be definitely avoided. Further on, a minimal (and problem-dependent) DI safe threshold taking into account also non-deterministic effects can be established, based on the most unfavorable (or, equivalently, most conservative) probabilistic evaluation of safety reduction produced by the expected disturbances. All of this allows to determine a safe region of *practical* stability, where the system can be confidently operated. It is: (i) certainly larger than that provided by conventional safety factors; (ii) such to not overlook specifically dangerous dynamic situations (e.g., nonlinear resonances); (iii) capable to exploit a system's nonlinear/global resources in an informed way, to the purpose of identifying a safe – but not too conservative – lower bound for design.

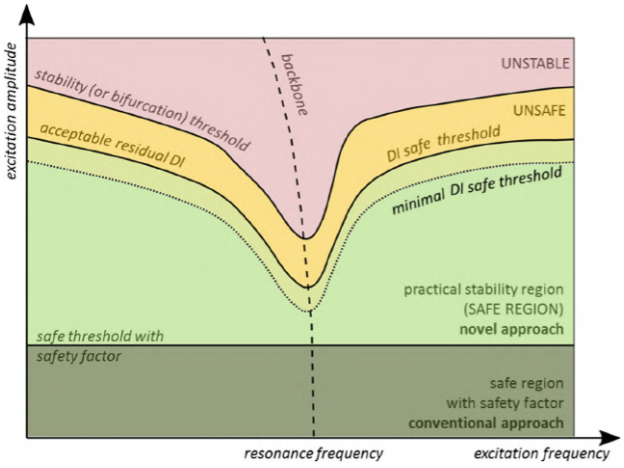


Fig. 11. Schematic safety chart for a generic softening system subjected to harmonic excitation, also accounting for non-deterministic effects. The light green–yellow–pink sequence of colors for increasing excitation amplitude corresponds to the DI strip sequence in Fig. 5a.

4 Global Dynamics of Multidimensional Systems

When dealing with the global dynamics of multidimensional models more reliably representing systems and structures of the real world, several theoretical and, mostly, computational challenges come into play. They are concerned both with the construction

(and also representation) of basins of attraction, and with the feasible accomplishment of a DI-based approach to systems' analysis, control and safety.

The first issue can be somewhat mitigated by implementing *high-performance computing* and a smart/full parallelization of the cell mapping procedures that allow to construct basins of attraction, even though demanding computational resources are all the way necessary mostly if operating in a non-deterministic framework [14]. The second issue is being dealt with by developing techniques capable of assessing solutions' robustness without computing basins of attraction. *Sample-based approaches* as the basin stability method fall into this scope [15]. It uses randomly chosen initial conditions and selected system parameters to account for all practically accessible dynamical states of a system with much lower computational effort than that required to construct full basins of attraction. However, quite limited information are obtained as regards basins' actual compactness and safety, with a substantial lack of knowledge about the overall robustness of the underlying solutions. This can be estimated confidently only through DI-based procedures, which are being made more rapid, and indeed less demanding, by using approximate and specifically tailored iterative procedures [16].

Anyway, independent of the further advancements which will certainly occur as to the feasible treatment of high-dimensional models, reduced order modeling remains of major importance for formulating refined and reliable ROMs of engineering systems, exhibiting the main phenomena of nonlinear modal interaction, and amenable to the extended computational analyses needed for safety evaluation and control via DI concepts and tools. In this last respect, the renewed interest towards *nonlinear model reduction* occurred in about the last decade within both the applied mathematics and the engineering environment, are to be adequately leveraged, along with the related advancements. These are concerned, on the theoretical side, with the developed mathematical theory of spectral sub-manifolds as exact invariant manifolds representing nonlinear continuations of linear normal modes [17] and, more generally, with the clear identification of two main different (graph and normal) styles of parametrization of invariant manifolds [18]. On the application side, with the meaningful advancements which are currently being made as regards the formulation of ROMs starting from outcomes of refined finite element models of high-dimensional systems (see, e.g., [19]). More generally, data-driven approaches using numerically generated data sets or experimental measurements are to be fully exploited both for formulating more comprehensive, yet more amenable, low-dimensional models [20], and for directly investigating the global properties of high-dimensional systems [21].

Acknowledgments. Credit is given to colleagues who collaborated to parts of the research work in the background of this overview, S. Lenci and V. Settimi, Polytechnic University of Marche, Italy, and K. Benedetti, Federal University of Goiás, Brazil.

References

1. Rega, G.: Nonlinear dynamics in mechanics and engineering: 40 years of developments and Ali H Nayfeh's legacy. Nonlinear Dyn. **99**(1), 11–34 (2020)

2. Rega, G.: Nonlinear dynamics in mechanics: state of the art and expected future developments. J. Comput. Nonlinear Dyn. **17**, 080802 (2022)
3. Rega, G., Settimi, V.: Global dynamics perspective on macro- to nano-mechanics. Nonlinear Dyn. **103**(2), 1259–1303 (2021)
4. Settimi, V., Rega, G., Saetta, E.: Avoiding/inducing dynamic buckling in a thermomechanically coupled plate: a local and global analysis of slow/fast response. Proc. Roy. Soc. A **474**, 20180206 (2018)
5. Thompson, J.M.T.: Chaotic phenomena triggering the escape from a potential well. Proc. Roy. Soc. A **421**, 195–225 (1989)
6. Rega, G., Lenci, S., Ruzziconi, L.: Dynamical integrity: a novel paradigm for evaluating load carrying capacity. In: Lenci, S., Rega, G. (eds.) Global Nonlinear Dynamics for Engineering Design and System Safety. CISM, vol. 588, pp. 27–112. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99710-0_2
7. Rega, G., Settimi, V.: Bifurcation, response scenarios and dynamic integrity in a singlemode model of noncontact atomic force microscopy. Nonlinear Dyn. **73**(1–2), 101–123 (2013)
8. Benedetti, K.C.B., Gonçalves, P.B., Lenci, S., Rega, G.: Global analysis of stochastic and parametric uncertainty in nonlinear dynamical systems: adaptative phase-space discretization strategy, with application to Helmholtz oscillator. Nonlinear Dyn. **111**, 15675–15703 (2023)
9. Rega, G., Lenci, S., Thompson, J.M.T.: Controlling chaos: the OGY method, its use in mechanics, and an alternative unified framework for control of non-regular dynamics. In: Thiel, M., Kurths, J., Romano, C., Moura A., Károlyi, G. (eds.) Nonlinear Dynamics and Chaos: Advances and Perspectives. UCS, pp. 211–269. Springer, Cham (2010). https://doi.org/10.1007/978-3-642-04629-2_11
10. Settimi, V., Rega, G.: Exploiting global dynamics of a noncontact atomic force microcantilever to enhance its dynamical robustness via numerical control. Int. J. Bifurcat. Chaos **26**(7), 1630018 (2016)
11. Lenci, S., Rega, G.: Optimal numerical control of single-well to cross-well chaos transition in mechanical systems. Chaos Solitons Fractals **15**, 173–186 (2003)
12. Lenci, S., Rega, G., Ruzziconi, L.: Dynamical integrity as a conceptual and operating tool for interpreting/predicting experimental behavior. Philos. Trans. Roy. Soc. Lond. A **371**, 20120423 (2013)
13. Settimi, V., Rega, G.: Local versus global dynamics and control of an AFM model in a safety perspective. In: Lenci, S., Rega, G. (eds.) Global Nonlinear Dynamics for Engineering Design and System Safety. CISM, vol. 588, pp. 229–286. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99710-0_5
14. Sun, J.-Q., Xiong, F.-R., Schütze, O., Hernández, C.: Cell Mapping Methods. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-0457-6
15. Brzeski, P., Perlikowski, P.: Sample-based methods of analysis for multistable dynamical systems. Arch. Comput. Methods Eng. **26**, 1515–1545 (2019)
16. Habib, G.: Dynamical integrity assessment of stable equilibria: a new rapid iterative procedure. Nonlinear Dyn. **106**(3), 2073–2096 (2021)
17. Haller, G., Ponsioen, S.: Nonlinear normal modes and spectral submanifolds: existence, uniqueness and use in model reduction. Nonlinear Dyn. **86**, 1–42 (2016)
18. Haro, A., Canadell, M., Figueras, J.-L., Luque, A., Mondelo, J.-M.: The Parameterization Method for Invariant Manifolds. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-29662-3
19. Opreni, A., Vizzaccaro, A., Touzé, C., Frangi, A.: High-order direct parametrisation of invariant manifolds for model order reduction of finite element structures: application to generic forcing terms and parametrically excited systems. Nonlinear Dyn. **111**, 5401–5447 (2023)

20. Cenedese, M., Axås, J., Yang, H., Eriten, M., Haller, G.: Data-driven nonlinear model reduction to spectral submanifolds in mechanical systems. Philos. Trans. Roy. Soc. Lond. A **380**, 20210194 (2022)
21. Li, Z., Jiang, J., Hong, L., Sun, J.-Q.: On the data-driven generalized cell mapping method. Int. J. Bifurcat. Chaos **29**(14), 1950204 (2019)

Keynote

Melnikov's Methods and Nonintegrability of Forced Nonlinear Oscillators

Kazuyuki Yagasaki(✉)

Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan
yagasaki@amp.i.kyoto-u.ac.jp

Abstract. Melnikov's methods, which enable us to detect the existence of transverse homo- and heteroclinic orbits and periodic orbits, and their bifurcations, are now ones of classical techniques to study forced nonlinear oscillators. Recently, similar approaches have been developed to show the nonintegrability of such systems, which are not necessarily Hamiltonian. In this overview article, I explain the approaches to prove the nonintegrability of forced nonlinear oscillators. I begin with standard results of Melnikov's methods for forced nonlinear oscillators, and give two definitions of integrability for Hamiltonian and more general systems. After briefly reviewing the classical work of Poincaré and Kozlov, the differential Galois theory and Morales-Ramis theory, I describe three new techniques to prove the nonintegrability of forced nonlinear oscillators and illustrate them for three types of the Duffing oscillators. Finally, I give some comments on related theoretical results and applications.

Keywords: Nonintegrability · Melnikov method · nonlinear oscillator · Morales-Ramis theory · differential Galois theory

1 Introduction

In this overview article we are mainly interested in forced nonlinear oscillators of the form

$$\dot{x} = J_1 \mathrm{D}H(x) + \varepsilon g(x, \omega t), \quad x \in \mathbb{R}^2, \tag{1}$$

where $0 < \varepsilon \ll 1$, $H : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$ are analytic, $g(x, \theta)$ is 2π-periodic in θ, and J_1 is the 2×2 symplectic matrix given by

$$J_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

When $\varepsilon = 0$, Eq. (1) becomes a Hamiltonian system

$$\dot{x} = J_1 \mathrm{D}H(x), \tag{2}$$

which can have various phase portraits as shown in Fig. 1: There may exist a homoclinic orbit, heteroclinic orbit or one-parameter family of periodic orbits. When $\varepsilon \neq 0$, they may persist or not.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 21–40, 2025.
https://doi.org/10.1007/978-3-031-72794-8_2

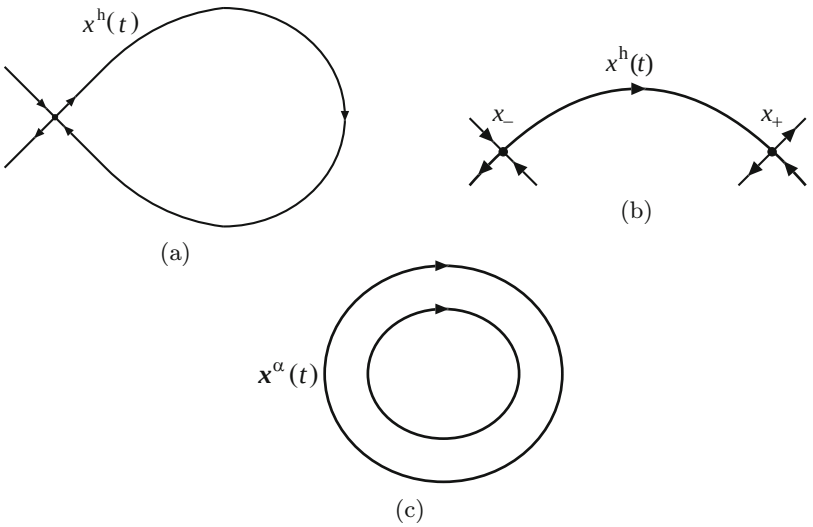


Fig. 1. Various phase portraits of the Hamiltonian system (2): (a) Homoclinic orbit; heteroclinic orbit; one-parameter family of periodic orbits.

Here we give two typical and important examples for (1). The first one is the forced Duffing oscillator

$$\ddot{y} + \varepsilon\delta\dot{y} + ay + by^3 = \varepsilon\beta\cos\omega t, \tag{3}$$

where $\delta, \beta, \omega > 0$ and a, b are constants. Equation (3) is rewritten in the form (1) with

$$H(x) = \tfrac{1}{2}x_2^2 + \tfrac{1}{2}ax_1^2 + \tfrac{1}{4}bx_1^4, \quad g(x,\theta) = \begin{pmatrix} 0 \\ -\delta x_2 + \beta\cos\theta \end{pmatrix}.$$

The second one is the forced damped pendulum

$$\ddot{\phi} + \varepsilon\delta\dot{\phi} + \sin\phi = \varepsilon\beta\cos\omega t, \quad \phi \in \mathbb{S}^1 = \mathbb{R}/2\pi\mathbb{Z}, \tag{4}$$

where $\delta, \beta, \omega > 0$ are constants. Equation (4) is rewritten in the form (1) with

$$H(x) = \tfrac{1}{2}x_2^2 - \cos x_1 \quad g(x,\theta) = \begin{pmatrix} 0 \\ -\delta x_2 + \beta\cos\theta \end{pmatrix}.$$

Without saying, such forced nonlinear oscillators are important in many fields of science and engineering [4,10,26,33]. It is believed that in general they are not "integrable" and their "general" solutions cannot be obtained analytically. So many approximate techniques have been developed to analyze these forced nonlinear oscillators [4,10,26,33]. Among them, *Melnikov's methods*, which are now classical, enable us to detect the existence of transverse homo- and heteroclinic orbits and periodic orbits, and their bifurcations [10,14,33,34,36,39].

On the other hand, for their nonintegrability, no proof has been given so far, unless they are Hamiltonian and have transverse homoclinic orbits. Note that Hamiltonian systems having transverse homoclinic orbits are analytically nonintegrable [21]. So we have the following questions on (1):

- Is it the reason we cannot obtain their "general" solutions that our knowledge is poor?
- Is it true that they are nonintegrable?

Note that it is impossible to obtain their "general" solutions if they are nonintegrable. As well-known, the problem of nonintegrability for Hamiltonian systems, especially for the restricted three-body problem, was studied by Poincaré [27,28] in the late nineteenth century. Subsequently, his result was refined by Kozlov [11,12] in the 80's.

In this article I present three methods developed recently by me and my coworker for proving the nonintegrability of forced nonlinear oscillators of the form (1) in an appropriate meaning. They are closely related to Melnikov's methods and to the classical results of Poincaré and Kozlov. See [24,25,41,45,47] for further details on the result. The problem of nonintegrability of dynamical systems has still attracted much attention and has been studied extensively. See also [50] and references therein for recent other developments on this topic.

The outline of the article is as follows. I begin with standard results of Melnikov's methods for forced nonlinear oscillators in Sect. 2. I next give two definitions of integrability for Hamiltonian and more general systems, and explain their meaning in Sect. 3. I briefly review the work of Poincaré and Kozlov in Sect. 4, and the Morales-Ramis theory [16,18], which is a modern theory for determining nonintegrability of dynamical systems based on the differential Galois theory [9,29], in a general setting in Sect. 5. In Sects. 6, 7 and 8 I explain the three new techniques to prove the nonintegrability of (1) and illustrate them for three types of the Duffing oscillator (3). The Morales-Ramis theory and its generalization are used in the second and third techniques while they are not in the first one. Finally, I give some comments on related theoretical results and applications in Sect. 9.

2 Melnikov's Methods

In this section, we consider forced nonlinear oscillators of the form (1) and review the standard results of Melnikov's methods [10,14,33,34].

2.1 Homoclinic Melnikov Method

We assume that the unperturbed system (2) has a hyperbolic saddle to which there exists a homoclinic orbit $x^{\mathrm{h}}(t)$, as shown in Fig. 1(a). Define the (*homoclinic*) *Melnikov function*

$$M(\theta) = \int_{-\infty}^{\infty} \mathrm{D}H(x^{\mathrm{h}}(t)) \cdot g(x^{\mathrm{h}}(t), \omega t + \theta)\mathrm{d}t, \tag{5}$$

where "$\cdot$" represents the inner product and $\theta \in \mathbb{S}^1$.

Theorem 1 (Melnikov [14], Guckenheimer & Holmes [10], Wiggins [33]). *If $M(\theta)$ has a simple zero, then there exist transverse homoclinic orbits.*

We have a similar result when the unperturbed system (2) has two hyperbolic saddles which are connected by a heteroclinic orbit $x^{\rm h}(t)$, as shown in Fig. 1(b). Theorem 1 is important since the existence of such orbits implies chaotic dynamics in (1) (see, e.g., [10,21,33]).

2.2 Subharmonic Melnikov Method

We next assume that the unperturbed system (2) has a one-parameter family of periodic orbits $x^\alpha(t)$, $\alpha \in (\alpha_1, \alpha_2)$, for some α_1, α_2 such that $\alpha_1 < \alpha_2$, as shown in Fig. 1(c). Let T^α denote the period of the periodic orbit $x^\alpha(t)$. Define the *subharmonic Melnikov function*

$$M^{m/n}(\theta) = \int_0^{mT} {\rm D}H(x^\alpha(t)) \cdot g(x^\alpha(t), \omega t + \theta) {\rm d}t, \tag{6}$$

where $\theta \in \mathbb{S}^1$, $T = 2\pi/\omega$ and the *resonance condition* $T^\alpha = mT/n$ holds for relatively prime integers $n, m > 0$.

Theorem 2 (Guckenheimer & Holmes [10], Wiggins [33]) *If ${\rm d}T^\alpha/{\rm d}\alpha \neq 0$ and $M^{m/n}(\theta)$ has a simple zero, there exist mT-periodic orbits near $x^\alpha(t)$ with $T^\alpha = mT/n$.*

The stability of periodic orbits detected in Theorem 2 can be easily determined by the sign of a Melnikov-type function

$$L^{m/n}(\theta) = \int_0^{mT} {\rm tr}\, {\rm D}_x g(x^\alpha(t), \omega t + \theta)\, {\rm d}t$$

and the Hopf and Bogdanov-Takens bifurcations of periodic orbits as well as saddle-node bifurcations can be detected by these functions. When ${\rm d}T^\alpha/{\rm d}\alpha = 0$ in Theorem 2, a little more complicated bifurcation behavior can be detected. See [34,36,39] for the details on the advanced results.

3 Definitions of Integrability

In this section I give two definitions of integrability for dynamical systems. The first one is for Hamiltonian systems, and the second one is for more general systems.

3.1 Liouville Integrability

Consider n-degree-of-freedom Hamiltonian system

$$\dot{x} = J_n \mathrm{D}H(x), \quad J_n = \begin{pmatrix} 0 & \mathrm{id}_n \\ -\mathrm{id}_n & 0 \end{pmatrix}, \quad x \in \mathbb{R}^{2n}, \tag{7}$$

where $H : \mathbb{R}^{2n} \to \mathbb{R}$ and id_n is the $n \times n$ identity matrix.

Definition 1 (Liouville). *The Hamiltonian system* (7) *is called* integrable *if there exist* n *scalar-valued functions* $F_1(x)(:= H(x)), F_2(x), \ldots, F_n(x)$ *such that*

(i) $\mathrm{D}F_1(x), \ldots, \mathrm{D}F_n(x)$ *are linearly independent almost everywhere (a.e) and their Poisson brackets are zero:*

$$\{F_j, F_k\}(x) := \mathrm{D}F_j(x) \cdot J_n \mathrm{D}F_k(x) \equiv 0, \quad j, k = 1, \ldots, n.$$

The functions $F_1(x), F_2(x), \ldots, F_n(x)$ *are called* first integrals.

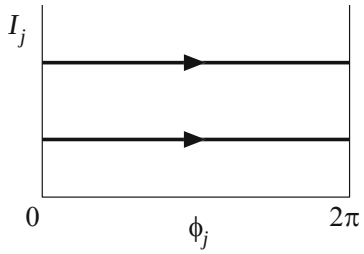


Fig. 2. Phase portraits of (8), where $\partial H/\partial I_j > 0$.

When $n = 1$, the Hamiltonian system (7) is always integrable. If the level set $\{x \in \mathbb{R}^{2n} | F_j(x) = c_j, j = 1, \ldots, n\}$, where c_j, $j = 1, \ldots, n$, are constants, is closed and bounded (i.e., compact in $\mathbb{R}^{2n}$), then there exist *action-angle coordinates* $(I_j, \phi_j) \in \mathbb{R} \times \mathbb{S}^1$, $j = 1, \ldots, n$, such that $H = H(I_1, \ldots, I_n)$ (see [2]):

$$\dot{I}_j = 0, \quad \dot{\phi}_j = \frac{\partial H}{\partial I_j}(I_1, \ldots, I_n), \quad j = 1, \ldots, n. \tag{8}$$

The phase portraits of (8) are very simple as displayed in Fig. 2.

3.2 Bogoyavlenskij Integrability

Consider a general n-dimensional system

$$\dot{x} = f(x), \quad x \in \mathbb{R}^n. \tag{9}$$

Definition 2 (Bogoyavlenskij [5]). *The system (9) is called* $(q, n-q)$-integrable *or simply* integrable *if there exist* q *vector fields* $f_1(x)$ $(:= f(x)), f_2(x), \ldots, f_q(x)$ *and* $n-q$ *scalar-valued functions* $F_1(x), \ldots, F_{n-q}(x)$ *such that*

(i) $f_1(x), \ldots, f_q(x)$ *are linearly independent a.e. and commutative, i.e.,*

$$[f_j, f_k](x) := \mathrm{D}f_k(x)f_j(x) - \mathrm{D}f_j(x)f_k(x) \equiv 0;$$

(ii) $\mathrm{D}F_1(x), \ldots, \mathrm{D}F_{n-q}(x)$ *are linearly independent a.e. and* $F_1(x), \ldots, F_{n-q}(x)$ *are first integrals of* $f_1, \ldots, f_q$*, i.e.,*

$$\mathrm{D}F_k(x) \cdot f_j(x) \equiv 0.$$

If the Hamiltonian system (7) is Liouville-integrable, then it is also (n, n)-integrable in the sense of Bogoyavlenskij. Actually, $J_n\mathrm{D}F_1(x), \ldots, J_n\mathrm{D}F_n(x)$ are commutative vector fields. Thus, Bogoyavlenskij-integrability is a generalization of Liouville-integrability. If the level set $\{x \in \mathbb{R}^{2n} \mid F_j(x) = c_j, j = 1, \ldots, n-q\}$, where c_j, $j = 1, \ldots, n-q$, are constants, is compact in $\mathbb{R}^{2n}$, then there exist action-angle coordinates $(I_1, \ldots, I_{n-q}, \phi_1, \ldots, \phi_q) \in \mathbb{R}^{n-q} \times \mathbb{T}^q$ with $\mathbb{T}^q = \prod_{j=1}^{q} \mathbb{S}^1$ such that

$$\dot{I}_j = 0, \quad \dot{\phi}_k = \Omega_k(I_1, \ldots, I_{n-q}), \quad j = 1, \ldots, n-q. \quad k = 1, \ldots, q, \tag{10}$$

like the Hamiltonian system (7) (see [5]). Moreover, an n-dimensional linear system with $f(x) = Ax$, $A \in \mathbb{R}^{n\times n}$, is $(n, 0)$-integrable. For instance, if A is diagonal, then $f_1(x) = Ax$ and $f_j(x) = x_j e_j$, $j = 2, \ldots, n$, where $e_1, \ldots, e_n \in \mathbb{R}^n$ are the standard basis, are commutative vector fields.

4 Classical Results of Poincaré and Kozlov

In this section I review the classical work of Poincaré [27,28] and Kozlov [11,12] on nonintegrability of dynamical systems.

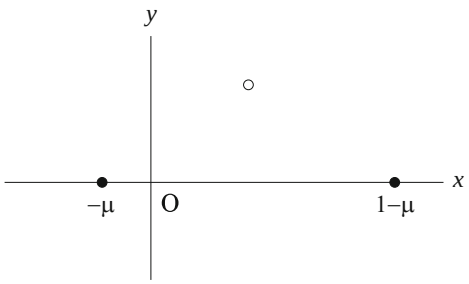


Fig. 3. Planar circular restricted three-body problem.

We begin with the planar circular restricted three-body problem (PCR3BP): Three particles move under mutual universal gravitation interaction, while one is massless and the other two are assumed to exhibit circular motions. See Fig. 3. The dimensionless equation of motion for the massless particle in the rotational frame where the other two are at rest is given by

$$\begin{aligned} \dot{x} &= p_x + y, \quad \dot{p}_x = p_y + \frac{\partial U}{\partial x}(x, y), \\ \dot{y} &= p_y - x, \quad \dot{p}_y = -p_x + \frac{\partial U}{\partial y}(x, y), \end{aligned} \tag{11}$$

where the two particles have masses $1-\mu$ and μ and are located at $(-\mu, 0)$ and $(1-\mu, 0)$, respectively, and

$$U(x, y) = \frac{\mu}{\sqrt{(x-1+\mu)^2 + y^2}} + \frac{1-\mu}{\sqrt{(x+\mu)^2 + y^2}}.$$

Poincaré [27,28] studied the PCR3BP (11) and showed the nonexistence of a first integral which is analytic in μ as well as the state variables x, y, p_x, p_y and functionally independent of the Hamiltonian

$$H(x, y, p_x, p_y) = \tfrac{1}{2}(p_x^2 + p_y^2) + (p_x y - p_y x) - U(x, y).$$

So he won the prize competition celebrating the 60th birthday of King Oscar II in 1889. To this end, he considered Hamiltonian systems of the general form

$$\dot{I} = -\varepsilon \mathrm{D}_\theta H_1(I, \theta; \varepsilon), \quad \dot{\theta} = \mathrm{D}H_0(I) + \varepsilon \mathrm{D}_I H_1(I, \theta; \varepsilon), \tag{12}$$

where $(I, \theta) \in \mathbb{R}^n \times \mathbb{T}^n$, $0 < |\varepsilon| \ll 1$ and H_0, H_1 are analytic in all the arguments. Recall that an integrable Hamiltonian system can be transformed into (12) with $\varepsilon = 0$ if the level set of first integrals is compact. The PCR3BP (11) is reduced to the Kepler problem and integrable when $\mu = 0$. Expand the perturbation term in the first equation of (12) into Fourier series as

$$\mathrm{D}_I H_1(I, \theta; 0) = \sum_{r \in \mathbb{Z}^n} \hat{h}_r(I) \exp(ir \cdot \theta),$$

where $\hat{h}_r(I)$, $r \in \mathbb{Z}^n$, are the Fourier coefficients. We refer to the set

$$\mathscr{P} = \{I \in \mathbb{R}^n \mid r \cdot \mathrm{D}_I H_0(I) = 0, \hat{h}_r(I) \neq 0\}$$

as a *Poincaré set*. We have the following theorem.

Theorem 3 (Poincaré [28], Kozlov [11]). *Suppose that $r \cdot \mathrm{D}_I H_0(I) \equiv 0$ only for $r = 0$, $\mathrm{D}_I H_0(I_0) \neq 0$ for some $I_0 \in \mathbb{R}^n$ and $\mathscr{P}$ is dense in a neighborhood of I_0. Then the system (12) has no first integral which is real-analytic in (I, θ, ε) and functionally independent of the Hamiltonian $H_0(I) + \varepsilon H_1(I.\theta; \varepsilon)$.*

In Theorem 3 the condition that $\mathscr{P}$ is dense can be weakened. See [11,12,24] for the details. Poincaré [28] applied his original version of Theorem 3 to (11) with $\varepsilon = \mu$. In particular, it was very hard to check the third condition of Theorem 3: He used 66 pages for it in the famous monograph [28]. Theorem 3 does not say about the possibility that the Hamiltonian system (12) is integrable for a specific value of ε. The result of Kozlov [11] contains stronger statements and was extended to non-Hamiltonian systems in his book [12].

5 Morales-Ramis Theory

In this section I review a generalized version due to Ayoul & Zung [3] of the Morales-Ramis theory [16,18], which is a modern theory for determining nonintegrability of dynamical systems based on the differential Galois theory [9,29].

5.1 Differential Galois Theory

We begin with the differential Galois theory. See the textbooks [9,29] for the details on this theory.

Consider a linear system on a (Riemann) surface Γ (e.g., a region in $\mathbb{C}$):

$$\dot{x} = A(t)x, \quad x \in \mathbb{C}^n, \quad t \in \Gamma, \quad A_{ij}(t) \in \mathbb{K}, \tag{13}$$

where $\mathbb{K}$ is a differential field (i.e., a field endowed with differentiation). Let $\Phi(t)$ be a fundamental matrix (of solutions) to (13). A differential field extension $\mathbb{L} \supset \mathbb{K}$ is called the *Picard-Vessiot extension* of (13) if $\Phi_{ij}(t) \in \mathbb{L}$. A field automorphism σ of $\mathbb{L}$ is called a $\mathbb{K}$-*automorphism* if $\sigma|_{\mathbb{K}} = \mathrm{id}$ and it commutes with the derivative. We easily see that $\sigma(\Phi(t))$ is also a fundamental matrix since

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma(\Phi(t)) = \sigma(\dot{\Phi}(t)) = \sigma(A(t)\Phi(t)) = A(t)\sigma(\Phi(t)).$$

Hence, by a fundamental result of linear differential equations (see, e.g., Chap. 3 of [8]), there exists a nonsingular matrix M_σ such that $\sigma(\Phi(t)) = \Phi(t)M_\sigma$. We see that

$$\mathrm{Gal}(\mathbb{L}/\mathbb{K}) = \{M_\sigma \mid \sigma \text{ is a } \mathbb{K}\text{-automorphism of } \mathbb{L}\}$$

is an (algebraic) group, and refer to it as the *differential Galois group* of (13). An algebraic group has a connected component containing the identity, which is called the *identity component*. In particular, we show that if the identity component of $\mathrm{Gal}(\mathbb{L}/\mathbb{K})$, which is denoted by $\mathrm{Gal}(\mathbb{L}/\mathbb{K})^0$, is conjugate to a triangular group, then Eq. (13) is solved by quadrature. See, e.g., Chapters 4 and 6 of [9], or Chapter 1 and Appendix A of [29] for the details.

Example 1. *Consider the linear system*

$$\dot{x} = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} - t & 2t \end{pmatrix} x, \quad x \in \mathbb{C}^2, \quad t \in \Gamma = \mathbb{C}.$$

Let $\mathbb{K} = \mathbb{C}(t)$*, which consists of all rational functions of* t *with complex coefficients. We easily see that*

$$\Phi(t) = \begin{pmatrix} e^t & 0 \\ \frac{1}{2}e^t & e^{t^2} \end{pmatrix}$$

is a fundamental matrix. We have $\mathbb{L} = \mathbb{C}(t, e^t, e^{t^2})$*, which consists of all rational functions of* t*,* e^t *and* e^{t^2} *with complex coefficients. For* $\sigma \in \mathrm{Gal}(\mathbb{L}/\mathbb{K})$ *we compute*

$$\frac{\frac{\mathrm{d}}{\mathrm{d}t}\sigma(e^t)}{\sigma(e^t)} = \sigma\left(\frac{\frac{\mathrm{d}}{\mathrm{d}t}e^t}{e^t}\right) = \sigma(1) = 1 = \frac{\frac{\mathrm{d}}{\mathrm{d}t}e^t}{e^t},$$

which means that $\log|\sigma(e^t)| = \log|e^t| + C$ *for some* $C \in \mathbb{C}$*. Hence, we obtain* $\sigma(e^t) = C_1 e^t$ *with* $C_1 = \pm e^C \in \mathbb{C}^* := \mathbb{C} \setminus \{0\}$*. Similarly,* $\sigma(e^{t^2}) = C_2 e^{t^2}$ *with* $C_2 \in \mathbb{C}^*$*. Thus,*

$$\sigma(\Phi(t)) = \begin{pmatrix} C_1 e^t & 0 \\ \frac{1}{2}C_1 e^t & C_2 e^{t^2} \end{pmatrix} = \Phi(t)\begin{pmatrix} C_1 & 0 \\ 0 & C_2 \end{pmatrix},$$

so that $\mathrm{Gal}(\mathbb{L}/\mathbb{K}) = \left\{ \begin{pmatrix} C_1 & 0 \\ 0 & C_2 \end{pmatrix} \middle| C_1, C_2 \in \mathbb{C}^* \right\} = \mathrm{Gal}(\mathbb{L}/\mathbb{K})^0$.

5.2 Morales-Ramis Theory

We turn to the generalized version of the Morales-Ramis theory for the general system (9). See [3,16,18] for the details on the theory.

Let $x = \phi(t)$ be a nonconstant solution to (9). The *variational equation* (*VE*) of (9) along $x = \phi(t)$ is given by

$$\dot{\xi} = \mathrm{D}f(\phi(t))\xi, \quad \xi \in \mathbb{C}^n. \tag{14}$$

We have the following theorem.

Theorem 4 (Morales-Ruiz & Ramis [16,18], Ayoul & Zung [3]). *If the system* (9) *is meromorphically integrable near* $x = \phi(t)$*, then the identity component of the differential Galois group is commutative.*

Here a function is called *meromorphic* if it is represented by the ratio of two holomorphic (analytic) functions such that its denominator is not constant zero. By Theorem 4, the identity component of the differential Galois group is not commutative, then the system (9) is not meromorphically integrable near $x = \phi(t)$.

Remark 1. *A higher-order theory in which higher-order VEs beyond* (14) *are used was also developed by Morales-Ruiz, Ramis and Simó* [20]*, and is called the* Morales-Ramis-Simó theory.

The Morales-Ramis and Morale-Ramis-Simó theories have been applied successfully to many dynamical systems including the Henon-Heiles system, general N-body problems ($N \geq 3$), the heavy top, homogeneous potentials, the Lorenz equation, the SEIR epidemic system and so on (see, e.g., [7,13,16,17,19,44]). In particular, parametrically forced nonlinear oscillators of the form

$$\ddot{y} + k_1(y, \dot{y}, \cos\omega t, \sin\omega t)y + k_2(y, \dot{y}, \cos\omega t, \sin\omega t)\dot{y} = 0,$$

where k_j, $j = 1, 2$, are rational functions of the arguments, were shown to be nonintegrable under a general condition in [23]. Moreover, the exact N-degree-of-freedom model for an undamped, straight beam

$$\ddot{u} + u'''' - \left(P + \kappa \int_0^\ell (u')^2 \mathrm{d}\zeta\right) u'' = 0$$

was shown to be meromorphically nonintegrable for any $N \geq 2$ in [38], where P, κ and ℓ, respectively, represent the initial tension, the stiffness and length of the beam, and the prime denotes differentiation with respect to $z \in (0, \ell)$. It is a key fact in application of the Morales-Ramis theory that when the identity component of the differential Galois group is triangularizable, the corresponding linear system is solved by quadrature, even if it is not commutative.

6 Method I: Real-Analytic Nonintegrability

In this and the next two sections I review three methods recently developed in [24,41,45,47] to prove nonintegrability of forced nonlinear oscillators of the form (1). These methods are closely related to Melnikov's methods in Sect. 2 and to the classical results of Poincaré and Kozlov in Sect. 4, as stated in Sect. 1.

We first consider real-analytic vector fields of the general form

$$\dot{x} = f(x; \varepsilon), \quad x \in \mathbb{R}^n, \tag{15}$$

where f is analytic in (x, ε) near $\varepsilon = 0$. We assume that the system (15) with $\varepsilon = 0$ is real-analytically $(q, n-q)$-integrable and that the level set $F^{-1}(c)$ of the $n-q$ first integrals $F(x) := (F_1(x), \ldots, F_{n-q}(x))$ is connected and compact for some $c \in \mathbb{R}^n$: Near the level set $F^{-1}(c)$ it is transformed into (10). We can show that F depends only on I. Let $\tilde{f}(I, \phi) = \mathrm{D}_\varepsilon f(\tilde{x}(I, \phi); 0)$, where $x = \tilde{x}(I, \phi)$ represents the transformation which recasts (15) into (10) when $\varepsilon = 0$. In this situation, we can prove the following.

Theorem 5 (Motonaga & Yagasaki [24]). *Suppose that $r \cdot \Omega(I) \equiv 0$ only for $r = 0$, $\operatorname{rank} \mathrm{D}_I \Omega(I^*) = n - q$ for some $I^* \in \mathbb{R}^{n-q}$ and the resonance condition $\dim_{\mathbb{Q}} \langle \Omega_1(I), \ldots, \Omega_q(I) \rangle = 1$ holds for a dense set of I in a neighborhood of I^* such that*

$$\mathscr{I}^I(\theta) := \int_0^{T^I} \mathrm{D}F(I) \cdot \tilde{f}(I, \Omega(I) + \theta) \mathrm{d}t \tag{16}$$

is not a constant function, where $\Omega_j(I)T^I = 0 \mod 2\pi$, $j = 1, \dots, q$. *Then the system* (15) *is not real-analytic integrable in a neighborhood of* $F^{-1}(c)$ *such that the first integrals and commutative vector fields also depend on* ε *real-analytically near* $\varepsilon = 0$.

Note that in Theorem 5 there exists a constant $T^I > 0$ such that $\Omega_j(I)T^I = 0 \mod 2\pi$, $j = 1, \dots, q$, if $\operatorname{rank} \mathrm{D}_I \Omega(I^*) = n - q$.

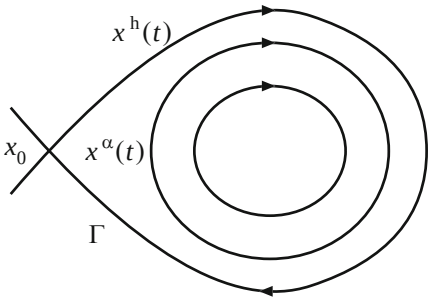


Fig. 4. Assumptions (M1)–(M3).

We turn to forced nonlinear oscillators of the form (1). Let H, g be real-analytic and assume the following when $\varepsilon = 0$:

- **(M1)** There exists a one-parameter family of periodic orbits $x^\alpha(t)$, $\alpha \in (\alpha_1, \alpha_2)$, with period $T^\alpha > 0$ for some $\alpha_1 < \alpha_2$;
- **(M2)** $x^\alpha(t)$ is real-analytic with respect to $\alpha \in (\alpha_1, \alpha_2)$.
- **(M3)** There exists a hyperbolic saddle x_0 with a homoclinic orbit $x^{\mathrm{h}}(t)$ such that $\lim_{\alpha \to \alpha_2} \sup_{t \in \mathbb{R}} d(x^\alpha(t), \Gamma) = 0$, where $\Gamma = \{x^{\mathrm{h}}(t) | t \in \mathbb{R}\} \cup \{x_0\}$ and $d(x, \Gamma) = \inf_{y \in \Gamma} |x - y|$. See Fig. 4.

Introducing $\phi_2 = \omega t \mod 2\pi$, we rewrite (1) as

$$\dot{x} = J_1 \mathrm{D}H(x) + \varepsilon g(x, \phi_2), \quad \dot{\phi}_2 = \omega. \tag{17}$$

When $\varepsilon = 0$, the system (17) is transformed to (10) with $n = 3$ and $q = 2$ (see also Eq. (18) below): The orbit $(x, \phi_2) = (x^\alpha(t), \omega t + \phi_{20})$ with a constant $\phi_{20} \in \mathbb{S}^1$ in (17) corresponds to an orbit in (17). Letting $F_1(x) = H(x)$ and $\mathrm{D}_\varepsilon f(x; 0) = g(x, \phi_2)$, we have $\mathscr{I}^I(\theta) = M^{m/n}(\theta)$. We apply Theorem 5 to obtain the following.

Theorem 6 (Motonaga & Yagasaki [24]). *Suppose that* $\mathrm{d}T^\alpha/\mathrm{d}\alpha \neq 0$ *at* $\alpha = \alpha_0$ *and* $M^{m/n}(\theta)$ *given by* (6) *is not constant in* θ *for a dense set of* $\alpha^{m/n}$ *in a neighborhood of* α_0 *under* (M1) *and* (M2). *Then the system* (17) *is not real-analytic integrable in a neighborhood of*

$$\{x^{\alpha_0}(t) \mid t \in [0, T^{\alpha_0})\} \times \mathbb{S}^1$$

such that the first integrals and commutative vector fields also depend on ε real-analytically near $\varepsilon = 0$.

Let $\hat{U} \subset \mathbb{R}^2$ be a region such that $\partial\hat{U} \supset \Gamma = \{x^{\rm h}(t) | t \in \mathbb{R}\} \cup \{x_0\}$ and $\hat{U} \supset \{x^\alpha(t) \mid t \in [0, T^\alpha)\}$ for some $\alpha \in (\alpha_1, \alpha_2)$. Since $\lim_{m\to\infty} M^{m/1}(\theta) = M(\theta)$, we use Theorem 5 to obtain the following.

Theorem 7 (Motonaga & Yagasaki [24]). *If $M(\theta)$ given by (5) is not constant, then the system (17) is not real-analytically integrable in $\hat{U} \times \mathbb{S}^1$ such that the first integrals and commutative vector fields also depend real-analytically on ε near $\varepsilon = 0$.*

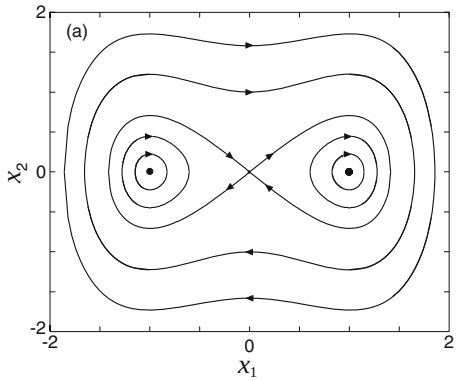


Fig. 5. Phase portraits of (3) with $a = -1$ and $b = 1$ when $\varepsilon = 0$.

Example 2. *Consider the periodically forced Duffing oscillator (3) with $a = -1$ and $b = 1$. This system was previously studied in many references, and in particular, it is one of early examples in which chaotic dynamics was proved to occur (see, e.g., [10,33]).*

When $\varepsilon = 0$, in the phase plane there exist a pair of homoclinic orbits

$$x^{\rm h}_\pm(t) = (\pm\sqrt{2}\,\mathrm{sech}\, t, \mp\sqrt{2}\,\mathrm{sech}\, t\, \tanh t),$$

a pair of one-parameter families of periodic orbits $x^k_\pm(t)$, $k \in (0,1)$, inside each of them, and a one-parameter family of periodic orbits $\tilde{x}^k(t)$, $k \in (1/\sqrt{2}, 1)$, outside of them, as shown in Fig. 5. Thus, assumptions (M1)–(M3) *hold for $x^k_\pm(t)$ and $x^{\rm h}_\pm(t)$. Moreover, assumptions* (M1) *and* (M2) *hold for $\tilde{x}^k(t)$ and*

$$\lim_{k\to 1} \sup_{t\in\mathbb{R}} d(\tilde{x}^k(t), \Gamma_+ \cup \Gamma_-) = 0, \quad \Gamma_\pm = \{x^{\rm h}_\pm(t) \mid t \in \mathbb{R}\} \cup \{(0,0)\}.$$

We compute the homoclinic Melnikov function (5) as

$$M_\pm(\theta) = -\frac{4}{3}\delta \pm \sqrt{2}\pi\omega\beta\,\mathrm{sech}\left(\frac{\pi\nu}{2}\right)\sin\theta.$$

Using Theorem 7 and its slight extension, we show that the system is real-analytically nonintegrable near $\Gamma_+ \cup \Gamma_-$ in the meaning of the theorem even if $M_\pm(\theta)$ have no zero, i.e., chaotic dynamics does not occur. See [24] *for more details.*

The real-analytic nonintegrability of the forced damped pendulum (4) near a pair of homoclinic orbits was proved in [25] by using Theorem 7 and its slight extension. Theorem 6 is not applicable for these examples, but it was improved very recently, so that their nonintegrability near periodic orbits can be discussed. See [22] for the details.

7 Method II: Complex-Meromorphic Nonintegrability

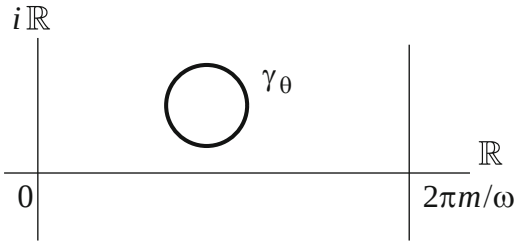


Fig. 6. Loop γ_θ

We consider (1) and assume that conditions (M1) and (M2) hold. Let

$$I = \frac{1}{2\pi}\int_{x^\alpha} x_2 \mathrm{d}x_1 = \frac{1}{2\pi}\int_0^{T^\alpha} x_2^\alpha(t)\dot{x}_1^\alpha(t)\mathrm{d}t.$$

We have the relation $\alpha = \alpha(I)$ as its inverse function. Using the transformation

$$x = x^{\alpha(I)}\left(\frac{\phi_1}{\Omega_1(I)}\right), \quad \Omega_1(I) = \frac{2\pi}{T^{\alpha(I)}},$$

and letting $\phi_2 = \omega t \mod 2\pi$, we rewrite (1) as

$$\dot{I} = \varepsilon h(I,\phi_1,\phi_2), \quad \dot{\phi}_1 = \Omega_1(I) + \varepsilon\Theta(I,\phi_1,\phi_2), \quad \dot{\phi}_2 = \omega, \tag{18}$$

where

$$h(I,\phi_1,\phi_2) = \frac{1}{\Omega_1(I)}\mathrm{D}H\left(x^{\alpha(I)}\left(\frac{\phi_1}{\Omega_1(I)}\right)\right)\cdot g\left(x^{\alpha(I)}\left(\frac{\phi_1}{\Omega_1(I)}\right),\phi_2\right),$$
$$\Theta(I,\phi_1,\phi_2) = J_1\frac{\partial}{\partial I}x^{\alpha(I)}\left(\frac{\phi_1}{\Omega_1(I)}\right)\cdot g\left(x^{\alpha(I)}\left(\frac{\phi_1}{\Omega_1(I)}\right),\phi_2\right).$$

Assume that $\Omega_1(I) = n\omega/m$, and add $\dot{\varepsilon} = 0$ to (18). The VE of the extended version of (18) along the resonant periodic orbit $(I, \theta_1, \theta_2, \varepsilon) = (I, \Omega_1(I)t + \theta_0, \omega t, 0)$ is given by

$$\begin{aligned} &\dot{\xi} = h(I, \Omega_1(I)t + \theta_0, \omega t)\chi, \quad \dot{\eta} = \mathrm{D}\Omega_1(I)\xi + \Theta(I, \Omega_1(I)t + \theta_0, \omega t)\chi, \\ &\dot{\zeta} = 0, \quad \dot{\chi} = 0, \end{aligned} \tag{19}$$

where $(\xi, \eta, \zeta, \chi) \in \mathbb{C}^4$. As a fundamental matrix of (19), we have

$$\Phi(t; \theta_0) = \begin{pmatrix} 1 & 0 & 0 & \Xi(t; \theta_0) \\ \mathrm{D}\Omega_1(I)t & 1 & 0 & \Psi(t; \theta_0) \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{20}$$

where

$$\begin{aligned} \Xi(t; \theta_0) &= \int_0^t h(I, \Omega_1(I)\tau + \theta_0, \omega\tau)\mathrm{d}\tau, \\ \Psi(t; \theta_0) &= \int_0^t \big(\mathrm{D}\Omega_1(I)\Xi(\tau; \theta_0) + \Theta(I, \Omega_1(I)\tau + \theta_0, \omega\tau)\big)\mathrm{d}\tau. \end{aligned}$$

Roughly speaking, we can compute the differential Galois group of (19) using the fundamental matrix (20). Define

$$\hat{\mathscr{I}}(\theta) = \int_{\gamma_\theta} \mathrm{D}H(x^\alpha(t)) \cdot g\,(x^\alpha(t), \omega t + \theta)\,\mathrm{d}t, \tag{21}$$

where γ_θ is a closed loop in a domain including $(0, 2\pi m/\omega)$ in $\mathbb{C}$ such that $\gamma_\theta \cap (i\mathbb{R} \cup (2\pi m/\omega + i\mathbb{R})) = \emptyset$ (see Fig. 6). The integrand in (21) is the same as in the subharmonic Melnikov function (6). Using Theorem 4, we can prove the following.

Theorem 8 (Yagasaki [41]). *Suppose that* $\mathrm{d}T^\alpha/\mathrm{d}\alpha \neq 0$ *and* $\hat{\mathscr{I}}(\theta) \neq 0$ *for some* $\theta = \theta_0 \in \mathbb{S}^1$. *Then the system* (17) *is not complex-meromorphically integrable near* $(x, \theta) = (x^\alpha(t), \omega t + \theta_0)$ *on any domain in* $\mathbb{C}/(2\pi m/\omega)\mathbb{Z}$ *containing* $\mathbb{R}/(2\pi m/\omega)\mathbb{Z}$ *and* γ_θ, *such that the first integrals and commutative vector fields depend complex-meromorphically on* ε *near* $\varepsilon = 0$.

A general result of [45] for systems of the form (12) was used in the proof of Theorem 8 in [41].

Example 3. *Consider the periodically forced Duffing oscillator* (3) *with* $a = 0$ *and* $b = 1$. *This system was one of early examples in which chaotic motions were observed in analog or numerical simulations* (*see, e.g.,* [31]).

Its phase portraits when $\varepsilon = 0$ *are shown in Fig. 7. In particular, there exists a one-parameter family of periodic orbits*

$$x^\alpha(t) = (\alpha\,\mathrm{cn}\,\alpha t, -\alpha^2\,\mathrm{sn}\,\alpha t\,\mathrm{dn}\,\alpha t), \quad \alpha \in (0, \infty),$$

and their period is given by $T^\alpha = 4K(1/\sqrt{2})/\alpha = 7.416\ldots/\alpha$, *where* sn, cn *and* dn *represent the Jacobi elliptic functions with the elliptic modulus* $k = 1/\sqrt{2}$, *and* $K(k)$ *is the complete elliptic integral of the first kind. See, e.g.,* [6,32] *for general information on elliptic functions. The integral* $\hat{\mathscr{I}}(\theta)$ *is computed as*

$$\hat{\mathscr{I}}(\theta) = -2\sqrt{2}\pi\omega\beta\left(\cosh\left(\frac{\pi m}{2n}\right)\sin\theta - i\sinh\left(\frac{\pi m}{2n}\right)\cos\theta\right).$$

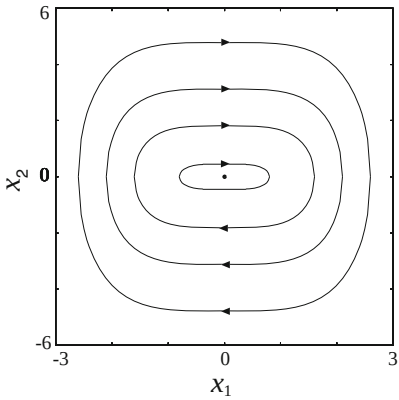


Fig. 7. Phase portraits of (3) with $a = 0$ and $b = 1$ when $\varepsilon = 0$.

See Appendix B of [41] *for the derivation of the above expression. Using Theorem 8, we show that the system is complex-meromorphic nonintegrable near the resonant periodic orbits satisfying*

$$mT^{\alpha} = \frac{2\pi l}{\nu}, \quad \textit{i.e.,} \quad \nu = \frac{\pi n \alpha}{2mK(1/\sqrt{2})}, \quad \textit{for } m, n \in \mathbb{N} \textit{ relatively prime}$$

in the meaning of the theorem even if the corresponding subharmonic Melnikov functions have no simple zero, i.e., the resonant periodic orbits do not persist. See [41] *for more details.*

The complex-meromorphic nonintegrability of the other cases of (3) and forced damped pendulum (4) near resonant periodic orbits was proved in [25,41] by using Theorem 8.

8 Method III: Real-Meromorphic Nonintegrability

We consider (1) and make the following assumptions for $\varepsilon = 0$:

- **(M1')** There exist two saddles at $x = x_{\pm}$ such that $J\mathrm{D}^2 H(x_{\pm})$ have a pair of real eigenvalues $\lambda_{\pm}, -\lambda_{\pm}$ $(\lambda_{\pm} > 0)$.
- **(M2')** $x = x_{\pm}$ are connected by a heteroclinic orbit $x^{\mathrm{h}}(t)$. See Fig. 1(b).
- **(M3')** $g(x, \omega t) = \sum_{j=-N}^{N} \hat{g}_j(x) e^{ij\omega t}$ for some $N \in \mathbb{N}$.

In assumption (M1') we allow $x_+ = x_-$, for which $x^{\mathrm{h}}(t)$ is a homoclinic orbit such as shown in Fig. 1(a), in (M2'). We first rewrite the system (1) as

$$\begin{aligned} &\dot{x} = J_1 \mathrm{D}H(x) + \varepsilon a_0(x) + \sum_{j=1}^{N} (a_j(x) u_j + b_j(x) v_j), \\ &\dot{\varepsilon} = 0, \quad \dot{u}_j = -j\omega v_j, \quad \dot{v}_j = j\omega u_j, \quad j = 1, \ldots, N, \end{aligned} \tag{22}$$

where

$$a_0(x) = \hat{g}_0(x), \quad a_j(x) = \hat{g}_j(x) + \hat{g}_{-j}(x),$$
$$b_j(x) = i(\hat{g}_j(x) - \hat{g}_{-j}(x)), \quad j = 1, \dots, N.$$

Letting $y_0 = \varepsilon$ and

$$y_j = \tfrac{1}{2}(u_j + iv_j), \quad y_{-j} = \tfrac{1}{2}(u_j - iv_j), \quad j = 1, \dots, N,$$

we express (22) as

$$\dot{x} = JDH(x) + \sum_{j=-N}^{N} \hat{g}_j(x) y_j, \quad \dot{y}_j = ij\omega y_j, \quad j = -N, \dots, N, \tag{23}$$

to which $(x, y) = (x^{\mathrm{h}}(t), 0) \in \mathbb{C}^2 \times \mathbb{C}^{2N+1}$ is a solution. We show that if the system (22) is real-meromorphically integrable near

$$\hat{\Gamma} = \{(x, \varepsilon, u, v) = (x^{\mathrm{h}}(t), 0, 0, 0) \in \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N \mid t \in \mathbb{R}\} \cup \{(x_\pm, 0, 0, 0)\},$$

then the system (23) is complex-meromorphically integrable near

$$\tilde{\Gamma} = \{(x, y) = (x^{\mathrm{h}}(t), 0) \in \mathbb{R}^2 \times \mathbb{R}^{2N+1} \mid t \in \mathbb{R}\} \cup \{(x_\pm, 0)\}$$

(see Lemma 3.1 of [47]). The VE of (23) along the solution is given by

$$\dot{\xi} = J_1 \mathrm{D}^2 H(x^{\mathrm{h}}(t))\xi + \sum_{j=-N}^{N} \hat{g}_j(x^{\mathrm{h}}(t))\eta_j, \quad \dot{\eta}_j = ij\omega\eta_j, \quad j = -N, \dots, N.$$

Using Theorem 4 based on an approach similar to those of [30,37,49], we can prove the following.

Theorem 9 (Yagasaki [47]). *Suppose that $M(\theta)$ given by (5) is not constant. Then the system (22) is not real-meromorphically integrable near $\hat{\Gamma}$ in $\mathbb{R}^{2N+3}$.*

Example 4. *Consider the periodically forced Duffing oscillator (3) with $a = 1$ and $b = -1$. See Fig. 8 for its phase portraits when $\varepsilon = 0$. In particular, there exists a pair of heteroclinic orbits*

$$x^{\mathrm{h}}(t) = \left(\pm \tanh(t/\sqrt{2}), \pm \operatorname{sech}^2(t/\sqrt{2})/\sqrt{2}\right)$$

connecting two equilibria $x = (-1, 0)$ and $(1, 0)$. We compute the Melnikov function (5) as

$$M_\pm(\theta) = -\frac{2\sqrt{2}}{3}\delta \pm \sqrt{2}\pi\omega\beta \operatorname{cosech}\left(\frac{\pi\omega}{\sqrt{2}}\right)\cos\theta.$$

Using Theorem 9, we show that the system is real-meromorphically nonintegrable near each of the heteroclinic orbits in the meaning of the theorem even if $M_\pm(\theta)$ have no zero. See [47] for more details.

The real-meromorphic nonintegrability of the case of $a = -1$ and $b = 1$, in which homoclinic orbits exist as shown in Fig. 5, was also discussed in [44].

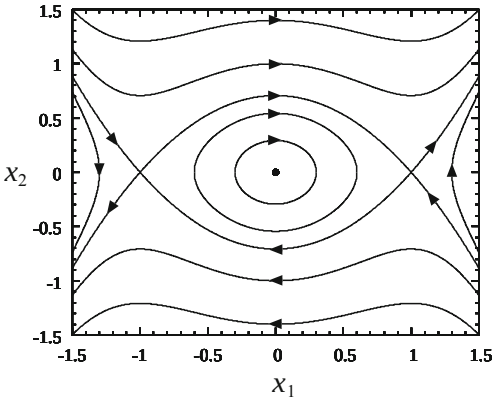


Fig. 8. Phase portraits of (3) with $a = 1$ and $b = -1$ when $\varepsilon = 0$.

9 Related Results

Finally, I give some comments on related theoretical results and applications.

First, the method of Sect. 7 was developed for more general systems written in the action-angle coordinates [41,45], which reduce to (12) when the systems are Hamiltonian. It was also used to show the nonintegrability of the CR3BP in the planar and spatial cases for any $\mu \in (0, 1)$ fixed in [45] as well as in a sense similar to those of Theorems 3 and 8 with $\varepsilon = \mu$ in [48]. A similar technique was developed for dissipative planar systems in [46].

Melnikov-type methods for two-degree-of-freedom Hamiltonian systems with saddle-center equilibria which have homo- or heteroclinic orbits on invariant planes were developed in [30,35]. These systems have one-parameter families of periodic orbits around the saddle-centers by the Lyapunov center theorem (see, e.g., Theorem 9.2.1 of [15]), and it is determined by the methods whether the stable and unstable manifolds of the periodic orbits intersect transversely. The relationship of the methods with the Morales-Ramis theory was discussed in [38,49] and the following were proven:

- When there exists a saddle-center having a homoclinic orbit, the sufficient condition obtained by the Morales-Ramis theory for the meromorphic nonintegrability near the homoclinic orbit is equivalent to one by the Melnikov method for transverse intersection between the stable and unstable manifolds of periodic orbits around it [38].
- When there exist a pair of saddle-centers that are connected by a heteroclinic orbit, the sufficient condition obtained by the Morales-Ramis theory for the meromorphic nonintegrability near the heteroclinic orbit is equivalent to one by the Melnikov method that the stable manifolds of periodic orbits around one of them intersect transversely or do not intersect the unstable manifolds of periodic orbits around the other [49].

The nonintegrability of normal forms of codimension-two bifurcations [10,33], which are one of the most interesting and important phenomena in dynamical systems, were discussed in [1,40,42,43]. The problem was reduced to the nonintegrability of planar

systems, on which the Morales-Ramis-Simó theory was used to discuss in [1,40,43] and not in [42]. In particular, the Rössler system and coupled van der Pol oscillators were shown to be analytically nonintegrable near equilibria, by using the result of [42].

Acknowledgments. This work was partially supported by the JSPS KAKENHI Grant Number JP22H01138. The author thanks the anonymous reviewers for pointing out some errors in the original version of the manuscript.

References

1. Acosta-Humánez, P.B., Yagasaki, K.: Nonintegrability of the unfoldings of codimension-two bifurcations. Nonlinearity **33**, 1366–1387 (2020). https://doi.org/10.1088/1361-6544/ab60d4
2. Arnold, V.I.: Mathematical Methods of Classical Mechanics, 2nd edn. Springer, Berlin (1989)
3. Ayoul, M., Zung, N.T.: Galoisian obstructions to non-Hamiltonian integrability. C. R. Math. Acad. Sci. Paris **348**, 1323–1326 (2010). https://doi.org/10.1016/j.crma.2010.10.024
4. Bogoliubov, N.N., Mitropolsky, Y.A.: Asymptotic Methods in the Theory of Nonlinear Oscillations. Gordon & Breach, New York (1961)
5. Bogoyavlenskij, O.I.: Extended integrability and bi-Hamiltonian systems. Comm. Math. Phys. **196**, 19–51 (1998). https://doi.org/10.1007/s002200050412
6. Byrd, P.F., Friedman, M.D.: Handbook of Elliptic Integrals for Engineers and Physicists. Springer, Berlin (1954)
7. Casale, G.: Morales-Ramis theorems via Malgrange pseudogroup. Ann. Inst. Fourier (Grenoble) **59**, 2593–2610 (2009). https://doi.org/10.5802/aif.2501
8. Coddington, E.A., Levinson, N.: Theory of Ordinary Differential Equations. McGraw-Hill, New York (1955)
9. Crespo, T., Hajto, Z.: Algebraic Groups and Differential Galois Theory. American Mathematical Society, Providence (2011)
10. Guckenheimer, J., Holmes, P.: Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields. Springer, New York (1983)
11. Kozlov, V.V.: Integrability and non-integrability in Hamiltonian mechanics. Russ. Math. Surv. **38**, 1–76 (1983). https://doi.org/10.1070/RM1983v038n01ABEH003330
12. Kozlov, V.V.: Symmetries. Topology and Resonances in Hamiltonian Mechanics. Springer, Berlin (1996)
13. Maciejewski, A., Przybylska, M.: Differential Galois theory and integrability. Int. J. Geom. Methods Mod. Phys **6**, 1357–1390 (2009). https://doi.org/10.1142/S0219887809004272
14. Melnikov, V.K.: On the stability of the center for time-periodic perturbations. Trans. Moscow Math. Soc. **12**, 1–56 (1963)
15. Meyer, K.R., Offin, D.C.: Introduction to Hamiltonian Dynamical Systems and the N-Body Problem, 3rd edn. Springer, Cham (2017)
16. Morales-Ruiz, J.J.: Differential Galois Theory and Non-integrability of Hamiltonian Systems. Birkhäuser, Basel (1999)
17. Morales-Ruiz, J.J.: Picard-Vessiot theory and integrability. J. Geom. Phys. **87**, 314–343 (2015). https://doi.org/10.1016/j.geomphys.2014.07.006

18. Morales-Ruiz, J.J., Ramis, J.P.: Galoisian obstructions to integrability of Hamiltonian systems. Methods Appl. Anal. **8**, 33–96 (2001). https://doi.org/10.4310/MAA.2001.v8.n1.a3
19. Morales-Ruiz, J.J., Ramis, J.P.: Integrability of dynamical systems through differential Galois theory: a practical guide. In: Acosta-Humánes, P.B., Marcellaán, F. (eds.) Differential Algebra, Complex Analysis and Orthogonal Polynomials, pp. 143–200. Mathematical Society, Providence (2010). https://doi.org/10.1090/conm/509/09980
20. Morales-Ruiz, J.J., Ramis, J.-P., Simó, C.: Integrability of Hamiltonian systems and differential Galois groups of higher variational equations. Ann. Sci. École Norm. Suppl. **40**, 845–884 (2007).https://doi.org/10.1016/j.ansens.2007.09.002
21. Moser, J.: Stable and Random Motions in Dynamical Systems. Princeton University Press, Princeton (1973)
22. Motonaga, S.: Subharmonic Melnikov functions and nonintegrability for autonomous and non-autonomous perturbations of single-degree-of-freedom Hamiltonian systems near periodic orbits. Phys. D **460**, 134088 (2024). https://doi.org/10.1016/j.physd.2024.134088
23. Motonaga, S., Yagasaki, K.: Nonintegrability of parametrically forced nonlinear oscillators. Regular Chaotic Dyn. **23**(3), 291–303 (2018). https://doi.org/10.1134/S156035471803005X
24. Motonaga, S., Yagasaki, K.: Obstructions to integrability of nearly integrable dynamical systems near regular level sets. Arch. Ration. Mech. Anal. **247**, 44 (2023). https://doi.org/10.1007/s00205-023-01880-x
25. Motonaga, S., Yagasaki, K.: Nonintegrability of forced nonlinear oscillators. Jpn. J. Ind. Appl. Math. **41**, 151–164 (2024). https://doi.org/10.1007/s13160-023-00592-9
26. Nayfeh, A.H., Mook, D.T.: Nonlinear Oscillations. Wiley, New York (1979)
27. Poincaré, H.: Sur le probléme des trois corps et les équations de la dynamique. Acta Math. **13**, 1–270 (1890). English translation: The Three-Body Problem and the Equations of Dynamics, 2017. Translated by D, Popp, Springer, Cham (2017)
28. Poincaré, H.: New Methods of Celestial Mechanics, vol. 1. AIP Press, New York (1992). (original 1892)
29. Van der Put, M., Singer, M.F.: Galois Theory of Linear Differential Equations. Springer, Berlin (2003)
30. Sakajo, T., Yagasaki, K.: Chaotic motion of the N-vortex problem on a sphere: I. Saddle-centers in two-degree-of-freedom Hamiltonians. J. Nonlinear Sci. **18**(5), 485–525 (2008). https://doi.org/10.1007/s00332-008-9019-9
31. Ueda, Y.: Random phenomena resulting from nonlinearity in the system described by Duffing's equation. Internat. J. Non-Linear Mech. **20**, 481–491 (1985). https://doi.org/10.1016/0020-7462(85)90024-1
32. Whittaker, E.T., Watson, G.N.: A Course in Modern Analysis, 4th edn. Cambridge University Press, Cambridge (1927)
33. Wiggins, S.: Introduction to Applied Nonlinear Dynamical Systems and Chaos. Springer, New York (1990)
34. Yagasaki, K.: The Melnikov theory for subharmonics and their bifurcations in forced oscillations. SIAM J. Appl. Math. **56**, 1720–1765 (1996). https://doi.org/10.1137/S00361399952813
35. Yagasaki, K.: Horseshoes in two-degree-of-freedom Hamiltonian systems with saddle-centers. Arch. Ration. Mech. Anal. **154**, 275–296 (2000). https://doi.org/10.1007/s002050000094

36. Yagasaki, K.: Melnikov's method and codimension-two bifurcations in forced oscillations. J. Differ. Equ. **185**, 1–24 (2002). https://doi.org/10.1006/jdeq.2002.4177
37. Yagasaki, K.: Galoisian obstructions to integrability and Melnikov criteria for chaos in two-degree-of-freedom Hamiltonian systems with saddle centres. Nonlinearity **16**, 2003–2012 (2003). https://doi.org/10.1088/0951-7715/16/6/307
38. Yagasaki, K.: Nonintegrability of an infinite-degree-of-freedom model for unforced and undamped, straight beams. Trans. ASME J. Appl. Mech. **70**, 732–738 (2003). https://doi.org/10.1115/1.1602483
39. Yagasaki, K.: Degenerate resonances in forced oscillators. Discrete Contin. Dyn. Syst. B **3**, 423–438 (2003). https://doi.org/10.3934/dcdsb.2003.3.423
40. Yagasaki, K.: Nonintegrability of the unfolding of the fold-Hopf bifurcation. Nonlinearity **31**, 341–350 (2018). https://doi.org/10.1088/1361-6544/aa92e8
41. Yagasaki, K.: Nonintegrability of nearly integrable dynamical systems near resonant periodic orbits. J. Nonlinear Sci. **32**, 43 (2022). https://doi.org/10.1007/s00332-022-09802-z
42. Yagasaki, K.: Nonintegrability of dynamical systems near degenerate equilibria. Comm. Math. Phys. **398**, 1129–1152 (2023). https://doi.org/10.1007/s00220-022-04545-0
43. Yagasaki, K.: Nonintegrability of truncated Poincaré-Dulac normal forms of resonance degree two. J. Differ. Equ. **373**, 526–563 (2023). https://doi.org/10.1016/j.jde.2023.07.017
44. Yagasaki, K.: Nonintegrability of the SEIR epidemic model. Phys. D **453**, 133820 (2023). https://doi.org/10.1016/j.physd.2023.133820
45. Yagasaki, K.: Non-integrability of the restricted three-body problem. Ergodic Theory Dyn. Syst. **44**, 3012–3040 (2024). https://doi.org/10.1017/etds.2024.4
46. Yagasaki, K.: Nonintegrability of dissipative planar systems. Phys. D **461**, 134106 (2024). https://doi.org/10.1016/j.physd.2024.134106
47. Yagasaki, K.: Nonintegrability of time-periodic perturbations of single-degree-of-freedom Hamiltonian systems near homo- and heteroclinic orbits. Phys. D **464**, 134189 (2024). https://doi.org/10.1016/j.physd.2024.134189
48. Yagasaki, K.: A new proof of Poincaré's result on the restricted three-body problem (Submitted)
49. Yagasaki, K., Yamanaka, S.: Heteroclinic orbits and nonintegrability in two-degree-of-freedom Hamiltonian systems with saddle-centers. SIGMA Symmetry Integrability Geom. Methods Appl. **15**, 049 (2019). https://doi.org/10.3842/SIGMA.2019.049
50. Zhang, X.: Integrability of Dynamical Systems: Algebra and Analysis. Springer, Singapore (2017)

Fundamental Mechanics

Intermodal Targeted Energy Transfer-Beyond Common Simplifications

Oleg V. Gendelman(✉)

Faculty of Mechanical Engineering, Technion - Israel Institute of Technology, 3200003 Haifa, Israel
ovgend@technion.ac.il

Abstract. Dynamics of transient strongly nonlinear phenomena develops rapidly, both at fundamental level and in view of multiple current and prospective applications. Intrinsic difficulty of the subject lies in the fact that many common notions and well-developed concepts of nonlinear dynamics, such as steady-state responses, bifurcations, nonlinear normal modes and many others, are not directly relevant, or at least hardly applicable in this sort of problems. Thus, the analysis and understanding of these phenomena require thinking beyond the aforementioned common simplifications. The paper offers a hands-on introduction to a novel topic within the broader subject of the transient strongly nonlinear phenomena. This is an emerging topic of intermodal targeted energy transfer for purposes of enhanced blast and earthquake mitigation. The paper presents simple introduction with basic developed examples. Brief descriptions of recent progress is accompanied by appropriate references.

Keywords: modal interactions · targeted energy transfer

1 Introduction

In most common designs in mechanical engineering and related fields in science and technology, one encounters primarily linear or weakly nonlinear (quasilinear) work regimes. This is easily understandable, since such regimes are well-explored and can be described by well-developed methods – e.g. averaging, multiple-scale expansions and other, often more advanced, asymptotic techniques. In the same time, strong (essential) nonlinearities occur in mechanical systems. Common reasons are clearances, impacts, friction, material, geometric or kinematic nonlinearities, external fields, scale effects and plasticity. In many applications such strongly nonlinear behavior is necessary, desirable and/or important for operational purposes, and widely studied [1–4]. In others, it is completely unwanted, but still profoundly reveals itself in the dynamics. Cracks in continuous structures are one such example [5]. Moreover, the accelerated need to predictively design lighter, faster, smaller, and ever more robust engineering systems with continuously expanding performance envelops, and operating in ever harsher environments dictates a paradigm-shifting approach, in the sense that nonlinear

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 43–57, 2025.
https://doi.org/10.1007/978-3-031-72794-8_3

effects and the ensuing cross-scale energy transfers may be intentionally used to achieve the design objectives. Among many applications, one can mention targeted energy transfer (TET) in essentially nonlinear systems with applications to energy absorption and harvesting [6–8], wave propagation, mitigation, energy redistribution and arrest in granular crystals and in systems characterized as "sonic vacua", i.e., with zero linear speed of sound [9–11], dynamic and acoustic non-reciprocity, as well as wave tailoring and control by means of acoustic metamaterials composed essentially nonlinear elements [12,13].

Recent observations and multiple results (see, e.g. [14]) allow one to claim that the intermodal energy transfers open an additional pathway for the mitigation of the undesired vibrations. Indeed, in typical excitation scenarios of large-scale structures (blast, earthquake etc.) commonly the low-frequency (long-wave) modes are excited. Then, if one succeeds to transfer substantial share of the energy to higher-frequency modes of the system, the intrinsic, and often relatively strong damping of these modes can substantially contribute to the overall mitigation. This concept is under intensive scrutiny now; it is referred as intermodal targeted energy transfer (IMTET).

The problem of IMTET exhibits certain salient features, absent in other common settings. First of all, it is profoundly nonlinear. The modes do not exchange energy in the linear system, therefore the IMTET effect is completely governed by the nonlinearity. Besides, this phenomenon is transient. The IMTET is intended to provide a mitigation for broadband and short-time excitations (like blast or earthquake). These features preclude the use of many concepts of nonlinear dynamics, such as periodic solutions, bifurcations, nonlinear normal modes etc., and puts the IMTET subject, to a large extend, beyond common concepts and approximations. Current paper introduces the reader into the IMTET subject. Section 2 is devoted to the resonant IMTET dynamics with closely spaced frequencies, Sect. 3 – to the non-resonant IMTET in various settings. Section 4 contains some concluding remarks.

2 Resonant IMTET - the Case of Weak Nonlinearity

The IMTET paradigm relies on enhancement of intermodal energy transfer inside the structure. To illustrate the principle, we consider a benchmark system with two degrees of freedom imbedded into large structure, as sketched in Fig. 1.

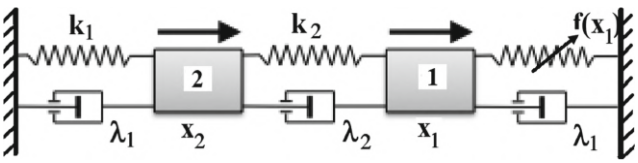


Fig. 1. Sketch of the benchmark 2DOF system with modal coupling.

The system consists of two oscillators with equal masses m, x_1 and x_2 denoting their displacements from equilibrium positions. Oscillator 1 is attached to the external system through a spring with nonlinear force characteristics $f(x_1)$ in parallel with linear damper with coefficient λ_1. Oscillator 2 is attached through similar damper, but in parallel with linear spring with stiffness coefficient k_1. The oscillators are coupled by linear spring with stiffness k_2 and additional linear damper with coefficient λ_2. For this simple benchmark model we assume that the underlying linear system is perfectly symmetric, but this symmetry in the system is broken by the nonlinear interactions. Therefore, the nonlinear force characteristics for Oscillator 1 is chosen in a form $f(x_1) = k_1 x_1 + a x_1^3$, where $a > 0$. Euations of motion of this system are written in the following way:

$$\begin{array}{l} m\frac{d^2x_1}{dt^2} + \lambda_1\frac{dx_1}{dt} + \lambda_2(\frac{dx_1}{dt} - \frac{dx_2}{dt}) + k_2(x_1 - x_2) + k_1x_1 + ax_1^3 = 0 \\ m\frac{d^2x_2}{dt^2} + \lambda_1\frac{dx_2}{dt} + \lambda_2(\frac{dx_2}{dt} - \frac{dx_1}{dt}) + k_2(x_2 - x_1) + k_1x_2 = 0 \end{array} \tag{1}$$

Equations (1) are brought to the non-dimensional form by the following redefinition of the constants and time units:

$$\tau = \omega t,\ \omega = \sqrt{k_1/m},\ \varepsilon\alpha = a/k_1,\ k_2/k_1 = \varepsilon\sigma,\ \lambda_1\Big/\sqrt{k_1 m} = \varepsilon\xi_1,\ \lambda_2\Big/\sqrt{k_1 m} = \varepsilon\xi_2 \tag{2}$$

With account of (2), Eq. (1) are re-written in the following form:

$$\begin{array}{l} \ddot{x}_1 + x_1 + \varepsilon\alpha x_1^3 + \varepsilon\sigma(x_1 - x_2) + \varepsilon\xi_1\dot{x}_1 + \varepsilon\xi_2(\dot{x}_1 - \dot{x}_2) = 0 \\ \ddot{x}_2 + x_2 + \varepsilon\sigma(x_2 - x_1) + \varepsilon\xi_1\dot{x}_2 + \varepsilon\xi_2(\dot{x}_2 - \dot{x}_1) = 0 \end{array} \tag{3}$$

The dot denotes differentiation with respect to the non-dimensional time τ. The notations in Eq. (3) reflect main assumptions concerning the model. First, we assume that the system works in weakly nonlinear regime, i.e. the non-dimensional bookkeeping parameter ε lies in the interval $0 < \varepsilon \ll 1$. Then, the coupling between the oscillators and all damping coefficients are small. Besides the small parameter, all coefficients and dependent variables are adopted to be of order unity.

To illustrate the considered phenomenon, we assume that the system is excited through fast strong perturbation of the external structure (for instance, a blast). It is reasonable to assume that such external excitation can be interpreted as simultaneous jump of the velocities of both oscillators to equal finite values. In other terms, Eq. (3) are simulated with the following initial conditions:

$$x_1(0) = 0, \dot{x}_1(0) = A, x_2(0) = 0, \dot{x}_2(0) = A \tag{4}$$

Due to the nonlinearity, the rate of energy absorption in the system will depend on the excitation level. To quantify this phenomenon, we define the *normalized energy* in the following way:

$$\begin{array}{l} \eta(\tau) = E(\tau)/E(0) \\ = \frac{1}{2A^2}\left[\dot{x}_1^2 + \dot{x}_2^2 + x_1^2 + x_2^2 + \varepsilon\sigma(x_1 - x_2)^2 + \frac{\varepsilon\alpha}{2}x_1^4\right] \end{array} \tag{5}$$

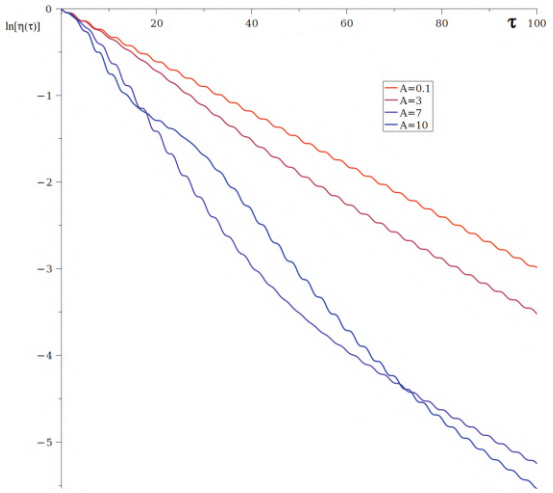


Fig. 2. Normalized energy versus time, semi-logarithmic scale, for initial conditions 4, with parameters $\varepsilon = 0.03, \sigma = 1, \xi_1 = 1, \xi_2 = 1, \alpha = 1$.

Here $E(\tau)$ is the energy stored in the system at given time instance. Time series for a logarithm of the normalized energy versus time for different excitation levels are presented in Fig. 2.

One can observe an interesting phenomenon: increase of the excitation level leads to substantially more rapid energy absorption in the system. In order to attain better understanding of the observed phenomenon, it is instructive to consider the model system in the modal space. To this end, we rely on linear modal substructure of the system, and introduce symmetric and antisymmetric modes by the following volume-preserving transformation:

$$v = (x_1 + x_2)\Big/\sqrt{2},\ w = (x_1 - x_2)\Big/\sqrt{2} \tag{6}$$

In terms of the modal coordinates, Eq. (6) are re-written in the following form:

$$\begin{aligned} &\ddot{v} + v + \varepsilon\xi_v\dot{v} + \tfrac{\varepsilon\alpha}{4}(v+w)^3 = 0 \\ &\ddot{w} + (1 + 2\varepsilon\sigma)w + \varepsilon\xi_w\dot{w} + \tfrac{\varepsilon\alpha}{4}(v+w)^3 = 0 \end{aligned} \tag{7}$$

Here coefficients $\xi_v = \xi_1$ and $\xi_w = \xi_1 + 2\xi_2$ characterize the proportional modal damping in the system. Initial conditions (4) in the modal coordinates (6) read simply as $v(0) = w(0) = \dot{w}(0) = 0, \dot{v}(0) = A\sqrt{2}$. Time series for the modal coordinates for the same parameter values as used for simulations in Fig. 2, are presented in Fig. 3.

Figure 3 highlights the internal mechanism of the observed phenomenon. For low excitation level (Fig. 3a) almost all energy remains in the initially excited symmetric mode, with relatively slow modal damping. For the high excitation level (Fig. 3b) the nonlinear coupling between the modes becomes significant, substantial amount of energy is transferred into the antisymmetric mode; the

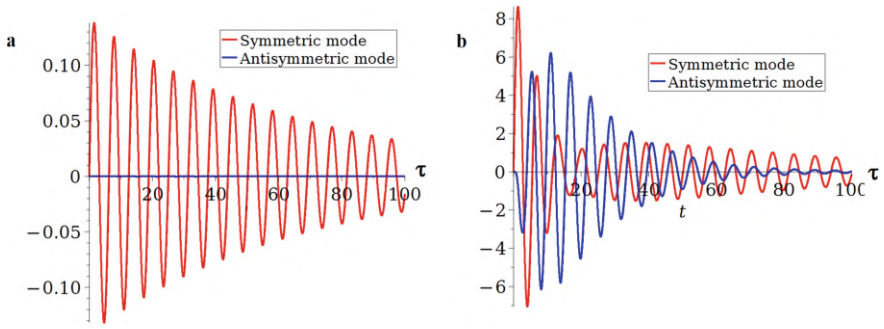


Fig. 3. Time series in the modal space for System 7, $\varepsilon = 0.03, \xi_v = 1, \xi_w = 3, \sigma = 1, \alpha = 1$. a) $A = 0.1$, b) $A = 7$. Red (thin) curve depicts the evolution $v(\tau)$ of the symmetric mode, and blue (thick) curve depicts the evolution $w(\tau)$ of the antisymmetric mode.

latter is characterized by substantially higher modal damping ratio. In Fig. 2 one can note that for low level of excitation (e.g. A = 0.1) the slope of the line is $-\xi_1$, and for higher levels of excitation (e.g. A = 0.7) it starts from values close to $-(\xi_1 + 2\xi_2)$ and returns back to $-\xi_1$ as the amplitude decreases. Thus, the enhancement of the damping properties of the system is realized through utilization of intrinsic modal structure, and excitation of additional oscillatory modes. This scenario supplies very interesting alternative to more common ways of mitigating the undesired vibrations, such as addition of degrees of freedom (e.g. tuned mass dampers or nonlinear energy sinks). In the current setting, the enhancement of the damping properties is achieved by design of the nonlinear intermodal interactions. The energy is absorbed by multiple modes, possibly with higher damping ratio. Appealing feature of this mechanism is that it does not require additional degrees of freedom. One can refer to this scenario as *intermodal targeted energy transfer* (IMTET). Major difference from common TET is that the energy is transferred into prescribed mode of the system, rather than to some physical fragment of the structure.

To quantify the IMTET phenomenon, we define for each set of parameters and initial conditions the *characteristic damping time* of the system τ^*, defined as

$$\eta(\tau^*) = e^{-1} \tag{8}$$

This definition is motivated by linear oscillator and is somewhat arbitrary. Still, it looks reasonable also in the nonlinear case, since it characterizes initial, most important and most dangerous stage of the blast mitigation. Later stages and lower energy levels are of less interest and concern. Dependence of the characteristic damping time on the modal excitation amplitude for select values of the modal damping coefficients is presented in Fig. 4. Values of the modal damping coefficients were chosen to illustrate various scenarios, in all of which both modes are underdamped, and the damping of the antisymmetric mode with higher frequency is also higher than that of the symmetric mode.

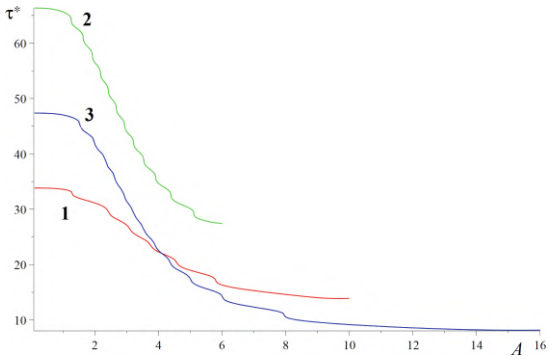


Fig. 4. Dependence of the characteristic damping time on the modal excitation amplitude, for $\varepsilon = 0.03, \sigma = 1, \alpha = 1$. Curve 1 (red) - $\xi_v = 1, \xi_w = 3$; curve 2 (green) - $\xi_v = 0.5, \xi_w = 1.5$; curve 3 (blue) - $\xi_v = 0.7, \xi_w = 5.7$.

The data presented in Fig. 4 lead to a number of interesting conclusions. First of all, the characteristic damping time can vary rather strongly as a function of the excitation amplitude (almost five-fold for Curve 3). So, the IMTET phenomenon allows very significant enhancement of the damping properties. Then, the transition to the regime of stronger damping is not sharp, and it is not possible to define strict threshold for the transition. Still, it is possible to see that the characteristic range of the amplitude, over which the transition takes place, significantly depends on the damping coefficients. In what follows, this characteristic range will be referred to as the IMTET threshold.

System 7 is formally quasilinear, and can be analyzed with the help of averaging method. To this end, we first perform a canonical transformation to action-angle variables:

$$\begin{array}{l} v = \sqrt{2I_v}\sin\theta_v;\ p_v = \dot{v} = \sqrt{2I_v}\cos\theta_v \\ w = \sqrt{\frac{2I_w}{\sqrt{1+2\varepsilon\sigma}}}\sin\theta_w;\ p_w = \sqrt{2I_w\sqrt{1+2\varepsilon\sigma}}\cos\theta_w \end{array} \tag{9}$$

Equations 7 can be re-written in canonical form:

$$\begin{array}{l} \dot{v} = \frac{\partial H}{\partial p_v};\ \dot{p}_v = -\xi_v p_v - \frac{\partial H}{\partial v};\ \dot{w} = \frac{\partial H}{\partial p_w};\ \dot{p}_w = -\xi_w p_w - \frac{\partial H}{\partial w} \\ H = \frac{p_v^2}{2} + \frac{p_w^2}{2} + \frac{v^2}{2} + \frac{(1+2\varepsilon\sigma)w^2}{2} + \frac{\varepsilon\alpha(v+w)^4}{16} \end{array} \tag{10}$$

General transformation to the action-angle variables is written in the following general form: $v = v(I_v, \theta_v)$; $p_v = p_v(I_v, \theta_v)$; $w = w(I_w, \theta_w)$; $p_w = p_w(I_w, \theta_w)$.

Equations 9 are transformed as follows:

$$\begin{pmatrix} \partial p_v/\partial I_v & \partial p_v/\partial \theta_v \\ \partial v/\partial I_v & \partial v/\partial \theta_v \end{pmatrix} \begin{pmatrix} \dot{I}_v \\ \dot{\theta}_v \end{pmatrix} = \begin{pmatrix} -\xi_v p_v - \frac{\partial H}{\partial v} \\ \frac{\partial H}{\partial p_v} \end{pmatrix}; \tag{11}$$
$$\begin{pmatrix} \partial p_w/\partial I_w & \partial p_w/\partial \theta_w \\ \partial w/\partial I_w & \partial w/\partial \theta_w \end{pmatrix} \begin{pmatrix} \dot{I}_w \\ \dot{\theta}_w \end{pmatrix} = \begin{pmatrix} -\xi_w p_w - \frac{\partial H}{\partial w} \\ \frac{\partial H}{\partial p_w} \end{pmatrix}$$

Due to canonicity of the action-angle transformation, both matrices in (11) have a unit determinant. Then, the following equations are obtained:

$$\begin{aligned} \dot{I}_v &= -\xi_v p_v \tfrac{\partial v}{\partial \theta_v} - \tfrac{\partial H}{\partial \theta_v}; \ \dot{I}_w = -\xi_w p_w \tfrac{\partial w}{\partial \theta_w} - \tfrac{\partial H}{\partial \theta_w} \\ \dot{\theta}_v &= \tfrac{\partial H}{\partial I_v} + \xi_v p_v \tfrac{\partial H}{\partial I_v}; \ \dot{\theta}_w = \tfrac{\partial H}{\partial I_w} + \xi_w p_w \tfrac{\partial H}{\partial I_w} \end{aligned} \tag{12}$$

The Hamiltonian is rewritten in the form of Fourier series:

$$H(I_v, I_w, \theta_v, \theta_w) = \sum_{k,l=-\infty}^{\infty} h_{kl}(I_v, I_w) \exp(i(k\theta_v + l\theta_w)), \ k, l \in Z \tag{13}$$

In the framework of approximation of the isolated resonance, we assume that the combination $\vartheta = \theta_v - \theta_w$ is the slow phase, and omit all other combinations of angles from expansion (13). Thus, one obtains the following expression for the *averaged Hamiltonian*:

$$\bar{H}(J_v, J_w, \vartheta) = \sum_{m=-\infty}^{\infty} h_{m,-m}(J_v, J_w) \exp(im\vartheta), \ m \in Z \tag{14}$$

Here J_v, J_w are average actions. While averaging Eq. (12), one should also take care of the damping terms in the form:

$$\begin{aligned} \left\langle p_v \tfrac{\partial v}{\partial \theta_v} \right\rangle &= \tfrac{1}{2\pi} \int_0^{2\pi} p_v \tfrac{\partial v}{\partial \theta_v} d\theta_v = J_v, \ \left\langle p_w \tfrac{\partial w}{\partial \theta_w} \right\rangle = J_w \\ \left\langle p_v \tfrac{\partial H}{\partial I_v} \right\rangle &= \left\langle p_w \tfrac{\partial H}{\partial I_w} \right\rangle = 0 \end{aligned} \tag{15}$$

Two first equations of (15) follow from the definition of action $I = \frac{1}{2\pi} \oint p dq$, and two second follow from quadratic dependence of the Hamiltonian on the canonic momenta – the trajectory of the single oscillator must be symmetric with respect to transformation p to $-p$.

Then, we note that $\frac{\partial \bar{H}}{\partial \theta_v} = -\frac{\partial \bar{H}}{\partial \theta_w} = \frac{\partial \bar{H}}{\partial \vartheta}$ and obtain the following generic averaged system:

$$\begin{aligned} \dot{J}_v &= -\xi_v J_v - \tfrac{\partial \bar{H}}{\partial \vartheta} \\ \dot{J}_w &= -\xi_w J_w + \tfrac{\partial \bar{H}}{\partial \vartheta} \\ \dot{\vartheta} &= \tfrac{\partial \bar{H}}{\partial J_v} - \tfrac{\partial \bar{H}}{\partial J_w} \end{aligned} \tag{16}$$

Substitution of (9) into Hamiltonian in (10) yields:

$$H = I_v + (1 + \varepsilon\sigma) I_w + \frac{\varepsilon\alpha(\sqrt{I_v}\sin\theta_v + \sqrt{I_w}\sin\theta_w)^4}{4} + O(\varepsilon^2) \tag{17}$$

Averaged Hamiltonian (15) is computed directly:

$$\bar{H} = J_v + (1+\varepsilon\sigma)J_w \qquad (18)$$
$$+\frac{\varepsilon\alpha}{4}\left(\begin{array}{l}\frac{3}{8}\left({J_v}^2+J_w^2\right)+\frac{3}{2}\left(J_v^{3/2}J_w^{1/2}+J_v^{1/2}J_w^{3/2}\right)\cos\vartheta\\+\frac{3}{2}J_vJ_w\left(1+\frac{1}{2}\cos 2\vartheta\right)\end{array}\right)+O(\varepsilon^2)$$

Substituting (18) into (16), one finally obtains:

$$\begin{aligned}J_v' &= -\xi_v J_v + \tfrac{3\alpha}{8}\left((J_v^{3/2}J_w^{1/2}+J_v^{1/2}J_w^{3/2})\sin\vartheta + J_vJ_w\sin 2\vartheta\right)\\ J_w' &= -\xi_w J_w - \tfrac{3\alpha}{8}\left((J_v^{3/2}J_w^{1/2}+J_v^{1/2}J_w^{3/2})\sin\vartheta + J_vJ_w\sin 2\vartheta\right)\\ \vartheta' &= -\sigma + \tfrac{3\alpha}{16}\left((J_w-J_v)(1+\cos 2\vartheta)+(J_w^{3/2}J_v^{-1/2}-J_v^{3/2}J_w^{-1/2})\cos\vartheta\right)\end{aligned} \qquad (19)$$

J_v and J_w are the average values of the actions for the symmetric and antisymmetric mode respectively. The apostrophe denotes differentiation with respect to the slow time scale $\tau_1 = \varepsilon\tau$. To obtain an introspection into the intrinsic IMTET dynamics, we present the numeric simulations of system (19) for typical set of parameters:

$$\xi_v = 1, \xi_w = 3, \sigma = 0.1, \alpha = 1 \qquad (20)$$

Results of simulations are presented in Fig. 5 (below the IMTET threshold) and Fig. 6 (in the IMTET range).

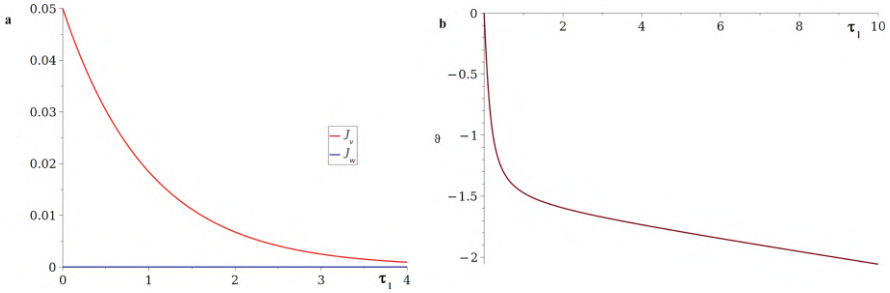


Fig. 5. Time series of averaged variables according to Eq. (19) with parameters (20) and initial conditions $J_v(0) = 0.05,\ J_w(0) = 10^{-7},\ \vartheta(0) = 0$.

Initial conditions for numeric simulation depicted in the preceding Figures were chosen to mimic as close as possible the initial excitation of the symmetric mode. Still one can see that the initial value for $J_w$is set to very small, but nonzero value. This is done to avoid the singularity in Eq. (19), but also has substantial physical meaning. Equations (19) describe the slow-flow dynamics on the RM. However, the system with generic initial conditions is captured into

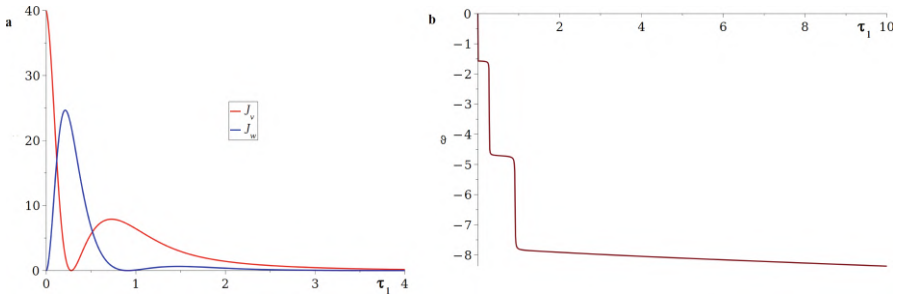


Fig. 6. Time series of averaged variables according to Eq. (19) with parameters (20) and initial conditions $J_v(0) = 40,\ J_w(0) = 10^{-7},\ \vartheta(0) = 0$.

the resonance and starts the evolution on the RM after certain non-resonance evolution. This process may be very short, but takes nonzero time. Far from the resonance, the action variable increases very slowly, and achieves very small, albeit nonzero, value. It is difficult to establish this value exactly, and its exact value does not have significant effect on the simulation results. We set it to be equal to 10^{-7}.

In Fig. 5a one observes the decay without beatings between the modal variables, and in Fig. 6a – final amount of beatings; in this case, both modes actively participate in the dynamics, and one can expect increased damping characteristics. Simulations of the phase difference yield an important insight. In the case without the beatings, one observes slow evolution of the phase difference. In the case of intense beatings, however, the phase difference remains almost constant ("locked") with jumps to the next locked states. The number of the locking periods is equal to the number of beating cycles. Moreover, one can observe that the locked states correspond to the phase difference values $\vartheta_{lock} \approx -\frac{\pi}{2} - \pi k,\ k = 0, 1, 2....$

Further understanding of system of equations (19) can be achieved with its certain modification. To this end, we define new variables:

$$J_v = P\cos^2(\gamma/2);\ J_w = P\sin^2(\gamma/2); \tag{21}$$

Here $P = J_v + J_w$ denotes the *intensity* of the excitation (not equal to the energy of the system!) and $\gamma = 2\arctan\sqrt{J_w/J_v}$ define the *intensity distribution* between the oscillators. One can restrict this angle in the interval $\gamma \in [0, \pi]$. Also, quite naturally, $P \in [0, \infty)$ and the phase difference can be restricted to $\vartheta \in [0, 2\pi)$. Thus, the set (P, γ, ϑ) describes the slow-flow dynamics in spherical coordinates with radius P, polar angle γ and azimuthal angle ϑ. In these coordinates, Eq. (19) are re-written in the following form:

$$\begin{aligned} P' &= -\tfrac{P}{2}\left(\xi_v + \xi_w + (\xi_v - \xi_w)\cos\gamma\right) \\ \gamma' &= \tfrac{1}{2}\left(\xi_v - \xi_w\right)\sin\gamma - \tfrac{3P\alpha}{8}\left(\sin\vartheta + \tfrac{1}{2}\sin\gamma\sin 2\vartheta\right) \\ \vartheta' &= -\left(\sigma + \tfrac{3P\alpha}{16}\left(\cos\gamma(1 + \cos 2\vartheta) + 2\cot\gamma\cos\vartheta\right)\right) \end{aligned} \tag{22}$$

Evolution of the intensity variable and the polar angle for the same initial conditions as used in Figs. 5 and 6 are presented in Figs. 7 and 8. These plots were computed by direct simulations of Eqs. (22).

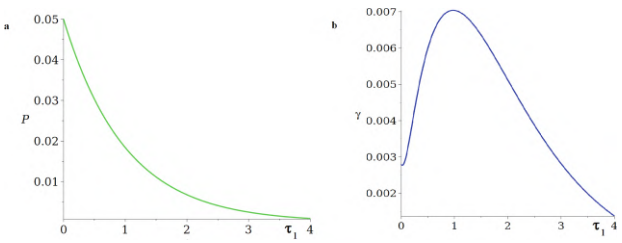


Fig. 7. Evolution of (a) intensity and (b) polar angle for Eq. (22) with parameters (20) and initial conditions $P(0) = 0.05, \gamma(0) = 0.0028, \vartheta(0) = 0$.

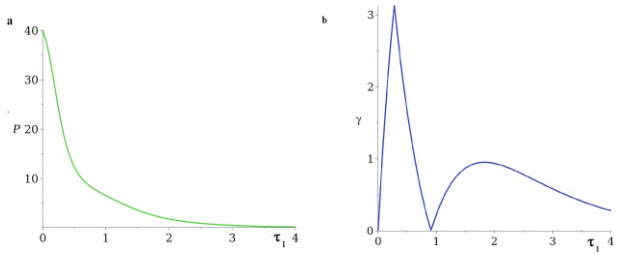


Fig. 8. Evolution of (a) intensity and (b) polar angle for Eq. (22) with parameters (20) and initial conditions $P(0) = 40, \gamma(0) = 10^{-4}, \vartheta(0) = 0$.

Most significant aspect of the RM dynamics is the behavior of polar angle γ. Below the IMTET threshold, it remains close to zero. Therefore, from the first equation of (22), it is obvious that the characteristic damping time for the intensity below the IMTET threshold can be estimated as

$$\tau_{below} \approx \frac{1}{\varepsilon \xi_v} \tag{23}$$

Note that we come back to physical time. Above the IMTET threshold, the polar angle rapidly oscillates between 0 and π, and therefore the characteristic damping time can be estimated as

$$\tau_{above} \approx \frac{2}{\varepsilon(\xi_v + \xi_w)} \tag{24}$$

The next, and most interesting question, is the estimation of the IMTET threshold in terms of initial intensity $P(0)$, directly related to the initial excitation amplitude. To this end, we assume that the dynamics is phase-locked at $\vartheta = -\pi/2$ and thus obtain

$$\begin{array}{l} P' \approx -\frac{P}{2}\left(\xi_v + \xi_w + (\xi_v - \xi_w)\cos\gamma\right) \\ \gamma' \approx \frac{1}{2}\left(\xi_v - \xi_w\right)\sin\gamma + \frac{3P\alpha}{8} \end{array} \tag{25}$$

The first term in the right-hand side of the second equation in (25) can be considered small, as P is large enough (note used numeric values). Then, we disregard it and, by dividing the first equation by the second, we obtain:

$$\begin{array}{r} \dfrac{dP}{d\gamma} \approx -\dfrac{4}{3\alpha}\left(\xi_v + \xi_w + (\xi_v - \xi_w)\cos\gamma\right) \\ \Rightarrow P(\gamma) \approx P(0) - \dfrac{4}{3\alpha}\left(\gamma(\xi_v + \xi_w) + (\xi_v - \xi_w)\sin\gamma\right) \end{array} \tag{26}$$

Here we set $\gamma(0) = 0$. Substituting expression (26) back into the second equation of (25) (with omitted first term), one obtains:

$$\gamma' \approx \frac{3\alpha}{8}\left(P(0) - \frac{4}{3\alpha}\left(\gamma(\xi_v + \xi_w) + (\xi_v - \xi_w)\sin\gamma\right)\right) \tag{27}$$

Equation (27) is still too difficult for analytic solution, but fortunately, it is not necessary for the crude estimation of the IMTET threshold. From Fig. 8b we can assume that the IMTET phenomenon occurs when γ can approach the value close to π. Then, at the maximum value, for $\gamma' = 0$, we can set $\gamma = \pi$, and obtain the following estimation for the critical IMTET threshold:

$$P_T \approx \frac{4}{3\alpha}\left(\pi(\xi_v + \xi_w) + (\xi_w - \xi_v)\right) = \frac{4}{3\alpha}\left((\pi - 1)\xi_v + (\pi + 1)\xi_w\right) \tag{28}$$

Keeping in mind changes of variables (6) and (9), one arrives to the following estimation for the threshold amplitude of excitation of the symmetric mode:

$$A_T \approx \sqrt{\frac{4}{3\alpha}\left((\pi - 1)\xi_v + (\pi + 1)\xi_w\right)} \tag{29}$$

Comparison between the analytic estimations are presented for different sets of parameters in three separate parts of Fig. 9.

One can observe that for all studied values of the parameters, the analytic approximations yield very accurate prediction for the characteristic damping time below the IMTET threshold (this is trivial), slightly overestimate this time above the IMTET threshold, and also, most importantly, yield very reasonable prediction for the characteristic threshold value.

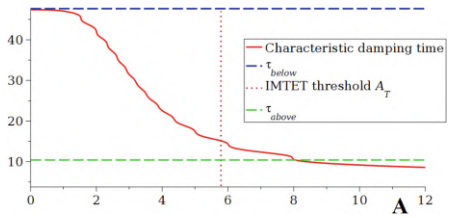


Fig. 9. Comparison between numeric dependence of the characteristic damping time on the modal excitation amplitude, versus theoretical predictions (23) (upper dashed line), (24) (lower dashed line) and (29) (vertical dotted line), for $\varepsilon = 0.03, \sigma = 1, \alpha = 1, \xi_v = 0.7, \xi_w = 5.7$.

3 Non-resonant IMTET

Cubic nonlinearity considered above allows efficient IMTET in the case of the modes with relatively close (narrowly spaced) frequencies. The case of widely spaced frequencies also can be handled, but requires stronger nonlinearity, going far beyond common quasilinear approaches. One simple solution is to apply, in a sense, the strongest possible nonlinearity - namely, the impact constraint. To this end, we modify the system sketched in Fig. 1, by leaving the right-hand spring linear (with stiffness coefficient c_1), but introducing rigid completely elastic impact barrier at clearance distance Δ from the right hand oscillator. The sketch of modified system is presented in Fig. 10.

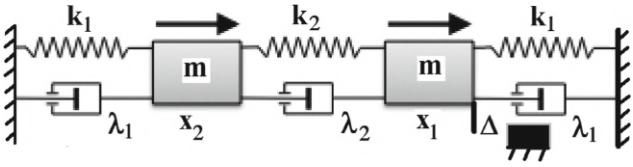


Fig. 10. Sketch of the benchmark 2DOF system with modal coupling and one-sided impact constraint.

Equations of motion in this case will read in the following form:

$$\begin{aligned} &m\frac{d^2x_1}{dt^2} + \lambda_1\frac{dx_1}{dt} + \lambda_2(\frac{dx_1}{dt} - \frac{dx_2}{dt}) + k_2(x_1 - x_2) + k_1x_1 = 0 \\ &m\frac{d^2x_2}{dt^2} + \lambda_1\frac{dx_2}{dt} + \lambda_2(\frac{dx_2}{dt} - \frac{dx_1}{dt}) + k_2(x_2 - x_1) + k_1x_2 = 0 \\ &x_1 < \Delta \end{aligned} \tag{30}$$

At the impact threshold, regular conditions of elastic Newtonian impact are assumed:

$$x_1(t_c) = \Delta; \quad \left.\frac{dx_1(t)}{dt}\right|_{t=t_c+0} = -\left.\frac{dx_1(t)}{dt}\right|_{t=t_c-0} \tag{31}$$

Here t_c is the collision time instant. For the simulation, we use the set of parameters $m = 1,\ k_1 = k_2 = 1,\ \lambda_1 = 0.05, \lambda_2 = 0.1, \Delta = 1$. Results of numeric simulations for modal amplitudes, for this set of parameters and initial conditions (4) are presented in Fig. 11.

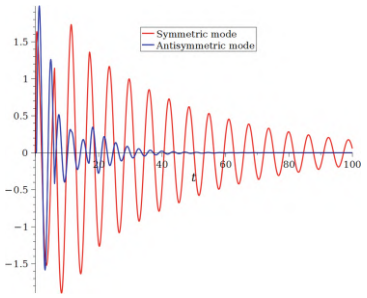


Fig. 11. Time series in the modal space for system with the impact constraint, A = 3. Red curve depicts the evolution $v(\tau)$ of the symmetric mode, and blue curve depicts the evolution $w(\tau)$ of the antisymmetric mode. The curves are denoted in the Figure.

One readily observes intense energy exchange between the modes; the plot for logarithm of normalized instantaneous energy as a function of time is presented in Fig. 12.

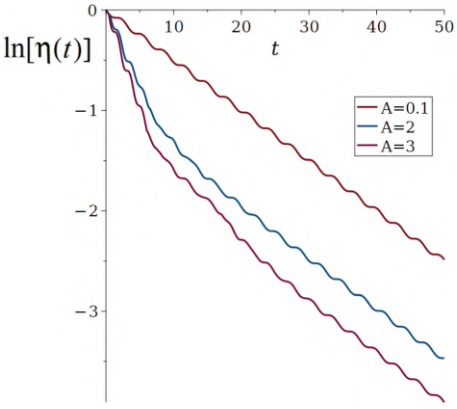


Fig. 12. Normalized energy versus time, semi-logarithmic scale, for the system with impact constraint.

The efficient damping in the system grows substantially also in the case of widely spaced modes. Further details on the non-resonant IMTET in a simple 2DOF model with a vibro-impact constraint under blast excitation are presented by [15]. The methodology has been expanded for multi-story buildings in [16].

Efficiency of the methodology for mitigation of multi-DOF systems in the case of seismic excitations has been demonstrated in a recent paper by [17]. In this paper, the condition of Newtonian impact has been relaxed to more realistic Hertzian contacts with the damping. The first experimental confirmation of the IMTET phenomenon has been presented by [18] for a simple cantilever beam with one-sided impact constraints. More general perspective on the IMTET and its relationship to other acute topics in nonlinear dynamics is outlined in feature paper [14].

4 Concluding Remarks

One can conclude that, despite all complications, some traditional concepts and methods of the nonlinear dynamics can keep their predictive force even in transient strongly nonlinear processes. Most important example is the notion of resonance - as it demonstrated above, the approximation of isolated resonance can be most instrumental also for the transient IMTET with close frequencies. In the same time, one also clearly observes the limitations of the resonant dynamics. The IMTET with widely spanned frequencies demonstrates that the performance can be achieved, and even substantially enhanced compared to the resonant interactions. This observation requires a lot of introspection. It turns out that in some sense the mechanism of the energy transport based on the internal and external resonances can be "substituted" by strongly nonlinear interactions that are not sensitive to exact frequency ratios. This peculiarity sharply distinguishes the IMTET from traditional resonance-based phenomena, as Sommerfeld effect [4] or the TET based on nonlinear energy sinks [6–8]. Such strongly nonlinear interactions are easily designed for mechanical systems (impacts, clearances etc.), but it is not clear whether one can find similar phenomena in other branches of science. In the same time, the resonant regimes are convenient for analysis, since they often allow substantial interrelated mathematical simplifications, such as averaging, multiple scales approach or isolation of the slow phase. The transient strongly nonlinear non-resonant regimes do not offer such options for the simplification and deep understanding, and to date the analysis relies primarily on numerics, especially for the systems with many degrees of freedom. It is a major challenge to reveal whether and how one can simplify the analysis of the dynamics in the case of non-resonant IMTET.

Acknowledgement. The author is very grateful to Israel Science Foundation (grant 2598/21) for support of this work.

References

1. Nayfeh, A.H., Balachandran, B.: Applied Nonlinear Dynamics: Analytical, Computational and Experimental Methods. Wiley, Hoboken (1995)
2. Lisyansky, A., Meimukhin, D., Starosvetsky, Y.: Primary wave transmission in the hexagonally packed, damped granular crystal with a spatially varying cross section. Commun. Nonlinear Sci. Numer. Simul. **27**, 193–205 (2015)

3. Lajimi, S.A.M., Friswell, M.I.: Dynamics of a non-linearly damped microresonator under parametric excitation and its application in developing sensitive inertial sensors with ultra-wide dynamic ranges. Int. J. Non-Linear Mech. **123**, 103491 (2020)
4. Fidlin, A.: Nonlinear Oscillations in Mechanical Engineering. Springer, Berlin (2006)
5. Babitsky, V.I., Hiwarkar, V.R.: Modelling of structures with developing discontinuity. J. Sound Vibr. **333**, 5917–5938 (2014)
6. Vakakis, A.F., Gendelman, O.V., Kerschen, G., Bergman, L.A., McFarland, D.M., Lee, Y.S.: Nonlinear Targeted Energy Transfer in Mechanical and Structural Systems, I and II. Springer, Heidelberg (2008)
7. Kremer, D., Liu, K.: A nonlinear energy sink with an energy harvester: transient responses. J. Sound Vib. **333**, 4859–4880 (2014)
8. Al-Shudeifat, M.A.: Highly efficient nonlinear energy sink. Nonlinear Dyn. **76**, 1905–1920–4880 (2014)
9. Nesterenko, V.F.: Dynamics of Heterogeneous Materials. Springer, New York (2001)
10. Boechler, N., Theocharis, G., Job, S., Kevrekidis, P.G., Porter, M.A., Daraio, C.: Discrete breathers in one-dimensional diatomic granular crystals. Phys. Rev. Lett. **104**, 244302 (2010)
11. Starosvetsky, Y.: Evolution of the primary pulse in one-dimensional granular crystals subject to on-site perturbations: analytical study. Phys. Rev. E **85**, 051306 (2012)
12. Boechler, N., Theocharis, G., Daraio, C.: Bifurcation-based acoustic switching and rectification. Nat. Mater. **10**, 665–668 (2011)
13. Vorotnikov, K., Starosvetsky, Y.: Nonlinear energy channeling in the 2D, locally resonant, unit cell model. Part I: high energy pulsations and routes to energy localization. Chaos **25**, 073106 (2015)
14. Vakakis, A.F., Gendelman, O.V., Bergman, L.A., Mojahed, A., Gzal, M.: Nonlinear targeted energy transfer: state of the art and new perspectives. Nonlinear Dyn. **108**, 711–741 (2022)
15. Gzal, M., Fang, B., Vakakis, A.F., Bergman, L.A., Gendelman, O.V.: Rapid nonresonant intermodal targeted energy transfer (IMTET) caused by vibro-impact nonlinearity. Nonlinear Dyn. **101**(4), 2087–2106 (2020). https://doi.org/10.1007/s11071-020-05909-8
16. Gzal, M., Vakakis, A.F., Bergman, L.A., Gendelman, O.V.: Extreme intermodal energy transfers through vibro-impacts for highly effective and rapid blast mitigation. Commun. Nonlinear Sci. Numer. Simulat. **103**, 106012 (2021)
17. Gzal, M., et al.: Seismic mitigation of a benchmark twenty-story steel structure based on intermodal targeted energy transfer (IMTET). Eng. Struct. **283**, 115868 (2023)
18. Tempelman, J.R., et al.: Experimental inter-modal targeted energy transfer in a cantilever beam undergoing vibro-impacts. J. Sound and Vibr. **539**, 117212 (2022)



Bias of the Steady-State Averaged Solutions of a Strongly Overdamped Particle in a Cosine Potential Under Harmonic Excitation

Attila Genda[1(✉)], Alexander Fidlin[1], and Oleg V. Gendelman[2]

[1] Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany
attila.genda@kit.edu
[2] Technion, Israel Institute of Technology, 3200003 Haifa, Israel
https://www.itm.kit.edu/dynamik/Mitarbeiter_Genda.php

Abstract. This study explores the dynamics of a strongly overdamped particle situated in a cosine potential when subjected to harmonic excitation. The primary focus is on the existence and stability of the particle's steady-state solutions. Using the harmonic balance method up to the first order and incorporating a constant bias term, our analytical approach reveals interesting dynamics influenced by the particle's mass. For a massless particle, steady-state solutions oscillate symmetrically around the potential well's minima or maxima. However, considering the particle's mass introduces a significant modification: new equilibrium positions emerge for the oscillation center, linking the extremities of the potential. This leads to a continuous biasing of the oscillation center solely by modulating the amplitude of the harmonic force. Our findings have potential implications for the precise control of particle positioning in various fields, including microscale and nanoscale applications.

Keywords: particle positioning · potential well · overdamped system · harmonic balance · stability analysis

1 Introduction

The precise and efficient control of particles at the microscale to the nanoscale is of growing interest, particularly for its rapid prototyping and biomedicine applications. Traditionally, optical manipulation has been a common approach in this domain. However, recent trends have shown an inclination towards acoustic forces for particle assembly enabling to avoid the need for chemical additives such as photoinitiators [1,2].

Moving beyond general techniques, our study focuses on the dynamics of a strongly damped particle in a cosine potential, a scenario akin to a damped-driven pendulum, a system investigated broadly in the literature. In his 1922 paper, *Hamel* [3] focused on the undamped forced oscillations of a pendulum

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 58–74, 2025.
https://doi.org/10.1007/978-3-031-72794-8_4

at finite amplitudes, analyzing the behavior and existence of periodic solutions. Subsequent research by *Hubbard* [4] examined the chaotic nature of the forced damped pendulum, shedding light on the intricacies and potential for control within these systems. The 21st century saw further developments, with *Coullet et al.* [5] analyzing the system's dynamics under a constant torque. *Boscaggin et al.* [6] contributed by investigating subharmonic solutions, and most recently, *Salas et al.* [7] provided closed-form solutions for forced damped rotational pendulum oscillators.

Investigation of the forced-damped pendulum allows us to explore fundamental aspects of particle behavior under harmonic forces in certain settings. We utilize the harmonic balance method, a well-established analytical tool in physics and engineering, to model and understand these dynamics. This method has been extensively covered in the literature and is applicable across a range of fields, from electrical circuits to fluid dynamics and mechanical systems (e.g., [8–12]).

The paper is structured as follows: Sect. 2 details the problem setting. Section 3 applies the harmonic balance method to our specific case. Section 4 presents validation of the theoretical results through numerical simulations. In Sect. 5, we discuss the implications of our findings and recognize the limitations of our approach. Finally, Sect. 6 summarizes the study and outlines directions for future research.

2 Problem Setting

We investigate the dynamics of a strongly damped particle in a cosine potential under harmonic excitation, as described by Eq. (1). The system is mechanically analogous to a damped pendulum in a gravitational field subject to periodic external momentum. The governing equation of motion is expressed as:

$$m\ddot{x} + k\dot{x} + c\sin x = f\sin(\omega\tau + \beta), \tag{1}$$

where τ [s] denotes the time, x [m] denotes the displacement and

- m [kg] is the mass parameter without further assumptions on its magnitude.
- k [kg/s] is a large damping coefficient that ensures an overdamped regime such that $k = \mathcal{O}(1)$.
- $c\sin x$ [kg $\cdot$ m/s^2] represents the restoring force of the potential.
- f [kg $\cdot$ m/s^2] denotes the amplitude of the external forcing and is also of the order $\mathcal{O}(1)$.

No specific conditions are imposed on the excitation frequency ω [rad/s] and the initial phase β [rad] (Fig. 1).

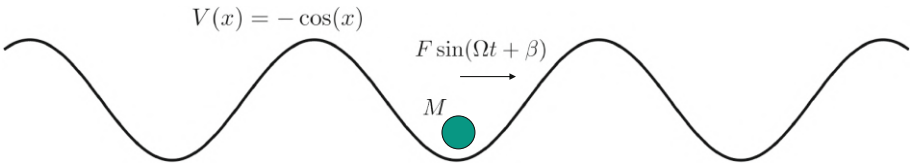
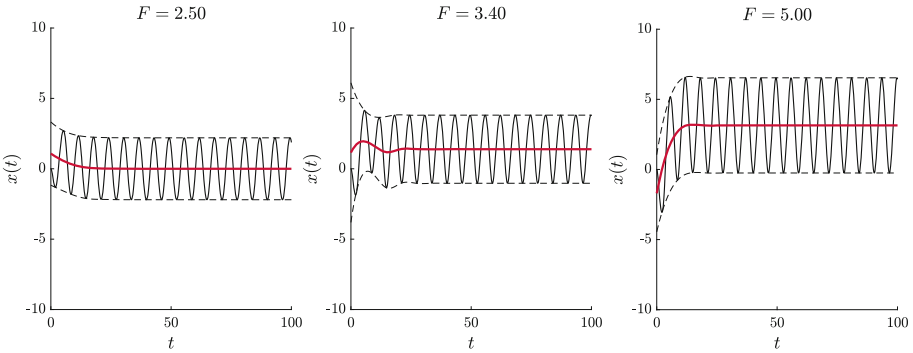


Fig. 1. Problem setting



(a) Bottom (b) Intermediate (c) Top

Fig. 2. Different kinds of solutions (black continuous curves) for parameter values $M = 1$, $\Omega = 1$, $\beta = \pi/2$ with homogeneous initial conditions. The center of the vibration (red curve) can be stable at the bottom, top, or any point of the cosine potential, depending on the chosen value of F. In these figures, the center of vibrations is estimated by the mean value of the upper and lower envelopes (dashed black lines) obtained from the numerical simulation data

We introduce a coordinate transformation $\tau = \omega_0 t$ (prime denotes a derivative with respect to τ) with $\omega_0 := c/k$ and the dimensionless frequency $\Omega := k\omega/c$, as well as the dimensionless mass and force parameters

$$M := \frac{cm}{k^2}, \qquad F := \frac{f}{k\omega_0}.$$

Hence, the differential equation becomes

$$Mx'' + x' + \sin x = F\sin(\Omega t + \beta). \tag{2}$$

Numerical simulations suggest that the steady-state solutions of this system are not limited to being centered at the bottom of the potential, as depicted in Fig. 2a. The solutions can also be centered around the top, given certain values of F, as illustrated in Fig. 2c.

Furthermore, solutions between these two points are also feasible, as highlighted in Fig. 2b. Hence, the symmetry-breaking bias can be observed in stationary solutions, which we will capture analytically in the next section.

3 Analytic Treatment

In the presence of strong damping, numerical simulations of Eq. (2) reveal only periodic solutions without any chaotic behavior or other anomalies such as period doubling. Consequently, for the upcoming analysis, it is assumed that after a transient phase, the solution stabilizes to a steady state, aligning its base frequency with that of the excitation. The subsequent section will identify these steady-state solutions and assess their stability. Therefore, the term "steady-state solutions" will be used to describe these types of long-term, periodic solutions.

3.1 Harmonic Balance

As only periodic solutions are observed at the steady state, the harmonic balance method is suitable for capturing the long-term dynamics of the system. To be able to address asymmetric solutions as well, we include a constant (bias) term besides the first harmonic term

$$x_0(t) = A_0 + A_1 \sin(\Omega t + \beta - \Psi).$$

Plugging this into the differential equation, given by Eq. (2), we get:

$$\begin{aligned} M(-\Omega^2 A_1 \sin(\Omega t + \beta - \Psi)) + A_1 \Omega \cos(\Omega t + \beta - \Psi) \\ + \sin(A_0 + A_1 \sin(\Omega t + \beta - \Psi)) = F \sin(\Omega t + \beta). \end{aligned} \tag{3}$$

The last term on the left-hand side of Eq. (3) needs further treatment since a sine function is nested in itself. The Jacobi-Anger expansion [13], as given by Eqs. (4–5), is suitable for solving this issue.

$$\sin(z \sin(\theta)) = 2 \sum_{n=1}^{\infty} J_{2n-1}(z) \sin\left[(2n-1)\theta\right], \tag{4}$$

$$\cos(z \sin(\theta)) = J_0(z) + 2 \sum_{n=1}^{\infty} J_{2n}(z) \cos(2n\theta), \tag{5}$$

where $J_n(z)$ is the Bessel function of the first kind of order n.

Applying Eqs. (4–5) to Eq. (3) and balancing the terms up to the first harmonic order in Ωt, we can derive three nonlinear algebraic equations by grouping the constant, the $\sin(\Omega t + \beta)$, and the $\cos(\Omega t + \beta)$ terms:

$$J_0(A_1) \sin A_0 = 0, \tag{6}$$

$$-A_1 \Omega^2 M \cos\Psi + A_1 \Omega \sin\Psi + 2 \cos A_0 J_1(A_1) \cos\Psi = F, \tag{7}$$

$$A_1 \Omega^2 M \sin\Psi + A_1 \Omega \cos\Psi - 2 \cos A_0 J_1(A_1) \sin\Psi = 0. \tag{8}$$

Equation (6) has three different kinds of solutions, corresponding to solution families (SFs): A_0 is either located at the bottom (bottom SF) or the top (top

SF) of the cosine potential; or the value of A_1 is a root of the Bessel function J_0 (intermediate SF), i.e.,

$$A_{0,1} = 2k\pi, \qquad k \in \mathbf{Z}, \tag{9}$$

$$A_{0,2} = (2k+1)\pi, \qquad k \in \mathbf{Z}, \tag{10}$$

$$A_{1,3} = J_0^{-1}(0), \qquad A_{1,3} > 0, \tag{11}$$

where we define $J_0^{-1}(0)$ as a set-valued expression to denote all positive roots of the Bessel function of the first kind of zeroth order. For the solutions given by Eqs. (9-10), the values of A_1 and Ψ still have to be determined from Eqs. (7–8), whereas in case of Eq. (11) the value of A_0 and Ψ must be determined using the remaining equations (7–8).

For a graphical representation of the SFs, see Fig. 4b. The following three SFs are given.

Bottom SF. Inserting Eq. (9) in Eqs. (7–8) and applying some algebra, we can separate Ψ and A_1 in two distinct equations

$$\sin\Psi_1 = \frac{A_{1,1}\Omega}{F},$$

$$\left(A_{1,1}\Omega^2 M - 2J_1(A_{1,1})\right)^2 + A_{1,1}^2\Omega^2 = F^2. \tag{12}$$

Equation (12) cannot be solved explicitly for $A_{1,1}$; however, an asymptotic solution can be given for large values of F when the force exerted by the cosine potential becomes negligible compared to the damping and inertial forces

$$A_{1,1} \approx \frac{F}{\Omega}\frac{1}{\sqrt{\Omega^2 M^2 + 1}}, \quad \text{for } F \gg 1.$$

Top SF. Similarly, by inserting Eq. (10) in Eqs. (7–8) and performing the same algebraic steps, we can separate Ψ and A_1:

$$\sin\Psi_2 = \frac{A_{1,2}\Omega}{F},$$

$$\left(A_{1,2}\Omega^2 M + 2J_1(A_{1,2})\right)^2 + A_{1,2}^2\Omega^2 = F^2, \tag{13}$$

thus, the equations for the top SF and the bottom SF only differ in the sign of $2J_1(A_{1,2})$.

Intermediate SF. The derivation of this family of solutions differs from the previous ones since the possible values of A_1 are already known; instead, A_0 is sought. Let us denote the k^{th} root of $J_0(A_1)$ by α_k. Using this notation, we can rewrite Eqs. (7–8) as

$$\left(2\cos A_0 J_1(\alpha_k) - \alpha_k\Omega^2 M\right)\cos\Psi + \alpha_k\Omega\sin\Psi = F,$$

$$-\left(2\cos A_0 J_1(\alpha_k) - \alpha_k\Omega^2 M\right)\sin\Psi + \alpha_k\Omega\cos\Psi = 0,$$

from which follows that

$$\sin \Psi_3 = \frac{\alpha_k \Omega}{F},$$
$$\left(2 \cos A_0 J_1(\alpha_k) - \alpha_k \Omega^2 M\right)^2 + \alpha_k^2 \Omega^2 = F^2. \tag{14}$$

Equation (14) can be resolved for the bias term as

$$A_0 = 2\pi l \pm \arccos\left(\frac{\alpha_k \Omega^2 M \pm \sqrt{F^2 - \alpha_k^2 \Omega^2}}{2J_1(\alpha_k)}\right), \quad l \in \mathbf{Z}. \tag{15}$$

Unlike the bottom and top SFs, the intermediate SF does not always exist. The solution exists if the arccos function has real arguments on $[-1, 1]$. To secure a real argument,

$$\alpha_k \Omega < F \tag{16}$$

has to be fulfilled. Furthermore, the arccos function will take real values if the inequalities

$$-1 \leq \frac{\alpha_k \Omega^2 M \pm \sqrt{F^2 - \alpha_k^2 \Omega^2}}{2J_1(\alpha_k)} \leq 1$$

are fulfilled. If $J_1(\alpha_k) > 0$ then we have

$$-2J_1(\alpha_k) - \alpha_k \Omega^2 M \leq \pm\sqrt{F^2 - \alpha_k^2 \Omega^2} \leq 2J_1(\alpha_k) - \alpha_k \Omega^2 M, \tag{17}$$

otherwise, we have

$$-2J_1(\alpha_k) - \alpha_k \Omega^2 M \geq \pm\sqrt{F^2 - \alpha_k^2 \Omega^2} \geq 2J_1(\alpha_k) - \alpha_k \Omega^2 M. \tag{18}$$

Whether the term with F should be taken with a plus or minus sign can be determined by case distinction. We only investigate the case $J_1(\alpha_k) > 0$, given by Eq. (17), since the procedure and results are analogous for Eq. (18).

The left term of Eq. (17) is always negative, and its right term is negative for

$$\frac{2J_1(\alpha_k)}{\alpha_k \Omega^2} < M,$$

in which case only $-\sqrt{F^2 - \alpha_k^2 \Omega^2}$ is feasible. Then, we have

$$\sqrt{\left(\alpha_k \Omega^2 M - 2J_1(\alpha_k)\right)^2 + \alpha_k^2 \Omega^2} \leq F \leq \sqrt{\left(\alpha_k \Omega^2 M + 2J_1(\alpha_k)\right)^2 + \alpha_k^2 \Omega^2},$$

that always fulfills inequality (16) as well. However, if

$$\frac{2J_1(\alpha_k)}{\alpha_k \Omega^2} > M, \tag{19}$$

then case $-\sqrt{F^2-\alpha_k^2\Omega^2}$ always fulfills the right hand side of inequality (17), but also $+\sqrt{F^2-\alpha_k^2\Omega^2}$ does so for

$$\alpha_k\Omega \le F \le \sqrt{\left(2J_1(\alpha_k)-\alpha_k\Omega^2M\right)^2+\alpha_k^2\Omega^2},$$

thus, in this case, the two kinds of solutions

$$A_0 = 2\pi l \pm \arccos\left(\frac{\alpha_k\Omega^2M \pm \sqrt{F^2-\alpha_k^2\Omega^2}}{2J_1(\alpha_k)}\right), \quad l \in \mathbf{Z},$$

exist. For

$$\sqrt{\left(\alpha_k\Omega^2M-2J_1(\alpha_k)\right)^2+\alpha_k^2\Omega^2} \le F \le \sqrt{\left(\alpha_k\Omega^2M+2J_1(\alpha_k)\right)^2+\alpha_k^2\Omega^2},$$

only

$$A_0 = 2\pi l \pm \arccos\left(\frac{\alpha_k\Omega^2M - \sqrt{F^2-\alpha_k^2\Omega^2}}{2J_1(\alpha_k)}\right), \quad l \in \mathbf{Z}, \tag{20}$$

exists.

The existence of two distinct kinds of solution within the intermediate SF in the parameter region given by Eq. (19) is an artifact of our chosen ansatz since we only included a constant term and the first harmonic to estimate the periodic steady-state solution of Eq. (2). Thus, the accuracy of the ansatz in the parameter region referred to above is insufficient, yielding a physically non-existent solution. Further (possibly infinitely many) terms in the harmonic balance ansatz would let this artifact disappear. In none of the numerical simulations, such double solutions could be observed (cf. Fig. 3a and 4a). In fact, by increasing the values of the amplitude of the force, the center of vibration A_0 shifts continuously and smoothly from the bottom of the potential to the top and vice versa.

As mentioned previously, in the case of $J_1(\alpha_k) < 0$, the calculation is very similar, and the artificial double solution parameter region emerges similarly.

3.2 Stability Analysis of the Steady-State Solutions

The periodic steady-state solutions for Eq. (2) have been identified. However, their stability still needs to be ascertained. We introduce a small perturbation δ in the steady-state solution x_0 as follows:

$$x(t) = x_0(t) + \delta(t),$$

and insert it in Eq. (2) to find

$$Mx_0'' + M\delta'' + x_0' + \delta' + \sin(x_0+\delta) = F\sin(\Omega t+\beta), \tag{21}$$

after linearizing the sine term around x_0, we have

$$M x_0'' + M\delta'' + x_0' + \delta' + \sin(x_0) + \cos(x_0)\delta = F\sin(\Omega t + \beta).$$

Given that x_0 satisfies the differential equation (2) within the first harmonic approximation, we can simplify Eq. (21) by letting the terms cancel each other out. We obtain the following.

$$M\delta'' + \delta' + \cos(x_0)\delta = 0. \tag{22}$$

Eq. (22) is a linear differential equation with parametric excitation because x_0 is a function of time. One way to obtain information on the stability of the solutions is to calculate the Hill's determinant. For this, we stretch the time coordinate by reusing the variable $\tau := \Omega t/2$, which yields

$$\frac{M\Omega^2}{4}\ddot{\delta} + \frac{\Omega}{2}\dot{\delta} + \cos\left(A_0 + A_1\sin\left(2\tau - \Psi\right)\right)\delta = 0, \tag{23}$$

where $\dot{\square} := \frac{\mathrm{d}\square}{\mathrm{d}\tau}$. Knowing that at the boundary of stable and unstable parameter regions, the solution is periodic with period π or 2π [14], we can write the solution as a Fourier series. In case the solution is π periodic, we have

$$\delta(t) = \sum_{n=0}^{\infty} a_n\cos(2n\tau) + \sum_{n=1}^{\infty} b_n\sin(2nt), \tag{24}$$

with unknown coefficients a_n and b_n. The time-dependent parameter of δ can be rewritten as

$$\cos\left(A_0 + A_1\sin\left(2\tau - \Psi\right)\right) = \sum_{n=0}^{\infty} c_n\cos(2n\tau) + \sum_{n=1}^{\infty} d_n\sin(2nt),$$

where the coefficients c_n and d_n are obtained by

$$c_0 = \frac{1}{\pi}\int_0^{\pi}\cos\left(A_0 + A_1\sin\left(2\tau - \Psi\right)\right)\mathrm{d}\tau, \tag{25}$$

$$c_n = \frac{2}{\pi}\int_0^{\pi}\cos\left(A_0 + A_1\sin\left(2\tau - \Psi\right)\right)\cos\left(2n\tau\right)\mathrm{d}\tau \quad \text{for } n \in \mathbb{N}^+, \tag{26}$$

$$d_n = \frac{2}{\pi}\int_0^{\pi}\cos\left(A_0 + A_1\sin\left(2\tau - \Psi\right)\right)\sin\left(2n\tau\right)\mathrm{d}\tau \quad \text{for } n \in \mathbb{N}^+. \tag{27}$$

To calculate Eq. (25–27), it is worth rewriting the integrand using the Jacobi-Anger expansion.

$$
\begin{aligned}
&\cos(A_0 + A_1 \sin(2\tau - \Psi)) = \cos A_0 \cos(A_1 \sin(2\tau - \Psi)) - \sin A_0 \sin(A_1 \sin(2\tau - \Psi)) \\
&\quad = \cos(A_0) \left(J_0(A_1) + 2 \sum_{n=1}^{\infty} J_{2n}(A_1) \cos(2n(2\tau - \Psi)) \right) \\
&\qquad - \sin(A_0) \left(2 \sum_{n=1}^{\infty} J_{2n-1}(A_1) \sin((2n-1)(2\tau - \Psi)) \right) \\
&\quad = \cos(A_0) \left(J_0(A_1) + 2 \sum_{n=1}^{\infty} J_{2n}(A_1) (\cos(4n\tau)\cos(2n\Psi) + \sin(4n\tau)\sin(2n\Psi)) \right) \\
&\qquad - \sin(A_0) \left(2 \sum_{n=1}^{\infty} J_{2n-1}(A_1) (\sin(2(2n-1)\tau)\cos((2n-1)\Psi) - \cos(2(2n-1)\tau)\sin((2n-1)\Psi)) \right).
\end{aligned}
$$

Since the orthogonality of the sines and cosines, the Fourier coefficients result in

$$
\begin{aligned}
c_0 &= \cos(A_0) J_0(A_1), \\
c_n &= \begin{cases} 2\sin(A_0) J_n(A_1) \sin(n\Psi) & \text{for } n \in \mathbb{N}^+ \text{ odd}, \\ 2\cos(A_0) J_n(A_1) \cos(n\Psi) & \text{for } n \in \mathbb{N}^+ \text{ even}, \end{cases} \\
d_n &= \begin{cases} -2\sin(A_0) J_n(A_1) \cos(n\Psi) & \text{for } n \in \mathbb{N}^+ \text{ odd}, \\ 2\cos(A_0) J_n(A_1) \sin(n\Psi) & \text{for } n \in \mathbb{N}^+ \text{ even}. \end{cases}
\end{aligned}
$$

Hence, insertion of Eq. (24) into Eq. (23) and comparing the coefficients for the $\cos(2n\tau)$ and $\sin(2n\tau)$ terms yields the following (infinite) linear equation system for a_n and b_n.

$$
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix} = \begin{bmatrix}
c_0 & \frac{c_1}{2} & \frac{d_1}{2} & \frac{c_2}{2} & \frac{d_2}{2} & \cdots \\
c_1 & c_0 + \frac{c_2}{2} - M\Omega^2 & \frac{d_2}{2} + \Omega & \frac{c_1}{2} + \frac{c_3}{2} & \frac{d_1}{2} + \frac{d_3}{2} & \cdots \\
d_1 & \frac{d_2}{2} - \Omega & c_0 - \frac{c_2}{2} - M\Omega^2 & -\frac{d_1}{2} + \frac{d_3}{2} & \frac{c_1}{2} - \frac{c_3}{2} & \cdots \\
c_2 & \frac{c_1}{2} + \frac{c_3}{2} & -\frac{d_1}{2} + \frac{d_3}{2} & c_0 + \frac{c_4}{2} - 4M\Omega^2 & \frac{d_4}{2} + 2\Omega & \cdots \\
d_2 & \frac{d_1}{2} - \frac{d_3}{2} & \frac{c_1}{2} - \frac{c_3}{2} & \frac{d_4}{2} - 2\Omega & c_0 - \frac{c_4}{2} - 4M\Omega^2 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ b_1 \\ a_2 \\ b_2 \\ \vdots \end{bmatrix}. \tag{28}
$$

In order to have a nontrivial solution, the determinant of the Hill matrix in Eq. (28) must disappear. Since calculating the infinite matrix's determinant is not feasible, we will approximate it by the determinant of its truncation, where we take only the first $(2n + 1)$ rows and columns. Indeed, for $n = 0$, we obtain only one value c_0, and thus the estimate for the boundary of the stable-unstable region is given by

$$
c_0 = \cos(A_0) J_0(A_1) = 0. \tag{29}
$$

Although it is possible to obtain the determinants for $n > 0$, we refrain from determining them due to the complexity of the calculations. A plausible argument that shows why the truncation at $n = 0$ should not result in large errors is

as follows. First, for nonnegative real arguments, the Bessel function of the first kind of any order is bounded by -1 and 1 [15]. Second, the values of $J_n(A_1)$ are close to zero for large n and moderate A_1. Third, the diagonal elements of the Hill matrix increase unboundedly as the term $-n^2M\Omega^2$ becomes increasingly significant. Consequently, the dominant term of the determinant is invariably the product of the diagonal elements. This product can be written as

$$D = c_0 \prod_{n=1}^{\infty} \left(c_0 + \frac{c_{2n}}{2} - n^2 M\Omega^2\right)\left(c_0 - \frac{c_{2n}}{2} - n^2 M\Omega^2\right).$$

Since the coefficients c_n are bounded, no other term in the product can vanish apart of c_0, which is equivalent to the stability boundary corresponding to the truncation of the Hill matrix with $n = 0$. In fact, when non-small frequency values are within the range of interest, numerical simulations indicate that the determinant, with increasing n, predicts a stability boundary that remains close to that suggested by $c_0 = 0$. For smaller values of Ω, the situation becomes more complex, as the matrix's diagonal terms do not increase as rapidly, allowing for other stability boundaries in addition to the one near that predicted by $c_0 = 0$. However, the investigation of this case is beyond the scope of this paper.

The reader can easily convince himself that the stability boundary corresponding to periodic solutions of 2π results in the same condition when truncated with $n = 0$. Interestingly, the same result as in Eq. (29) follows from the naive approach when one tries to determine the average effect of the time-dependent parameter in Eq. (23), that is,

$$\langle \cos(A_0 + A_1 \sin(2\tau - \Psi)) \rangle = \cos A_0 J_0(A_1),$$

where $\langle f(t) \rangle$ stands for the average of a periodic function $f(t)$ with period T over one period, i.e., $\langle f(t) \rangle = 1/T \int_0^T f(t)\mathrm{d}t$. Introducing the averaged disturbance $\tilde{\delta} = \langle \delta \rangle$, the average of Eq. (23) becomes

$$\frac{M\Omega^2}{4}\ddot{\tilde{\delta}} + \frac{\Omega}{2}\dot{\tilde{\delta}} + \cos A_0 J_0(A_1)\tilde{\delta} = 0. \tag{30}$$

Taking the values of A_0 and A_1 corresponding to the different SFs, we can use a simple linear stability analysis to determine the stability of the steady-state solutions found in Secs. 3.1–3.1. For asymptotic stability, $\cos A_0 J_0(A_1) > 0$ must be fulfilled.

Stability of the Bottom SF When the center of vibration, A_0, is at the bottom of the potential well, we have $\cos A_0 = 1$. Thus, the corresponding solution is stable if $J_0(A_1)$ has a positive sign. Since the change of A_1 is continuous in F, two consecutive roots of $J_0(A_1) = 0$ define the stability boundaries. Inserted in Eq. (12), these roots provide the critical forcing values where the stability of the steady-state solution changes:

$$F_{\mathrm{crit},1,k} = \sqrt{\left(2J_1(\alpha_k) - \alpha_k \Omega^2 M\right)^2 + \alpha_k^2 \Omega^2}, \tag{31}$$

with α_k being the k^{th} root of $J_0(A_1)$. Hence, the steady-state solution is stable for the following conditions.

$$\sqrt{\left(2J_1(\alpha_{2k}) - \alpha_{2k}\Omega^2 M\right)^2 + \alpha_{2k}^2\Omega^2} < F < \sqrt{\left(2J_1(\alpha_{2k+1}) - \alpha_{2k+1}\Omega^2 M\right)^2 + \alpha_{2k+1}^2\Omega^2}, \tag{32}$$

with $k \in \mathbf{N}$ and $\alpha_0 := 0$.

Stability of the Top SF. When the center of vibrations is at the top of the potential, the situation is similar to the previous one with the difference that now $\cos A_0 = -1$; thus, the solution at the top of the potential becomes stable for negative values of $J_0(A_1)$. Since A_1 is the solution of Eq. (13), the critical forcing values, where the stability of the steady-state solution changes, are determined by

$$F_{\text{crit},2,k} = \sqrt{\left(2J_1(\alpha_k) + \alpha_k\Omega^2 M\right)^2 + \alpha_k^2\Omega^2}, \tag{33}$$

with α_k being the k^{th} root of $J_0(A_1)$. Hence, a steady-state solution is stable for

$$\sqrt{\left(2J_1(\alpha_{2k+1}) + \alpha_{2k+1}\Omega^2 M\right)^2 + \alpha_{2k+1}^2\Omega^2} < F < \sqrt{\left(2J_1(\alpha_{2k+2}) + \alpha_{2k+2}\Omega^2 M\right)^2 + \alpha_{2k+2}^2\Omega^2}, \tag{34}$$

with $k \in \mathbf{N}$.

The critical force values in Eq. (31) and (33) are not identical due to the difference in sign under the square roots. Thus, the stable regions of the top and bottom SFs are disjoint and do not cover all possible values of F (cf. purple regions in Fig. 4b). There are parameter regions of F, where neither the solution with a vibration center at the bottom of the potential nor at the top of the potential is stable. Since two unstable equilibria cannot exist next to each other, an intermediate stable solution must be located between the top and bottom solutions.

Stability of the Intermediate SF. The linear stability analysis does not give any statement regarding the stability of the intermediate SF because the values of A_1 are determined by the roots of $J_0(A_1)$. Consequently, the term $\cos A_0 J_0(A_1)$ invariably equals zero, leading to an eigenvalue in Eq. (30) that has a real part of zero. Due to the problem's linearization, it is impossible to determine stability from this conclusively. However, it is noteworthy that both topological considerations and results from thousands of numerical simulations suggest that the intermediate SF given by Eq. (20) is stable for

$$\sqrt{\left(2J_1(\alpha_{2k+1}) - \alpha_{2k+1}\Omega^2 M\right)^2 + \alpha_{2k+1}^2\Omega^2} < F < \sqrt{\left(2J_1(\alpha_{2k+1}) + \alpha_{2k+1}\Omega^2 M\right)^2 + \alpha_{2k+1}^2\Omega^2} \tag{35}$$

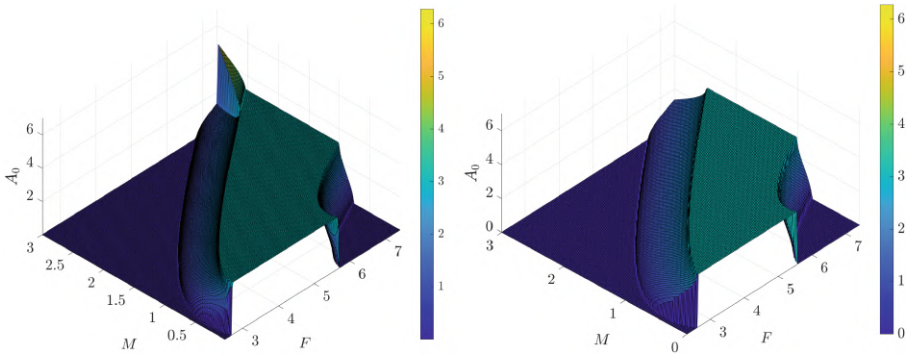
and

$$\sqrt{(2J_1(\alpha_{2k+2}) + \alpha_{2k+2}\Omega^2 M)^2 + \alpha_{2k+2}^2\Omega^2} < F < \sqrt{(2J_1(\alpha_{2k+2}) - \alpha_{2k+2}\Omega^2 M)^2 + \alpha_{2k+2}^2\Omega^2}, \tag{36}$$

with $k \in \mathbf{N}$ and $\alpha_0 := 0$. Note that the sign of $J_1(\alpha_k)$ alternates for consecutive values of k, making Eqs. (35–36) consistent with Eqs. (32) and (34).

4 Validation

In Fig. 3a, we display the numerically derived steady-state center of vibration as a function of parameters F and M. A more detailed view of the parameter region $F = 2.65 \ldots 2.71$ and $M = 0.001 \ldots 0.05$ is provided in Fig. 4a. The analytical estimate for the center of steady-state oscillations is shown in Fig. 3b. There is a clear qualitative agreement between the analytic and numerical results. However, some discrepancies between them can be observed; for example, the first predicted value, where the stability of the steady-state solution changes when $M = 0$, is numerically obtained at $F_{\mathrm{crit,num}}(M = 0) = 2.677$. At the same time, the analytic prediction is $F_{\mathrm{crit,anal}}(M = 0) = 2.619$. Similarly, the transition region from $A_0 = 0$ to $A_0 = \pi$ is offset by some amount for $M \neq 0$.



(a) Numerically obtained center of vibration A_0. Bifurcation in the upper corner due to Eq. (20) prescribing a family of solutions. The actual solution depends on the excitation phase and initial conditions

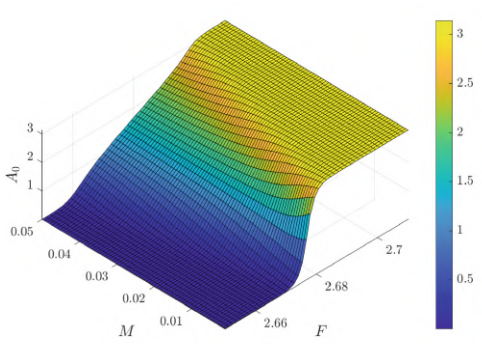
(b) Analytically obtained center of vibration A_0. The function in the bottom corner is not displayed correctly due to the artificial double-solution of the harmonic balance ansatz. Simulations show the continuous transition of A_0 from 0 to π

Fig. 3. Position of the center of steady state vibrations A_0 against the values of F and M with $\Omega = 1$, $\beta = \pi/2$ and homogeneous initial conditions, i.e., $x(0) = 0$ and $\dot{x}(0) = 0$. The shade of color represents the value of A_0. The jump in the upper corner of Fig. 3a occurs due to the change in the location of the center of the steady-state solution within the same SF. This is possible since Eqs. (9), (10) and (15) only determine the stable SFs, but not the exact member of these families, where the steady-state will finally settle

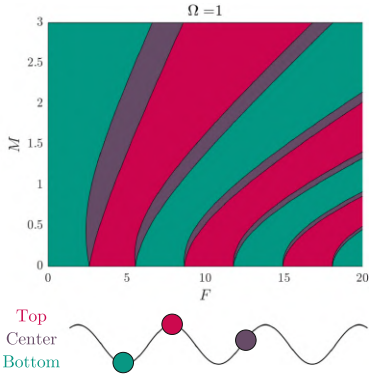
Enhancing the harmonic balance approach by incorporating even a single additional term, like:

$$x_0(t) = A_0 + A_1 \sin(\Omega t - \Psi_1) + A_2 \sin(2\Omega t - \Psi_2),$$

significantly complicates the problem to a point where analytic solutions become nearly unmanageable.



(a) High-resolution view of the bottom corner of Fig. 3a. The color shade represents the value of A_0



(b) Stable steady-state solution families

Fig. 4. Comparison of numerical and analytic results against parameters F and M. The left subfigure shows the numerical results for small mass values, while the right presents the families of stable steady-state solutions obtained analytically: bottom (green), top (red) and intermediate (purple). In the left figure, for $M = 0$, the function converges to a step function at $F_{\text{crit}} = 2.677$ that changes from $A_0 = 0$ to $A_0 = \pi$

Nevertheless, numerically, we can observe that the intermediate solution breaks the symmetry not only by its biased vibration center but also by its non-vanishing even Fourier terms. To demonstrate this, in Fig. 5, two examples are given, where the Fourier coefficients of the periodic steady-state solution are determined by the numerical evaluation of Eq. (37)

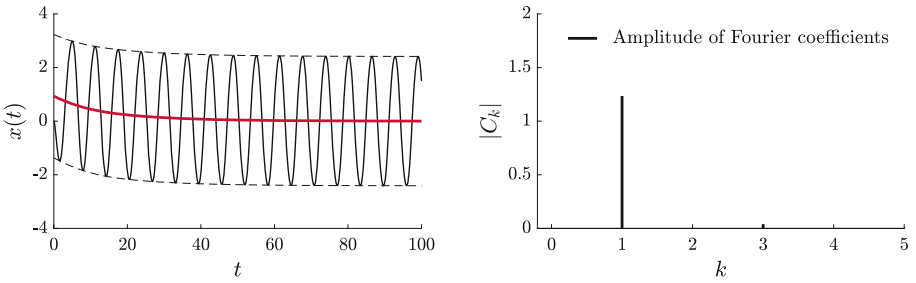
$$C_k = \frac{1}{T} \int_{t_\infty}^{t_\infty + T} x(t) e^{-i\Omega k t} \mathrm{d}t, \tag{37}$$

where $T = 2\pi/\Omega$; and t_∞ denotes a sufficiently late time instance after which the transient motion might be neglected. With these coefficients, the steady-state solution can be represented as

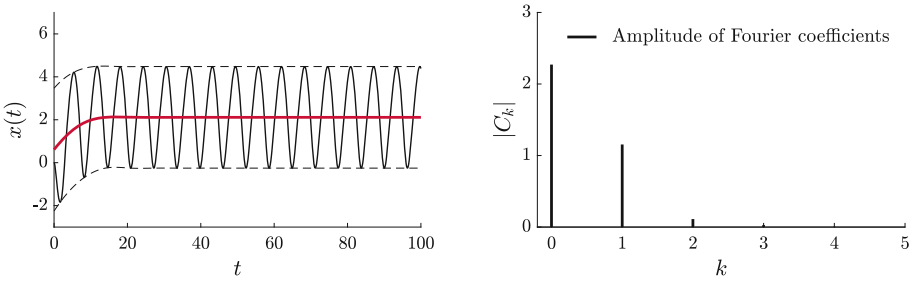
$$x(t) = \sum_{k=-\infty}^{\infty} C_k e^{i\Omega k t}.$$

As the value of F increases, the solution at the bottom of the potential loses stability by Eq. (32). However, by Eq. (34), solutions centered around the top

of the potential do not gain stability immediately. A solution with a biased center achieves stability when the underlying potential is symmetric, which occurs exclusively at multiples of π. Consequently, a symmetric solution is expected solely around these values, which implies a solution that exhibits identical shapes on both sides of its center, thereby incorporating only terms with odd frequencies in its Fourier series (cf. Fig. 5a). Solutions with a biased center, lacking symmetry in the restoring force, are not expected to be symmetric; that is, the shape of the solution differs around the center. Therefore, the solution's Fourier series will include not only odd but also even frequencies (cf. Fig. 5b).



(a) Symmetric steady-state solution with $F = 2.5$



(b) Asymmetric steady-state solution with $F = 3$

Fig. 5. Solution $x(t)$ decaying to steady state (on the left) and the amplitudes of its spectral decomposition $|C_k|$ (on the right) with parameter values $M = 0.5$, $F = \{2.5, 3\}$, $\Omega = 1$ and $\beta = -\pi/2$ with ICs $x(0) = 0$ and $\dot{x}(0) = 0$. The red line denotes the center of vibration, while the dashed line is the slowly evolving vibration envelope

5 Discussion

This paper investigated steady-state solutions of an overdamped particle in a cosine potential under harmonic excitation. Depending on the particle's mass, the exciting force amplitude, and the excitation frequency, the center of the steady-state oscillation might differ. One can observe that by increasing the value of F, at a certain level, the center of vibration starts to shift continuously from

the bottom position $2k\pi$ to the neighboring top position $(2k \pm 1)\pi$, where after reaching a certain critical force level, the center of oscillation settles. After the excitation force is further increased, a continuous shift of the center of oscillation starts back to the bottom position. This process repeats itself infinitely, although the transition from the bottom to the top of the potential always happens faster by increasing the values of F. The observed continuous shift is an effect of the particle's mass; in the case of a massless particle, the continuous shift cannot be observed; it is reduced to a discontinuous jump in the value of the center of steady vibrations.

Furthermore, the initial conditions and the starting phase of the excitation have a minor impact on the steady-state outcome of the system. They determine the location of the vibration center but do not change the stable solution family itself, as shown in Fig. 3a.

At first glance, it seems that the particle's mass is responsible for the surprising asymmetric dynamics of the system. However, we should realize that the system dynamics is naturally asymmetrical, and neglecting the mass results in degenerate symmetric dynamics, as revealed by Eqs. (7–8). A rather wordy narrative of what has been derived analytically in the paper is as follows.

For a given excitation, the resulting steady-state vibration amplitude is primarily determined by the damping and, eventually, by the particle's mass. However, the underlying force of the potential also contributes to the steady-state amplitude value. This contribution is based on the average force acting on the particle during one period of its supposed steady-state movement. Depending on the magnitude of the exciting force, the potential's average force can either increase or decrease the amplitude of the steady-state solution compared to the case of a damped free particle (without potential). In most cases, this will either result in a bottom- or top-stabilized vibration center. However, for the same excitation, the steady-state vibration amplitude at the bottom always differs from the top. Thus, when the steady-state solution loses its stability around one of the positions, the other one still has not gained stability. Therefore, any stable periodic solution in such a case must be centered at an intermediate position between the bottom and the top of the potential.

In contrast, when a massless particle is considered, the system's structure degenerates: it becomes a first-order system, and as such, it is no "classical" mechanical system anymore. Thus, our intuition, accustomed to second-order mechanical systems, is deceived. In the model reduction with no mass, may it only be an artifact, the steady-state vibration amplitude at the top and at the bottom are equal. Thus, by the loss of stability at one of them, stability is immediately gained at the other. To investigate the transition $M \to 0$, applying the singular perturbation method could be a promising technique. However, that analysis is beyond the scope of the current work.

6 Conclusions and Scope for Future Research

The existence of a stable biased steady-state solution in a seemingly symmetric problem emphasizes the critical role of an ansatz in research methodology.

An ansatz, essentially a well-informed assumption, is key to simplifying complex problems, allowing researchers to focus on crucial variables and underlying relationships. Its correct selection can lead to significant breakthroughs by revealing hidden patterns in complex scenarios.

However, choosing an appropriate ansatz is challenging, particularly in nonlinear systems with hidden asymmetries. An incorrect ansatz can mislead the analysis, as demonstrated in our study. Specifically, selecting an ansatz that appears redundant but permits symmetry-breaking in symmetric problems can lead to an accurate description of the underlying physics. Notable examples include the Kármán vortex street, characterized by asymmetric flow around a perfectly cylindrical object and turbulent flow in a cylindrical pipe [16].

The existence of stable steady-state solutions with an intermediate center of vibration has practical implications as well. The overdamped particle M might be a molecule actuated by an electromagnetic or acoustic field. Its position might be controlled only by changing the value of the excitation amplitude, and there is no need to control the excitation phase.

These results have potential applications in ultrasound-driven particle systems and in the design of microelectromechanical systems.

A pressing question from this study is about the nature of potentials that allow the vibration center to shift continuously just by increasing the excitation force's amplitude. Identifying the conditions for such potentials remains a valuable area for further research.

Acknowledgment. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project number: 508244284. This support is greatly appreciated.

References

1. Melde, K., Choi, E., Wu, Z., Palagi, S., Qiu, T., Fischer, P.: Acoustic fabrication via the assembly and fusion of particles. Adv. Mater. **30**, 1704507 (2018)
2. Ma, Z., et al.: Acoustic holographic cell patterning in a biocompatible hydrogel. Adv. Mater. **32**, 1904181 (2020)
3. Hamel, G.: über erzwungene schwingungen bei endlichen amplituden. Mathematische Annalen (1922). Accessed 18 May 2000
4. Hubbard, J.: The forced damped pendulum: chaos, complication and control. Amer. Math. Monthly **106**(8), 741–758 (1999)
5. Coullet, P., Gilli, J.M., Monticelli, M., Vanderberghe, N.: A damped pendulum forced with a constant torque. Am. J. Phys. **73** (2005)
6. Boscaggin, A., Ortega, R., Zanolin, F.: Subharmonic solutions of the forced pendulum equation: a symplectic approach. Arch. Math. **102**(5), 459–468 (2014). https://doi.org/10.1007/s00013-014-0644-2
7. Salas, A., AbuHammad, M., Alotaibi, B.M., El-Sherif, L., El-Tantawy, S.: Closed-form solutions to a forced damped rotational pendulum oscillator. Mathematics **10**(21), 4000 (2022)
8. Nayfeh, A.H., Mook, D.T.: Nonlinear Oscillations, vol. 1979. Wiley, New York (1979)

9. Gilmore, R.J., Steer, M.B.: Nonlinear circuit analysis using the method of harmonic balance-a review of the art. Part I. Introductory concepts. Int. J. Microw. Millimeter-Wave Comput. Aided Eng. **1**(1), 22–37 (1991)
10. Hall, K.C., Thomas, J.P., Clark, W.S.: Computation of unsteady nonlinear flows in cascades using a harmonic balance technique. AIAA J. **40**(5), 879–886 (2002)
11. Sarrouy, E., Sinou, J.J.: Non-linear periodic and quasi-periodic vibrations in mechanical systems-on the use of the harmonic balance methods. Intech (2011)
12. Krack, M., Gross, J.: Harmonic Balance for Nonlinear Vibration Problems. Mathematical Engineering, 1st edn. Springer, Cham (2019). Number of Pages: XII, 159. Number of Illustrations: 21 b/w illustrations, 35 illustrations in colour. Topics: Engineering Mathematics, Solid Mechanics, Fourier Analysis, Multibody Systems and Mechanical Vibrations
13. Olver, F.W.J., et al. (eds.): NIST Digital Library of Mathematical Functions: Jacobi-Anger Expansions (2023). https://dlmf.nist.gov/search/search?q=Jacobi%E2%80%93Anger%20expansions
14. Magnus, W., Winkler, S.: Hill's Equation. Dover Publications (2004). Dover Books on Mathematics
15. Abramowitz, M., Stegun, I.: Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover, New York (1964). Ninth dover printing, tenth gpo printing edition
16. Von Karman, T.: Aerodynamics: Selected Topics in the Light of Their Historical Development. Dover Publications Inc. (2004)

Nonlinear Oscillations of a Bistable Composite Shell: Experimental Studies

Jerzy Warminski[1(✉)], Andrzej Mitura[1], Lukasz Kloda[1], Danyil Melnyk[1], Matteo Brunetti[2], and Francesco Romeo[3]

[1] Department of Applied Mechanics, Mechanical Engineering Faculty, Lublin University of Technology, Nadbystrzycka 36, Lublin, Poland
j.warminski@pollub.pl

[2] Polytechnic Department of Engineering and Architecture, University of Udine, via delle Scienze 206, Udine, Italy

[3] Department of Structural and Geotechnical Engineering, University of Roma La Sapienza, via Eudossiana 18, Rome, Italy

Abstract. The dynamics of a cantilever bistable composite shell is investigated in the paper. The shell shape is derived from a region of a cone surface and is composed of eight composite layers. The proposed shell's geometry and the asymmetric configurations of the layers give rise to the structure with bistability. A prototype of the shell is tested experimentally to detect nonlinear dynamic phenomena in the vicinity of each equilibrium point (in-well dynamics) and to study global (cross-well) oscillations. The experimental campaign shows strong softening resonance characteristics around I-shape and almost linear resonance curves for C-shape. Upon increasing the amplitude of excitation reversible snap-through and nonlinear chaotic oscillations are observed.

Keywords: Bistable Shell · Nonlinear Vibrations · Composite · Experiment

1 Introduction and Problem Formulation

In mechanical engineering a system or device that has two or more stable equilibria is known as bistable or multi-stable. Such bistable system can assume two distinct, stable configurations and can switch between these states under the influence of external forces or thermal conditions. In practical use, multi-stability is a valuable feature, namely after a minor disturbance, the system returns to its original state. However, for a snap to another equilibrium, an external impulse or harmonic excitation is needed to pass through unstable equilibrium (a critical deformation). Once the external force is removed, the system can settle into a new equilibrium position. This property provides stability when needed and enables controlled transitions between states [6].

In recent decades, a highly developing area within multi-stable structures are composite structures. When lamina is arranged on a flat surface in the desired

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 75–87, 2025.
https://doi.org/10.1007/978-3-031-72794-8_5

asymmetric configuration of the plies, it may acquire morphing properties under the influence of curing process [4,7]. Experimental research on these laminates in a flat orientation before curing was replicated and then expanded, encompassing increasingly complex shapes subjecting them to comparative tests based on deviation analysis and linear dynamic response in the vicinity of a stable equilibrium [10]. In the review papers on this topic [5,12], the practical significance of such structures was outlined e.g. for a bistable trailing edge box of an airfoil section, energy harvesting or bio-inspired Venus flytrap robot. The latter case presented prestresses caused by the clamping of one of the edges leading to significantly different stable configurations [11,13]. This prompted the community to extend the analysis of a cantilever structure to include the curvilinear geometry of the composite, i.e. not flat mould's shape, before initiating the autoclave process [1,2,9].

In recent years the attention was focused on pseudo-conical structures clamped at one end and then subjected to kinematic excitation. In a recent paper [3], two separate nonlinear frequency response curves corresponding to *in-well* dynamics were shown and discussed. Due to equipment limitations, experimental observation was not feasible for global dynamics. Afterwards, in [9] nonlinear dynamics within two potential wells as well as *cross-well* were shown, although the clamp did not introduce prestress in the structure and stable configurations differed only slightly.

The main goal of this paper is to investigate nonlinear dynamics of a bistable shell with a geometry based on a conical shape and with an asymmetric configuration of the composite layers. The shell prototype is tested experimentally to determine nonlinear resonance characteristics around local "in-well" as well as global "cross-well" oscillations. Differently from the previous studies related to cylindrical shells, in this work a pseudo-conical shell with a dedicated composite layers ply alignement is proposed. It allows to obtain two stable configurations, the so called I- and C-shape, significantly different from each other in their geometry and natural frequency. The proposed shell design maximises a difference in the stored elastic energy between both considered equilibria. The experimental results explain the mechanism leading to a snap-through effect and will serve as a base for a derivation of a reduced model which to the best of authors knowledge is not yet available in the literature.

2 Geometry of Bistable Composite Shell

The considered structure belongs to a class of pseudo-conical shells with the same sign of curvature. The shell is made of composite material AS4-GP-12K_40gsm-300 mm-ThinPreg135EP, sequenced in $[45, -45_2, 45, -45, 45_2, -45]$, with stacking angle measured from the longitudinal direction of the shell. Properties of the single ply are given in Table 1.

Table 1. Physical parameters of carbon-epoxy composite [8].

Parameter (unit)	Value
E_1 (GPa)	129.58
E_2 (GPa)	9.67
G_{12} (GPa)	13.19
ν_{21} (-)	0.32
ν_{12} (-)	0.02
ρ (kg/m^3)	1532 ± 2
t_ℓ (mm)	0.038
t (mm)	0.304

The original design of the shell's geometry, is taken from published paper [2,3], and is defined by

$$w_0(x_1, x_2) = \left[\frac{1}{r_0} + \left(\frac{1}{r_f} - \frac{1}{r_0}\right)\frac{x_1}{L_1}\right]\frac{x_2^2}{2} + \Delta w, \tag{1}$$

where $\Delta w = 0.0005$ m represents technological offset. Pseudoradii $r_0 = 0.114$ m and $r_f = 0.07$ m define the curvatures of the original configuration at $x_1 = 0$ and $x_1 = L_1 = 0.45$ m, respectively. The width of the shell is $L_2 = 0.15$ m and falls in the range $x_2 = \pm 0.075$ m. Due to the inability to trigger a jump in experimental tests, it was necessary to decrease the working length of the cantilever shell. In present study, the shell was clamped at length $L_0 = 0.205$ m from the free edge, with the same width $L_2 = 0.15$ m which gives $L_0/L_2 = 1.36(6)$ instead of $L_1/L_2 = 3$ of the original design. In Fig. 1 a change of original geometry to the new one is presented. The bistability of the proposed new shell is assessed in the same way as in [8]. Based on the reduced analytical model the static stable configuration can be computed as function of normalised r_0 and r_f. Figure 2 displays the new prototype at C and I configurations. This procedure significantly affected the system dynamics, elevating the first natural vibration frequencies from 1.59 Hz to 5.5 Hz for configuration C and from 2.35 Hz to 9.5 Hz for configuration I.

3 Experimental Analysis of Nonlinear Shell Dynamics

3.1 Experimental Setup

Experimental tests are performed on an electrodynamic shaker with large range of excitation displacements. The shaker is supported by a system for excitation amplitude and frequency control. The shell is clamped at the edge of the smaller curvature at r_0^*, as presented in Fig. 2. Dynamic response is measured by a strain gauge placed on the bottom surface of the shell (not gauged). The measurement is made in x_3 direction and is defined as a function of voltage $\epsilon(V)$.

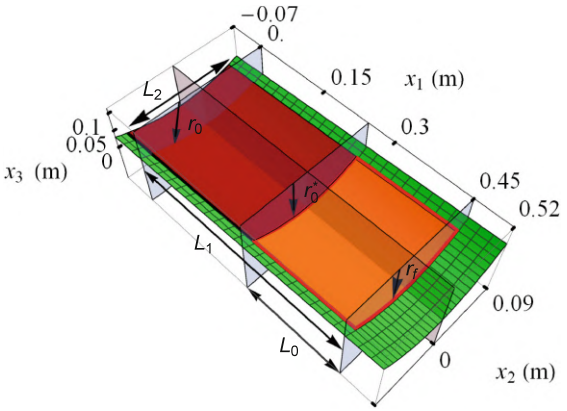


Fig. 1. The top surface of the mold (green checkered), the shell original dimensions (red and orange) and the case of present study (orange).

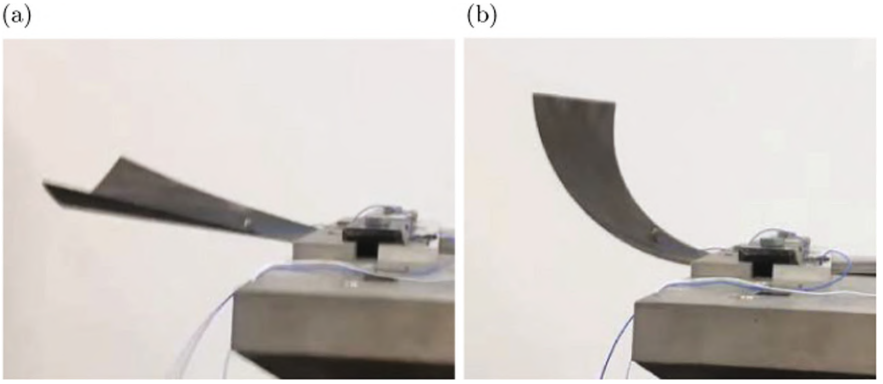


Fig. 2. Clamped prototype of composite shell for (a) I-shape equilibrium and (b) C-shape equilibrium.

The experimental tests are divided into two stages. In the first stage, the resonance characteristics are recorded. For this purpose, the so-called Sine Control excitation is applied. In this variant, the shaker generates sinusoidal signal with constant acceleration amplitude a_s and time-varying frequency $f(t)$. The slow sweep is implemented by linearly varying frequency, increasing for forward and decreasing for backward sweeps. The rate of frequency change is equal 0.01 Hz/s.

To observe various dynamic phenomena the tests are repeated for the initial I-shape and C-shape equilibria. The measured responses are automatically analysed in the TestLab software by the periodic estimator which allows obtaining experimental resonance characteristics. The following resonance curves are marked taking into account the initial configuration of the tested shell: $\mathrm{I}_{initial}$

for initial I-shape or $C_{initial}$ for initial C-shape and a for various directions of the frequency sweep $f_{forward}$ or $f_{backward}$.

In the second stage, for the selected cases of shaker acceleration a_s and frequency of shaker's excitation f long series of the shell's response are recorded. Then, the time series are subjected to additional qualitative dynamics analysis, e.g. Poincaré maps or Fast Fourier Transform (FFT) are computed to identify the nature of oscillations.

3.2 Nonlinear Shell Dynamics

To be close to linear shell dynamics the analysis begins from small excitation level. The resonance curves for relatively small values of amplitude of excitation ($a_s = 0.1$ g and $a_s = 0.25$ g) are presented in Fig. 3(a) and (b), respectively. The full dots denote sweep forward with red colour for initial configuration of C-shape and black colour of I-shape. The circles denote sweep backward with green colour for initial C-shape and blue colour for initial I-shape configuration. This notation is applied for resonance curves plotted on all figures.

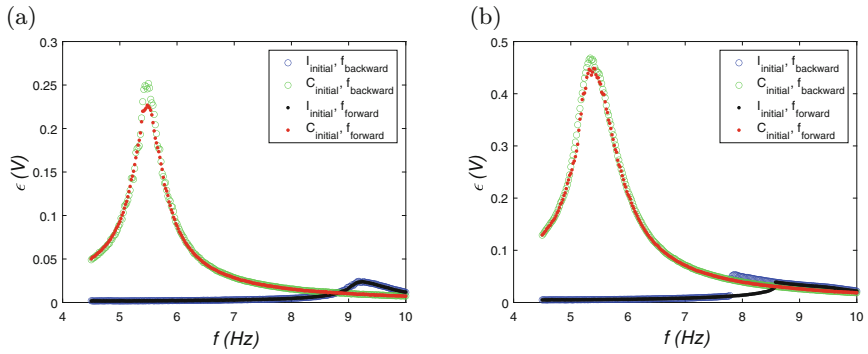


Fig. 3. Resonance curves for sweep forward and sweep backward activated from initial I or C shapes for small amplitude of excitation (a) $a_s = 0.1$ g and (b) $a_s = 0.25$ g.

For very small oscillations the resonance curves presented in Fig. 3(a) are almost linear. Around frequency 5.5 Hz the shell oscillates around C-shape configuration and just slightly change amplitude for sweep forward or backward (red dots and green circles). Around frequency 9.2 Hz oscillations maintain the initially fixed I-shape with amplitude almost the same for forward and backward sweep (black dots and blue circles). However, the nonlinear softening effect is slightly visible for I-shape oscillations. If the excitation amplitude is increased up to $a_s = 0.25$ g strong softening occurs with a wide hysteresis of amplitude starting from about 7.8 Hz up to 8.7 Hz. The shell vibrates locally around I-shape equilibrium but with two stable solution existing in this frequency domain. In contrast, oscillations around C-shape remain linear in spite of the fact that amplitude essentially increased.

A strong nonlinear effect for I-shape equilibrium is clearly visible for moderately large oscillations obtained for $a_s = 0.5$ g, $a_s = 0.6$ g, $a_s = 0.75$ g and $a_s = 0.9$ g as presented in Fig. 4(a), (b), (c) and (d), respectively. Oscillations around C-shape equilibrium continue to have linear nature but the I-shape resonance curve becomes stronger and stronger nonlinear and it is approaching C-shape resonance. For $a_s = 0.9$ g (Fig. 4d) both curves are very close each other with similar amplitudes. However, for all considered cases the shell oscillates in the mode which has been initially set.

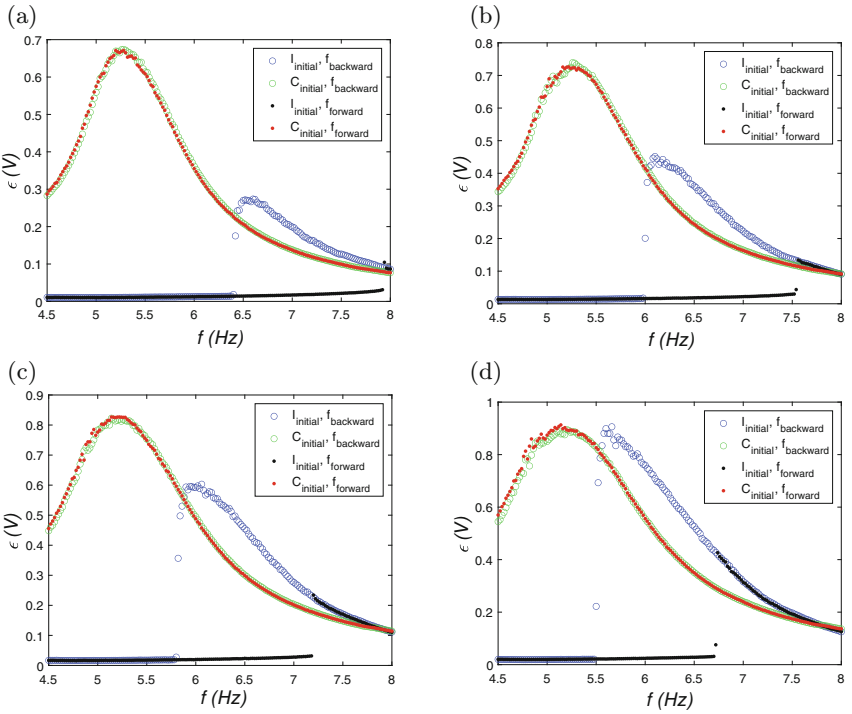


Fig. 4. Resonance curves for sweep forward and sweep backward activated from initial I or C shapes for moderate amplitude of excitation (a) $a_s = 0.5$ g, (b) $a_s = 0.6$ g, (c) $a_s = 0.75$ g and (d) $a_s = 0.9$ g.

In Fig. 5 the resonance curve corresponding to the I-shape oscillations overlaps the resonance curve of C-shape getting higher amplitude, however, without a change of the vibration mode. For $a_s = 1.25$ g dynamics changes qualitatively, a jump from I-shape oscillation to C-shape oscillation takes place for sweep backward, as presented in Fig. 5(b). It means that we obtained the one-way snap-through effect as a snap from I to C shape. The reverse effect has not been observed.

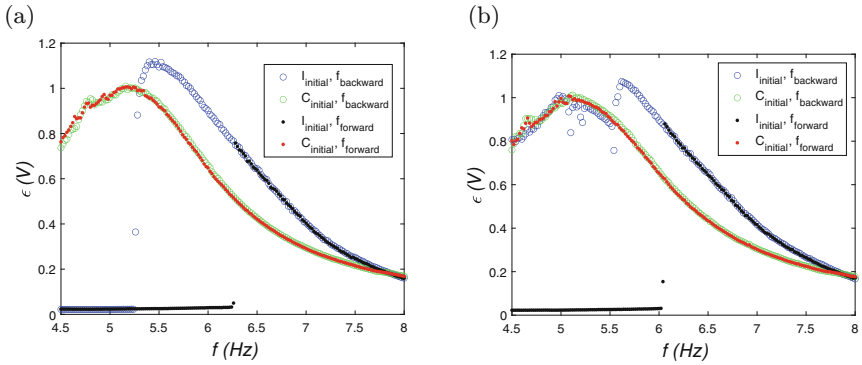


Fig. 5. Resonance curves for sweep forward and sweep backward activated from initial I or C shapes for large amplitude of excitation (a) $a_s = 1.1\,\mathrm{g}$ and (b) $a_s = 1.25\,\mathrm{g}$.

Further increase of excitation up to $a_s = 1.4$ g and $a_s = 1.6$ g leads to two way snap thorough effect when the I resonance curves reaches zone close to the C-shape natural frequency around 5.2 Hz. Starting from C-configuration and going forward or backward the snap-through effect and global cross-well oscillations are obtained in the frequency domain 5.0–5.5 Hz, as indicated by red dots and green circles in Fig. 6(a). Starting from I-configuration and going forward (black dots) about 5.6 Hz the system jumps to C-configuration and remains on it. For sweep backward (blue circles) about 5.6 Hz system jumps to C-configuration and about 5.5 Hz after snap-though goes to cross-well oscillations. This means that for the considered case (Fig. 6a) the transition to global oscillations and snap-though occurs only starting from C-configuration. The exit from snap-though oscillations to I-shape motion occurs for about 5 Hz. The transition from I-shape to global dynamics has not been observed. Larger amplitude of excitation $a_s = 1.6$ g, creates more complex dynamics. In the vicinity of C-shape natural frequency as shown in Fig. 6(b) system may transit to snap-through motion both either from initial C-shape or I-shape and may return from global vibrations either to C or I shapes.

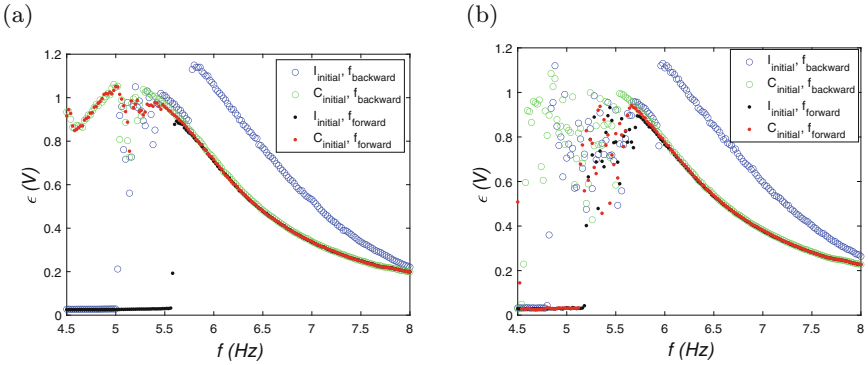


Fig. 6. Resonance curves for sweep forward and sweep backward activated from initial I or C shapes for large amplitude of excitation and transition to chaotic global dynamics (a) $a_s = 1.4\,\mathrm{g}$ and (b) $a_s = 1.6\,\mathrm{g}$.

In summary we conclude that for large amplitude oscillations the strong softening effect is observed for I-shape resonance curve. At the beginning it overlaps C-shape oscillations and eventually destroys C-resonance curve leading to the one way snap-though effect (from C state to snap-through) and then to the two-way snap-through, from C or I shape to cross-well oscillations).

3.3 Qualitative Analysis of Shell's Motion

Analysis of resonance curves shows the possibility of complex system dynamics. It is found that three types of motion may occur: regular vibration around C equilibrium (C-shape vibration), regular vibration around I equilibrium (I-shape vibration) and global vibration around both equilibria (I-C shape vibration). Thus, on the basis of Fig. 6(a), three excitation frequencies are selected 5.1 Hz, 5.5 Hz, 5.9 Hz, to examine different motion types. Therefore, for selected cases long time series are recorded (Fig. 7, 8 and 9) to check the three mentioned above types of motion: I, C and I-C vibration.

The qualitative analysis begins from standard Fast Fourier Transform of recorded data series. The FFT computed for C-shape vibrations presented in Fig. 10(b) demonstrates a clear periodic response with just two dominating harmonics. The FFT of I-shape confirms periodicity of the vibration but in this case response is more complex because several harmonics occur in the frequency spectrum in Fig. 10(c). This fact can be explained by the specific shape of I mode. As presented in Fig. 2(a) the shell stiffness essentially differs while moving in up or down directions, moving down shell stiffness is very high and the response reminds impact. After a change of the velocity sign the soft motion in up direction takes place. The model of I-shape vibrations requires a separate detailed analysis. The FFT for I-C mode vibrations presented in Fig. 10(a) is dominated by one harmonic and a wide (infinite) spectrum of other harmonics

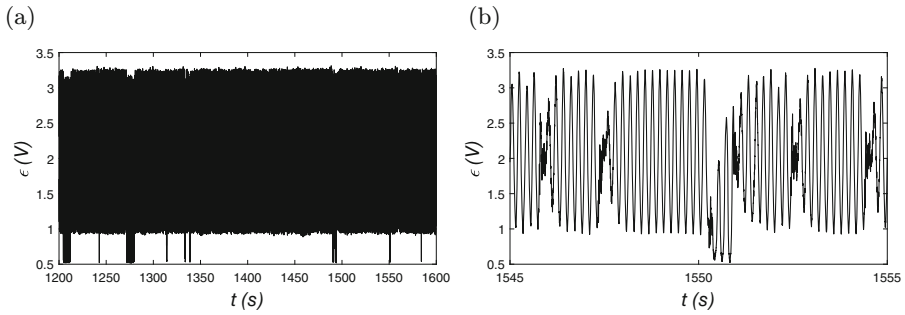


Fig. 7. Time series of I-C vibration for excitation amplitude $a_s = 1.4\,\mathrm{g}$ and frequency 5.1 Hz, (a) long time history and (b) zoom time window.

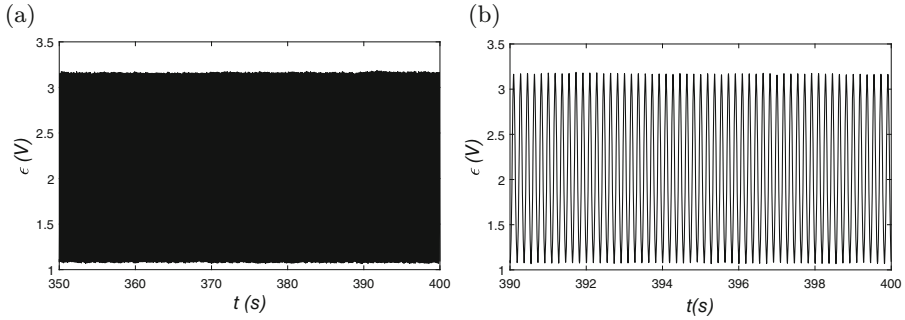


Fig. 8. Time series of C-shape vibration for excitation amplitude $a_s = 1.4\,\mathrm{g}$ and frequency 5.5 Hz, (a) long time history and (b) zoom time window.

located around. This kind of frequency spectrum is characteristic for chaotic motion.

The tools for qualitative motion analysis are phase portraits and Poincaré maps. Because just a single signal from the strain gauge is measured the missing second state variable (velocity) is reconstructed. For periodic motion with one harmonic, both state variables can be written in the form $\varepsilon(t) = A\sin\omega t$, $\dot{\varepsilon}(t) = A\omega\cos\omega t$. Velocity can be transformed into $\dot{\varepsilon}(t) = A\omega\sin\omega(t+\tau)$, where $\tau = \pi/2\omega$ is time delay. Velocity is reconstructed as time delay of displacement: $\dot{\varepsilon}(t)/\omega = \varepsilon(t+\tau)$.

Poincaré maps reconstructed from time series are plotted in Fig. 11. In fact C-shape vibration in Fig. 11(b) and I-shape vibration in Fig. 11(c) are represented by a single point which confirms their periodicity. In contrast, global oscillation of I-C type which includes the snap-through effect are presented in Fig. 11(a) by a cloud of points creating a strange chaotic attractor.

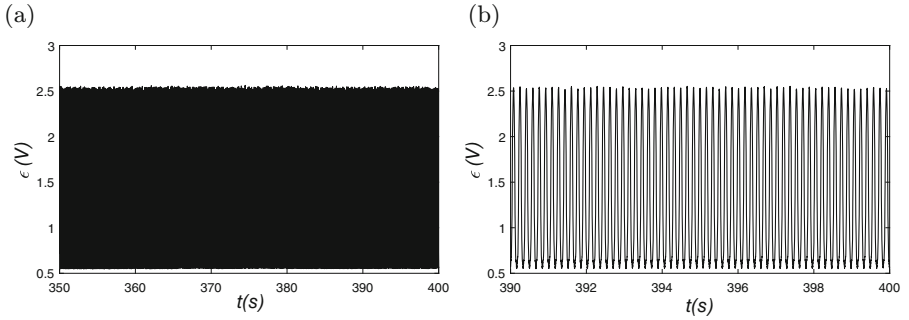


Fig. 9. Time series of I-shape vibration and for excitation amplitude $a_s = 1.4\,\mathrm{g}$ and frequency 5.9 Hz, (a) long time history and (b) zoom time window.

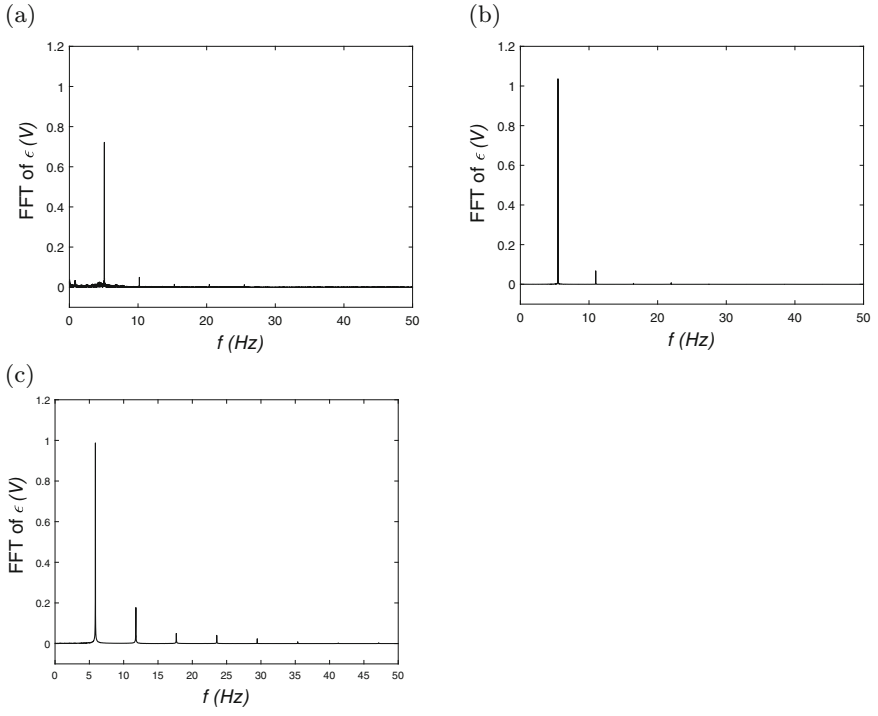


Fig. 10. Fast Fourier transforms from experimental data obtained for (a) I-C vibration $f = 5.1$ Hz, (b) C-shape vibration $f = 5.5$ Hz, (c) I-shape vibration $f = 5.9$ Hz; $a_s = 1.4\,\mathrm{g}$.

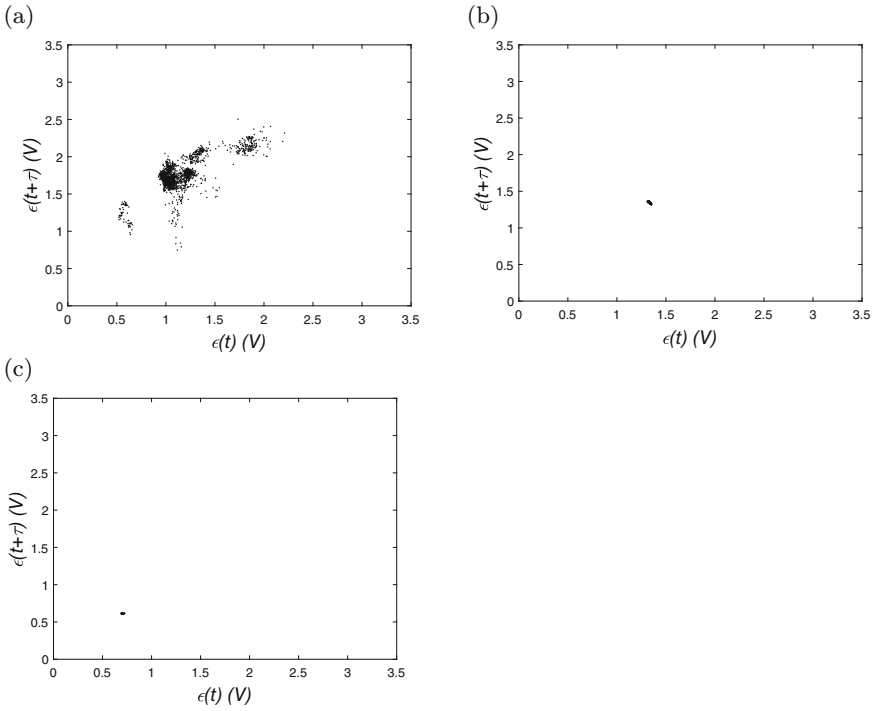


Fig. 11. Reconstructed Poincaré maps for (a) $f = 5.1$ Hz, (b) $f = 5.5$ Hz, (c) $f = 5.9$ Hz; a_s=1.4 g.

4 Conclusions

A dedicated design of composite shell with asymmetric configuration is presented in this paper. In contrast to most of the published papers, the proposed structure exhibits two stable equilibrium configurations: I-shape and C-shape equilibrium which differs essentially. The experimental investigations demonstrate strongly nonlinear softening effect for I-shape vibrations. The nonlinear behaviour is observed even for small oscillations and increases essentially for moderate oscillations. Furthermore, due to the huge difference in upward and downward stiffness, the motion for I-shape configuration reminds one-way impact oscillations. For the C-shape configuration the response is almost linear and remains almost linear even for moderate and large oscillations. As the excitation amplitude increases the I-shape resonance curve gets closer to C-shape resonance curve and for moderate vibration the one-way snap from I to C-shape vibration takes place. Further increase of excitation leads the I and C resonance curves to overlap. Then, a strong interaction between both vibration modes, C and I, occur and the two-way snap-through effect is activated creating cross well global dynamics. Detailed signal analysis by FFT and Poincaré maps application shows three different vibration types: periodic C-shape vibration composed of two har-

monics, periodic I-shape oscillation composed of several harmonics and global I-C chaotic vibration composed of the snap-through effect and represented by the strange chaotic attractor on Poincaré map. The domain of parameters where chaotic dynamics has been detected is attractive to design an effective system for vibration energy harvesting. For this purpose ongoing research activity is devoted to derive reduced order models enabling to perform parametric study of the structure, to understand complex nonlinear dynamics, including mode interactions as well as to optimise the shell's performance for energy harvesting.

Acknowledgment. This research was funded in whole or in part by National Science Centre, Poland 2021/41/B/ST8/03190. For the purpose of Open Access, the author has applied a CC-BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

References

1. Brunetti, M., Favata, A., Vidoli, S.: Enhanced models for the nonlinear bending of planar rods: localization phenomena and multistability. Proc. Math. Phys. Eng. Sci. **476**(2242), 20200455 (2020). https://doi.org/10.1098/rspa.2020.0455.
2. Brunetti, M., Kloda, L., Romeo, F., Warminski, J.: Multistable cantilever shells: Analytical prediction, numerical simulation and experimental validation. Compos. Sci. Technol. **165**(14), 397–410 (2018). https://doi.org/10.1016/j.compscitech.2018.06.021
3. Brunetti, M., Mitura, A., Romeo, F., Warminski, J.: Nonlinear dynamics of bistable composite cantilever shells: an experimental and modelling study. J. Sound Vib. **526**(2), 116779 (2022). https://doi.org/10.1016/j.jsv.2022.116779
4. Dano, M.L., Hyer, M.W.: Thermally-induced deformation behavior of unsymmetric laminates. Int. J. Solids Struct. **35**(17), 2101–2120 (1998). https://doi.org/10.1016/S0020-7683(97)00167-4
5. Emam, S.A., Inman, D.J.: A review on bistable composite laminates for morphing and energy harvesting. Appl. Mech. Rev. **67**(6), 060803 (2015). https://doi.org/10.1115/1.4032037
6. Harne, R.L., Wang, K.W.: Harnessing Bistable Structural Dynamics. For Vibration Control, Energy Harvesting and Sensing. Wiley, Chichester (2017)
7. Hyer, M.W.: Some observations on the cured shape of thin unsymmetric laminates. J. Compos. Mater. **15**(2), 175–194 (1981). https://doi.org/10.1177/002199838101500207
8. Mitura, A., Brunetti, M., Kloda, L., Romeo, F., Warminski, J.: Experimental nonlinear dynamic regimes for energy harvesting from cantilever bistable shells. Mech. Syst. Signal Process. **206**(9), 110890 (2024). https://doi.org/10.1016/j.ymssp.2023.110890
9. Ren, L.L., Zhang, W., Zhang, Y.F.: Dynamic snap-through and nonlinear vibrations of bistable asymmetric cross-ply composite laminated cantilever shell under external excitation. Mech. Syst. Signal Process. **195**(1), 110193 (2023). https://doi.org/10.1016/j.ymssp.2023.110193
10. Xu, Y., Hao, Y., Zhang, W., Zhang, Y.: Free vibration analysis of trapezoidal bistable laminates supported at the elastic midpoint of the median line. Mathematics **11**(15), 3326 (2023). https://doi.org/10.3390/math11153326

11. Zhang, Z., Chen, D., Wu, H., Bao, Y., Chai, G.: Non-contact magnetic driving bioinspired Venus flytrap robot based on bistable anti-symmetric CFRP structure. Compos. Struct. **135**(6), 17–22 (2016). https://doi.org/10.1016/j.compstruct.2015.09.015
12. Zhang, Z., et al.: Bistable morphing composite structures: a review. Thin-Walled Struct. **142**(6), 74–97 (2019). https://doi.org/10.1016/j.tws.2019.04.040
13. Zhang, Z., Li, X., Yu, X., Chai, H., Li, Y., Wu, H., Jiang, S.: Magnetic actuation bionic robotic gripper with bistable morphing structure. Compos. Struct. **229**(7024), 111422 (2019). https://doi.org/10.1016/j.compstruct.2019.111422



The Effect of Accuracy of Initial Velocity Discretisations on Discrete Energy in Variational Integration

Mamoru Gunji, Yusuke Ono, and Linyu Peng(✉)

Department of Mechanical Engineering, Faculty of Science and Technology, Keio University, Yokohama 223-8522, Japan
{mg,yuu555yuu}@keio.jp, l.peng@mech.keio.ac.jp
http://www.peng.mech.keio.ac.jp

Abstract. Energy fluctuation is regarded as a fundamental characteristic of symplectic integration, even in long-term simulations. A systematic approach for developing symplectic integration involves discretising a variational principle rather than the corresponding Euler–Lagrange equations, which are the equations of motion. The resulting numerical method is referred to as a variational integrator (VI). In the current paper, we investigate how the selection of initial values and energy discretisation in VIs can affect the behaviour of energy fluctuation. Numerical simulation of several conservative systems suggests that the combination of energy and initial velocity discretisations, both of higher and the same order of accuracy, can amount to numerical energy behavior with smaller error.

Keywords: variational integration · initial values · energy fluctuation · Kepler problem

1 Introduction

In the study of numerical methods for solving differential equations, there has been significant attention drawn to their effectiveness in accurately modeling physical phenomena. In the exploration of these numerical schemes, our focus extends beyond considerations of simulation cost, and stability. We also delve into preserving specific geometric or algebraic structures, with an emphasis on preserving symplectic structures, symmetries, volume, and conserved quantities; see, for instance, [1–4]. In particular in long-term simulations of conservative mechanical systems, it is crucial to maintain the symplectic structure. For instance, the forward Euler method and the Runge–Kutta method, lacking in this preservation, can lead to substantial deviations from the true solution (see, e.g., [2,5]). For the systematic development of symplectic integration in Lagrangian mechanics, a typical approach involves discretizing the variational principle and deriving a numerical scheme from discrete Euler–Lagrange equations using discrete variational calculus. Numerical methods constructed through

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 88–98, 2025.
https://doi.org/10.1007/978-3-031-72794-8_6

this process are referred to as variational integrators (VIs) [5–7], and they have been extensively applied (e.g., [8–12]).

Energy fluctuation is a well-recognised feature of symplectic integration, including VIs, particularly when applied to conservative mechanical systems. Conventionally, the initial velocity has often been discretised using a first-order Taylor expansion. In this study, we aim to investigate how the accuracy of initial velocity discretisations and the representations of discrete energy in VIs affect the behaviour of energy fluctuation. To be more precise, we consider the conventional forward difference, the central difference, and the fourth-order central difference as methods for discretising the initial velocity and kinetic energy, respectively. These methods are then applied to illustrative examples from classical mechanics: the simple harmonic oscillator, the simple pendulum, and the Kepler problem.

This paper is structured as follows. In Sect. 2, we provide an overview of VIs and summarise the discretisation methods for initial values and energy. In Sect. 3, we briefly review the simple harmonic oscillator, the simple pendulum and the Kepler problem as examples of conservative systems. In Sect. 4, we show the results of numerical simulations for the simple harmonic oscillator, the simple pendulum and the Kepler problem. Additionally, we analyse and compare the results for the Kepler problem in different step sizes. Finally, we draw our conclusions and address potential avenues for future research in Sect. 5.

2 Variational Integration

In this section, we briefly overview variational calculus in Lagrangian mechanics and its discrete counterpart particularly for deriving VIs as discrete Euler–Lagrange equations.

Let $t \in [0, T]$. Consider a smooth curve $q : [0, T] \to \mathcal{M}$ on an n-dimensional manifold $\mathcal{M}$ and its tangent vector at $q(t)$, i.e., $\dot{q} \in T_q\mathcal{M}$. A Lagrangian L is defined in the tangent bundle $T\mathcal{M}$ coordinated $(q, \dot{q})$ and the corresponding action functional reads

$$\mathscr{L}[q] = \int_0^T L\left(q, \dot{q}\right) \mathrm{d}t. \tag{1}$$

Consider variation of the curve q: $q + \varepsilon\delta q$ $(\varepsilon \in \mathbb{R})$ with fixed endpoints $\delta q(0) = \delta q(T) = 0$. Variational calculus (e.g., [13]) gives

$$\begin{aligned} 0 = \delta\mathscr{L}[q] &:= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} \mathscr{L}[q + \varepsilon\delta q] \\ &= \int_0^T \left(\frac{\partial L\left(q, \dot{q}\right)}{\partial q^i} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L\left(q, \dot{q}\right)}{\partial \dot{q}^i} \right) \delta q^i \mathrm{d}t \\ &\quad + \left[\frac{\partial L(q, \dot{q})}{\partial \dot{q}^i} \right]_{t=T} \delta q^i(T) - \left[\frac{\partial L(q, \dot{q})}{\partial \dot{q}^i} \right]_{t=0} \delta q^i(0). \end{aligned} \tag{2}$$

Due to the arbitrarity of the variation δq and the fixed-endpoint conditions, the equations of motion, namely, the Euler–Lagrange equations read

$$\frac{\partial L\left(q,\dot{q}\right)}{\partial q^i}-\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L\left(q,\dot{q}\right)}{\partial \dot{q}^i}=0,\quad i=1,2,\ldots,n, \tag{3}$$

where $\frac{\mathrm{d}}{\mathrm{d}t}$ denotes the total derivative with respect to t, and the Einstein summation convention is consistently utilised throughout this paper.

Now we consider discrete variational calculus and derive the discrete Euler–Lagrange equations. Let t_j $(j=0,1,\ldots,N)$ be the discretised time, and $h:=t_{j+1}-t_j$ be the fixed time step. Consider a discrete curve $q_j:=q(t_j)$ on the manifold $\mathcal{M}$. A discrete Lagrangian $L^{\Delta}:\mathcal{M}\times\mathcal{M}\to\mathbb{R}$ is defined as $L^{\Delta}(q_j,q_{j+1})$ and the corresponding discrete action functional is given by

$$\mathscr{L}^{\Delta}\left[q\right]:=\sum_{j=0}^{N-1}L^{\Delta}\left(q_j,q_{j+1}\right). \tag{4}$$

Similarly, we consider variation of the discrete curve q_j: $q_j+\varepsilon\delta q_j$ $(\varepsilon\in\mathbb{R})$ with fixed endpoints $\delta q_0=\delta q_N=0$. Discrete variational calculus gives

$$\begin{aligned}0=\delta\mathscr{L}^{\Delta}[q]&:=\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0}\mathscr{L}^{\Delta}[q+\varepsilon\delta q]\\&=\sum_{j=1}^{N-1}\left(\mathrm{D}_1L^{\Delta}(q_j,q_{j+1})+\mathrm{D}_2L^{\Delta}(q_{j-1},q_j)\right)\cdot\delta q_j\\&\quad+\mathrm{D}_1L^{\Delta}(q_0,q_1)\cdot\delta q_0+\mathrm{D}_2L^{\Delta}(q_{N-1},q_N)\cdot\delta q_N,\end{aligned} \tag{5}$$

where D_1 and D_2 denote gradients with respect to the first and second arguments, respectively. Due to fixed-endpoint conditions and the arbitrarity of δq_j $(j=1,2,\ldots,N-1)$, we obtain the discrete Euler–Lagrange equations as follows

$$\mathrm{D}_1L^{\Delta}(q_j,q_{j+1})+\mathrm{D}_2L^{\Delta}(q_{j-1},q_j)=0,\quad j=1,2,\ldots,N-1. \tag{6}$$

In this paper, for simplicity, we assume that the Lagrangian is the difference of the kinetic energy and the potential energy, namely

$$L(q,\dot{q})=\frac{1}{2}\dot{q}^{\mathrm{T}}M\dot{q}-V(q), \tag{7}$$

where $(\cdot)^{\mathrm{T}}$ represents the transpose of a vector or a matrix, M denotes the mass matrix, and $V(q)$ is the potential energy. By substituting (7) into (3), the corresponding Euler–Lagrange equations become

$$M\ddot{q}=-\nabla V(q), \tag{8}$$

where ∇ denotes the gradient of a function. This is exactly Newton's second law of motion in classical mechanics.

Define the discrete Lagrangian $L_j^{\triangle} := L^{\triangle}(q_j, q_{j+1})$ as one of the simplest discretisations of (7) as follows

$$L_j^{\triangle} = \frac{1}{2}\left(\frac{q_{j+1}-q_j}{h}\right)^{\mathrm{T}} M \left(\frac{q_{j+1}-q_j}{h}\right) - V\left(q_j\right). \tag{9}$$

The discrete Euler–Lagrange equations can be directly calculated:

$$\begin{aligned} 0 &= \mathrm{D}_1 L_j^{\triangle} + \mathrm{D}_2 L_{j-1}^{\triangle} \\ &= -M\left(\frac{q_{j+1}-q_j}{h^2}\right) - \nabla V\left(q_j\right) + M\left(\frac{q_j - q_{j-1}}{h^2}\right). \end{aligned} \tag{10}$$

Assuming M being invertible, this gives the VI

$$q_{j+1} = 2q_j - q_{j-1} - h^2 M^{-1} \nabla V\left(q_j\right). \tag{11}$$

It is equivalent to the Störmer–Verlet method by eliminating the velocities, which is also referred to as the leapfrog method [14].

In the VI (11), initial values q_0 and q_1 are required, which are determined by the given initial conditions $q_0 = q(0)$ and $\dot{q}_0 = \dot{q}(0)$. Conventionally, q_1 is determined by the following discretisation of $\dot{q}_0$, denoted as VI1:

$$\text{VI1}\colon\ q_1 = q_0 + h\dot{q}_0 + \mathcal{O}\left(h^2\right). \tag{12}$$

To investigate its efficiency, we propose the following discretisations for the initial velocity $\dot{q}_0$:

$$\begin{aligned} &\text{VI2}\colon\ q_1 = q_0 + h\dot{q}_0 - \frac{h^2}{2} M^{-1} \nabla V\left(q_0\right) + \mathcal{O}\left(h^3\right), \\ &\text{VI3}\colon\ q_1 = q_0 + h\dot{q}_0 - \frac{h^2}{2} M^{-1} \nabla V\left(q_0\right) - \frac{h^3}{6} M^{-1} \operatorname{Hess}(V(q_0))\dot{q}_0 + \mathcal{O}\left(h^4\right), \end{aligned} \tag{13}$$

where $\operatorname{Hess}(V(q_0))$ denotes the Hessian of $V(q)$ at $q = q_0$. The approximations VI1, VI2, and VI3 exhibit first-order, second-order and third-order accuracy, respectively.

The total energy corresponding to the Lagrangian (7) reads

$$E(q, \dot{q}) = \frac{1}{2}\dot{q}^{\mathrm{T}} M \dot{q} + V(q), \tag{14}$$

which is a constant of motion with respect to the system (8). When restricted to the discrete curve q_j, the total energy becomes

$$E(q_j, \dot{q}_j) = \frac{1}{2}\dot{q}_j^{\mathrm{T}} M \dot{q}_j + V(q_j), \tag{15}$$

where the discrete 'velocity' $\dot{q}_j$ at q_j is to be approximated by using the series $\{q_j\}_{j=0,1,\ldots,N}$. To access the energy fluctuation behaviour along the discrete curve $\{q_j\}_{j=0,1,\ldots,N}$, which satisfies the VI (11), we propose the following discretisation methods for the discrete velocity $\dot{q}_j$:

$$\begin{aligned}
&\text{VI}\alpha\text{:}\ \ \dot{q}_j = \frac{q_{j+1} - q_j}{h} + \mathcal{O}(h),\ j \geq 1,\\
&\text{VI}\beta\text{:}\ \ \dot{q}_j = \frac{q_{j+1} - q_{j-1}}{2h} + \mathcal{O}\left(h^2\right),\ j \geq 1,\\
&\text{VI}\gamma\text{:}\ \ \dot{q}_j = \frac{-q_{j+2} + 8q_{j+1} - 8q_{j-1} + q_{j-2}}{12h} + \mathcal{O}\left(h^4\right),\ j \geq 2,
\end{aligned} \tag{16}$$

where VIα, VIβ, and VIγ denote the conventional forward difference, the central difference, and a fourth-order central difference discretisations, respectively.

In the rest of the paper, we will explore the influence of the discretisations, namely VI1, VI2, and VI3, applied to the initial velocity, as well as the representations of discrete energy, namely VIα, VIβ, and VIγ, on the behaviour of energy fluctuation. This investigation will be carried out through a numerical analysis of several illustrative examples.

3 Examples of Conservative Systems

In this section, we will provide a brief overview of the simple harmonic oscillator, the simple pendulum, and the Kepler problem. All of them are conservative systems, serving as the illustrative examples of our study.

The Simple Harmonic Oscillator. The simple harmonic oscillator can be depicted as a mass attached to one end of a spring, with the assumption that there is no friction. When it is displaced from its equilibrium position, this system encounters a restoring force that is directly proportional to the displacement, denoted as $x \in \mathbb{R}$, and acts in the direction opposite to the displacement. The Lagrangian reads

$$L(x, \dot{x}) = \frac{1}{2} m\dot{x}^2 - \frac{1}{2} kx^2 \tag{17}$$

and the equation of motion is

$$m\ddot{x} = -kx, \tag{18}$$

where k is the spring constant and m is the mass. The total energy

$$E(x, \dot{x}) = \frac{1}{2} m\dot{x}^2 + \frac{1}{2} kx^2 \tag{19}$$

remains constant along solutions of the equation of motion.

The Simple Pendulum. A simple pendulum consists of a mass m, attached to a lightweight string which can swing back and forth under the action of gravity. When the mass is displaced away from its equilibrium position (the vertical hanging direction), it is acted upon by gravity which tries to restore it to this equilibrium. This causes the pendulum to oscillate. The equation of motion for a small-angle simple pendulum with no air resistance or friction is

$$l\ddot{\theta} = -g \sin \theta, \tag{20}$$

where $\theta \in \mathcal{S}^1$ is the angle between the string and the vertical direction, g is the gravitational acceleration, l is the length of the pendulum, and $\ddot{\theta}$ is the angular acceleration. The Lagrangian reads

$$L(\theta, \dot{\theta}) = \frac{1}{2} ml^2 \dot{\theta}^2 - mgl(1 - \cos \theta), \tag{21}$$

and the total energy is given by

$$E(\theta, \dot{\theta}) = \frac{1}{2} ml^2 \dot{\theta}^2 + mgl(1 - \cos \theta). \tag{22}$$

At the extremes of the swing, all energy is potential, while at the bottom of the swing, all energy is kinetic.

The Kepler Problem. The Kepler problem refers to the motion of one body with mass m_1 orbiting around another body with usually a bigger mass m_2 under gravity, such as a planet revolving around the sun. Let $x \in \mathbb{R}^3$ be the relative position, and the Lagrangian reads

$$L(x, \dot{x}) = \frac{1}{2} \dot{x}^{\mathrm{T}} M \dot{x} + \frac{G m_1 m_2}{|x|}, \tag{23}$$

where G is the gravitational constant, and

$$M = \frac{m_1 m_2}{m_1 + m_2} I_3 \tag{24}$$

is the relative mass matrix with I_3 the 3×3 identity matrix. The equation of motion in the Kepler problem reads (see, e.g., [15])

$$\ddot{x} = -\frac{G(m_1 + m_2)}{|x|^3} x. \tag{25}$$

The total energy is

$$E(x, \dot{x}) = \frac{1}{2} \dot{x}^{\mathrm{T}} M \dot{x} - \frac{G m_1 m_2}{|x|}. \tag{26}$$

The Kepler problem is completely integrable according to Liouville's theorem [16]. In particular, $SO(3)$ invariance of the Lagrangian (and hence the variational problem) amounts to the conservation of angular momenta from Noether's theorem, namely $x \times \dot{x}$ is a constant vector of motion that is fully determined by the initial conditions; this implies that the motion is planetary, on a plane orthogonal to $x \times \dot{x}$, coordinated by $q \in \mathbb{R}^2$ [16,17]. Kepler's first law states that the trajectory of a body with negative energy in a classical gravitational potential is an elliptical orbit, with one of the foci positioned at the gravitational center. In the current paper, we will use the planar coordinates $q \in \mathbb{R}^2$ instead of $x \in \mathbb{R}^3$ to conduct the numerical analysis. Namely, the Lagrangian reads

$$L(q, \dot{q}) = \frac{1}{2}\dot{q}^{\mathrm{T}}\widetilde{M}\dot{q} + \frac{k}{|q|}, \quad q \in \mathbb{R}^2, \tag{27}$$

where $\widetilde{M}$ is a 2×2 mass matrix and k is a constant. The corresponding total energy is

$$E(q, \dot{q}) = \frac{1}{2}\dot{q}^{\mathrm{T}}\widetilde{M}\dot{q} - \frac{k}{|q|}. \tag{28}$$

Several integrators for the Kepler problem have been suggested in prior research, such as those presented in [8,18,19]. In this study, our primary focus is on assessing the impact of initial velocity discretisation and discrete energy on the evaluation of VIs. To simplify our analysis, we will employ the VI (11).

4 Numerical Simulations

In this section, we perform numerical simulations to investigate the influence of velocity discretisation accuracy in the initial conditions and discrete energy on the energy fluctuation behaviour. We consider the three systems introduced in the previous section: the simple harmonic oscillator, the simple pendulum, and the reduced Kepler problem with an elliptical orbit. The parameters used in each numerical simulation are summarised in Table 1 in which I_2 denotes the 2×2 identity matrix.

Table 1. Parameters and initial conditions used in the numerical analysis

	Parameters	Potentials	Initial positions & velocity
Harmonic oscillator	$m = 1, k = 1$	$\frac{1}{2}kx^2$	$x_0 = 1,\ \dot{x}_0 = 1$
Simple pendulum	$m = 1, g = 9.8, l = 1$	$mgl(1 - \cos\theta)$	$\theta_0 = 1,\ \dot{\theta}_0 = 1$
Kepler problem	$\widetilde{M} = I_2, k = 1$	$-\frac{k}{\lvert q \rvert}$	$q_0 = (1, 0)^{\mathrm{T}},\ \dot{q}_0 = (0, 1)^{\mathrm{T}}$

The maximum relative error of the total energy

$$\max_j \left| \frac{E(q_j, \dot{q}_j) - E(q_0, \dot{q}_0)}{E(q_0, \dot{q}_0)} \right| \tag{29}$$

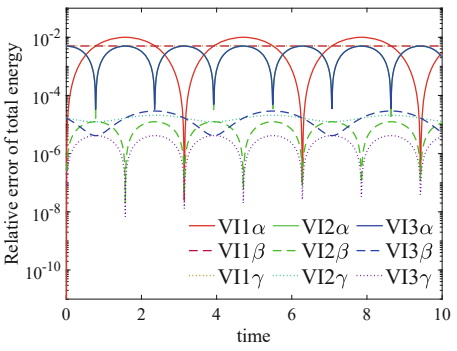
for longer time ($T = 1000$) with time step $h = 0.01$ is summarised in Table 2 and it happens to be similar as that for shorter time ($T = 10$). The shorter time results are shown in Fig. 1 (the column on the left) by semi-log graphs. The discrete energy in all simulations fluctuates within a certain range, and longer time calculations will not cause the energy range to exceed the current maximum range according to Table 2. In all cases, VI3γ is the most accurate. The simulations reveal that employing both higher-order accuracy in discrete energy and initial velocity discretization can result in a smaller error from the discrete energy perspective. The maximum error of the position

$$\max_j |q_j - \widehat{q}_j| \,. \tag{30}$$

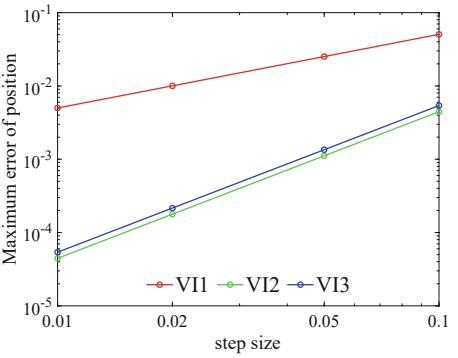
in a short time range ($0 \le t \le 10$) for time steps $h = 0.1, 0.05, 0.02, 0.01$ is shown in Fig. 1 (the column on the right) by log-log graphs. Here, $\widehat{q}_j$ is calculated using the ode89 in Matlab of 8th-order accuracy with the options given by odeset ('Reltol', 1e−13, 'AbsTol', 1e−14), namely, with relative error tolerance 1e−13 and absolute error tolerance 1e-14; the initial velocity is defined using VI3 with $h = 10^{-5}$. Simulations for the illustrative systems show that VI1 is of first-order accuracy, and VI2 and VI3 are of second-order accuracy.

Table 2. Maximum relative error of the total energy in longer time ($T = 1000$) and for step size $h = 0.01$

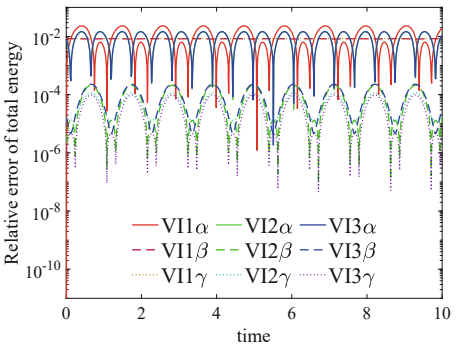
	VI1α	VI1β	VI1γ	VI2α	VI2β	VI2γ	VI3α	VI3β	VI3γ
Harmonic oscillator	1.0e−2	5.0e−3	5.4e−3	5.0e−3	1.3e−5	2.1e−5	5.0e−3	2.9e−5	4.2e−6
Simple pendulum	2.3e−2	8.4e−3	9.1e−3	1.5e−2	2.1e−4	1.2e−4	1.5e−2	2.3e−4	9.8e−5
Kepler problem	1.0e−4	2.5e−5	5.9e−5	2.5e−5	2.5e−9	3.6e−5	8.4e−6	3.3e−5	1.4e−9



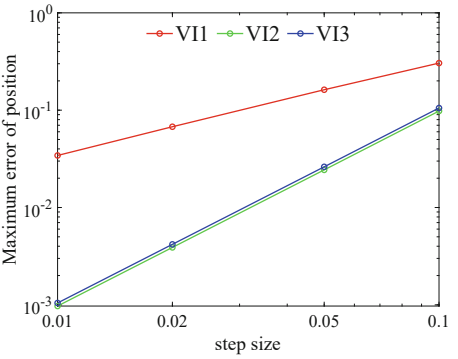
(a) Relative error of total energy: Harmonic oscillator



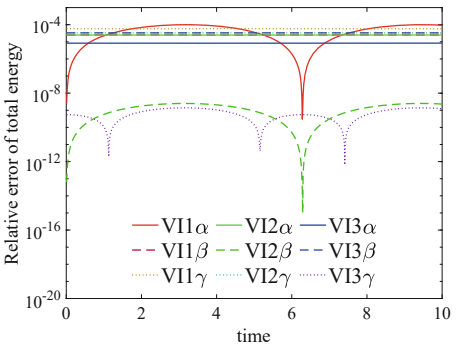
(b) Maximum error of position: Harmonic oscillator



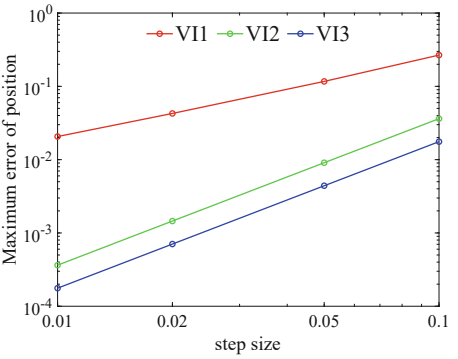
(c) Relative error of total energy: Simple pendulum



(d) Maximum error of position: Simple pendulum



(e) Relative error of total energy: Kepler problem



(f) Maximum error of position: Kepler problem

Fig. 1. Figures (a), (c), (e): Relative error of the total energy in shorter time ($T = 10$) with $h = 0.01$. The longer time results are similar. Figures (b), (d), (f): The maximum error of the position in short time ($T = 10$). (a). Relative error of total energy: Harmonic oscillator (b). Maximum error of position: Harmonic oscillator. (c). Relative error of total energy: Simple pendulum. (d). Maximum error of position: Simple pendulum. (e). Relative error of total energy: Kepler problem. (f). Maximum error of position: Kepler problem.

5 Conclusion

In this paper, the effect of accuracy of initial velocity discretisations on the discrete energy of VIs was investigated through numerical simulations. We introduced three methods for discretising the initial value and for discretising the total energy with different orders of accuracy respectively, and applied these methods to the 1-dimensional harmonic oscillator, the 1-dimensional simple pendulum, and the reduced Kepler problem in $\mathbb{R}^2$. From the numerical simulation results using the simplest Störmer–Verlet method, it can be noticed that both the way the discrete energy is defined and the accuracy for discretising the initial velocity can greatly affect the evaluation of a VI. In conclusion, it tends to perform better by enhancing the accuracy of both of them based on the results of the current paper. The best can often be achieved by aligning the higher-order of accuracy in both the discrete energy and the initial velocity discretisation. However, enhancing the accuracy of only one of these aspects without ensuring a corresponding match in their order of accuracy might not necessarily result in improved error behavior. This is evident from the VI1β and VI1γ results presented in Table 2.

Ge–Marsden's theorem [20] states that a symplectic integrator cannot exactly preserve the exact energy without computing the exact solution. In a recent study [21], the moving frame was applied to construct VIs that preserve the total energy exactly but with varying time steps. Results of the current paper may provide new insights in reducing error of the energy behaviour for symplectic integrators with fixed time steps, although further studies and verifications are still needed.

Acknowledgement. YO was supported by JST-SPRING grant number JPMJSP 2123. LP was partially supported by JSPS KAKENHI grant number JP20K14365, JST-CREST grant number JPMJCR1914, and the Fukuzawa Fund and KLL of Keio University.

References

1. Hairer, E.: Backward analysis of numerical integrators and symplectic methods. Ann. Numer. Math. **1**, 107–132 (1994)
2. Hairer, E., Lubich, C., Wanner, G.: Geometric Numerical Integration: Structure-preserving Algorithms for Ordinary Differential Equations. Springer, Heidelberg (2006)
3. Yoshida, H.: Recent progress in the theory and application of symplectic integrators. Celest. Mech. Dyn. Astron. **56**, 27–43 (1993)
4. McLachlan, R.I., Quispel, G.R.W.: Splitting methods. Acta Numer. **11**, 341–434 (2002)
5. Marsden, J.E., West, M.: Discrete mechanics and variational integrators. Acta Numer. **10**, 357–514 (2001)
6. Veselov, A.: Integrable discrete-time systems and difference operators. Funct. Anal. Appl. **22**(2), 83–93 (1988)
7. Veselov, A.P.: Integrable Lagrangian correspondences and the factorization of matrix polynomials. Funct. Anal. Appl. **25**(2), 112–122 (1991)

8. Vermeeren, M.: Numerical precession in variational discretizations of the Kepler problem. In: Discrete Mechanics, Geometric Integration and Lie–Butcher Series, pp. 333–348. Springer (2018)
9. Lee, T., McClamroch, N., Leok, M.: A Lie group variational integrator for the attitude dynamics of a rigid body with applications to the 3D pendulum. In: Proceedings of 2005 IEEE Conference on Control Applications, CCA 2005, pp. 962–967 (2005)
10. Bou-Rabee, N., Owhadi, H.: Stochastic variational integrators. IMA J. Numer. Anal. **29**(2), 421–443 (2009)
11. Marsden, J.E., Patrick, G.W., Shkoller, S.: Multisymplectic geometry, variational integrators, and nonlinear PDEs. Commun. Math. Phys. **199**, 351–395 (1998)
12. Lew, A., Marsden, J.E., Ortiz, M., West, M.: Asynchronous variational integrators. Arch. Ration. Mech. Anal. **167**, 85–146 (2003)
13. Marsden, J.E., Ratiu, T.S.: Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems. Springer, Heidelberg (2013)
14. Leimkuhler, B., Reich, S.: Simulating Hamiltonian Dynamics. Cambridge University Press, Cambridge (2004)
15. Kibble, T., Berkshire, F.H.: Classical Mechanics. World Scientific Publishing Company (2004)
16. Arnol'd, V.I.: Mathematical Methods of Classical Mechanics. Springer, Heidelberg (2013)
17. Goldstein, H., Poole, C., Safko, J.: Classical Mechanics. Addison-Wesley, New York (2002)
18. Chin, S.A.: Physics of symplectic integrators: perihelion advances and symplectic corrector algorithms. Phys. Rev. E **75**(3), 036701 (2007)
19. Wanner, G.: Kepler, Newton and numerical analysis. Acta Numer **19**, 561–598 (2010)
20. Zhong, G., Marsden, J.E.: Lie-Poisson Hamilton-Jacobi theory and Lie-Poisson integrators. Phys. Lett. A **133**(3), 134–139 (1988)
21. Mansfield, E.L., Rojo-Echeburua, A., Hydon, P.E., Peng, L.: Moving frames and Noether's finite difference conservation laws I. Trans. Math. Appl. **3**(1), tnz004 (2019)

A Harmonic-Balance Power Flow Analysis Method for Nonlinear Vibration Systems

Jian Yang(✉), Cui Chao, and Yuhao Liu

University of Nottingham Ningbo China, Ningbo 315100, China
jian.yang@nottingham.edu.cn

Abstract. This paper presents the harmonic-balance-based power flow analysis (HB-PFA) method for nonlinear vibrating systems. A general power flow formulation for dynamics analysis of systems with various nonlinearities is firstly introduced. The formulation includes the derivation of energy and power balance equations, with each term representing a specific energy or power quantity. For periodic responses, approximate analytical solutions for power flow quantities are derived by the harmonic balance method. To illustrate the effectiveness of the method, a case study on the power flow of systems modelled by the Van der Pol oscillator is presented.

Keywords: harmonic balance method · power flow analysis, energy flow · Van der Pol oscillator · nonlinear vibrations

1 Introduction

The power flow analysis (PFA) approach is widely used as an efficient method that can be employed to assess the dynamic characteristics of complex structures and interconnected systems [1]. In Ref. [1], Goyder and White presented the fundamental principles and concepts of vibratory power flow. The main idea is to use the product of velocity and force as an index to evaluate the vibration transfer level within dynamical systems. Any attempt to control vibration by purely looking at force transmission or dynamic response may not be successful. Instead, successful vibration suppression can be guaranteed by using the power transmission between subsystems within an integrated structure. Many PFA methods, including the mobility method and dynamic stiffness method were used to investigate linear metallic structures. In recent years, there has been an increasing interest in developing PFA methods for nonlinear mechanical systems [2]. Studies were reported on PFA of nonlinear systems, including systems represented by the Duffing oscillator [3], interactive oscillators [4], impact oscillators [5,6], frictional systems [7] and systems with nonlinear joints [8]. Vibrational power flow has been used as an index to evaluate the performance of nonlinear vibration isolators [9], nonlinear dynamic vibration absorbers [10], and composite structures [11].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 99–107, 2025.
https://doi.org/10.1007/978-3-031-72794-8_7

The governing equation describing the dynamics of a nonlinear system with m degrees of freedom is

$$\mathbf{M}\ddot{\mathbf{X}} + \mathbf{C}\dot{\mathbf{X}} + \mathbf{K}\mathbf{X} = \mathbf{f}_e(t). \tag{1}$$

The vector $\mathbf{X} = \{x_1, x_2, \cdots, x_j, \cdots, x_m\}^{\mathrm{T}}$ denotes the coordinates, where each element x_j represents a coordinate. The mass, damping, and stiffness matrices $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K}$, respectively, are of dimension of m by m. The right-hand-side term $\mathbf{f}_e(t)$ denotes the external force vector.

The nonlinearity of the system introduces variability in these matrices, causing them to be dependent on the displacement and/or velocity variables, which can be represented by $\mathbf{X}$ and/or $\dot{\mathbf{X}}$.

Pre-multiply Eq. (1) by the velocity vector $\dot{\mathbf{X}}^{\mathrm{T}}$, the power balance equation that governs the system can be derived:

$$\dot{\mathbf{X}}^{\mathrm{T}}\mathbf{M}\ddot{\mathbf{X}} + \dot{\mathbf{X}}^{\mathrm{T}}\mathbf{C}\dot{\mathbf{X}} + \dot{\mathbf{X}}^{\mathrm{T}}\mathbf{K}\mathbf{X} = \dot{\mathbf{X}}^{\mathrm{T}}\mathbf{f}_e(t). \tag{2}$$

The instantaneous dissipated and input powers are defined by

$$\bar{p}_d = \dot{\mathbf{X}}^{\mathrm{T}}\mathbf{C}\dot{\mathbf{X}}, \tag{3a}$$

$$\bar{p}_{in} = \dot{\mathbf{X}}^{\mathrm{T}}\mathbf{f}_e(t), \tag{3b}$$

respectively. In many engineering applications, the time-averaged behavior in terms of power flows is a critical consideration. When employing a time period t_p, which is usually set as a period of a harmonic excitation, the average dissipation power and power inputted can be determined:

$$\bar{p}_d = \frac{1}{t_p}\int_{t_i}^{t_i+t_p} \dot{\mathbf{X}}^{\mathrm{T}}\mathbf{C}\dot{\mathbf{X}}\mathrm{d}t, \tag{4a}$$

$$\bar{p}_{in} = \frac{1}{t_p}\int_{t_i}^{t_i+t_p} \dot{\mathbf{X}}^{\mathrm{T}}\mathbf{f}_e(t)\mathrm{d}t, \tag{4b}$$

respectively, where t_i denotes the initial time from which the averaging process begins. It is worth emphasizing that when there is a periodic response, there will be zero net change in mechanical energy of the system. This means that the input power by external forces is dissipated entirely by damping, resulting in equal time-averaged dissipated and input powers. For nonlinear systems, their steady-state responses can be non-periodic in the forms such as chaotic or quasi-periodic responses. For non-periodic response, the averaging time can be set much larger, with consideration of the major response component of the lowest frequency. Studies on the influence of averaging time and initial conditions on the time-averaged power associated with a chaotic response can be found [3].

Solving Eq. (1) is essential for understanding and predicting the power flow properties of nonlinear vibration systems. Analytical approximation methods, including the harmonic balancing and averaging methods, can be applied to obtain periodic responses. The governing equations typically require numerical

solutions to attain responses of quasi-periodic or chaotic nature. Subsequent sections focus on the application of these techniques to analyze the power flow in nonlinear dynamical systems.

2 PFA Based on Harmonic Balance Method

Harmonic balancing (HB) is a valuable technique for studying nonlinear dynamical systems in the frequency domain (see, [12]). The primary idea behind the HB method is to use harmonic series to express the nonlinear terms and response in the governing equation. Then, to create a set of algebraic equations, the coefficients of the terms with the same frequency component are balanced. An iterative process is then used to find the solutions to these equations. To acquire power flow variables through the harmonic balancing approach, Eq. (1) is split into two parts, which are named the linear component and the nonlinear part. A generic representation of the solution process is provided with the transformed governing equation written as

$$\mathbf{M}_l\ddot{\mathbf{X}} + \mathbf{C}_l\dot{\mathbf{X}} + \mathbf{K}_l\mathbf{X} = \mathbf{f}_e(t) + \mathbf{f}_{\text{non}}(\dot{\mathbf{X}}, \ddot{\mathbf{X}}), \tag{5}$$

where $\mathbf{M}_l$, $\mathbf{C}_l$ and $\mathbf{K}_l$ denote the linear components of the corresponding matrices such that they do not depend on the response denoted by $\mathbf{X}$, $\dot{\mathbf{X}}$ and $\ddot{\mathbf{X}}$. On the other hand, the nonlinear component, $\mathbf{f}_{\text{non}}(\dot{\mathbf{X}}, \ddot{\mathbf{X}})$ encompasses nonlinear terms due to geometry, friction, damping, and other factors. If the primary oscillation frequency of the system response is ω_1, then the jth coordinate's time histories can be approximated using simplified harmonic series with N harmonics:

$$x_j(t) = \sum_{n=1}^{N}(\hat{x}_{(j,2n)}\sin(n\omega_1 t) + \hat{x}_{(j,2n-1)}\cos(n\omega_1 t)) + \hat{x}_{(j,0)} = \sum_{n=0}^{N}\tilde{R}_{(j,n)}\mathrm{e}^{\mathrm{i}(n\omega_1 t)}. \tag{6}$$

In this equation, $\tilde{R}_{(j,n)}$ represents a complex variable that denotes the jth coordinate with nth harmonic. The coefficients of the nth harmonic term are denoted as $\hat{x}_{(j,2n)}$ and $\hat{x}_{(j,2n-1)}$. Note that to deal with the dynamic equation of a self-excited system without the inhomogeneous term, the fundamental frequency is determined by solving the algebraic equations derived using harmonic balance method. It is also noted that $\tilde{R}(j,0)$ corresponds to $\hat{x}_{(j,0)}$ and if $n = 0$, it is a real number. However for $n > 0$, the following relationship holds:

$$\begin{aligned}\tilde{R}_{(j,n)} = &\Big(\hat{x}_{(j,2n)}\sin(n\omega_1 t) + \hat{x}_{(j,2n-1)}\cos(n\omega_1 t)\Big)\cos(n\omega_1 t)\\ &- \mathrm{i}\Big((\hat{x}_{(j,2n)}\sin(n\omega_1 t) + \hat{x}_{(j,2n-1)}\cos(n\omega_1 t))\sin(n\omega_1 t)\Big).\end{aligned} \tag{7}$$

In the same way, the corresponding component in the nonlinear term $\mathbf{f}_{\text{non}}(\dot{\mathbf{X}}, \ddot{\mathbf{X}})$, and the jth component in the external force term $\mathbf{f}_e(t)$ are expanded as

$$f_{e,j}(t) = \hat{s}_{(j,0)} + \sum_{n=1}^{N}(\hat{s}_{(j,2n)}\sin(n\omega_1 t) + \hat{s}_{(j,2n-1)}\cos(n\omega_1 t)) = \sum_{n=0}^{N}\tilde{S}_{(j,n)}\mathrm{e}^{\mathrm{i}(n\omega_1 t)}, \tag{8}$$

$$f_{non,j}(t) = \hat{q}_{(j,0)} + \sum_{n=1}^{N} (\hat{q}_{(j,2n)} \sin(n\omega_1 t) + \hat{q}_{(j,2n-1)} \cos(n\omega_1 t)) = \sum_{n=0}^{N} \tilde{Q}_{(j,n)} \mathrm{e}^{\mathrm{i}(n\omega_1 t)}, \tag{9}$$

respectively. By substituting Eqs. (7), (8) and (9) into Eq. (5) and equalising the nth harmonic coefficients, it gives

$$\Big(-(n\omega_1)^2 \mathbf{M}_l + \mathrm{i}(n\omega_1)\mathbf{C}_l + \mathbf{K}_l \Big) \tilde{\mathbf{R}}_\mathbf{n} = \tilde{\mathbf{S}}_\mathbf{n} + \tilde{\mathbf{Q}}_\mathbf{n}, \tag{10}$$

where

$$\tilde{\mathbf{Q}}_\mathbf{n} = \{\tilde{Q}_{(0,n)}, \cdots, \tilde{Q}_{(j,n)}, \cdots, \tilde{Q}_{(m,n)}\}^{\mathrm{T}}, \tag{11}$$

$$\tilde{\mathbf{R}}_\mathbf{n} = \{\tilde{R}_{(0,n)}, \cdots, \tilde{R}_{(j,n)}, \cdots, \tilde{R}_{(m,n)}\}^{\mathrm{T}}, \tag{12}$$

$$\tilde{\mathbf{S}}_\mathbf{n} = \{\tilde{S}_{(0,n)}, \cdots, \tilde{S}_{(j,n)}, \cdots, \tilde{S}_{(m,n)}\}^{\mathrm{T}}. \tag{13}$$

When ensuring the equilibrium between the real and imaginary components, it is crucial to emphasize that the system will have a total of m equations for $\tilde{R}_0$. However, when $0 < n \leq N$, there will be $2m$ equations for $\tilde{R}_n$.

Introducing a dynamic stiffness matrix

$$\mathbf{D}_n = -(n\omega_1)^2 \mathbf{M}_l + \mathrm{i}(n\omega_1)\mathbf{C}_l + \mathbf{K}_l. \tag{14}$$

By considering the nth harmonic and balancing the formulas, it gives

$$\begin{pmatrix} \mathbf{D}_0 & 0 & \cdots & 0 \\ 0 & \mathbf{D}_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{D}_N \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{R}}_\mathbf{0} \\ \tilde{\mathbf{R}}_\mathbf{1} \\ \vdots \\ \tilde{\mathbf{R}}_\mathbf{N} \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{S}}_\mathbf{0} + \tilde{\mathbf{Q}}_\mathbf{0} \\ \tilde{\mathbf{S}}_\mathbf{1} + \tilde{\mathbf{Q}}_\mathbf{1} \\ \vdots \\ \tilde{\mathbf{S}}_\mathbf{N} + \tilde{\mathbf{Q}}_\mathbf{N} \end{pmatrix}. \tag{15}$$

With the same number of unknowns, the given equation is converted into an equal set of $(2N+1)m$ nonlinear equations. Iterative solutions can be achieved when the number is large by using techniques like the Newton-Raphson method.

Note that nonlinearity is often prominent in a small number of coordinates in practical scenarios. For instance, local nonlinear elements can associate linear continuous systems with one another. Because of this trait, the problem can be made simpler by avoiding the need for a harmonic expansion of the linear degrees of freedom and lowering the computing expense by utilizing a sub-structure approach.

The estimation of the instantaneous input power can be achieved by:

$$p_{in} = \sum_{j=1}^{m} f_{e,j}(t) \dot{x}_j(t), \tag{16}$$

where the velocity is calculated by:

$$\dot{x}_j(t) = \sum_{n=1}^{N} n\omega_1 (\hat{x}_{(j,2n)} \cos(n\omega_1 t) - \hat{x}_{(j,2n-1)} \sin(n\omega_1 t)). \tag{17}$$

The averaging input power throughout a vibration period can be approximately determined using the HB method:

$$\bar{p}_{in} = \frac{\omega_1}{2\pi} \int_0^{2\pi/\omega_1} \sum_{j=1}^{m} f_{e,j}(t)\dot{x}_j(t)\,\mathrm{d}t. \tag{18}$$

The expressions for $f_{e,j}(t)$ and $\dot{x}_j(t)$ shown by Eq. (8) and Eq. (17) are recalled. For Eq. (18), the time-averaged input power becomes

$$\bar{p}_{in} = \frac{1}{2}\omega_1 \sum_{j=1}^{m} \Big(\sum_{n=1}^{N} n\big(-\hat{s}_{(j,2n)}\hat{x}_{(j,2n-1)} + \hat{s}_{(j,2n-1)}\hat{x}_{(j,2n)}\big)\Big). \tag{19}$$

The trigonometric functions' properties are used to analyze the system. A similar procedure can be used to attain the expression of both instantaneous and time-averaged dissipated (or transmitted) power by multiplying the relevant velocities by the Fourier expanded damping (or transmitted) forces.

3 Case Study

In this study, the harmonic balance power flow analysis method is employed to investigate the Van der Pol oscillator, which is used to model physical systems such as electric circuits and other mechanical systems [13–15]. The governing equation for a Van der Pol oscillator based system with harmonic excitation is given by:

$$\ddot{x} + \alpha(x^2 - 1)\dot{x} + \beta x = f\cos(\omega t). \tag{20}$$

Throughout this study, the parameter α is set to be positive. The variables ω and f represent the frequency and amplitude of the external force, respectively.

Multiply both sides of Eq. (20) by velocity $\dot{x}$, the power balance equation is found, that is,

$$\dot{x}\ddot{x} + \alpha\dot{x}(x^2 - 1)\dot{x} + \beta\dot{x}x = \dot{x}f\cos(\omega t), \tag{21}$$

from which we can establish definitions for the instantaneous change of the kinetic energy, as well as the input and dissipated powers. When $x \neq 0$ and $|x| > 1$, the dissipated power $p_d = \alpha(x^2-1)\dot{x}\dot{x}$ is positive. As with the behaviour of linear viscous damping, the so-called 'damping' term $\alpha(x^2 - 1)\dot{x}$ dissipates vibration energy when the displacement is considerably large. When the displacement is moderate (i.e., when $|x| < 1$) and the velocity is non-zero ($\dot{x} \neq 0$), the value of the product p_d becomes negative. This negative value indicates the total input power by the nonlinear force term $\alpha(x^2 - 1)\dot{x}$, implying that the corresponding force acts as an energy source rather than a sink. This property is unique to the Van der Pol oscillator different from that of linear vibration systems. The nonlinear term significantly influences the dynamic behavior of the system.

For a periodic response, the expression of a periodic response is assumed to be

$$x = r_1\cos(\omega t + \phi), \tag{22}$$

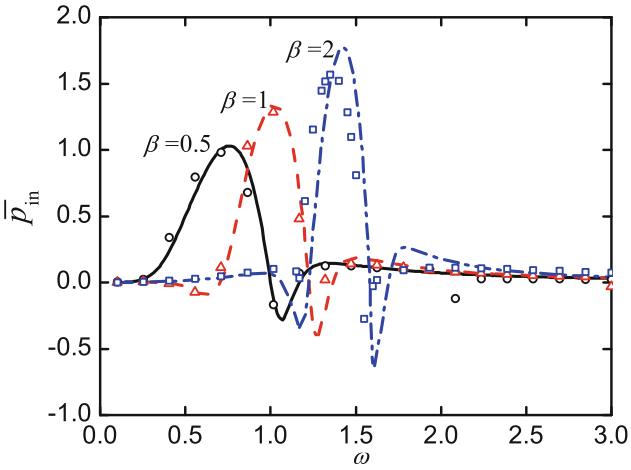


Fig. 1. The influence of the parameter β on the input power is examined for a system with $f = 1.0$ and $\alpha = 0.5$. The results obtained by the harmonic balance (HB) approximation: The dash-doted line for $\beta = 2$, the dashed line for $\beta = 1$, and the solid line for $\beta = 0.5$. The results obtained by numerical simulation: squares for $\beta = 2$, triangles for $\beta = 1$ and circles for $\beta = 0.5$.

where r_1 is the amplitude of displacement response and ϕ is the phase angle. The averaged input power can be obtained:

$$\bar{p}_{in} = \frac{1}{8}\alpha\omega^2 r_1^2(r_1^2 - 4). \tag{23}$$

The oscillator can also exhibit quasi-periodic responses, as shown in Ref. [2]. In such cases, the steady-state response is expressed as

$$x = z_1 \cos(\omega_p t + \theta) + r_1 \cos(\omega t + \phi). \tag{24}$$

where z_1 is the amplitude for the response component at ω_p. The vibration power flow quantities can be obtained following the proposed HB-PFA approach presented in the previous section.

The influence of parameters β on the averaging input power is studied and shown in Fig. 1. To enhance understanding, the figure does not display the unstable branches of the analytical approximation for periodic movements where $r_1^2 < 2$. Figure 1 demonstrates the impact of the stiffness coefficient β on the system. It is apparent that an increase in β causes a shift of the peak in each curve towards higher frequencies. This observation holds true for a system with fixed values of $\alpha = 0.5$ and $f = 1.0$. In addition, as β increases, the minimum value of the time-averaged input power ($\bar{p}_{in}$) decreases while the peak value increases, because of the increase in natural frequency. It is worth noting that as β becomes larger, a greater disparity between the analytical and numerical results becomes apparent. The reason for the difference is the comparatively few

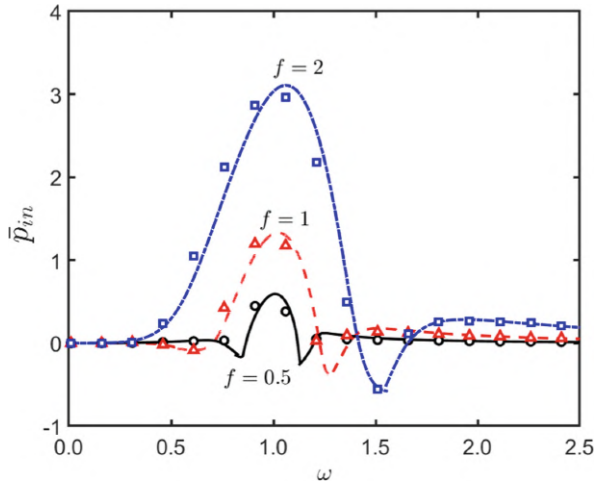


Fig. 2. The influence of the force amplitude f on the input power is examined for a system with $\beta = 1.0$ and $\alpha = 0.5$. The results obtained by the harmonic balance (HB) approximation: The dash-doted line for $f = 2$, the dashed line for $f = 1$, and the solid line for $f = 0.5$. The results obtained by numerical simulation: squares for $f = 2$, triangles for $f = 1$ and circles for $f = 0.5$.

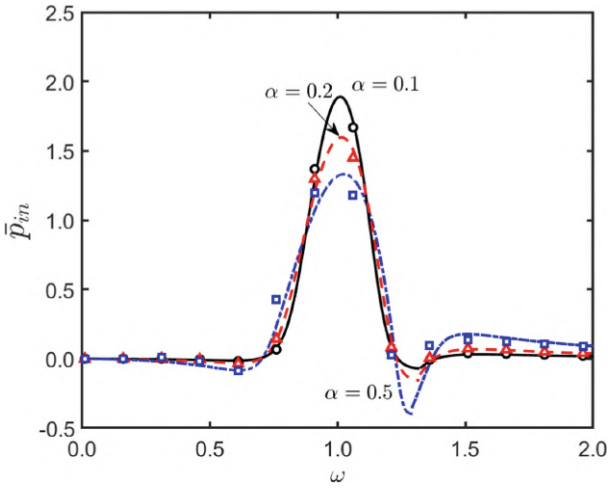


Fig. 3. The influence of the force amplitude α on the input power is examined for a system with $f = 1.0$ and $\beta = 1.0$. The results obtained by the harmonic balance (HB) approximation: the dash-doted line for $\alpha = 0.5$, the dashed line for $\alpha = 0.2$, and the solid line for $\alpha = 0.1$. The results obtained by numerical simulation: squares for $\alpha = 0.5$, triangles for $\alpha = 0.2$ and circles for $\alpha = 0.1$.

Fourier terms used in the HB-PFA method. The figure shows that ($\bar{p}_{in}$) can be negative. For physical systems, the negative power by the external force means that the nonlinear term $\alpha(x^2 - 1)\dot{x}$ provides negative energy dissipation, meaning actual energy input. In the case of vortex-induced self-excited vibrations of structures such as bridges, the aerodynamic force (by the vortices) can provide energy input other than energy dissipation. The use of more harmonic terms can reduce the differences between the analytical approximations and numerical integration results. Furthermore, the time-averaged input power demonstrates reduced sensitivity to changes in β within the high-frequency range.

The effects of excitation amplitude f and nonlinear term α on the power inputted are analyzed in Fig. 2 and Fig. 3, respectively. The figures reveal that increasing the excitation amplitude f leads to an increase in the peak of $\bar{p}_{in}$. As the excitation frequency increases beyond 2, the curve becomes flatter. On the other hand, varying the parameter α demonstrates the ability to suppress the peak in the $\bar{p}_{in}$ curve. Near the peak frequency, increasing the nonlinear coefficient increases the system's damping effect. However, the input power becomes larger as α increases in the high-frequency range. That is consistent with the findings in Ref. [16], which is an application of the Van der Pol oscillator in the circuit. Additionally, as the value of α increases, the difference between the numerical simulation and first-order HB approximation also increases, because the natural frequency of the system decreases with the nonlinear parameter α.

4 Conclusions

The paper presented a power flow analysis method based on the harmonic balancing method for nonlinear dynamical systems. The study aimed to demonstrate the analysis approach and the influence of nonlinearity in damping on power flows. The approach was then demonstrated via a PFA of the forced Van der Pol oscillator. Both numerical simulations and analytical approximations were utilized to investigate the system.

In the research, the effects of parameters on the vibration energy input were examined. It was observed that at large displacements, the so-called dissipated power by the nonlinear term $\alpha(x^2 - 1)\dot{x}$ is positive. In comparison at small displacements, the related power can be either positive or negative. Furthermore, the behavior of the forced system depends on the excitation frequency ω in relation to the corresponding linear system's natural frequency. When ω is approaching the natural frequency, the system exhibits periodic motion. In contrast, when ω is close to the low or high value range, the system may exhibit quasi-periodic motion.

The harmonic balancing approach was employed to analytically define the time-averaged input and dissipated powers. The findings highlight that, within a range of excitation frequencies, the averaging input energy of the external force can be negative. Notably, an analytical expression of the input power showed that its value was directly related to the damping parameter α and the excitation amplitude.

References

1. Goyder, H.G.D., White, R.G.: Vibrational power flow from machines into built-up structures, part I: introduction and approximate analyses of beam and plate-like foundations. J. Sound Vib. **68**, 59–75 (1980). https://doi.org/10.1016/0022-460X(80)90452-6
2. Yang, J.: Power flow analsyis of nonlinear dynamical systems. Ph.D. thesis, University of Southampton (2013)
3. Yang, J., Xiong, Y.P., Xing, J.T.: Nonlinear power flow analysis of the Duffing oscillator. Mech. Syst. Signal Process. **45**, 563–578 (2014). https://doi.org/10.1016/j.ymssp.2013.11.004
4. Yang, J., Shi, B., Rudd, C.: On vibration transmission between interactive oscillators with nonlinear coupling interface. Int. J. Mech. Sci. **137**, 238–251 (2018)
5. Dai, W., Yang, J., Shi, B.: Vibration transmission and power flow in impact oscillators with linear and nonlinear constraints. Int. J. Mech. Sci. **168**, 105234 (2020)
6. Dai, W., Yang, J.: Vibration transmission and energy flow of impact oscillators with nonlinear motion constraints created by diamond-shaped linkage mechanism. Int. J. Mech. Sci. **194**, 106212 (2021)
7. Dai, W., Yang, J., Wiercigroch, M.: Vibration energy flow transmission in systems with Coulomb friction. Int. J. Mech. Sci. **214**, 106932 (2022)
8. Shi, B., Yang, J., Wiercigroch, M.: Vibrational energy transfer in coupled mechanical systems with nonlinear joints. Int. J. Mech. Sci. **260**, 108612 (2023)
9. Yang, J., Xiong, Y.P., Xing, J.T.: Dynamics and power flow behaviour of a nonlinear vibration isolation system with a negative stiffness mechanism. J. Sound Vib. **332**(1), 167–183 (2013)
10. Yang, J., Xiong, Y., Xing, J.T.: Power flow behaviour and dynamic performance of a nonlinear vibration absorber coupled to a nonlinear oscillator. Nonlinear Dyn. **80**, 1063–1079 (2015)
11. Zhu, C., Yang, J., Rudd, C.: Vibration transmission and power flow of laminated composite plates with inerter-based suppression configurations. Int. J. Mech. Sci. **190**, 106012 (2021)
12. von Groll, G., Ewins, D.: The harmonic balance method with arc-length continuation in rotor/stator contact problems. J. Sound Vib. **2001**(241), 223–233 (2001)
13. Urasaki, S., Yabuno, H.: Identification method for backbone curve of cantilever beam using van der Pol-type self-excited oscillation. Nonlinear Dyn. **103**(4), 3429–3442 (2021)
14. Yabuno, H., Kunitho, Y., Kashimura, T.: Analysis of the van der Pol system with Coulomb friction using the method of multiple scales. J. Vibr. Acoust. **130**(4), 041008 (2008)
15. Yabuno, H., Kaneko, H., Kuroda, M., Kobayashi, T.: Van der Pol type self-excited micro-cantilever probe of atomic force microscopy. Nonlinear Dyn. **54**, 137–149 (2008)
16. Kpomahou, Y.J.F., Miwadinou, C.H., Agbokpanzo, R.G., Hinvi, L.A.: Nonlinear dynamics of a RLC series circuit modeled by a generalized Van der Pol oscillator. Int. J. Nonlinear Sci. Numer. Simul. **22**(3–4), 479–494 (2021)

Stochastic Dynamics

Asymptotic Approximation of the Maximal Lyapunov Exponent of Moore-Greitzer PDE Model with Small Multiplicative Noise Close to Stall Bifurcation

Yiming Meng[1(✉)], N. Sri Namachchivaya[1], and Nicolas Perkowski[2]

[1] Department of Applied Mathematics, Waterloo University, Waterloo, ON, Canada
{yiming.meng,nsnamachchivaya}@uwaterloo.ca
[2] Institut für Mathematik, Freie Universität Berlin, Berlin, Germany
perkowski@math.fu-berlin.de

Abstract. The Moore-Greitzer partial differential equation (PDE) is a commonly used parameter-dependent mathematical model for capturing flow disturbances in axial-flow jet engine compressors. For a certain compressor configuration, as the throttle coefficient (the parameter) decreases, inlet flow disturbances dominate, and stall occurs. Instabilities place fundamental limits on jet engine operating range and thus limit the design space. In the presence of stochastic multiplicative noise, in contrast to the original deterministic PDE model, almost-sure asymptotic stability is not well understood. To control and mitigate dynamic instabilities in modern jet engine compressors, the detection of the critical point where the system loses stability becomes essential. In this paper, we take advantage of recent advances in utilizing multiscale analysis to approximate the maximal Lyapunov exponent, which can readily serve as an indicator of stability change, and apply the results to study the impact of small multiplicative noise on compression system instabilities. A comprehensive review of related works is also provided.

Keywords: Moore-Greitzer PDE model · multiplicative noise · SPDE · almost-sure asymptotic stability · Hopf bifurcation · maximal Lyapunov exponent · multiscale approximation

1 Introduction

Jet engine compressors may experience instabilities in the vicinity of their optimal operating range, leading to performance degradation and potential safety risks. Among these instabilities, rotating stall is particularly concerning. Rotating stall is defined by the circulation of a gas wave around the compressor's annulus and occurs when a non-axisymmetric flow disturbance develops around the rotor's annulus [18]. This phenomenon leads to a notable reduction in compressor efficiency. The research is motivated to develop theories and methods for

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 111–124, 2025.
https://doi.org/10.1007/978-3-031-72794-8_8

detecting, controlling, and mitigating stall in modern jet engines, which will be essential for increasing compressor efficiency, preventing damage or failure, and extending the lifespan of engine components.

As a crucial preliminary step before implementing control measures, the identification of bifurcations is necessary to gain an understanding of the system's properties and the optimal operational range of the engine. To achieve this quantitative objective, we employ the commonly used Moore-Greitzer nonlinear partial differential equation (PDE) model to investigate the evolution, particularly the long-term behavior, of the turbulent flow determining the stall phenomena. This paper addresses stability analysis based on the Moore-Greitzer PDE model with stochastic uncertainties. Such a modified model naturally captures the average phenomena and satisfies the prior belief of the space-time turbulence.

We conduct a comprehensive stochastic multi-scale analysis in the vicinity of the quasi-static stall bifurcation point of the unperturbed system. From this analysis, we derive an asymptotic approximation for the maximal Lyapunov exponent. This approximation can be used as an indicator of stability changes and, consequently, to determine the robust operational range of the jet engine compressor. This robust result is essential for improving feedback controls to manage compressor instabilities, ultimately contributing to lighter and more efficient jet engine designs.

The rest of this paper is organized as follows. In Sect. 2, we review the unperturbed Moore-Greitzer Model and define the functional space where the solutions exist. In Sect. 3, we investigate how stochastic noise can be integrated to the original model and utilize a generalized solution space to ensure a more robust analysis. We also show that this perturbed model falls in a general category of the frame work developed in [11], from which the analytical result can be utilized. In Sect. 4, we review the main theoretical results from [11], allowing us to leverage the analytical results from that framework. In particular, we provide a review of the derivation of the Furstenberg–Khasminskii formula for the maximal Lyapunov exponent. We apply the general results to a simplified stochastic Moore-Greitzer PDE model with multiplicative noise in Sect. 5. A complete derivation is provided. The conclusions follow in Sect. 6.

2 Unperturbed Moore-Greitzer Model

2.1 Modelling Introduction

A turbo-jet engine consists of three main parts: the axial flow compressor where air is compressed, the plenum where air undergoes combustion and expands rapidly, and the turbine through which the air is released [13]. The compressor is composed of an entrance duct, an inlet guide vane, multiple stages of stator-rotor pairs, and an exit duct leading towards the plenum (see Fig. 1). A stator consists of static blades, while a rotor comprises revolving blades.

At the rear of the plenum, there is a throttle that controls the average mass flow through the system. The instability arises from two primary factors: (stall) non-uniform disturbances originating upstream lead to a locally increased angle

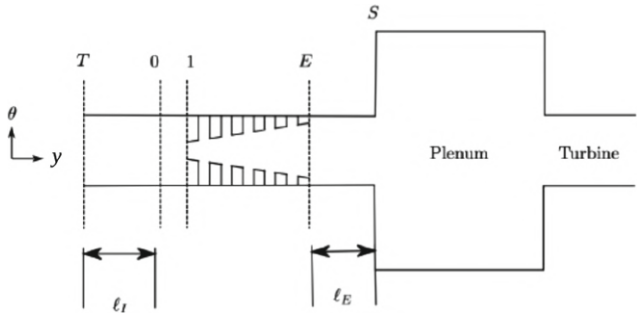


Fig. 1. Compression system.

of attack and move through the blade row without mitigation; (surge) there is a continuous oscillation of the average flow and pressure, which results in the formation of standing waves [6]. The deterministic Moore-Greitzer model captures the dynamic evolution of the above states, and is given as [19]:

$$\frac{\partial}{\partial t}\begin{bmatrix}\mathrm{g}\\ \Phi\\ \Psi\end{bmatrix} = \begin{bmatrix}\mathrm{K}^{-1}\left(\frac{\nu}{2}\frac{\partial^2}{\partial\theta^2} - \frac{1}{2}\frac{\partial}{\partial\theta}\right) & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\mathrm{g}\\ \Phi\\ \Psi\end{bmatrix} + \begin{bmatrix}\mathfrak{a}\,\mathrm{K}^{-1}(\psi_c(\Phi+\mathrm{g}) - \overline{\psi}_c(\Phi,\mathrm{g}))\\ \frac{1}{\mathfrak{l}_c}(\overline{\psi}_c(\Phi,\mathrm{g}) - \Psi)\\ \frac{1}{4\mathfrak{l}_c\mathfrak{B}^2}(\Phi - \gamma\sqrt{\Psi})\end{bmatrix}, \tag{1}$$

where θ is the angular coordinate inside the jet engine, which is assumed to be cylindrical; the states $[\mathrm{g}(t), \Phi(t), \Psi(t)]^T$ belongs to a product space $\mathscr{H} := \mathcal{H} \times \mathbb{R}^2$ with $\mathcal{H} := \mathrm{L}^2([0, 2\pi])$ as an infinite-dimensional separable Hilbert space. The physical meaning of the states is explained as follows: $\mathrm{g}(t, \theta)$ represents the velocity of upstream disturbance along the axial direction at the duct entrance, $\Phi(t)$ is the averaged mean flow rate, $\Psi(t)$ is the averaged pressure. To fulfill the physical meaning of g, we also require that $\mathrm{g}(t, 0) = \mathrm{g}(t, 2\pi)$, $\mathrm{g}_\theta(t, 0) = \mathrm{g}_\theta(t, 2\pi)$ and $\int_0^{2\pi} \mathrm{g}(t, \theta)d\theta = 0$. Thus, we can write g as a Fourier expansion $\mathrm{g}(t, \theta) = \sum_{k\in\mathbb{Z}_0} \mathrm{g}_k(t)e^{ik\theta}$. The operator K is defined as a Fourier multiplier

$$\mathrm{K}(\mathrm{g}) = \sum_{k\in\mathbb{Z}_0}\left(1 + \frac{\mathfrak{a}\mathfrak{m}}{|k|}\right)\mathrm{g}_k(t)e^{ik\theta}, \tag{2}$$

where $\mathfrak{a}$ (that also appears in (1)) is the internal compressor lag and $\mathfrak{m}$ is the duct parameter. The compressor characteristic ψ_c is given in a cubic form $\psi_c(\Phi) = \psi_{c_0} + \iota\left[1 + \frac{3}{2}\left(\frac{\Phi}{M} - 1\right) - \frac{1}{2}\left(\frac{\Phi}{M} - 1\right)^3\right]$, where ψ_{c_0}, ι and M are real-valued parameters that are defined by the compressor configuration. We also define the quantity $\overline{\psi}_c(\Phi, \mathrm{g}) := \frac{1}{2\pi}\int_0^{2\pi} \psi_c(\Phi + \mathrm{g}(\theta))d\theta$.

In the nonlinear term of (1), one can observe the coupling of Φ and g through the compressor characteristic function $\psi_c(\Phi + \mathrm{g})$, as well as the average $\overline{\psi}_c(\Phi, \mathrm{g})$ along the θ direction. The explicit form of these couplings can be found in [12].

The meaning of the other parameters in (1) is as follows: $\mathfrak{l}_c > 0$ is the compressor length, $\mathfrak{B} > 0$ is the plenum-to-compressor volume ratio, $\nu > 0$ is

the viscous coefficient. The parameter $\gamma \in \mathbb{R}$ represents the throttle coefficient, the decrease of which will cause the stability change.

We determine the type of Hopf bifurcation based on the sign of the indicator

$$\Delta := \frac{\psi_{c_0} + \iota\left[1 + \frac{3}{2}\sqrt{1-\frac{\nu M}{3\mathfrak{a}\iota}} - \frac{1}{2}\left(\sqrt{1-\frac{\nu M}{3\mathfrak{a}\iota}}\right)^3\right]}{M\left(1+\sqrt{1-\frac{\nu M}{3\mathfrak{a}\iota}}\right)} - \frac{\mathfrak{a}}{4\mathfrak{B}^2\nu},$$

which is solely influenced by the compressor geometry [19]. More specifically, surge bifurcation happens in $\mathbb{R}^2$ when $\Delta > 0$, stall bifurcation happens in $\mathcal{H}$ when $\Delta < 0$. When $\Delta = 0$, then the bifurcations occur in both of the $\mathcal{H}$ and $\mathbb{R}^2$ subspaces.

Assumption 1. *In this paper, we only focus on the analysis for compressors whose configuration parameters make* $\Delta < 0$.

2.2 Abstract Form and Solution Space

For the consideration of bifurcation analysis and control, we can write (1) as an abstract evolution equation $\partial_t u = Au + f_\gamma(u)$, where $u = [\mathrm{g}, \Phi, \Psi]^T \in \mathscr{H}$, as well as

$$A = \begin{bmatrix} \mathrm{K}^{-1}\left(\frac{\nu}{2}\frac{\partial^2}{\partial\theta^2} - \frac{1}{2}\frac{\partial}{\partial\theta}\right) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } f_\gamma(u) = \begin{bmatrix} \mathfrak{a}\,\mathrm{K}^{-1}(\psi_c(\Phi+\mathrm{g}) - \overline{\psi}_c(\Phi,\mathrm{g})) \\ \frac{1}{\mathfrak{l}_c}(\overline{\psi}_c(\Phi,\mathrm{g}) - \Psi) \\ \frac{1}{4\mathfrak{l}_c\mathfrak{B}^2}(\Phi - \gamma\sqrt{\Psi}) \end{bmatrix}.$$

We consider a fixed point of the form $u_e(\gamma) = [0, \Phi_e(\gamma), \Psi_e(\gamma)]^T$ at which the vector field $f_\gamma(u_e(\gamma)) = \mathbf{0}$ for each γ. In particular $(\Phi_e(\gamma), \Psi_e(\gamma))$ is determined by the intersection of the compressor characteristic $\Psi = \psi_c(\Phi)$ and the throttle characteristic $\Phi = \gamma\sqrt{\Psi}$ (see [19, Figure 4] for details). Linearization about $u_e(\gamma)$ for each γ results in the following linear operator $\mathcal{A}(\gamma) := A + Df_{u_e}(\gamma)$, where $Df_{u_e}(\gamma)$ is the Fréchet derivative at $u_e(\gamma)$.

As proved in [19], the linear operator $\mathcal{A}(\gamma)$ exhibits a discrete spectrum and other nice spectral properties. Particularly, it demonstrates a decoupling effect of the eigenspaces $\mathcal{H}$ and $\mathbb{R}^2$, which makes the linearized flow of g and (Φ, Ψ) invariant respectively under the semigroups $e^{\mathcal{A}|_{\mathcal{H}}(\gamma)t}$ and $e^{\mathcal{A}|_{\mathbb{R}^2}(\gamma)t}$. This fact can also be intuitively explained by the block representation of the matrix operator

$$\mathcal{A}(\gamma) = \begin{bmatrix} \left(\frac{\nu}{2}\frac{\partial^2}{\partial\theta^2} - \frac{1}{2}\frac{\partial}{\partial\theta} + \mathfrak{a}(\psi_c'(\Phi_e(\gamma)) - \overline{\psi_c}')\right)\mathrm{K}^{-1} & 0 & 0 \\ 0 & \frac{1}{\mathfrak{l}_c}\overline{\psi_c}' & -\frac{1}{\mathfrak{l}_c} \\ 0 & \frac{1}{4\mathfrak{B}^2\mathfrak{l}_c} & \frac{1}{4\mathfrak{B}^2\mathfrak{l}_c}\frac{\gamma}{2\sqrt{\Psi_e(\gamma)}} \end{bmatrix}. \quad (3)$$

Based on Assumption 1, the subspace $\mathcal{H}$ is decoupled from $\mathbb{R}^2$ when the state variable in $\mathbb{R}^2$ quickly reaches the steady state. This allows us to consider the evolution, and hence the bifurcation within $\mathcal{H}$ separately. To make the analysis

within $\mathcal{H}$ more precise, we work on the following refined subspace. Note that the solution of $\mathrm{g}(t)$ lies in an infinite-dimensional Hilbert space $\mathcal{H} := \{f \in \mathrm{L}^2 : \int_0^{2\pi} f(\theta)d\theta = 0\}$ (with periodicity) equipped with the inner product $\langle \cdot, \cdot \rangle := \langle \cdot, \mathrm{K}\cdot \rangle$ as well as the induced norm $\|\cdot\|$. In general, due to the spatial periodicity and the zero-average property of $\mathrm{g}(t)$, we can expect the solution to be at least in a Sobolev space

$$\mathbf{H}^2_{\mathrm{per}} := \left\{ \mathrm{g} \in \mathbf{H}^2 : \mathrm{g}(0) = \mathrm{g}(2\pi),\ \mathrm{g}_\theta(0) = \mathrm{g}_\theta(2\pi), \int_0^{2\pi} \mathrm{g}(\theta)d\theta = 0 \right\},$$

where $\mathbf{H}^2 \subseteq \mathcal{H}$ is the Sobolev Hilbert space of L^2 functions with weak derivatives of order up to 2, and $\mathbf{H}^2_{\mathrm{per}} \subseteq \mathbf{H}^2$ is the subspace of periodic functions.

Working on the above mentioned refined Sobolev space, active stall control strategies have been introduced to stabilize g while also mitigating the potential decline in the compressor's transient performance efficiency caused by recycling flow. In this regard, Banaszuk et al. in [2] designed a back stepping control, Birnir and Hauksson in [4] went one step further to construct a control strategy using the knowledge of asymptotic dynamics. Wen et al. in [18] studied the local feedback stabilization of Hopf bifurcation for infinite-dimensional nonlinear systems where one pair of the modes are at the risk of being unstable. The above investigations all require the knowledge of the critical stall bifurcation point γ_c.

In practice, to incorporate noise into the system and develop a robust active stall control, an investigation of the impact of multiplicative noise is essential to determine the operating point at which the stochastic system loses stability.

3 Stochastically Perturbed Moore-Greitzer PDE Model and Almost-Sure Asymptotic Stability

The viscous Moore-Greitzer equation is based on the Navier-Stokes equation and a non-rigorous stochastic homogenization theory of fluids [8]. In other words, the quantity g captures the mean of the actual stochastically excited turbulent flow $g(t,\theta) := \mathrm{g}(t,\theta) + \varepsilon \dot{W}(t,\theta)$, where $W(t,\theta)$ is a space-time noise with mean $\mathbf{0}$ and $\dot{W}(t,\theta) := \frac{\partial W}{\partial t}(t,\theta)$ denotes its (distributional) derivative of time; $\varepsilon > 0$ represents the small intensity of the noise. Based on Assumption 1 and the above modelling perspective, we can verify that g satisfies the following SPDE

$$dg(t) = Ag(t)dt + f_\gamma(g(t), \Phi_e(\gamma), \Psi_e(\gamma))dt + \varepsilon \frac{3\iota}{M} g(t)dW_t + \varepsilon \frac{3\iota}{M} \Phi_e(\gamma)dW_t \quad (4)$$

for all $\gamma \in \mathbb{R}$, where the multiplicative term $g(t)dW_t$ and the additive term[1] $\Phi_e(\gamma)dW_t$ are generated through the nonlinear term f_γ. It is clear that $\mathrm{g} = \mathbf{E}[g]$ satisfies the original equation. Note that we have investigated the impact of the additive noise in [12]. As a vital step towards understanding the combined effects of the additional stochastic terms in (4), this paper examines the influence of

[1] Recall that $\phi_e(\gamma)$ at any γ is a steady point.

the multiplicative noise on stability. We suppose that $\epsilon := \varepsilon \frac{3\iota}{M} \ll 1$ and consider the system

$$dg(t) = Ag(t)dt + f_\gamma(g(t), \Phi_e(\gamma), \Psi_e(\gamma))dt + \epsilon g(t)dW_t. \tag{5}$$

3.1 Extended Solution Space and Modeling of Disturbances

A vast body of literature exists on how external forces impact the regularity, especially the space-time differentiability, of solutions to semilinear PDEs. Given that the sample paths of space-time noise typically present as "rough", we consider fractional Sobolev spaces relative to $\mathbf{H}^2_{\text{per}}$. A concept closely related to this is the fractional power spaces (see Definition 1 below). They demonstrates the spatial weak differentiabilities from the spectral perspective of the linearized operator $\mathcal{A}(\gamma)$. A thorough investigation of the connection between the above mentioned fractional spaces are provided in [5,7,12]. To study the long-term behavior of the system, we prefer using the fractional power space to determine the solution spaces, due to the nice spectral spectral properties of $\mathcal{A}(\gamma)$.

Definition 1 (Fractional Power Space). *For $\alpha \in \mathbb{R}$, given the linear operator $\mathcal{A}(\gamma)$, define the interpolation fractional power (Hilbert) space [15] $\mathcal{H}_\alpha := \text{dom}(\mathcal{A}^\alpha(\gamma))$ endowed with inner product $\langle u, v\rangle_\alpha = \langle \mathcal{A}^\alpha u, \mathcal{A}^\alpha v\rangle$ and corresponding induced norm $\|\cdot\|_\alpha := \|\mathcal{A}^\alpha \cdot\|$. Further more, we denote the dual space of $\mathcal{H}_\alpha$ by $\mathcal{H}_{-\alpha}$ w.r.t. the inner product in $\mathcal{H}$.*

Remark 1. *For all $\alpha \geq \beta$, we have the inclusion $\mathcal{H}_\alpha \subseteq \mathcal{H}_\beta$.* □

Building on the aforementioned regularity concept, we model the space-time noise W as a (generalized) Q-Winner process. This is a generalized Gaussian process in an infinite-dimensional space, characterized by the covariance operator Q and the appropriate fractional power space wherein the Gaussian measure can be properly defined. We kindly refer readers keen on the rigorous mathematical formulation to [5] for comprehensive details and to [10,12] for a concise summary.

For the Moore-Greitzer model, based on the prior belief that the noise periodically excites the g variable and hence contributes to (5), we construct W as

$$W(t) = \sum_{k\in\mathbb{Z}^+} \sqrt{q_k}(\beta_k(t) + i\beta_{-k}(t))e^{ik\theta} + \sum_{k\in\mathbb{Z}^-} \sqrt{q_k}(\beta_{-k}(t) - i\beta_k(t))e^{ik\theta},$$

where $q_k = |k|^{-(4\alpha+1)}$, β_k are i.i.d. Brownian motions. Then the process $W(t)$ belongs to $\mathcal{H}_\alpha$ a.s.. Note that,

1. For $\alpha > -\frac{1}{4}$, q_k decays as k increases, then Q is a trace class operator (i.e. $\text{tr}(Q) = \sum_{k\in\mathbb{Z}_0} q_k < \infty$ [5]) in $\mathcal{H}_\alpha \subseteq \mathcal{H}$. In this case, $\dot{W}(t)$ is a noise that is white in time and colored in space.
2. For $\alpha > 1$, sample paths of W demonstrate sufficient spatial smoothness relative to the solution g of the original equation.

3. For $\alpha \leq -\frac{1}{4}$, W only exists in a functional space $\mathcal{H}_\alpha \supseteq \mathcal{H}$, and renders a distributional (functional) space-time noise.

To make the analysis more practical, we consider $\alpha \in (0, 1]$. This setting also makes the $\mathcal{H}_\alpha$ space an algebra, which implies that the multiplication $g(t)dW_t$ is also within the $\mathcal{H}_\alpha$ space given that each individual quantity belongs to $\mathcal{H}_\alpha$. More details about the multiplications within fractional power or Sobolev spaces can be found in [1,3].

Remark 2. *Let* $\mathcal{W} = \sum_{k\in\mathbb{Z}^+} (\beta_k + i\beta_{-k})e^{ik\theta} + \sum_{k\in\mathbb{Z}^-} (\beta_{-k} - i\beta_k)e^{ik\theta}$, *and let* $\Sigma(u)$ *be such that*

$$\Sigma(u)h = \sum_{k\in\mathbb{Z}_0} \langle e^{-ik\theta}, uQ^{1/2}h\rangle e^{ik\theta} \tag{6}$$

where Q *is a trace-class operator with eigenvalues* $q_k = |k|^{-(4\alpha+1)}$ *for* $\alpha > 0$. *For each* $\alpha \in (0, 1]$, *it can be verified that* $\Sigma(u)$ *is a Hilbert-Schmidt operator from* $\mathcal{H}_{-1/4}$ *to* $\mathcal{H}_\alpha$, *i.e.,* $\Sigma : \mathcal{H}_\alpha \to \mathcal{L}_2(\mathcal{H}_{-1/4}, \mathcal{H}_\alpha)$, *where* $\mathcal{L}_2(E, K)$ *denotes the set of all Hilbert-Schmidt operators from* E *to* K *for any separable Hilbert spaces* E *and* K. *We can further verify that* $\Sigma(u)$ *is linear (in* u*), Fréchet-differentiable, and is invertible for each* $u \in \mathcal{H}_\alpha \setminus \{\mathbf{0}\}$.

Upon closer observation, it is evident that $g(t)dW_t = \Sigma(g(t))d\mathcal{W}_t$. *We prefer to use the form* $\Sigma(g(t))d\mathcal{W}_t$ *as it offers convenient and robust operator-wise calculation. Unless otherwise specified, we will use this form of expression for the rest of this paper.* □

3.2 The Study of Almost-Sure Asymptotic Stability

Despite the presence of small multiplicative noise in (4), it is clear that $g = \mathbf{0}$ remains a trivial steady state for systems operating at all γ. To observe the stochastic bifurcation, it is imperative to determine the operational point γ where the stochastically perturbed g loses stability around the trivial solution $\mathbf{0}$.

For parameter-dependent semilinear equations operating at non-critical points, it is important to note that the almost-sure convergence/divergence rate with respect to $\mathbf{0}$ is exponential and is primarily governed by the linearized operator. Analogous to the finite-dimensional case, this almost-sure exponential growth/decay rate is determined by the maximal Lyapunov exponent λ^γ. The almost-sure asymptotic stability of $g = \mathbf{0}$ is then determined by the sign of λ^γ. Specifically, as γ varies in a way that λ^γ changes sign from negative to positive, the trivial solution loses its almost-sure asymptotic stability.

Since the intensity ϵ of the noise is assumed to be small, it is reasonable to only look at the small neighborhood of the quasi-static stall bifurcation point γ_c (of the unperturbed system) and search for the new critical point of (5). We then investigate the almost-sure asymptotic stability through the linearization of the (5) around the equilibrium point for $\gamma = \gamma_c + \epsilon^2\mathfrak{q}$ with some $\mathfrak{q} \in \mathbb{R}$. Ignoring the high order terms, we obtain the following abstract form of linear parabolic

stochastic PDE (recall Remark 2 for the expression of the noise term),

$$dg(t) = [\mathcal{A}(\gamma_c) + \varepsilon^2 \mathfrak{q}\mathcal{A}'(\gamma_c)]g(t)dt + \epsilon\Sigma(g(t)) \cdot d\mathcal{W}_t, \quad \left(' = \frac{d}{d\gamma}\right). \tag{7}$$

By Assumption 1, the linearized operator is such that

$$\mathcal{A}(\gamma_c) = [A + Df_{u_e}(\gamma_c)]|_{\mathcal{H}} = \left(\frac{\nu}{2}\frac{\partial^2}{\partial\theta^2} - \frac{1}{2}\frac{\partial}{\partial\theta} + \mathfrak{a}\psi_c'\right)\mathrm{K}^{-1},$$

and $\mathcal{A}'(\gamma_c) = (a\psi_c''\Phi_c)\,\mathrm{K}^{-1}$, where

$$\psi_c' := \psi_c'(\Phi_e(\gamma_c)) = \frac{3\iota}{2M}\left[1 - \left(\frac{\Phi_e(\gamma_c)}{M} - 1\right)^2\right],$$

$$\psi_c'' := \psi_c''(\Phi_e(\gamma_c)) = -\frac{3\iota}{M^2}\left(\frac{\Phi_e(\gamma_c)}{M} - 1\right),$$

and $\Phi_c = \Phi_e'(\gamma_c)$. The point spectrum $\{\rho_k(\gamma)\}_{k\in\mathbb{Z}_0}$ of $\mathcal{A}(\gamma)$ is complex-valued and analytic, where $\rho_k(\gamma) := a_k(\gamma) + ib_k(\gamma) \in \mathbb{C}$ and $\rho_{-k}(\gamma) = \overline{\rho}_k(\gamma)$ for all $k \in \mathbb{Z}_0$. The corresponding eigenvectors $\{e^{ik\theta}\}_{k\in\mathbb{Z}_0}$ form a complete orthonormal basis of $\mathcal{H}$. For $\gamma = \gamma_c + \epsilon\mathfrak{q}$, it can be verified that

$$\rho_k = \frac{\mathfrak{a}|k|}{|k| + \mathfrak{a}\mathfrak{m}}\left(\psi_c' - \frac{\nu k^2}{2\mathfrak{a}} + \epsilon^2\mathfrak{q}\psi_c''\Phi_c\right) + i\frac{|k|}{2\mathfrak{a}}, \quad \forall k \in \mathbb{Z}_0.$$

Particularly, by the settings of jet engine parameters, we have that $\rho_{\pm1} = \epsilon^2\frac{\mathfrak{a}}{1+\mathfrak{a}\mathfrak{m}}(\mathfrak{q}\psi_c''\Phi_c) + i\frac{1}{2\mathfrak{a}}$, $\alpha_k(\gamma) < 0$ for all $k \in \mathbb{Z}_0 \setminus \{\pm1\}$, and $a_1(\gamma) > a_2(\gamma) \geq \cdots \geq a_k(\gamma) \geq \ldots$. Due to this natural spectral gap, we define the following projection operators.

Definition 2. *The critical projection operator is defined as* $P_c(\cdot) := \langle e^{-i\theta}, \cdot\,\rangle e^{i\theta} + \langle e^{i\theta}, \cdot\,\rangle e^{-i\theta}$. *The stable projection operator is* $P_s = I - P_c$. *For simplicity, we introduce shorthand notation* $\mathcal{A}_c(\gamma) := P_c\mathcal{A}(\gamma)$. *We define* $\mathcal{A}_s(\gamma)$, Σ_c, Σ_s, $\mathcal{H}_c$, $\mathcal{H}_s$, $\mathcal{H}_{\alpha,c}$ *and* $\mathcal{H}_{\alpha,s}$ *in a similar way.*

For simplicity, we also introduce the following notations.

Definition 3 (Other notations for $\mathcal{A}(\gamma)$). *To this end, we use*

(1) $\mathbb{Z}_c := \{\pm1\}$, $\mathbb{Z}_s := \mathbb{Z}_0 \setminus \{\pm1\}$.
(2) $\mathcal{A}_c^{\mathfrak{c}} := \mathcal{A}_c(\gamma_c)$, $\mathcal{A}_c^{\mathfrak{q}} := \mathfrak{q}\mathcal{A}_c'(\gamma_c)$; *the associated eigenvalues of* $\mathcal{A}_c^{\mathfrak{c}}$ *and* $\mathcal{A}_c^{\mathfrak{q}}$ *are respectively denoted by* $\rho_c^{\mathfrak{c}} = ib_c^{\mathfrak{c}} := ib_1(\gamma_c)$, $\overline{\rho}_c^{\mathfrak{c}} = -ib_c^{\mathfrak{c}} := -ib_1(\gamma_c)$, *and* $\rho_c^{\mathfrak{q}} = a_c^{\mathfrak{q}} + ib_c^{\mathfrak{q}} := \mathfrak{q}(a_1'(\gamma_c) + ib_1'(\gamma_c))$.
(3) $\mathcal{A}_s := \mathcal{A}_s(\gamma)$ *for all* γ.

Remark 3. *Based on the formula of* ρ_k, *we have that* $a_c^{\mathfrak{q}} = \frac{\mathfrak{a}\mathfrak{q}}{1+\mathfrak{a}\mathfrak{m}}\psi_c''\Phi_c$, $b_c^{\mathfrak{c}} = \frac{1}{2\mathfrak{a}}$, *and* $b_c^{\mathfrak{q}} = 0$.

Given the above properties of (7), we can leverage the results in [11] to estimate the maximal Lyapunov exponent in determining the almost-sure asymptotic stability. Before delving further, we review the methodology and its key results.

4 Review of Methodology and Main Results

Fundamental stability analysis and the formula for the maximal Lyapunov exponent have been well-studied for finite-dimensional stochastic systems [14,16]. As for infinite-dimensional case, the almost-sure stability of scalar stochastic delay differential equation (driven by real noise) has been studied in [17]. The formula was obtained by an asymptotic approximation. We adopt a similar strategy for the asymptotic approximation of the invariant measure as well as the maximal Lyapunov exponent as in [17]. In contrast to the real noise excitation, the generator of SPDEs driven by multiplicative Gaussian-type noise only exists under certain assumptions. The stability analysis is conducted by assuming necessary conditions, under which the invariant measure is guaranteed to exist. The ergodicity of the invariant measure, however, relies on the property related to Σ. The analytical approximation of Furstenberg-Khasminskii formula for the maximal Lyapunov exponent can be further compared with industrial experiment data. The technical details will be summarized in this section, where the notation in [11] has been recast to make it consistent with the contexts in this paper.

4.1 Multiscale Decomposition and the Furstenberg–Khasminskii Formula for the Maximal Lyapunov Exponent

The work in [11] is based on the equation of the form (7), but with more generalized space setups as follows.

1. The orthonormal basis of $\mathcal{H}$ is some $\{e_k\}_{k\in\mathbb{Z}_0}$ such that $\mathcal{A}(\gamma)e_k = \rho_k(\gamma)e_k$, $\langle e_{-k}, e_k\rangle = 1$ for all $k \in \mathbb{Z}_0$, and $\langle e_i, e_j\rangle = 0$ for all $i + j \neq 0$.
2. $\mathcal{V}$ is a separable Hilbert space with orthonormal basis $\{\mathfrak{z}_k\}_{k\in\mathbb{Z}_0}$ and $\mathcal{W}$ is a $\mathcal{V}$-valued cylindrical Wiener process.
3. For each $\alpha \in (0, 1]$ and for all $u \in \mathcal{H}_\alpha$, $\Sigma : \mathcal{H}_\alpha \to \mathcal{L}_2(\mathcal{V}, \mathcal{H}_\alpha)$ is a Hilbert-Schmidt operator and satisfies the conditions in Remark 2.

Taking advantage of the natural separation of temporal scales in the neighborhood of the trivial solution for $\gamma = \gamma_c + \epsilon^2\mathfrak{q}$ with some unfolding parameter $\mathfrak{q} \in \mathbb{R}$, we introduce the notion of solution under re-scaled time: $z(t) = \langle e_1^*, g(\epsilon^{-2}t)\rangle$ and $\bar{z}(t) = \langle e_{-1}^*, g(\epsilon^{-2}t)\rangle$ as the complex amplitudes of the critical mode, and $y(t) = P_s g(\epsilon^{-2}t)$. Then the solution g of (7) can be written as $g(t) = z(\epsilon^2 t)e_1 + \bar{z}(\epsilon^2 t)e_{-1} + y(\epsilon^2 t)$. We denote the real part and imaginary part of z as $z_1 = \mathrm{Re}(z)$ and $z_2 = \mathrm{Im}(z)$, respectively. Omitting the high-order terms with insignificant effect, the dynamics of z and y are as follows.

$$d\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 0 & -\epsilon^{-2}b_c^{\mathfrak{c}} \\ \epsilon^{-2}b_c^{\mathfrak{c}} & 0 \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} a_c^{\mathfrak{q}} & -b_c^{\mathfrak{q}} \\ b_c^{\mathfrak{q}} & a_c^{\mathfrak{q}} \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} \Sigma_c^R(ze_1 + \bar{z}e_{-1} + y) \\ \Sigma_c^I(ze_1 + \bar{z}e_{-1} + y) \end{bmatrix} d\mathcal{W}_t, \tag{8a}$$

$$dy = -\epsilon^{-2}\mathcal{A}_s y dt + \Sigma_s(ze_1 + \bar{z}e_{-1} + y)d\mathcal{W}_t, \tag{8b}$$

where $\mathcal{W}(t) \sim \epsilon^{-1}\mathcal{W}(\epsilon^{-2}t)$[2] and $\hat{\Sigma}_1(u)w := \langle e_1^*, \Sigma(u)w\rangle$ $\forall u \in \mathcal{H}_\alpha$ and $\forall w \in \mathcal{V}$, $\Sigma_c^R(u) = \frac{\hat{\Sigma}_1(u)+\overline{\hat{\Sigma}_1(u)}}{2}$, and $\Sigma_c^I(u) = \frac{\hat{\Sigma}_1(u)-\overline{\hat{\Sigma}_1(u)}}{2}$.

The maximal Lyapunov exponent is defined as[3] $\lambda^{\mathfrak{q},\epsilon} = \lim_{t\to\infty} \frac{1}{t}\ln|z(t)|$. To derive the Furstenberg-Khasminskii formula, one can perform a direct change of coordinate, i.e., let $\mathfrak{p}(t) = \frac{1}{2}\ln(z_1^2(t) + z_2^2(t))$, $z_1(t) = e^{\mathfrak{p}(t)}\cos(\phi(t))$, $z_2(t) = e^{\mathfrak{p}(t)}\sin(\phi(t))$, and $\eta_t = e^{-\mathfrak{p}(t)}y_t$, where $\phi \in [0, 2\pi]$ is the phase angle in the unit sphere [9] satisfying $z_1 = |z|\cos(\phi)$ and $z_2 = |z|\sin(\phi)$. Then, by Itô's formula,

$$d\mathfrak{p} = a_c^{\mathfrak{q}}dt + \Xi(\phi,\eta)dt + [\Sigma_c^R(\phi,\eta)\cos\phi + \Sigma_c^I(\phi,\eta)\sin\phi]d\mathcal{W}_t, \tag{9a}$$

$$d\phi = (\epsilon^{-2}b_c^{\mathfrak{c}} + b_c^{\mathfrak{q}})dt + \Gamma(\phi,\eta)dt - \Sigma_c^{\phi}(\phi,\eta)d\mathcal{W}_t, \tag{9b}$$

$$d\eta = \epsilon^{-2}\mathcal{A}_s\eta dt + \Sigma_s(\phi,\eta)d\mathcal{W}_t, \tag{9c}$$

where

$$\Sigma_s(\phi,\eta) := \Sigma_s(\cos(\phi)e_1 + \sin(\phi)e_{-1} + \eta),$$
$$\Sigma_c^{\phi}(\phi,\eta) := \Sigma_c^R(\phi,\eta)\sin\phi - \Sigma_c^I(\phi,\eta)\cos\phi,$$

as well as

$$\Xi(\phi,\eta) := -\frac{\cos(2\phi)}{2}\,\mathrm{tr}^{(1)}(\phi,\eta) - \frac{\sin(2\phi)}{2}\,\mathrm{tr}^{(2)}(\phi,\eta)$$

and

$$\Gamma(\phi,\eta) := \frac{\sin(2\phi)}{2}\,\mathrm{tr}^{(1)}(\phi,\eta) - \frac{\cos(2\phi)}{2}\,\mathrm{tr}^{(2)}(\phi,\eta),$$

for $\mathrm{tr}^{(1)} = \mathrm{tr}[\Sigma_c^R(\Sigma_c^R)^* - \Sigma_c^I(\Sigma_c^I)^*]$ and $\mathrm{tr}^{(2)} = \mathrm{tr}[\Sigma_c^R(\Sigma_c^I)^* + \Sigma_c^I(\Sigma_c^R)^*]$.

We also name $\Sigma_c^{\mathfrak{p}}(\phi,\eta) := \Sigma_c^R(\phi,\eta)\cos\phi + \Sigma_c^I(\phi,\eta)\sin\phi$ for future references.

Notice that $\mathfrak{p}(t)$ only depends on $\phi(t)$ and $\eta(t)$. [11, Section 3] has rigorously proved the existence and uniqueness of the invariant measure μ^ϵ for the product process $(\phi(t), \eta(t)) \in [0, 2\pi] \times \mathcal{H}_s$. The top Lyapunov exponent of $g(\varepsilon^{-2}t)$ can be determined by the Furstenberg-Khasminskii formula:

$$\lambda^{\mathfrak{q},\epsilon} = \lim_{t\to\infty}\frac{1}{t}\ln|z(t)| = \int_{[0,2\pi]\times\mathcal{H}_s} \mathcal{Q}(\phi,\eta)\mu^\epsilon(d\phi, d\eta) =: \langle\mathcal{Q}, \mu^\epsilon\rangle, \tag{10}$$

where $\mathcal{Q}(\phi,\eta) := a_c^{\mathfrak{q}} + \Xi(\phi,\eta)$.

Even though η decays exponentially, it is not decoupled from ϕ. We are not able to explicitly solve the invariant measure for $\{\eta(t)\}_{t\geq 0}$ by considering the transitions separately and treat μ^ϵ as a product probability measure. This, however, motivates us to approximate the μ^ϵ with a disintegrated form.

[2] We abuse the notation to avoid redundancy.

[3] Since we have specified $\gamma = \gamma_c + \epsilon^2\mathfrak{q}$, the notation λ^γ is recast as $\lambda^{\mathfrak{q},\epsilon}$.

4.2 Asymptotic Approximation of the Top Lyapunov Exponent

Due to the existence of multi-scale, consider an asymptotic expansion of the infinitesimal generator [5, Theorem 9.25] of (9) $\mathfrak{L}^{\mathfrak{q},\epsilon} = \frac{1}{\epsilon^2}\mathfrak{L}_0 + \mathfrak{L}_1^{\mathfrak{q}}$, where $\mathfrak{L}_0(\cdot) = \left[b_c^{\mathfrak{c}}\frac{\partial}{\partial\phi} + \mathcal{A}_s\eta\frac{\partial}{\partial\eta}\right](\cdot)$ and

$$\mathfrak{L}_1^{\mathfrak{q}}(\cdot) = \left[(b_c^{\mathfrak{q}} + \Gamma(\phi,\eta)\frac{\partial}{\partial\phi}\right](\cdot) + \frac{1}{2}\,\mathrm{tr}\left[\frac{\partial^2(\cdot)}{\partial\eta^2}\Sigma_s\Sigma_s^* + \frac{\partial^2(\cdot)}{\partial\phi^2}\Sigma_c^{\phi}(\Sigma_c^{\phi})^*\right](\phi,\eta).$$

For any test function $f \in C^2([0,2\pi]\times\mathcal{H}_s)$, one should have $\langle\mathfrak{L}^{\epsilon}f,\mu^{\epsilon}\rangle = 0$. We expand μ^{ϵ} as $\mu^{\epsilon} = \mu_0 + \epsilon^2\mu_1 + \mathcal{O}(\epsilon^3)$, then, on the level $\mathcal{O}(\epsilon^{-2})$ and $\mathcal{O}(1)$,

$$\langle\mathfrak{L}_0 f,\mu_0\rangle = 0,\quad \langle\mathfrak{L}_0 f,\mu_1\rangle = -\langle\mathfrak{L}_1 f,\mu_0\rangle. \tag{11}$$

In [12, Proposition 4.1 and 4.3], the solution of (11) is analyzed. We encourage readers to verify the results. It is worth noting that, under the restriction that μ_0 and μ_1 are probability measures, the solution is indeed unique. Consequently, $\lambda^{\mathfrak{q},\epsilon} = \langle\mathcal{Q},\mu^{\epsilon}\rangle = \langle\mathcal{Q},\mu_0\rangle + \epsilon^2\langle\mathcal{Q},\mu_1\rangle + r(\epsilon)$, where $r(\epsilon)$ represents the remainder. It has been shown that $r(\epsilon) = \mathcal{O}(\epsilon^3)$ [12, Proposition 4.4]. The final approximation is as follows:

$$\lambda^{\mathfrak{q},\epsilon} = \frac{1}{2\pi}\int_0^{2\pi}\mathcal{Q}(\phi,0)d\phi + \frac{\epsilon^2}{2\pi}\int_0^{2\pi}[\mathcal{Q}(\phi,0)\kappa(\phi) + \mathcal{Q}''(\phi,0;\chi(\phi),h)]d\phi + \mathcal{O}(\epsilon^3), \tag{12}$$

where

$$\kappa(\phi) = 1 - \frac{1}{b_c^{\mathfrak{c}}}\Gamma(\phi,0) - \frac{1}{b_c^{\mathfrak{c}}}\left(\frac{1}{2}\,\mathrm{tr}[G_c^{\phi}(D_{\phi}G_c^{\phi})^* + (D_{\phi}G_c^{\phi})(G_c^{\phi})^*](\phi,0)\right),$$

$$h = \sum_{k\in\mathbb{Z}_s}\frac{1}{2^{|k|+2}(1-\rho_k)}e_k,$$

$$\chi(\phi) = \sum_{k\in\mathbb{Z}_s}(\chi_k^R(\phi) + i\chi_k^I(\phi))e_k,$$

and

$$\begin{bmatrix}\chi_k^R\\ \chi_k^I\end{bmatrix}(\phi) = \begin{bmatrix}a_k-1 & -b_k\\ b_k & a_k-1\end{bmatrix}^{-1}\cdot\left(e^{\vartheta_k}\begin{bmatrix}\cos(w_k) & -\sin(w_k)\\ \sin(w_k) & \cos(w_k)\end{bmatrix} - \begin{bmatrix}G_k^R\\ G_k^I\end{bmatrix}\right)$$

with $\vartheta_k := \frac{(a_k-1)\phi}{b_c^{\mathfrak{c}}}$ and $w_k := \frac{(b_k-1)\phi}{b_c^{\mathfrak{c}}}$.

5 Impact of Multiplicative Noise of the Almost-Sure Asymptotic Stability to the Jet Engine Compressor

We examine the $\lambda^{\mathfrak{q},\epsilon}$ of the re-scaled solution $g(\epsilon^{-2}t)$ of (5) using the formula (12). Taking advantage of the discrete spectrum and the Fourier basis $\{e_k\}_{k\in\mathbb{Z}_0}$

with $e_k = e^{ik\theta}$, we calculate the related quantities through direct discrete expansion and multiplication as follows.

$$\Sigma_c^R(g) = \sum_{k\in\mathbb{Z}_+} \frac{\hat{\Sigma}_1(g) + \overline{\hat{\Sigma}_1(g)}}{2} \text{ and } \Sigma_c^I(g) = \frac{\hat{\Sigma}_1(g) - \overline{\hat{\Sigma}_1(g)}}{2},$$

where $\hat{\Sigma}_1(g)(\cdot) = \sqrt{q_{k+1}} g_{-k} \langle e_{-k-1}, \cdot\rangle e_{k+1}$ and $g_k = \langle e_{-k}, g\rangle$. Consequently, letting $\eta_k = \langle e_{-k}, \eta\rangle$ and $\mathfrak{Y} := \sum_{k\in\mathbb{Z}_+\setminus\{1\}} q_{k+1}\left(\frac{\eta_{-k}^2+\eta_k^2}{2}\right)$, we can obtain

$$\Xi(\phi,\eta) = -\frac{\sin(2\phi)}{2} q_2 - \frac{\sin(2\phi)}{2}\mathfrak{Y} \text{ and } \Gamma(\phi,\eta) = -\frac{\cos(2\phi)}{2} q_2 - \frac{\cos(2\phi)}{2}\mathfrak{Y}.$$

Hence, $\Xi(\phi,0) = -\frac{\sin(2\phi)}{2} q_2$, $\Gamma(\phi,0) = -\frac{\cos(2\phi)}{2} q_2$, $\mathcal{Q}(\phi,0) = a_c^{\mathfrak{q}} + \Xi(\phi,0)$. In addition,

$$\sum_{j\in\mathbb{Z}_s} \{[G_s G_s^*]_{kj}\}_{\eta=0} = \frac{q_k}{2},\ \forall k \in \mathbb{Z}_s \setminus \{\pm 2\},$$

and $\sum_{j\in\mathbb{Z}_s} \{[G_s G_s^*]_{kj}\}_{\eta=0} = 0$ for $k = \pm 2$. We can also verify that $\kappa(\phi) = \frac{1}{b_c^{\mathfrak{c}}}\left(\frac{\cos(2\phi)}{2} q_2 - \sin(2\phi) q_2\right) + 1$, and

$$\begin{aligned}
&\mathcal{Q}''\left(\phi, 0; \sum_{k\in\mathbb{Z}_s} (\chi_k^R(\phi) + i\chi_k^I(\phi)) e_k, \sum_{k\in\mathbb{Z}_s} \frac{1}{2^{|k|+2}(1-\rho_k)} e_k\right) \\
=& -\frac{\sin(2\phi)}{2} \sum_{k\in\mathbb{Z}_+\setminus\{1\}} q_{k+1} \cdot \frac{\chi_k^R(\phi) + i\chi_k^I(\phi)}{2^{|k|+2}(1-\rho_{-k})} < \infty.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\lambda^{\mathfrak{q},\epsilon} =& a_c^{\mathfrak{q}} + \frac{\epsilon^2}{2\pi}\int_0^{2\pi} \mathcal{Q}(\phi,0)\kappa(\phi) d\phi \\
&+ \frac{\epsilon^2}{2\pi}\int_0^{2\pi} \mathcal{Q}''\left(\phi, 0; \sum_{k\in\mathbb{Z}_s} (\chi_k^R(\phi) + i\chi_k^I(\phi)) e_k, \sum_{k\in\mathbb{Z}_s} \frac{1}{2^{|k|+2}(1-\rho_k)} e_k\right) d\phi + \mathcal{O}(\epsilon^3) \\
=& a_c^{\mathfrak{q}} + \frac{\epsilon^2}{2\pi}\int_0^{2\pi} \left[\frac{a_c^{\mathfrak{q}} q_2}{b_c^{\mathfrak{c}}}\left(\frac{\cos(2\phi)}{2} - \sin(2\phi)\right) - \frac{q_2^2}{b_c^{\mathfrak{c}}}\frac{\sin(2\phi)}{2}\left(\frac{\cos(2\phi)}{2} - \sin(2\phi)\right)\right] d\phi \\
&+ \frac{\epsilon^2}{2\pi}\int_0^{2\pi} \left(a_c^{\mathfrak{q}} - \frac{\sin(2\phi) q_2}{2}\right) d\phi \\
&- \frac{\epsilon^2}{4\pi}\left(\sum_{k\in\mathbb{Z}_+\setminus\{1\}} q_{k+1}\frac{\chi_k^R(\phi) + i\chi_k^I(\phi)}{2^{|k|+2}(1-\rho_{-k})}\right)\int_0^{2\pi} \sin(2\phi) d\phi + \mathcal{O}(\epsilon^3) \\
=& a_c^{\mathfrak{q}} + \frac{\epsilon^2}{2\pi}\int_0^{2\pi}\left[a_c^{\mathfrak{q}} + \frac{q_2^2}{2b_c^{\mathfrak{c}}}\left(\frac{1-\cos(4\phi)}{2}\right)\right] d\phi + \mathcal{O}(\epsilon^3) = a_c^{\mathfrak{q}} + \epsilon^2\left(a_c^{\mathfrak{q}} + \frac{q_2^2}{4b_c^{\mathfrak{c}}}\right) + \mathcal{O}(\epsilon^3)
\end{aligned}$$

As $\mathfrak{q}$ converges to 0 (i.e. γ converges to the deterministic Hopf bifurcation point γ_c), the variable $a_c^{\mathfrak{q}} \to 0$. Since $b_c^{\mathfrak{c}} > 0$ based on the choice of parameters, in this particular case, the multiplicative noise does not stabilize the system given small values of ϵ. For engineering operations, to avoid rotating stall, one should carefully evaluate the safety range of $\mathfrak{q}$ so that $\lambda^{\mathfrak{q},\epsilon}$ is negative.

6 Conclusion

In this paper, we provide a comprehensive review of the Moore-Greitzer Model and investigate the integration of stochastic multiplicative noise into the original model. Specifically, we explore a family of fractional Sobolev spaces that facilitate more robust analysis, and we use them to model the stochastic noise. We also conduct a review of the main theoretical results from [11] and employ the analytical findings, particularly the Furstenberg–Khasminskii formula for the maximal Lyapunov exponent, to assess the almost-sure asymptotic stability of the stochastically perturbed model. The quantitative result demonstrates that the presence of multiplicative noise destabilizes the original system and narrows the optimal safe operation range of the jet engine compressor. Our future direction is to explore how the stability changes affect invariant measures in nonlinear systems and their impact on probabilistic stability.

Acknowledgement. The authors acknowledge partial support for this work from National Sciences and Engineering Research Council (NSERC) Discovery grant 50503-10802. The authors also acknowledge Dr. Alyssa Novelia, who provided a thorough review from the modeling perspective of the Moore-Greitzer model.

References

1. Bahouri, H., Chemin, J.Y., Danchin, R.: Fourier Analysis and Nonlinear Partial Differential Equations, vol. 343. Springer (2011)
2. Banaszuk, A., Hauksson, H.A., Mezic, I.: A backstepping controller for a nonlinear partial differential equation model of compression system instabilities. SIAM J. Control. Optim. **37**(5), 1503–1537 (1999)
3. Behzadan, A., Holst, M.: Multiplication in sobolev spaces, revisited. Ark. Mat. **59**(2), 275–306 (2021)
4. Birnir, B., Hauksson, H.A.: Basic control for the viscous Moore-Greitzer partial differential equation. SIAM J. Control Optim. **38**(5), 1554–1580 (2000)
5. Da Prato, G., Zabczyk, J.: Stochastic Equations in Infinite Dimensions. Cambridge University Press, Cambridge (2014)
6. Gravdahl, J.T.: Modeling and control of surge and rotating stall in compressors (1998)
7. Hairer, M.: An introduction to stochastic PDEs. arXiv preprint arXiv:0907.4178 (2009)
8. Hou, S.: Solutions of multidimensional hyperbolic systems of conservation laws by discontinuous Galerkin methods and a derivation of the Moore-Greitzer equation using homogenization. University of California, Santa Barbara (2002)
9. Khas' minskii, R.Z.: Necessary and sufficient conditions for the asymptotic stability of linear stochastic systems. Theory Prob. Appl. **12**(1), 144–147 (1967)
10. Meng, Y.: Bifurcation and robust control of instabilities in the presence of uncertainties. Ph.D. Thesis (2022)
11. Meng, Y., Namachchivaya, N., Perkowski, N.: Almost sure asymptotic stability of parabolic SPDEs with small multiplicative noise. arXiv preprint arXiv:2310.19255 (2023)

12. Meng, Y., Namachchivaya, N.S., Perkowski, N.: Hopf bifurcations of Moore-Greitzer PDE model with additive noise. J. Nonlinear Sci. **33**(5), 74 (2023)
13. Novelia, A., Aydogdu, Y., Ravichandran, T., Namachchivaya, N.: Compressed compressor. arXiv preprint arXiv:2107.08938 (2021)
14. Pardoux, E., Wihstutz, V.: Lyapunov exponent and rotation number of two-dimensional linear stochastic systems with small diffusion. SIAM J. Appl. Math. **48**(2), 442–457 (1988)
15. Pazy, A.: Semigroups of Linear Operators and Applications to Partial Differential Equations, vol. 44. Springer, Cham (2012)
16. Sri Namachchivaya, N., Vedula, L.: Stabilization of linear systems by noise: application to flow induced oscillations. Dyn. Stab. Syst. **15**(2), 185–208 (2000)
17. Sri Namachchivaya, N., Wihstutz, V.: Almost sure asymptotic stability of scalar stochastic delay equations: finite state Markov process. Stochastics Dyn. **12**(01), 1150,010 (2012)
18. Wen, J., Xiao, M., Xu, J.: Control of hopf bifurcations of nonlinear infinite-dimensional systems: application to axial flow engine compressor. Appl. Anal. **91**(11), 1959–1980 (2012)
19. Xiao, M.: Quantitative characteristic of rotating stall and surge for Moore-Greitzer PDE model of an axial flow compressor. SIAM J. Appl. Dyn. Syst. **7**(1), 39–62 (2008)

Nonlinear Stochastic Global Dynamics of an Imperfect, Electrically Actuated Microcantilever Through an Operator Approach

Kaio C. B. Benedetti[1], Paulo B. Gonçalves[2(✉)], Stefano Lenci[3], and Giuseppe Rega[4]

[1] School of Civil and Environmental Engineering, Federal University of Goiás, Goiânia, Brazil
[2] Department of Civil and Environmental Engineering, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil
paulo@puc-rio.br
[3] Department of Civil and Building Engineering and Architecture, Polytechnic University of Marche, Ancona, Italy
[4] Department of Structural and Geotechnical Engineering, Sapienza University of Rome, Rome, Italy

Abstract. The aim of this work is to analyze the nondeterministic global dynamics of an electrically actuated microbeam under harmonic excitation with added noise, via a phase-space operator approach allowing a unified description for both deterministic and nondeterministic cases. In this context, the stochastic basins of attraction and attractors' distributions replace the usual basin and attractor concepts. For the numerical analysis, an adaptative phase-space discretization strategy based on the classical Ulam method is adopted. The results highlight the importance of randomness and time-dependency of stochastic response on the global dynamics of systems with competing attractors. A dynamic integrity measure based on the obtained probability distributions is employed to confidently quantify the safety of the dynamical system. Finally, sample results on the influence of a parameter uncertainty on global dynamics are presented.

Keywords: Electrically Actuated Microbeam · Operator Approach · Stochastic Global Dynamics · Dynamic Integrity

1 Introduction

The study of microelectromechanical systems (MEMs) encompasses a broad range of disciplines, from structural mechanics to electrodynamics [1–3]. Regarding uncertainty, Vig and Kim [4] described several possible noise sources in MEMs, while Agarwal and Aluru tested the effects of stochastic variations in various design parameters [5]. Recently, we investigated both noise and damping ratio uncertainty in a microarch electrically actuated [6]. It was evident the importance of nondeterministic effects, affecting noticeably the expected outcomes.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 125–140, 2025.
https://doi.org/10.1007/978-3-031-72794-8_9

The methodology in [6] was later explained in more detail in [7], with application to the Helmholtz oscillator. It was an extension of classic global analysis [8] of the Ulam method [9] and the Hsu generalized cell-mapping [10, 11] to nondeterministic systems, where numerical diffusion was addressed through an adaptative refinement of the phase-space. Even the deterministic analyses had their overall computational cost reduced by this methodology.

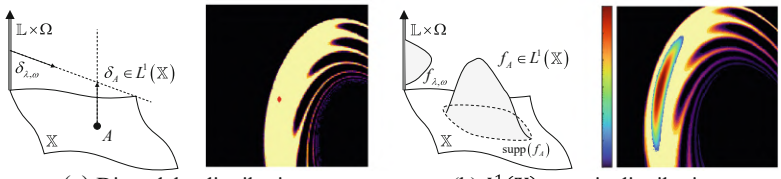
The aim of this work is to study the influence of added noise and uncertainties on the global dynamics of an electrically actuated microcantilever, whose modeling and nonlinear response through classical generalized cell mapping were previously presented in [12], by also considering nondeterministic effects. Here, we employ the adaptative Ulam method [7], a general strategy based on transfer operators for deterministic, noise-driven and uncertain parameter dynamics, which allows more systematic and reliable investigations, along with the detection of the significant influence of noise on the system long-term dynamics. In Sect. 2 a summary of concepts/definitions of stochastic global dynamics is reported, within the framework of operators' approach and its discretization. Section 3 deals extensively with the global dynamics of the microcantilever. Final remarks are given in Sect. 4.

2 Definitions and Discretization of Attractors, Basins and Global Operators with Noise and Parametric Uncertainty

The key elements in global dynamics are attractors and their basins. Nondeterminism affects these global structures, which must be redefined. Attractors are usually understood in the pushforward sense, that is, set of states towards which a system evolves as time increases. However, in nondeterministic dynamics, attractors are dependent on the random source and cannot be uniquely located in phase-space or, for noise-driven dynamics, can be time-dependent [7]. Thus, it must be investigated how attractors are distributed over the phase-space, applying the ergodic theory and measurable dynamics [13]. Indeed, attractors' distributions f_A are densities of probability measures over $L^1(\mathbb{X})$, that is, the integral of f_A over a region in $\mathbb{X}$ is the probability of the attractor A being in such region. Schematics and examples of distribution f_A are depicted in Fig. 1, for a deterministic fixed-point attractor, and for a diffused attractor, which is common in noise-driven problems. Fixed-point attractor is marked as a red dot in the phase-space plane, while diffused attractor is depicted through the colormap, from blue (low density distribution) to red (high density distribution).

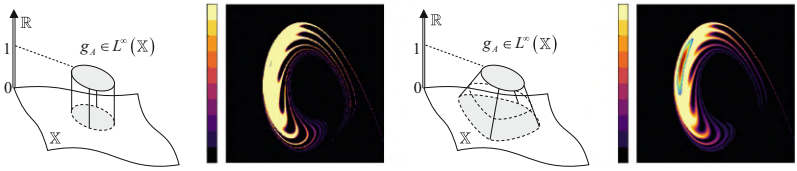
Informally, a basin of attraction is the set of initial conditions converging to the corresponding attractor. This basic definition loses meaning in the nondeterministic case, since the attractor becomes dependent on the random source. As discussed in [7], the basin definition of Lindner and Hellmann [14] is well-suited to such cases. They propose a map $g_A : [0,1] \times \mathbb{X} \to [0,1]$ of probabilities of convergence to A over the phase-space $\mathbb{X}$. For deterministic cases, this map is just an indicator function of the classical basin definition. For noise-driven dynamics, it is necessary to think of it as the probability of convergence to the attractor after a given time, which is the time-horizon $1/\varepsilon$ within which a trajectory converges to the attractor: for $\varepsilon = 1$, the function $g_A(1, \cdot)$ gives the probability of immediate convergence, and for $\varepsilon \to 0$, the function $g_A(0, \cdot)$

gives the probability of convergence for time at infinity. Schematic and examples of both deterministic and nondeterministic basins are depicted in Fig. 2. The colormap varies from black (probability 0) to yellow (probability 1). Deterministic case is approximated by a step function over the basin, with almost the entire basin marked as yellow, while nondeterministic case diffuses over the phase-space, with intermediate colors and values.



(a) Dirac delta distribution (b) $L^1(\mathbb{X})$ generic distribution

Fig. 1. Distributions of (a) a fixed-point attractor and (b) a diffused attractor with its color map. Probability spaces for random parameters $\mathbb{L}$ and noise Ω are orthogonal to the phase-space $\mathbb{X}$



(a) deterministic basin (b) nondeterministic basin

Fig. 2. Basins of attraction functions g_A of (a) deterministic and (b) nondeterministic dynamics, with their color map. The image [0,1] is over the real line $\mathbb{R}$, which is depicted orthogonal to the phase-space $\mathbb{X}$

The evolution of observable complex functions of the phase-space $g : \mathbb{X} \to \mathbb{R}^+, L^\infty(\mathbb{X})$ is governed by the Koopman operator $\mathcal{K}$. Since $g_A(\varepsilon, \cdot)$ is also an observable g, the Koopman operator governs the evolution of basins of attraction too. Its definition can be extended to consider a dynamical system $\varphi(\omega, \lambda) : \mathbb{X} \to \mathbb{X}$ with a noise $\omega \in \Omega$ (Ω representing a stochastic space $(\Omega, \mathfrak{F}, P_\omega)$ – a sample space Ω with a σ-algebra $\mathfrak{F}$ and a probability measure P_ω), and a random parameter $\lambda \in \mathbb{L}$ ($\mathbb{L}$ representing a random space $(\mathbb{L}, \mathfrak{S}, P_\lambda)$ – a sample space $\mathbb{L}$ with a σ-algebra $\mathfrak{S}$ and a probability measure P_λ), resulting in [7]

$$\mathcal{K}(\lambda)\big[g(\boldsymbol{x})\big] = \int_\Omega g \circ \varphi(\omega, \lambda)\boldsymbol{x} dP_\omega, \tag{1}$$

where $\boldsymbol{x} \in \mathbb{X}$. The dual or adjoint operator has different names and definitions depending on whether the underlying dynamical system is deterministic or random, and on the type of randomness. Nevertheless, they all are transfer operators $\mathcal{P}$ that govern the evolution of complex distributions $f \in L^1(\mathbb{X})$. Since f_A is also a complex distribution in $L^1(\mathbb{X})$, it evolves according to a transfer operator, too. By considering noise $\omega \in \Omega$ and random

parameter $\lambda \in \mathbb{L}$, a general transfer operator $\mathcal{P}$ is defined as [7]

$$\begin{aligned} &\int_B \mathcal{P}(\lambda)\big[f(\boldsymbol{x})\big]d\boldsymbol{x} = \int_{\mathbb{X}} \left\{ \int_{\Omega_{\boldsymbol{x}}(\lambda,B)} dP_\omega \right\} f(\boldsymbol{x})d\boldsymbol{x}, \\ &\Omega_{\boldsymbol{x}}(\lambda, B) = \{\omega \in \Omega : \varphi(\omega, \lambda)\boldsymbol{x} \in B\} \end{aligned} \tag{2}$$

where $\Omega_{\boldsymbol{x}}(\lambda, B) \subseteq \Omega$ is the set of all ω-values for which the dynamical system φ is in $B \subset \mathbb{X}$, for any $\lambda \in \mathbb{L}$.

Attractor's distributions f_A and basins g_A (at infinite time $\varepsilon \to 0$) are in the fixed space of $\mathcal{P}$ and $\mathcal{K}$, respectively, being their eigenfunctions with corresponding eigenvalue 1. By expanding the notation, the effects of random parameters λ are considered: both are rewritten as $f_A(\boldsymbol{x}; \lambda)$ and $g_A(0, \boldsymbol{x}; \lambda)$, respectively, being solutions to the eigenvalue problems

$$\begin{aligned} \mathcal{K}(\lambda)\big[g_A(0, \boldsymbol{x}; \lambda)\big] &= g_A(0, \boldsymbol{x}; \lambda), \\ \mathcal{P}(\lambda)\big[f_A(\boldsymbol{x}; \lambda)\big] &= f_A(\boldsymbol{x}; \lambda). \end{aligned} \tag{3}$$

Averaged functions over the parameter space $\mathbb{L}$ are obtained through a simple mean over the corresponding probability space $(\mathbb{L}, \mathfrak{S}, P_\lambda)$. For basins and attractors in (3), the averaged functions are simple integrals [15], which, for a general ε-valued basin observable $g_A(\varepsilon, \boldsymbol{x}; \lambda)$, are

$$\begin{aligned} \overline{g}_A(\varepsilon, \boldsymbol{x}) &= \int_{\mathbb{L}} g_A(\varepsilon, \boldsymbol{x}; \lambda) dP_\lambda, \\ \overline{f}_A(\boldsymbol{x}) &= \int_{\mathbb{L}} f_A(\boldsymbol{x}; \lambda) dP_\lambda. \end{aligned} \tag{4}$$

These integrals can be discretized through polynomial chaos [15] or by a weighted sum [16]. Other important properties of $\mathcal{P}$ and $\mathcal{K}$ are their linearity, the definition of general infinity-dimensional phase-spaces $\mathbb{X}$, unitary spectral radius, and multistability implying multiple distributions with the same unitary eigenvalue. For a modern treatment of the theory of Koopman and transfer operators, refer to [17].

The discretization methodology developed in [7] is based on the Ulam method, which is a zeroth order discretization of the space of distributions over the phase-space, $L^1(\mathbb{X})$. The phase-space is partitioned in a set of disjoint cells $\mathbb{B} = \{b_1, \ldots, b_i\}$, with the transfer operator being discretized into a row-stochastic matrix p_{ij}, where each entry gives the probability of jumping from cell b_i to cell b_j for a fixed time interval Δt, usually of one forcing period T. The procedure to construct the discretized operators is extensively discussed in [7], and can be summarized in the following expression,

$$\begin{aligned} p_{ij}(\lambda) &= \tfrac{1}{P_{\mathbb{X}}(b_i)} \int_{b_i} \left\{ \int_{\Omega_{\boldsymbol{x}}(\lambda, b_j)} dP_\omega \right\} d\boldsymbol{x}, \\ \Omega_{\boldsymbol{x}}\big(\lambda, b_j\big) &= \big\{\omega \in \Omega : \varphi(\omega, \lambda)\boldsymbol{x} \in b_j\big\}, \end{aligned} \tag{5}$$

where $P_{\mathbb{X}}(b_i)$ is the measure of cell b_i. A succinct description of Eq. (5) is

$$p_{ij}(\lambda) \approx \frac{\#\text{states in } b_j \text{ due to } \varphi(\omega, \lambda) \text{ and with i.c. in } b_i}{\#\text{i.c. in } b_i}. \tag{6}$$

Discretized attractors $f_i(\lambda)$ are solutions of the fixed space problem

$$f_i(\lambda)\delta_{ij} = f_i(\lambda) p_{ij}(\lambda), \tag{7}$$

and discretized basins of attraction $g_j(\varepsilon;\lambda)$ are given by

$$\left[\delta_{ij} - (1-\varepsilon)p_{ij}(\lambda)\right]g_j(\varepsilon;\lambda) = \varepsilon \mathrm{id}_{A(\lambda)}, \tag{8}$$

where δ_{ij} is the Kronecker delta and $\mathrm{id}_{A(\lambda)}$ is the indicator function of the attractor $A(\lambda)$, with ones in entries $b_i \subset A(\lambda)$. Both $f_i(\lambda)$ and $g_j(\varepsilon;\lambda)$ are vectors with values in [0; 1], where each entry is

$$g_j(\varepsilon;\lambda) = \frac{1}{P_{\mathbb{X}}(b_j)}\int_{b_j} g_j(\varepsilon, \boldsymbol{x};\lambda)d\boldsymbol{x}, \tag{9}$$

$$f_i(\lambda) = \int_{b_i} f(\boldsymbol{x};\lambda)d\boldsymbol{x}.$$

The parameter $\varepsilon \in (0,1]$ controls until when it is expected for trajectories to stay in the attractor $A(\lambda)$. For $\varepsilon = 1$, the basin of attraction is only the attractor itself, at time zero. For $\varepsilon \to 0$, the basin of attraction converges to the response at infinity time, similar to the continuous map definition $g_A(\varepsilon, \boldsymbol{x};\lambda)$. For more details, see Lindner and Hellmann [14]. Equations (5), (6), (7), (8), and (9) are a generalization of the Ulam method for noise-driven dynamics with parametric uncertainty, containing all cases.

3 Electrically Actuated Microcantilever Analysis

Here, a slender, clamped-free, imperfect, electrically actuated microbeam is considered. The planar transverse imperfect Rayleigh beam equation is a model previously addressed in [12], resumed in Fig. 3. The electrical actuation load is composed of a direct current voltage V_{dc} and an alternating current voltage V_{ac}, each contributing to the static and dynamic components of the vertical displacement, $w = w_d + w_s$. Therefore, there are two main problems, a static equilibrium due to V_{dc} and a dynamic response due to V_{ac}. In both cases, the axial displacement u is a dependent variable due to the inextensibility assumption of cantilevers. Rotational inertia is taken into account.

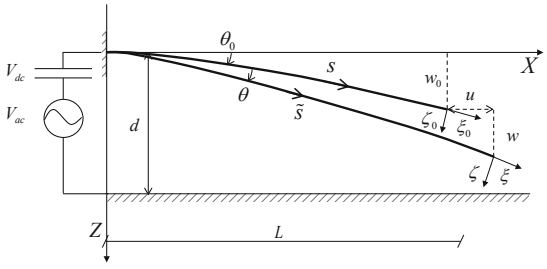


Fig. 3. Imperfect undeformed and deformed beam and coordinate systems

The vertical displacement w_s and imperfection w_0 are expanded in Taylor series up to the third order. The resulting nondimensional static equilibrium equation is [12]

$$\frac{b\varepsilon V_{dc}^2}{2d^3(1-\overline{w}_s)^2} = d^2\left[\overline{w}_s' w_s''\left(\overline{w}_s'' + w_0''\right) + w_s' w_0''^2 + w_s' w_0'''\left(w_0' + \frac{w_s'}{2}\right) + w_s'''\left(\overline{w}_s'^2 + \frac{1}{d^2}\right)\right]', \tag{10}$$

where $\overline{w}_s = w_s + w_0$. The additional dynamic displacement w_d is assumed as a perturbation of the static equilibrium position w_s. The resulting nondimensional equation of motion is

$$\begin{aligned}
&\ddot{w}_d - J_\eta \ddot{w}_d'' + c_w \dot{w}_d + w_d^{iv} \\
&= \left\{-d^2\left[w_d''' \overline{w}'^2 + w_d' \overline{w}_s''^2 + \overline{w}' w_d''\left(\overline{w}'' + \overline{w}_s''\right)\right.\right. \\
&\left. + \frac{1}{2} w_d'\left(\overline{w}' + \overline{w}_s'\right)\left(\overline{w}_s''' + w_s'''\right)\right] + d^2 J_\eta\left[\overline{w}' \dot{w}_d'^2 + \ddot{w}_d' \overline{w}'^2\right] \\
&\left. - d^2 \overline{w}' \int_1^s \int_0^s \left(\dot{w}_d'^2 + \overline{w}' \ddot{w}_d'\right) ds ds\right\}' \\
&+ \frac{b\varepsilon}{d^3}\left[\frac{2w_d^3 V_{dc}^2}{(1-\overline{w}_s)^5} + \frac{w_d^2\left(3V_{dc}^2 - 2V_{dc}V_{ac}\right)}{2(1-\overline{w}_s)^4} + \frac{w_d(V_{ac}+V_{dc})^2}{(1-\overline{w}_s)^3}\right. \\
&\left. + \frac{2V_{dc}V_{ac} + V_{ac}^2}{2(1-\overline{w}_s)^2}\right],
\end{aligned} \tag{11}$$

where $\overline{w} = w_d + w_s + w_0$. The sum of zeroth-order terms is equal to zero since they correspond to the static equilibrium position, Eq. (10). For the definition of the nondimensional parameters, refer to [12].

In the present work, a first clamped-free Rayleigh beam mode F_1 [18], is considered in the Galerkin expansion for both the geometric imperfection w_0, the static deflection w_s, and the dynamic deflection w_d, leading to a sdof reduced-order model (ROM). Hereafter, the symbols for the modal amplitudes will be adopted as the same as those adopted for the state variables, that is, w_0, w_s, and w_d, simplifying the notation. The equilibrium equation of the static deflection w_s is obtained by multiplying Eq. (10) by the load denominator, $(1-\overline{w}_s)^2$, following with the Galerkin projection and resulting in Eq. (12) [1]. Given a static deflection w_s resulted from Eq. (12) with a direct actuation V_{dc} and an imperfection w_0, the modal equation of w_d is obtained by the Galerkin projection onto the first mode F_1 of Eq. (11), resulting in Eq. (13). The first mode was chosen because it is the dominant one regarding geometric imperfection and static deflection. As for the dynamic deflection, the excitation frequency is adopted close to the first natural frequency, exciting mainly the first mode as well. Additionally, an additive white noise term $\sigma \dot{W}$ is considered in Eq. (13), where $\dot{W}$ is the generalized derivative of the standard

Wiener process and σ is a scaling parameter.

$$\begin{aligned}
&\frac{b\varepsilon V_{dc}^2}{2d^3}\int_0^1 F_1 ds \\
&= \int_0^1 d^2 F_1 (1-\overline{w}_s F_1)^2 \left\{ \frac{w_s}{d^2} F_1^{iv} + \left(w_s w_0^2 + \overline{w}_s^2 w_s + \overline{w}_s w_s w_0 \right) \left[F_1' F_1''^2 \right]' \right. \\
&\left. + \left(w_s w_0^2 + \frac{w_s^2 w_0}{2} + \overline{w}_s^2 w_s \right) \left[F_1'^2 F_1''' \right]' \right\} ds, \qquad (12)
\end{aligned}$$

$$\begin{aligned}
&(\ddot{w}_d + c_w \dot{w}_d) \int_0^1 F_1^2 ds - J_\eta \ddot{w}_d \int_0^1 F_1'' F_1 ds + w_d \int_0^1 F_1^{iv} F_1 ds \\
&= \left\{ -\left(d^2 w_d \overline{w}^2 + w_d (\overline{w} + \overline{w}_s)(\overline{w}_s + w_s) \right) \int_0^1 \left[F_1'^2 F_1''' \right]' F_1 ds \right. \\
&\left. - \left(w_d \overline{w}_s^2 + \overline{w}^2 w_d + \overline{w}_s \overline{w} w_d \right) \int_0^1 d^2 \left[F_1' F_1''^2 \right]' F_1 ds \right\} \\
&+ \left\{ J_\eta \left(\overline{w} \dot{w}_d^2 + \overline{w}^2 \ddot{w}_d \right) \int_0^1 d^2 \left[F_1'^3 \right]' F_1 ds \right. \\
&\left. - \overline{w} \left(\dot{w}_d^2 + \overline{w} \ddot{w}_d \right) \int_0^1 d^2 \left(F_1' \int_1^s \int_0^s F_1'^2 ds ds \right)' F_1 ds \right\} \qquad (13) \\
&+ \frac{b\varepsilon}{d^3} \left\{ 2 V_{dc}^2 w_d^3 \int_0^1 \frac{F_1^4}{(1-\overline{w}_s F_1)^5} ds \right. \\
&+ \left(\frac{3}{2} V_{dc}^2 - V_{dc} V_{ac} \right) w_d^2 \int_0^1 \frac{F_1^3}{(1-\overline{w}_s F_1)^4} ds \\
&+ (V_{ac} + V_{dc})^2 w_d \int_0^1 \frac{F_1^2}{(1-\overline{w}_s F_1)^3} ds \\
&\left. + \left(V_{dc} V_{ac} + \frac{V_{ac}^2}{2} \right) \int_0^1 \frac{F_1}{(1-\overline{w}_s F_1)^2} ds \right\} + \sigma \dot{W}.
\end{aligned}$$

The geometric and material parameters adopted are the same employed in [2, 3, 12]. The nondimensional parameters are summarized in Table 1.

For the static response of the imperfect microbeam, see [12] and [16]. The nonlinear response of the imperfect microbeam reduced-order model under increasing direct current voltage V_{dc} exhibits two limit points along the nonlinear equilibrium path, Fig. 4, which are rather sensitive to the imperfection magnitude and sign, decreasing with positive imperfections and increasing with negative imperfections. The pull-in instability is present in all cases, being the pull-in voltage particularly sensitive to the imperfection level and sign. The perfect system obtained pull-in voltage is 65.341 V, being in good agreement with that in [1], where the pull-in voltage is 67.443 V for the constants employed in [2, 3, 12].

For the global dynamic analysis of the microbeam reduced-order model, a direct current voltage $V_{dc} = 45$ V is adopted together with a periodic alternate voltage between the beam and the substrate, $V_{ac} = \overline{V}_{ac} \cos(\Omega t)$, where $\overline{V}_{ac}$ is the forcing magnitude and

Table 1. Microbeam geometric and material parameters.

Parameters	Symbol	Values
Width	b	0.25
Nominal gap	d	0.0046
Free space permittivity $\left(V^{-2}\right)$	ε	$2.9446\ 10^{-10}$
Damping	c_w	0.05
Rotational inertia	J_η	$6.7685\ 10^{-7}$

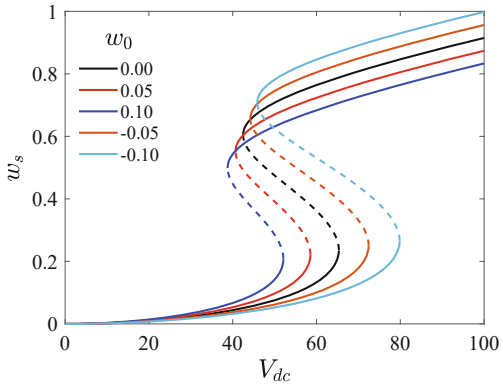


Fig. 4. Nonlinear equilibrium paths of the microbeam under DC actuation for selected levels of geometric imperfection and cusp catastrophe surface

Ω is the forcing frequency. Stochastic cases are addressed by controlling the scaling parameter σ.

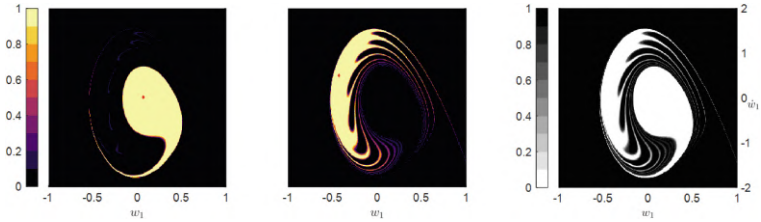
3.1 Global Dynamics for Added White Noise

The global dynamic analysis of the microcantilever with added noise was previously addressed through the classical generalized cell mapping [12]. Although the results were satisfactory, this strategy exhibited numeric diffusion [19], which demanded a well-refined phase-space partition at a high computational cost. To balance these two factors and improve the analysis, we apply the adaptative phase-space refinement of stochastic global operators, which results in an irregular partition of the phase-space, with smaller cells in regions of interest, i.e., attractors and basins' boundaries [7]. It allows their correct identification, necessary to understand the effects of noise on global dynamics.

The phase-space window $\mathbb{X} = [-1, 1] \otimes [-2, 2]$ is adopted, and the boundaries are assumed as of the absorbing type. The phase-space is initially discretized into 32×32 cells. Then, by applying the refinement algorithm, each box is subdivided into two new ones, up to eight times. The total of integrated initial conditions in each cell starts at 122

collocation points, reducing at each interaction attaining 42, in the final refinement step. The time integration scheme is the fourth order Runge-Kutta for deterministic systems or its trivial stochastic version for cases with noise, both with time-step $\Delta t = T/4000$. For the stochastic cases, each initial condition is integrated ten times. At each refinement step, a new stochastic matrix p_{ij} is constructed. Probability density distributions and (stochastic) basins of attraction are calculated through Eqs. (7) and (8), respectively. Finally, the phase-space is subdivided, and the process is repeated until the final discretization level is reached.

Consider a perfect case, $w_0 = 0$, with nonresonant and resonant solutions. Figure 5 displays the deterministic case, with attractors marked in red and basins of attraction in yellow. Regions are marked with values between 0 and 1, that is, with a probability of 0 up to 1 to converge to the corresponding solution. Basins' boundaries are identified in detail, as expected, with corresponding probability values in between zero and one being only due to numerical diffusion. The escape solution is depicted for completeness, in Fig. 5c, with dark regions corresponding to probability 1 to escape.



(a) nonresonant attractor (b) resonant attractor (c) escape solution

Fig. 5. Microcantilever global dynamics obtained after the application of the refinement algorithm for $V_{dc} = 45$ V, $\overline{V}_{ac} = 5$ V, $\Omega = 2.8$, $w_0 = 0$, $\sigma = 0$. Attractors marked in red. Color bars: probability of convergence.

Stochastic cases are presented in Fig. 6, for the nonresonant and resonant attractor, depicting attractors' distributions and basins. The attractors spread around the deterministic position, where higher densities are observed, diffusing over the phase-space as noise increases. Basins' boundaries diffusion is also observed, with increasing areas with probability value in between 0 and 1. The resonant solution is more impacted by noise, with a larger diffusion if compared to the nonresonant attractor.

The transient analysis for $\sigma = 0.013$ and $\sigma = 0.025$ is detailed in Fig. 7 and Fig. 8, respectively. Again, basins are depicted as a map of values between 0 and 1. As the time-horizon increases (ε decreases), nonresonant and resonant basins increase from initial regions close to the corresponding attractors, for both noise intensities. However, while the former ends up to a robust final shape, the latter vanishes after a long transient, upon which the resonant region is absorbed by the escape one. In the case $\sigma = 0.013$, this becomes evident for time-horizons $1/\varepsilon > 10^{10}T$. This is an interesting outcome due to the localized refinement. The previous investigation in [12], with a regular partition, only observed this behavior for $\sigma \geq 0.015$, with the resonant and nonresonant regions not being separated from each other. The Monte-Carlo experiments in [12] were conducted only until $10^4 T$, missing the long-transient characteristic of

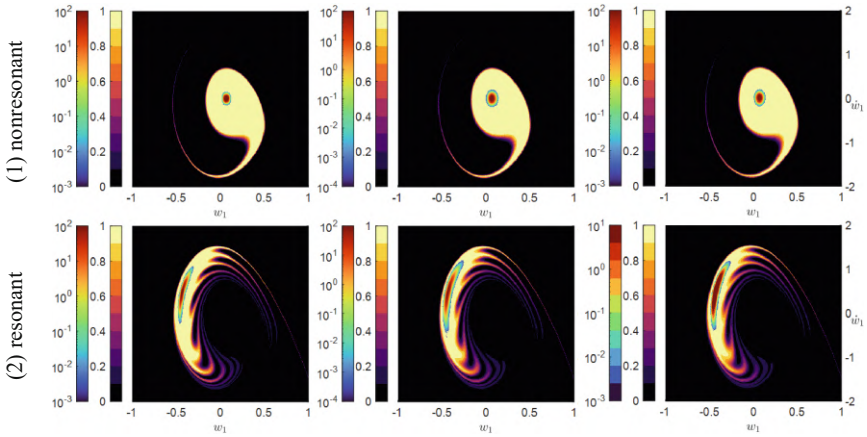


Fig. 6. Microcantilever's noise-driven global dynamics for $V_{dc} = 45$ V, $\overline{V}_{ac} = 5$ V, $\Omega = 2.8$, $w_0 = 0, \varepsilon = 10^{-8}$. First color bar: attractor's density, second color bar: probability of convergence.

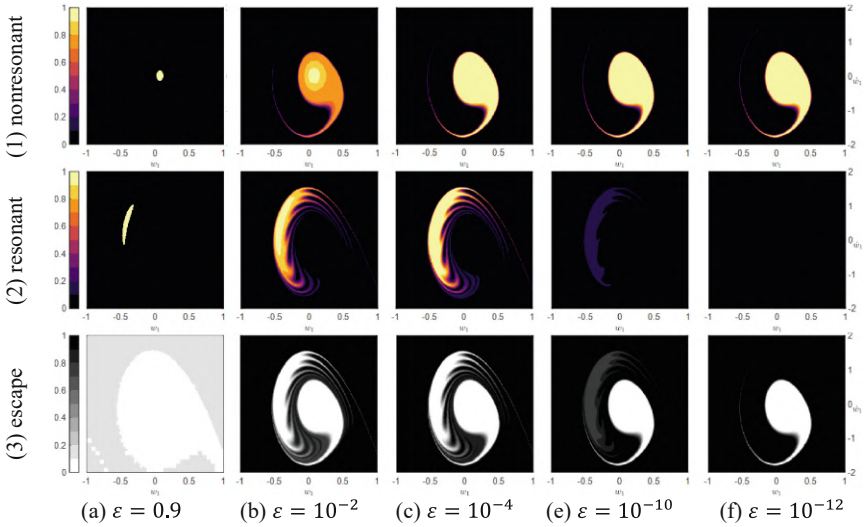


Fig. 7. Dependency of the stochastic basin of attraction (color bars) on the final time-horizon $1/\varepsilon$ for $V_{dc} = 45$ V, $\overline{V}_{ac} = 5$ V, $\Omega = 2.8$, $w_0 = 0$, $\sigma = 0.013$.

this solution. This clarification is only possible if the resonant and nonresonant regions are properly separated, which would demand a very refined phase-space, whereas the present localized subdivision strategy can do so without significantly increasing the computational cost. The case with $\sigma = 0.025$ shows that the resonant region vanishes sooner, for $1/\varepsilon = 10^4 T$, with some solutions leading to the nonresonant attractor, see Fig. 8(1.c), while the majority leads to escape, see Fig. 8(3.c). This agrees with the Monte-Carlo experiment in [12], where the same outcome is observed.

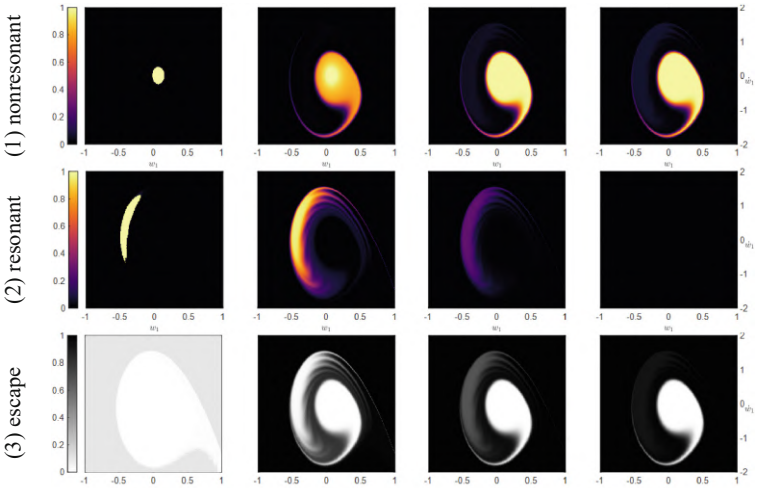
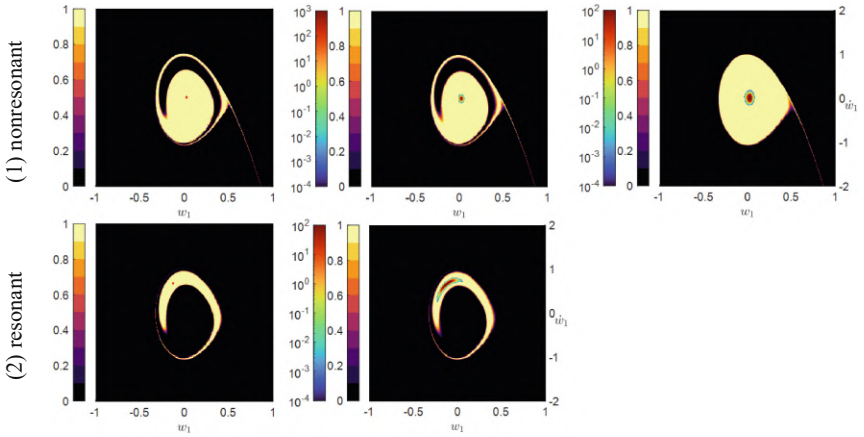


Fig. 8. Dependency of the stochastic basin of attraction (color bars) on the final time-horizon $1/\varepsilon$ for $V_{dc} = 45$ V, $\overline{V}_{ac} = 5$ V, $\Omega = 2.8$, $w_0 = 0$, $\sigma = 0.025$

A case of large imperfection value, $w_0 = 0.05$, is discussed next. The deterministic case is presented in Fig. 9a. Compared to previous results in [12], it is observed that the refinement strategy shows the basins' boundaries in much more detail. Regions with probability values between 0 and 1 are small, showing that the numerical diffusion is indeed mitigated. Figure 9b shows the results for the stochastic case with $\sigma = 0.005$. The diffusion of the attractors' distributions is demonstrated, with the resonant attractor being more affected by it. The basins' boundaries diffusion is not large, being localized to specific regions in phase-space. For $\sigma = 0.01$ (Fig. 9c), only the nonresonant attractor remains. This result seems to be in contradiction with the Monte-Carlo experiment in [12]. However, the Monte-Carlo was conducted only until $t = 10^5 T$, while here the outcome for larger time-horizon is depicted. This will be made clear in the following transient analysis.

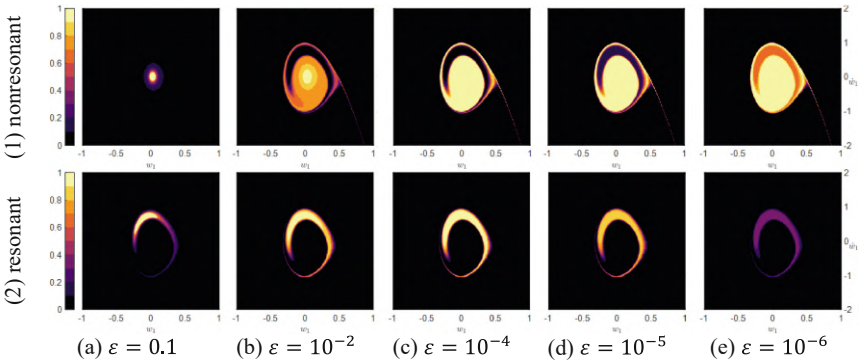
The transient analysis is presented in Fig. 10 for $\sigma = 0.01$. It is observed that for time-horizons $1/\varepsilon \leq 10^4 T$, increasing basins are obtained up to the usual ones. The vanishing of the resonant region is only observed for larger time-horizons, and is almost complete for $1/\varepsilon = 10^6 T$. This result confirms the Monte-Carlo experiment in [12], which was conducted only until $t = 10^5 T$, a time for which a separation between the basins is still observed. The transient basin analysis is only possible because the boundaries between basins are well refined. The adoption of a crude refinement in such regions would result in wrong time-horizon values of basin decay, with prevalent numerical diffusion due to poor discretization. These results demonstrate the effectiveness of an adaptative phase-space refinement, allowing a correct description of long transients with a small increase in the computational cost.

The previous results clarify the importance of noise in the evaluation of the dynamic integrity of dynamical systems. Figure 11 shows, for $\varepsilon = 10^{-8}$, the variation of the



(a) $\sigma = 0.000$ (b) $\sigma = 0.005$ (c) $\sigma = 0.010$

Fig. 9. Microcantilever's noise-driven global dynamics for $V_{dc} = 45\ rmV$, $\overline{V}_{ac} = 1$ V, $\Omega = 2.8$, $w_0 = 0.05$, $\varepsilon = 10^{-8}$. First color bar: attractor's density, second color bar: probability of convergence.



(a) $\varepsilon = 0.1$ (b) $\varepsilon = 10^{-2}$ (c) $\varepsilon = 10^{-4}$ (d) $\varepsilon = 10^{-5}$ (e) $\varepsilon = 10^{-6}$

Fig. 10. Dependency of the stochastic basin of attraction (color bars) on the final time-horizon $1/\varepsilon$ for $V_{dc} = 45$ V, $\overline{V}_{ac} = 1$ V, $\Omega = 2.8$, $w_0 = 0.05$, $\sigma = 0.010$

basins' area as a function of the noise intensity σ for the perfect case, $w_0 = 0$, and the imperfect case, $w_0 = 0.05$, for selected probability thresholds. For each attractor, the measure is computed as [7]

$$\frac{\int_{\mathbb{X}} \mathrm{id}_{[p,1]}(g) g d\boldsymbol{x}}{\int_{\mathbb{X}} d\boldsymbol{x}}, \tag{14}$$

where g is a stochastic basin of attraction, $\mathrm{id}_{[p,1]}(g)$ an indicator function, equal to 1 if $g \in [p, 1]$ and zero otherwise, and p the assumed probability threshold, between 0 and 1. In the deterministic limit (no uncertainty or noise, and infinite resolution), the function g is an indicator function of the basin and Eq. (14) reduces to the global integrity measure (GIM) [20]. For the perfect case, the basin area of the nonresonant

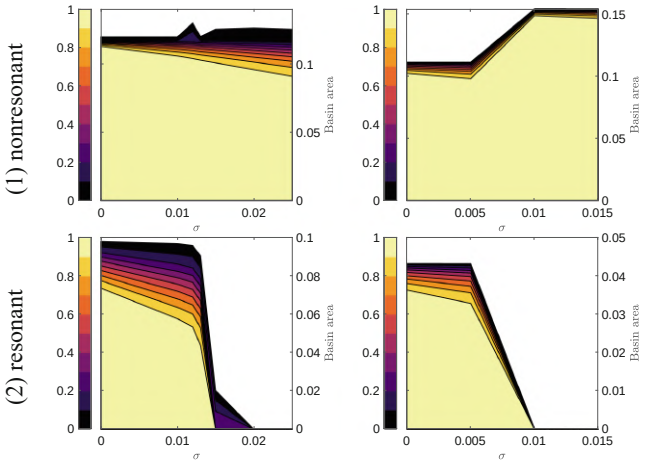


Fig. 11. Variation of the microcantilever basins area as a function of the noise intensity σ for $\varepsilon = 10^{-8}$, $V_{dc} = 45$ V, $\Omega = 2.8$, showing various probability thresholds (color bar).

attractor is almost constant, with only a small increase due to the erosion of the resonant basin as σ increases, thus demonstrating resilience against noise. As noise increases, a divergence of the various probability thresholds is observed. For the imperfect case, the vanishing of the resonant basin results in an increase in the basin area of the nonresonant attractor, jumping from 10% of the total area to almost 15%, for all probability levels. The marked sensitivity to noise of the resonant response is correctly depicted in both perfect and imperfect cases, showing that this measure can be confidently used for a reliable quantification of dynamic integrity of systems under noise.

3.2 Global Dynamics for Parametric Uncertainty

To exemplify the influence of uncertainty in global dynamics, the forcing amplitude $\overline{V}_{ac}$ is considered as a random variable uniformly distributed over a real continuous interval $[a; b]$, $a < b$. That is, it is a sample of the interval $[a; b]$, each value with the same probability of occurring. Referring to the previous notation in the text, the forcing amplitude is a random variable $\overline{V}_{ac} = \lambda$, sampled from the space $\mathbb{L} = [a; b]$, which is equipped with an σ-algebra $\mathfrak{S}$ (the Borel algebra) and measure P_λ with constant density

$$dP_\lambda = p(\lambda) = \frac{1}{b-a}. \tag{15}$$

Which means that $\overline{V}_{ac}$ can have any value in $[a; b]$, each with equal probability. The level of uncertainty depends on the value $b-a$. A large (resp. small) value implies a large (resp. small) uncertainty. Finally, the usual notation which summarize these concepts is $\overline{V}_{ac} \sim U(a, b)$ [15].

A 10-point Gauss-Legendre quadrature is adopted for the integrals in Eq. (4), a common choice for the discretization of uniform random variables [15]. Two cases were

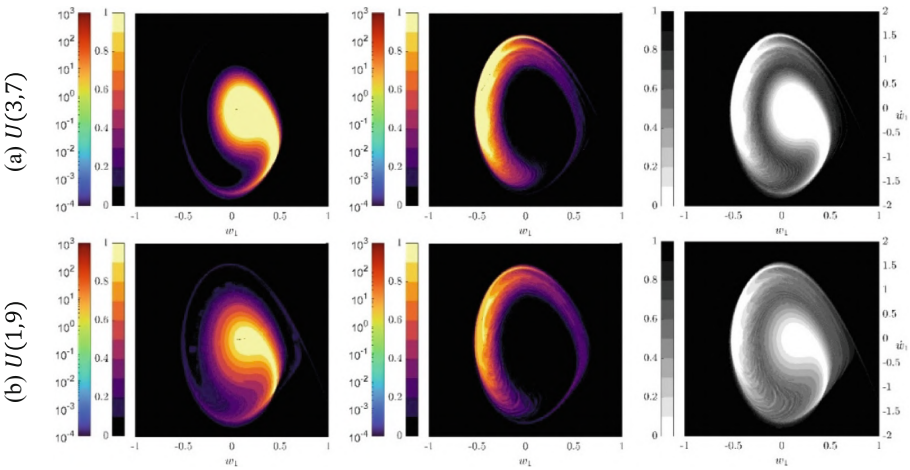


Fig. 12. Microcantilever's global dynamics for $V_{dc} = 45$ V, $\Omega = 2.8$, $w_0 = 0.00$, and uncertainty in $\overline{V}_{ac}$. First color bar: attractor's density, second color bar: probability of convergence.

investigated, with $U(3, 7)$ and $U(1, 9)$. The former represents a mild uncertainty, while the latter is a case with strong uncertainty. The same phase-space window and numeric parameters of the previous analysis are adopted here.

Both uncertainty cases are depicted in Fig. 12, for all attractors and basins. Instead of diffusing over the phase-space, the attractors remain over a line, reminiscent of the bifurcation branch for the forcing amplitude $\overline{V}_{ac}$ [7]. Nevertheless, the basins change drastically from mild to strong uncertainty, Fig. 12a and Fig. 12b, respectively. They diffuse over the phase-space, with decreasing regions of certainty (value 1 in the color scale). All solutions display this pattern, although the resonant attractor seems to be again more affected by uncertainty. Regions of certainty of convergence for this attractor are almost inexistent for $\overline{V}_{ac} \sim U(1, 9)$. Even the escape basin exhibits an increasing uncertainty, although its overall size increases. This result suggests further investigation of the effects of parametric uncertainty in this structure.

4 Final Remarks

The analysis of MEMs dynamics is a multidisciplinary effort, and nondeterministic effects are unavoidable at their size. To address these issues, we revisited and applied a methodology defined in [7] for the analysis of random forcing in a previously explored microcantilever [12]. Two cases were considered, an additive white noise, whose effects have been investigated in detail, and a random forcing amplitude of the alternating current voltage. In the first case, solutions lose stability depending on the noise intensity, becoming long transients with basins fusing with each other or disappearing. A loss of dynamical integrity was verified through the evolution of basins' areas with noise intensity. For the second case, attractors stretch over the bifurcation paths, while their

basins diffuse and the regions with certainty of convergence shrink. Future works will focus on the analysis of uncertainties in other relevant dynamical problems.

References

1. Younis, M.I.: MEMS Linear and Nonlinear Statics and Dynamics. Springer New York (2011). https://doi.org/10.1007/978-1-4419-6020-7
2. Dai, H.L., Wang, L.: Size-dependent pull-in voltage and nonlinear dynamics of electrically actuated microcantilever-based MEMS: a full nonlinear analysis. Commun. Nonlinear Sci. Numer. Simul. **46**, 116–125 (2017). https://doi.org/10.1016/j.cnsns.2016.11.004
3. Hu, Y.C., Chang, C.M., Huang, S.C.: Some design considerations on the electrostatically actuated microstructures. Sens. Actuators A Phys. **112**, 155–161 (2004). https://doi.org/10.1016/J.SNA.2003.12.012
4. Vig, J.R., Kim, Y.: Noise in microelectromechanical system resonators. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 1558–1565 (1999). https://doi.org/10.1109/58.808881
5. Agarwal, N., Aluru, N.R.: Stochastic analysis of electrostatic MEMS subjected to parameter variations. J. Microelectromech. Syst. **18**, 1454–1468 (2009). https://doi.org/10.1109/JMEMS.2009.2034612
6. Benedetti, K.C.B., Gonçalves, P.B., Lenci, S., Rega, G.: Parameter uncertainty and noise effects on the global dynamics of an electrically actuated microarch. J. Micromech. Microeng. **33**, 064001 (2023). https://doi.org/10.1088/1361-6439/acceb0
7. Benedetti, K.C.B., Gonçalves, P.B., Lenci, S., Rega, G.: Global analysis of stochastic and parametric uncertainty in nonlinear dynamical systems: adaptative phase-space discretization strategy, with application to Helmholtz oscillator. Nonlinear Dyn. **111**, 15675–15703 (2023). https://doi.org/10.1007/s11071-023-08667-5
8. Lenci, S., Rega, G. (eds.): Global Nonlinear Dynamics for Engineering Design and System Safety. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-99710-0
9. Ulam, S.M.: Problems in Modern Mathematics. Wiley, New York (1964)
10. Hsu, C.S.: A generalized theory of cell-to-cell mapping for nonlinear dynamical systems. J. Appl. Mech. **48**, 634–642 (1981). https://doi.org/10.1115/1.3157686
11. Hsu, C.S.: A theory of cell-to-cell mapping dynamical systems. J. Appl. Mech. **47**, 931–939 (1980). https://doi.org/10.1115/1.3153816
12. Benedetti, K.C.B., Gonçalves, P.B.: Nonlinear response of an imperfect microcantilever static and dynamically actuated considering uncertainties and noise. Nonlinear Dyn. **107**, 1725–1754 (2022). https://doi.org/10.1007/s11071-021-06600-2
13. Lasota, A., Mackey, M.C.: Chaos, Fractals, and Noise. Springer, New York (1994). https://doi.org/10.1007/978-1-4612-4286-4
14. Lindner, M., Hellmann, F.: Stochastic basins of attraction and generalized committor functions. Phys. Rev. E **100**, 022124 (2019). https://doi.org/10.1103/PhysRevE.100.022124
15. Le Maître, O.P., Knio, O.M.: Spectral Methods for Uncertainty Quantification. Springer, Dordrecht (2010). https://doi.org/10.1007/978-90-481-3520-2
16. Benedetti, K.C.B.: Global Analysis of Stochastic Nonlinear Dynamical Systems: An Adaptative Phase-Space Discretization Strategy. PhD thesis, Pontifical Catholic University of Rio de Janeiro (2022)
17. Brunton, S.L., Budišić, M., Kaiser, E., Kutz, J.N.: Modern Koopman theory for dynamical systems. SIAM Rev. **64**, 229–340 (2022). https://doi.org/10.1137/21M1401243
18. Han, S.M., Benaroya, H., Wei, T.: Dynamics of transversely vibrating beams using four engineering theories. J. Sound Vibr. **225**, 935–988 (1999). https://doi.org/10.1006/jsvi.1999.2257

19. Koltai, P.: Efficient approximation methods for the global long-term behavior of dynamical systems - theory, algorithms and examples. PhD thesis, Technischen Universität München (2010)
20. Benedetti, K.C.B., Gonçalves, P.B., da Silva, F.M.A.: Nonlinear oscillations and bifurcations of a multistable truss and dynamic integrity assessment via a Monte Carlo approach. Meccanica **55**, 2623–2657 (2020). https://doi.org/10.1007/s11012-020-01202-5

Definition and Data-Driven Reconstruction of Asymptotic Phase and Amplitudes of Stochastic Oscillators via Koopman Operator Theory

Shohei Takata[1], Yuzuru Kato[2(✉)], and Hiroya Nakao[1]

[1] Department of Systems and Control Engineering, Tokyo Institute of Technology, Tokyo 152-8552, Japan
nakao@sc.e.titech.ac.jp
[2] Department of Complex and Intelligent Systems, Future University Hakodate, Hakodate, Hokkaido 041-8655, Japan
katoyuzu@fun.ac.jp

Abstract. Asymptotic phase and amplitudes are fundamental concepts in the analysis of limit-cycle oscillators. In this paper, we briefly review the definition of these quantities, particularly a generalization to stochastic oscillatory systems from the viewpoint of Koopman operator theory, and discuss a data-driven approach to estimate the asymptotic phase and amplitude functions from time-series data of stochastic oscillatory systems. We demonstrate that the standard Extended Dynamic Mode Decomposition (EDMD) can successfully reconstruct the phase and amplitude functions of the noisy FitzHugh-Nagumo neuron model only from the time-series data.

Keywords: limit-cycle oscillators · phase-amplitude reduction · Koopman operator · stochastic dynamics · dynamic mode decomposition

1 Introduction

Spontaneous rhythmic oscillations and synchronization are widely observed in the real world. Such rhythmic oscillators are typically modeled as nonlinear dynamical systems possessing stable limit-cycle solutions [1–4]. The *asymptotic phase* and *amplitudes* are fundamental quantities for analyzing and controlling the dynamics of limit-cycle oscillators [2,5–9]. Recently, it has been shown that the asymptotic phase and amplitudes can be systematically defined based on Koopman operator theory [10,11] by using the Koopman eigenfunctions of the limit-cycle oscillator [12], and this definition has further been extended to stochastic oscillators [13,14]. In this paper, we briefly review the definition of the asymptotic phase and amplitudes and discuss data-driven estimation of these quantities from time-series data. We apply the standard Extended Dynamic

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 141–153, 2025.
https://doi.org/10.1007/978-3-031-72794-8_10

Mode Decomposition (EDMD) [15–17] to stochastic oscillators, which reconstructs the Koopman eigenvalues and eigenfunctions from time-series data, and show that the asymptotic phase and amplitudes of the noisy FitzHugh-Nagumo neuron model can be successfully reproduced.

2 Asymptotic Phase and Amplitudes for Deterministic Limit-Cycle Oscillators

2.1 Asymptotic Phase and Amplitudes

First, we provide a brief review of the concepts of asymptotic phase and amplitudes, along with their definitions from the Koopman operator viewpoint.

The asymptotic phase has been a fundamental concept in the analysis of limit-cycle oscillators [1,2,6,18]. Let us consider a limit-cycle oscillator, i.e., a dynamical system possessing a stable limit-cycle solution described by

$$\frac{d\boldsymbol{x}(t)}{dt} = \boldsymbol{f}(\boldsymbol{x}(t)), \tag{1}$$

where $\boldsymbol{x}(t) \in \mathbb{R}^N$ is the state of the system at time t and $\boldsymbol{f} : \mathbb{R}^N \to \mathbb{R}^N$ is a smooth vector field describing the system dynamics. We assume that Eq. (1) has an exponentially stable limit-cycle solution $\boldsymbol{x}_0(t)$, satisfying $\boldsymbol{x}_0(t+T) = \boldsymbol{x}_0(t)$, where T is the period of the limit cycle. We define the frequency as $\omega = 2\pi/T$. We denote the limit cycle by $\chi = \{\boldsymbol{x}_0(t) \mid 0 \le t < T\}$ and assign a phase θ $(0 \le \theta < 2\pi)$ to each state $\boldsymbol{x}_0(t = \theta/\omega)$ on χ, where 0 and 2π are identified. The linear stability of χ is characterized by the Floquet exponents $\lambda_m \in \mathbb{C}$ $(m = 1, ..., N-1)$ [19], whose real parts are negative as χ is assumed to be stable. We sort the exponents λ_m in decreasing order of their real parts.

In the basin of attraction $B \subset \mathbb{R}^N$ of the limit cycle χ, the asymptotic phase function $\Theta : B \to [0, 2\pi)$, which maps a system state $\boldsymbol{x} \in B$ to a phase value in $[0, 2\pi)$, is defined such that

$$\boldsymbol{f}(\boldsymbol{x}) \cdot \nabla\Theta(\boldsymbol{x}) = \omega \tag{2}$$

is satisfied for any $\boldsymbol{x} \in B$, where $\cdot$ represents the dot product of two vectors and $\nabla = \partial/\partial\boldsymbol{x}$ is the gradient of a scalar function of $\boldsymbol{x}$. It can then easily be shown that, for any state $\boldsymbol{x}(t) \in B$, the phase $\theta(t) = \Theta(\boldsymbol{x}(t))$ of $\boldsymbol{x}(t)$ evolves with time as

$$\frac{d\theta(t)}{dt} = \left.\frac{\partial\Theta}{\partial\boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}(t)} \cdot \frac{d\boldsymbol{x}(t)}{dt} = \boldsymbol{f}(\boldsymbol{x}(t)) \cdot \nabla\Theta(\boldsymbol{x}(t)) = \omega \tag{3}$$

by the chain rule of differentiation. Thus, the asymptotic phase $\theta(t) = \Theta(\boldsymbol{x}(t))$ of the system state $\boldsymbol{x}(t)$ always increases with a constant frequency ω for any trajectory in B. Since all states in B are eventually attracted to the limit cycle χ, all states on the same level set of the asymptotic phase Θ converge to the same state on χ, because their phase values are always kept equal by Eq. (3). From a

geometrical viewpoint, the above definition of the asymptotic phase is equivalent to assigning the same phase value to the set of system states that eventually converge to the same phase on χ. Such level sets are called *isochrons* [1,2,20,21]. The asymptotic phase has played a fundamental role in analyzing the dynamics of limit-cycle oscillators [1,2,4,6,18].

In a similar manner to the asymptotic phase function, we can also introduce the amplitude functions $R_m : B \to \mathbb{C}$ $(m = 1, ..., N-1)$, which map a system state $\boldsymbol{x} \in B$ to (generally complex) amplitudes such that

$$\boldsymbol{f}(\boldsymbol{x}) \cdot \nabla R_m(\boldsymbol{x}) = \lambda_m R_m(\boldsymbol{x}) \tag{4}$$

holds for any $\boldsymbol{x} \in B$, where $\lambda_m \in \mathbb{C}$ are the Floquet exponents of χ. As we explain in the next subsection, these $R_m(\boldsymbol{x})$ are the Koopman eigenfunctions of Eq. (1) associated with the Koopman eigenvalues λ_m. It can then easily be shown that, for an arbitrary state $\boldsymbol{x}(t) \in B$, each amplitude $r_m(t) = R_m(\boldsymbol{x}(t))$ of the state $\boldsymbol{x}(t)$ evolves with time as

$$\frac{dr_m(t)}{dt} = \left.\frac{\partial R_m(\boldsymbol{x})}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}(t)} \cdot \frac{d\boldsymbol{x}(t)}{dt} = \boldsymbol{f}(\boldsymbol{x}(t)) \cdot \nabla R_m(\boldsymbol{x}(t)) = \lambda_m r_m(t). \tag{5}$$

Since any $\boldsymbol{x}(t) \in B$ converges to χ, $R_m(\boldsymbol{x}_0(t)) = 0$ should hold for any $\boldsymbol{x}_0(t) \in \chi$. Thus, the amplitude $r_m(t) = R_m(\boldsymbol{x}(t))$ characterizes the deviation of the state $\boldsymbol{x}(t)$ from χ, which always decays linearly with a constant rate λ_m determined by the Floquet exponent for any trajectory. The level sets of R_m are called *isostables* [12].

The asymptotic phase function Θ and the amplitude functions R_m define a nonlinear coordinate transformation from B to $[0, 2\pi) \times \mathbb{C}^{N-1}$, which gives a globally linearized representation, $\dot{\Theta} = \omega$ and $\dot{R}_m = \lambda R_m$, of the original nonlinear dynamics $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$ in B. Moreover, by eliminating some of the fast-decaying amplitudes, i.e., assuming $R_m = 0$ for some ranges of m corresponding to sufficiently small Re λ_m, we can choose only the slowly decaying amplitudes and approximately reduce the dimensionality of the system. Applications of the reduced phase-amplitude equations for analyzing and controlling of limit cycles have attracted attention recently [5,6,8,22–25].

2.2 Koopman Eigenvalues and Eigenfunctions

The concepts of asymptotic phase and amplitudes are closely related to Koopman eigenfunctions. Indeed, the amplitude functions have been explicitly defined as the eigenfunctions of the Koopman operator only recently in [12], despite similar concepts of amplitudes having often been used in the past [2,7,26–29].

We denote the flow of the limit-cycle oscillator described by Eq. (1) as S^τ, satisfying $\boldsymbol{x}(t+\tau) = S^\tau \boldsymbol{x}(t)$, where $\tau \geq 0$. The Koopman operator U^τ of Eq. (1) is then defined as

$$(U^\tau g)(\boldsymbol{x}) = g(S^\tau \boldsymbol{x}) \tag{6}$$

for $\boldsymbol{x} \in B$, where $g : B \to \mathbb{C}$ is an observable of the system that assigns a (generally complex) value $g(\boldsymbol{x})$ to the system state $\boldsymbol{x}$. The infinitesimal generator of U^τ is defined as

$$Ag(\boldsymbol{x}) = \lim_{\tau \to +0} \frac{U^\tau g(\boldsymbol{x}) - g(\boldsymbol{x})}{\tau} = \lim_{\tau \to +0} \frac{g(S^\tau \boldsymbol{x}) - g(\boldsymbol{x})}{\tau} = \boldsymbol{f}(\boldsymbol{x}) \cdot \nabla g(\boldsymbol{x}), \quad (7)$$

where we assumed that $g(\boldsymbol{x})$ can be expanded in $\boldsymbol{x}$ and used that $S^\tau \boldsymbol{x} = \boldsymbol{x} + \boldsymbol{f}(\boldsymbol{x})\tau + O(\tau^2)$ and $g(S^\tau \boldsymbol{x}) = g(\boldsymbol{x}) + \boldsymbol{f}(\boldsymbol{x})\tau \cdot \nabla g(\boldsymbol{x}) + O(\tau^2)$ for sufficiently small $\tau \geq 0$. It can easily be shown that U^τ and A are linear operators even if $\boldsymbol{f}(\boldsymbol{x})$ is a nonlinear function of $\boldsymbol{x}$ [10,11]. The eigenvalue $\tilde{\lambda} \in \mathbb{C}$ and associated eigenfunction $\phi_{\tilde{\lambda}}(\boldsymbol{x}) : B \to \mathbb{C}$ of A satisfying

$$A\phi_{\tilde{\lambda}}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}) \cdot \nabla \phi_{\tilde{\lambda}}(\boldsymbol{x}) = \tilde{\lambda}\phi_{\tilde{\lambda}}(\boldsymbol{x}) \quad (8)$$

are called the Koopman eigenvalue and Koopman eigenfunction, respectively.

For an exponentially stable limit cycle, it is known that the set of Koopman eigenvalues includes $i\omega$ and the Floquet exponents λ_m $(m = 1, ..., N-1)$, which are called the principal Koopman eigenvalues, and analytic observables can be expanded by using the associated principal Koopman eigenfunctions and their products [30]. From Eqs. (2) and (4), the complex exponential of the asymptotic phase, $\Phi(\boldsymbol{x}) = e^{i\Theta(\boldsymbol{x})}$, and the amplitudes R_m $(m = 1, ..., N-1)$ satisfy

$$A\Phi(\boldsymbol{x}) = i\omega\Phi(\boldsymbol{x}), \quad AR_m(\boldsymbol{x}) = \lambda_m R_m(\boldsymbol{x}). \quad (9)$$

Thus, the asymptotic phase $\Theta(\boldsymbol{x})$ is the argument of the Koopman eigenfunction $\Phi(\boldsymbol{x})$ with the eigenvalue $i\omega$, i.e., $\Theta(\boldsymbol{x}) = \mathrm{Arg}\ \Phi(\boldsymbol{x})$, and the amplitudes $R_m(\boldsymbol{x})$ are nothing but the Koopman eigenfunctions with the principal Koopman eigenvalues λ_m given by the Floquet exponents.

Thus, recent developments in Koopman operator theory for dissipative dynamical systems have clarified that the asymptotic phase of the limit cycle, originally introduced from a geometrical viewpoint [20], can be defined using the Koopman eigenfunction, and also enabled to define the amplitudes of limit cycle explicitly as eigenfunctions of the Koopman operator in a natural and unified manner. This motivates us to define the asymptotic phase and amplitudes for stochastic oscillatory systems also from the Koopman operator viewpoint.

3 Asymptotic Phase and Amplitudes for Stochastic Oscillatory Systems

3.1 Stochastic Oscillators

In the real world, there exist many examples of stochastic oscillatory systems in which noise plays essential roles in their dynamics. In this section, we review the definitions of asymptotic phase and amplitudes for stochastic oscillators that we recently proposed based on Koopman operator theory [14].

We consider a stochastic oscillator described by the following Itô stochastic differential equation (SDE):

$$d\boldsymbol{x}(t) = \boldsymbol{A}(\boldsymbol{x}(t))dt + \boldsymbol{B}(\boldsymbol{x}(t))d\boldsymbol{W}(t), \tag{10}$$

where $\boldsymbol{x}(t) \in \mathbb{R}^N$ represents the system state at time t, $\boldsymbol{A}(\boldsymbol{x}) \in \mathbb{R}^N$ is a vector field representing the deterministic part of the dynamics, $\boldsymbol{B}(\boldsymbol{x}) \in \mathbb{R}^{N\times N}$ is a matrix representing the noise intensity, and $\boldsymbol{W}(t) \in \mathbb{R}^N$ is a N-dimensional Wiener process [31,32]. We assume that $\boldsymbol{A}(\boldsymbol{x})$ and $\boldsymbol{B}(\boldsymbol{x})$ satisfy appropriate conditions on smoothness and boundedness [32]. Since we consider stochastic oscillators, we assume that the deterministic part $\boldsymbol{A}(\boldsymbol{x})$ possesses an exponentially stable limit-cycle solution as before. By the effect of noise characterized by $\boldsymbol{B}(\boldsymbol{x})$, the average properties of the stochastic oscillator, e.g., the period, shape of the trajectory, response properties, can differ from the deterministic case.

As an example, we will consider a stochastic FitzHugh-Nagumo model of a spiking neuron [14] described by the following Itô SDE:

$$\begin{aligned} dx &= (x - a_1x^3 - y)dt + \sqrt{D_x}dW_x, \\ dy &= \eta_1(x + b_1)dt + \sqrt{D_y}dW_y, \end{aligned} \tag{11}$$

where x and y represent real variables, a_1, b_1, and η_1 are parameters of the system, W_x and W_y represent independent Wiener processes, and D_x and D_y are the intensities of the noise, respectively. The FitzHugh-Nagumo model is a two-dimensional model of spiking neurons simplified from the more realistic Hodgkin-Huxley model with four variables [4]. We set the parameters as $(a_1, b_1, \eta_1, D_x, D_y) = (1/3, 0.5, 0.5, 0.2, 0.2)$ [14], with which the deterministic part of the system possesses a stable limit-cycle solution.

Figure 1 shows an example of the stochastic trajectory of Eq. (11), simulated by the Euler-Maruyama method with a time step 10^{-3} and sampled with a time interval $\tau = 0.1$. Due to relatively strong noise, the trajectory of the oscillator shows strong fluctuations.

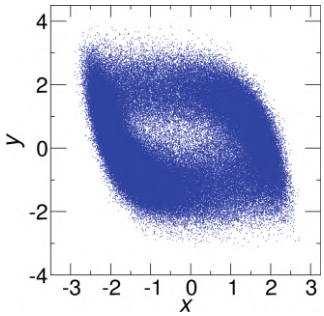


Fig. 1. Time-series data obtained from the stochastic FitzHugh-Nagumo model.

3.2 Stochastic Koopman Operator

The Koopman operator for the stochastic oscillator described by Eq. (10) is introduced as follows. The transition probability density $p(\boldsymbol{x}, t|\boldsymbol{y}, s)$ $(t \geq s)$ of Eq. (10) obeys the forward and backward Fokker-Planck equations [31],

$$\frac{\partial}{\partial t}p(\boldsymbol{x}, t|\boldsymbol{y}, s) = L_{\boldsymbol{x}}p(\boldsymbol{x}, t|\boldsymbol{y}, s), \quad \frac{\partial}{\partial s}p(\boldsymbol{x}, t|\boldsymbol{y}, s) = -L_{\boldsymbol{y}}^{*}p(\boldsymbol{x}, t|\boldsymbol{y}, s), \tag{12}$$

where the forward and backward Fokker-Planck operators are given by

$$L_{\boldsymbol{x}} = -\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{A}(\boldsymbol{x}) + \frac{1}{2}\frac{\partial^2}{\partial \boldsymbol{x}^2}\boldsymbol{D}(\boldsymbol{x}), \quad L_{\boldsymbol{x}}^{*} = \boldsymbol{A}(\boldsymbol{x})\frac{\partial}{\partial \boldsymbol{x}} + \frac{1}{2}\boldsymbol{D}(\boldsymbol{x})\frac{\partial^2}{\partial \boldsymbol{x}^2}. \tag{13}$$

Here, $\boldsymbol{D}(\boldsymbol{x}) = \boldsymbol{B}(\boldsymbol{x})\boldsymbol{B}(\boldsymbol{x})^{\mathsf{T}} \in \mathbb{R}^{N\times N}$ is a matrix of diffusion coefficients with T representing the matrix transposition. The forward and backward operators $L_{\boldsymbol{x}}$ and $L_{\boldsymbol{x}}^{*}$ are mutually adjoint, i.e., $\langle L_{\boldsymbol{x}}G(\boldsymbol{x}), H(\boldsymbol{x})\rangle_{\boldsymbol{x}} = \langle G(\boldsymbol{x}), L_{\boldsymbol{x}}^{*}H(\boldsymbol{x})\rangle_{\boldsymbol{x}}$, where the inner product is defined as $\langle G(\boldsymbol{x}), H(\boldsymbol{x})\rangle_{\boldsymbol{x}} = \int \overline{G(\boldsymbol{x})}H(\boldsymbol{x})d\boldsymbol{x}$ for two functions $G(\boldsymbol{x}), H(\boldsymbol{x}) : \mathbb{R}^N \to \mathbb{C}$ with the overline indicating complex conjugate, and the integration is taken over the whole range of $\boldsymbol{x}$.

The linear differential operators $L_{\boldsymbol{x}}$ and $L_{\boldsymbol{x}}^{*}$ have a biorthogonal eigensystem $\{\mu_k, P_k, Q_k\}_{k=0,1,2,\ldots}$ of eigenvalues μ_k and eigenfunctions $P_k(\boldsymbol{x})$ and $Q_k(\boldsymbol{x})$ $(k = 0, 1, 2, ...)$, satisfying

$$L_{\boldsymbol{x}}P_k(\boldsymbol{x}) = \mu_k P_k(\boldsymbol{x}), \quad L_{\boldsymbol{x}}^{*}Q_k(\boldsymbol{x}) = \overline{\mu_k}Q_k(\boldsymbol{x}), \quad \langle P_k(\boldsymbol{x}), Q_l(\boldsymbol{x})\rangle_{\boldsymbol{x}} = \delta_{kl}, \tag{14}$$

where $k, l = 0, 1, 2, \ldots$ and δ_{kl} is the Kronecker delta [31]. We sort the eigenvalues μ_k in decreasing order of their real parts. Among the eigenvalues, one eigenvalue is zero, $\mu_0 = 0$, which is associated with the stationary probability density function $P_0(\boldsymbol{x})$ of the system satisfying $L_{\boldsymbol{x}}P_0(\boldsymbol{x}) = 0$, and all other eigenvalues have negative real parts.

For Eq. (10), the stochastic Koopman operator U_{st} for an observable $g : \mathbb{R}^N \to \mathbb{C}$ is defined as [33]

$$U_{st}^{\tau}g(\boldsymbol{x}) = \mathbb{E}[g(S_{st}^{\tau}\boldsymbol{x})] = \int p(\boldsymbol{y}, \tau|\boldsymbol{x}, 0)g(\boldsymbol{y})d\boldsymbol{y}, \tag{15}$$

where S_{st}^{τ} is a stochastic flow of Eq. (10) and $\mathbb{E}$ represents the expectation over realizations of S_{st}^{τ}, and the infinitesimal generator of U_{st}^{τ} is defined as

$$A_{st}g(\boldsymbol{x}) = \lim_{\tau\to+0}\frac{U_{st}^{\tau}g(\boldsymbol{x}) - g(\boldsymbol{x})}{\tau} = \lim_{\tau\to+0}\frac{\mathbb{E}[g(S_{st}^{\tau}\boldsymbol{x})] - g(\boldsymbol{x})}{\tau}. \tag{16}$$

It can be shown that the backward Fokker-Planck operator $L_{\boldsymbol{x}}^{*}$ in Eq. (13) is the infinitesimal generator of the stochastic Koopman operator [14,17], i.e.,

$$A_{st} = L_{\boldsymbol{x}}^{*} = \boldsymbol{A}(\boldsymbol{x})\frac{\partial}{\partial \boldsymbol{x}} + \frac{1}{2}\boldsymbol{D}(\boldsymbol{x})\frac{\partial^2}{\partial \boldsymbol{x}^2}. \tag{17}$$

Indeed, from the Itô formula [31], $g(\boldsymbol{x}(t))$ obeys

$$dg(\boldsymbol{x}(t)) = \left(\boldsymbol{A}(\boldsymbol{x})\frac{\partial g}{\partial \boldsymbol{x}} + \frac{1}{2}\boldsymbol{D}(\boldsymbol{x})\frac{\partial^2 g}{\partial \boldsymbol{x}^2}\right)dt + \left(\frac{\partial g}{\partial \boldsymbol{x}}\right)^{\top}\boldsymbol{B}(\boldsymbol{x})d\boldsymbol{W}(t), \tag{18}$$

and, using $U_{st}^{dt}g(\boldsymbol{x}) = g(S_{st}^{dt}\boldsymbol{x}) = g(\boldsymbol{x}) + dg(\boldsymbol{x})$, we obtain

$$A_{st}g(\boldsymbol{x}) = \lim_{\tau\to+0}\frac{\mathbb{E}[g(\boldsymbol{x}) + dg(\boldsymbol{x})] - g(\boldsymbol{x})}{\tau} = \boldsymbol{A}(\boldsymbol{x})\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}} + \frac{1}{2}\boldsymbol{D}(\boldsymbol{x})\frac{\partial^2 g(\boldsymbol{x})}{\partial \boldsymbol{x}^2}. \tag{19}$$

We note that $\mathbb{E}$ operates only on $dg(\boldsymbol{x})$. If the noise is absent, $\boldsymbol{D}(\boldsymbol{x}) = 0$, the generator A_{st} will take the form $\boldsymbol{A}(\boldsymbol{x})\frac{\partial}{\partial \boldsymbol{x}} = \boldsymbol{A}(\boldsymbol{x})\cdot\nabla$, which is the Koopman operator A for the deterministic case.

3.3 Asymptotic Phase and Amplitudes for Stochastic Oscillators

How to define the asymptotic phase for stochastic oscillators has been discussed in several studies [13,34]. Recently, we proposed general definitions of the asymptotic phase and amplitudes for stochastic oscillators from the Koopman operator viewpoint [14], where the deterministic definitions in Eq. (9) using the Koopman eigenfunctions were naturally generalized to stochastic oscillators. In [35,36], we further developed similar ideas for quantum nonlinear oscillators.

As we are considering stochastic oscillators, we assume that the eigenvalues of $L_{\boldsymbol{x}}$ and $L_{\boldsymbol{x}}^*$ with the largest non-zero real part appear as a complex conjugate pair, denoted by μ_1 and $\mu_2 = \overline{\mu_1}$, where the real part Re μ_1 characterizes the decay rate and the imaginary part $\Omega = \mathrm{Im}\ \overline{\mu_1}$ characterizes the oscillation frequency (note that $\mu_0 = 0$). We may also take Im μ_1 as Ω, which reverses the direction of the phase. We focus on this slowest-decaying oscillatory mode of the transition probability density and define the system's phase with respect to this mode. In general, we obtain a branch of complex-conjugate eigenvalues of $L_{\boldsymbol{x}}^*$ in addition to the above slowest-decaying fundamental mode, but we do not consider those modes because their imaginary parts are essentially integer multiples of Ω and are not independent from the fundamental mode. For the amplitude, we choose a non-zero eigenvalue μ_r with the largest real part that is not included in the above branch as the independent, second-slowest decaying mode, which corresponds to the slowest amplitude in the noiseless deterministic limit [25]. We note that μ_r may be a complex value when $N \geq 3$.

In [14], focusing on these Koopman eigenvalues μ_1 and μ_r of $A_{st} = L_{\boldsymbol{x}}^*$, we proposed to use the associated eigenfunctions $Q_1(\boldsymbol{x})$ and $Q_r(\boldsymbol{x})$ to define the asymptotic phase and the slowest-decaying amplitude of a stochastic oscillator, respectively. That is, we consider Arg $Q_1(\boldsymbol{x})$ as the asymptotic phase, i.e., $Q_1(\boldsymbol{x})$ as the complex exponential of the asymptotic phase, and $Q_r(\boldsymbol{x})$ as the amplitude, in a similar manner to the deterministic case. We note that we consider only the asymptotic phase and the slowest-decaying amplitude here.

We can then show that the averaged asymptotic phase and the slowest-decaying amplitude of the system at time t, given by

$$\theta(t) = \mathrm{Arg}\ \mathbb{E}[Q_1(\boldsymbol{x})], \quad r(t) = \mathbb{E}[Q_r(\boldsymbol{x})], \tag{20}$$

satisfy

$$\frac{d\theta(t)}{dt} = \Omega, \quad \frac{dr(t)}{dt} = \mu_r r(t), \tag{21}$$

where $\Omega = \mathrm{Im}\,\overline{\mu_1}$ and μ_r characterize the average oscillation frequency and decay rate of the phase and amplitude, respectively. In deriving the above equations, we used the fact that the eigenfunctions $Q_k(\boldsymbol{x})$ of $A_{st} = L^*_{\boldsymbol{x}}$ generally satisfy

$$\frac{d}{dt}\mathbb{E}[Q_k(S^t_{st}\boldsymbol{x})] = \mu_k \mathbb{E}[Q_k(S^t_{st}\boldsymbol{x})] \tag{22}$$

for $k = 1, 2, 3, ...$, which can easily be shown by taking the time derivative of the expectation $\mathbb{E}[Q_k(S^t_{st}\boldsymbol{x})] = \int p(\boldsymbol{y}, \tau|\boldsymbol{x}, 0)Q_k(\boldsymbol{y})d\boldsymbol{y}$ [14].

Thus, we proposed definitions of the asymptotic phase that, on average, increases with a constant frequency Ω and of the amplitude that, on average, decay linearly with a constant decay rate μ_r with the stochastic evolution of $\boldsymbol{x}$, for stochastic oscillatory systems. They can be interpreted as natural generalizations of the corresponding definitions for deterministic oscillators. The Koopman eigenfunctions $Q_1(\boldsymbol{x})$ and $Q_r(\boldsymbol{x})$ for the stochastic oscillator play essentially similar roles to the Koopman eigenfunctions $\phi_0(\boldsymbol{x})$ and $\phi_1(\boldsymbol{x})$ for the deterministic oscillator. Our definition of the asymptotic phase and amplitudes based on the Koopman eigenfunctions coincides with the definition based on the first-passage time problem by Thomas and Lindner [13] and Pérez-Cervera *et al.* [37], respectively. Recently, Pérez-Cervera *et al.* [38] also introduced a description of stochastic oscillators using a complex Koopman eigenfunction $Q_1(\boldsymbol{x})$.

4 Data-Driven Estimation of the Asymptotic Phase and Amplitude Functions

4.1 Extended Dynamic Mode Decomposition

In this section, we aim to reconstruct the asymptotic phase and amplitude functions in a data-driven manner using the Extended Dynamic Mode Decomposition (EDMD), which is a standard method for obtaining the leading Koopman eigenvalues and eigenfunctions from time-series data [15–17,39].

First, we briefly review the algorithm of EDMD. We consider a dictionary consisting of N_k basis functions (observables) ψ_1, ..., $\psi_{N_k} : \mathbb{R}^N \to \mathbb{R}$ and arrange them in the vector form as

$$\boldsymbol{\psi}(\boldsymbol{x}) = [\psi_1(\boldsymbol{x}), \psi_2(\boldsymbol{x}), \cdots, \psi_{N_k}(\boldsymbol{x})]^{\mathsf{T}} \in \mathbb{R}^{N_k \times 1}. \tag{23}$$

The action of the stochastic Koopman operator U^τ_{st} on $\boldsymbol{\psi}(\boldsymbol{x})$ is

$$U^\tau_{st}\boldsymbol{\psi}(\boldsymbol{x}) = \mathbb{E}[\boldsymbol{\psi}(S^\tau_{st}\boldsymbol{x})]. \tag{24}$$

We seek the optimal matrix $K_{st} \in \mathbb{R}^{N_k \times N_k}$ that gives the best finite-dimensional approximation to $U^\tau_{st}\boldsymbol{\psi}(\boldsymbol{x})$, i.e.,

$$U^\tau_{st}\boldsymbol{\psi}(\boldsymbol{x}) \approx K_{st}\boldsymbol{\psi}(\boldsymbol{x}). \tag{25}$$

The bias of the above equation is $\mathbb{E}[\boldsymbol{\psi}(S_{st}^{\tau}\boldsymbol{x})]-K_{st}\boldsymbol{\psi}(\boldsymbol{x}) = \mathbb{E}[\boldsymbol{\psi}(S_{st}^{\tau}\boldsymbol{x})-K_{st}\boldsymbol{\psi}(\boldsymbol{x})]$, where $\mathbb{E}$ operates only on S_{st}^{τ}.

To find the matrix K_{st} from observed time-series data, we introduce two matrices $Y, Y' \in \mathbb{R}^{N_k \times M-1}$ consisting of $M-1$ data points sampled with an interval of $\tau > 0$,

$$Y = [\boldsymbol{\psi}(\boldsymbol{x}_1), \boldsymbol{\psi}(\boldsymbol{x}_2), \cdots, \boldsymbol{\psi}(\boldsymbol{x}_{M-1})], \quad Y' = [\boldsymbol{\psi}(\boldsymbol{x}_2), \boldsymbol{\psi}(\boldsymbol{x}_3), \cdots, \boldsymbol{\psi}(\boldsymbol{x}_M)], \quad (26)$$

where $\boldsymbol{x}_k = \boldsymbol{x}(k\tau)$. Since $\boldsymbol{x}_{k+1} = S_{st}^{\tau}\boldsymbol{x}_k$, the mean-squared approximation error in Eq. (25) can be estimated from the time-series data as

$$\begin{aligned}\mathbb{E}\left[\|\boldsymbol{\psi}(S_{st}^{\tau}\boldsymbol{x}) - K_{st}\boldsymbol{\psi}(\boldsymbol{x})\|^2\right] &\approx \frac{1}{M-1}\sum_{k=1}^{M-1}\|\boldsymbol{\psi}(\boldsymbol{x}_{k+1}) - K_{st}\boldsymbol{\psi}(\boldsymbol{x}_k)\|^2 \\ &= \frac{1}{M-1}\|Y' - K_{st}Y\|_F^2, \end{aligned} \quad (27)$$

where the Frobenius norm of a matrix $C \in \mathbb{R}^{N_k \times (M-1)}$ is defined as $\|C\|_F = \sqrt{\sum_{j=1}^{N_k}\sum_{k=1}^{M-1} C_{jk}^2}$. Thus, the optimal matrix K_{st} is obtained as

$$K_{st} = \arg\min_{\tilde{K}} \left\|Y' - \tilde{K}Y\right\|_F^2, \quad (28)$$

which is explicitly given by

$$K_{st} = Y'Y^{+}. \quad (29)$$

Here, $Y^{+} = Y^{\mathsf{T}}(YY^{\mathsf{T}})^{-1}$ represents the Moore-Penrose pseudoinverse of Y. In practice, we also add a small regularization term proportional to $\|K\|_F^2$ in Eq. (28) to avoid overfitting and obtain the optimal matrix K by

$$K_{st} = \arg\min_{\tilde{K}} \left\|Y' - \tilde{K}Y\right\|_F^2 + \alpha\left\|\tilde{K}\right\|_F^2, \quad (30)$$

where α is a regularization parameter [40]. We note that $\boldsymbol{x}_{k+1} = S_{st}^{\tau}\boldsymbol{x}_k$ represents the stochastic evolution of $\boldsymbol{x}_k$ determined by the Itô SDE (10). Thus, K_{st} estimated by Eq. (30) corresponds to the stochastic Koopman operator U_{st}^{τ} and the infinitesimal generator A_{st}, i.e., the backward Fokker-Planck operator L_x^*.

The approximate Koopman eigenfunctions can be obtained as follows. Using the basis functions $\boldsymbol{\psi}(\boldsymbol{x})$, we approximately represent the Koopman eigenfunction $Q_k(\boldsymbol{x})$ associated with the eigenvalue Λ_k as

$$Q_k(\boldsymbol{x}) \approx \sum_{j=1}^{N_k} w_j^{(k)}\psi_j(\boldsymbol{x}) = \boldsymbol{w}_k\boldsymbol{\psi}(\boldsymbol{x}), \quad (31)$$

where $\boldsymbol{w}_k = [w_1^{(k)}, w_2^{(k)}, \cdots, w_{N_k}^{(k)}] \in \mathbb{R}^{1\times N_k}$ represents the expansion coefficients. Plugging the above expansion into the eigenvalue equation $U_{st}^{\tau}Q_k(\boldsymbol{x}) = \Lambda_k Q_k(\boldsymbol{x})$ for the stochastic Koopman operator, we obtain

$$U_{st}^{\tau}Q_k(\boldsymbol{x}) \approx U_{st}^{\tau}\{\boldsymbol{w}_k\boldsymbol{\psi}(\boldsymbol{x})\} \approx \boldsymbol{w}_k K_{st}\boldsymbol{\psi}(\boldsymbol{x}) \approx \Lambda_k\boldsymbol{w}_k\boldsymbol{\psi}(\boldsymbol{x}). \quad (32)$$

Thus, we can approximately obtain Λ_k and $\boldsymbol{w}_k$ from the eigenvalue equation for K_{st},

$$\boldsymbol{w}_k K_{st} = \Lambda_k \boldsymbol{w}_k. \tag{33}$$

The eigenvalue Λ_k approximates the eigenvalue of U_{st}^{τ}, from which we approximately obtain the Koopman eigenvalue $\mu_k = (1/\tau) \ln \Lambda_k$ of $A_{st} = L_{\boldsymbol{x}}^*$ and the associated Koopman eigenfunction by Eq. (31).

Thus, the EDMD gives a finite-dimensional approximation to the Koopman operator, eigenvalues, and eigenfunctions projected on the basis $\boldsymbol{\psi}(\boldsymbol{x})$. The estimated eigenvalue $\overline{\mu_1}$ gives the frequency Ω of the stochastic oscillation, the eigenvalue μ_r gives the decay rate of the slowest amplitude, and the associated eigenfunctions $Q_1(\boldsymbol{x})$ and $Q_r(\boldsymbol{x})$ give the estimates of the asymptotic phase function and amplitude function, respectively.

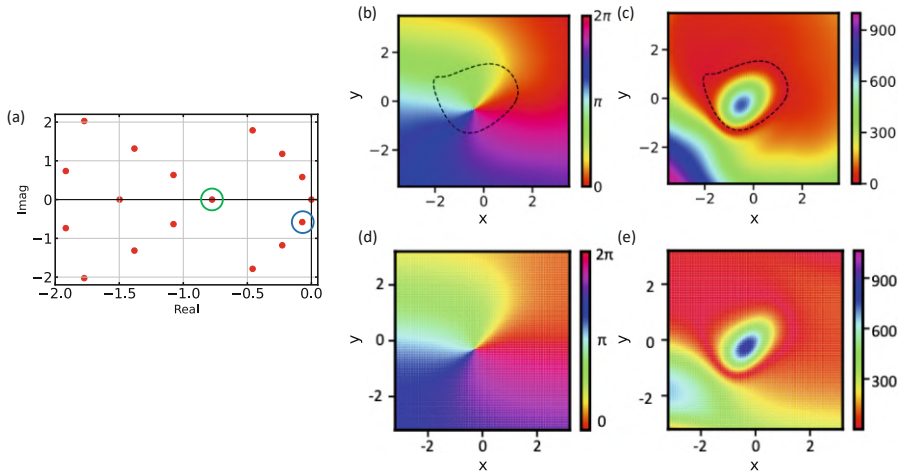


Fig. 2. Reconstruction of the asymptotic phase and the slowest-decaying amplitude functions of the stochastic FitzHugh-Nagumo model. (a) Eigenvalues of A_{st}. (b) Phase function Arg $Q_1(\boldsymbol{x})$, and (c) modulus of the amplitude function $|Q_r(\boldsymbol{x})|$ reconstructed by EDMD. The zero-level set of the amplitude function $Q_r(\boldsymbol{x})$ representing the stochastic periodic orbit of the system is drawn by the black-dotted line. (d) Phase function Arg $Q_1(\boldsymbol{x})$ and (e) modulus of the amplitude function $|Q_r(\boldsymbol{x})|$ obtained from the Koopman operator [14].

4.2 Example: Noisy FitzHugh-Nagumo Oscillator

As an example, we consider a stochastic FitzHugh-Nagumo model, Eq. (11), and reconstruct the asymptotic phase and amplitude functions, i.e., Arg $Q_1(\boldsymbol{x})$ and $Q_r(\boldsymbol{x})$, from time-series data by EDMD. We observed a single trajectory of

Eq. (11) with a sampling interval of $\tau = 10^{-1}$ (see Fig. 1). Following [15], we used 1000 radial basis functions (RBFs) $\psi_j(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{c}_j\|^2 \ln(\|\boldsymbol{x} - \boldsymbol{c}_j\| + \varepsilon)$ ($j = 1, \ldots, 1000$) as the observation functions, where the centers $\boldsymbol{c}_j$ are determined by k-mean clustering of the sampled data and $\epsilon = 10^{-4}$, to construct the two data matrices in Eq. (26) (see [15] for the details of the RBFs). The regularization parameter in Eq. (30) was set as $\alpha = 10^{-3}$. In the numerical analysis, we confirmed that 10^6 data points were enough for reconstructing the phase and amplitude functions with appropriate accuracy and that reasonably accurate reconstruction was also possible with only 10^5 data points.

Figure 2(a) shows the eigenvalues of the infinitesimal generator of the Koopman operator $A_{st} = L_{\boldsymbol{x}}^*$ estimated from the time series, where μ_1 and μ_r are indicated by blue and green circles, respectively. We can observe that the Koopman eigenvalues with the largest non-zero real parts are mutually complex conjugate as assumed, which correspond to μ_1 and $\overline{\mu_1}$, respectively. We choose the real eigenvalue μ_r belonging to the next branch for the slowest-decaying amplitude. The average frequency Ω of the phase is given by the imaginary part of $\overline{\mu_1}$ and estimated as $\Omega \approx 0.581$, and the decay rate of the amplitude μ_r is given by the largest negative real eigenvalue and estimated as $\mu_r \approx -0.775$. The reconstructed phase and amplitude functions are shown in Fig. 2(b) and (c). In these figures, the zero-level set of the amplitude function is shown by a black-dotted curve, which can be regarded as an effective periodic trajectory of the stochastic oscillator.

For comparison, the frequency and decay rate obtained by direct numerical evaluation of the eigenvalues and eigenfunctions of $L_{\boldsymbol{x}}^*$ are $\Omega \approx 0.582$ and $\mu_r \approx -0.778$ [14], and the associated asymptotic phase and amplitude functions are plotted in Fig. 2(d) and (e). We can confirm that the results obtained from the time-series data by EDMD agree well with the results evaluated directly from $L_{\boldsymbol{x}}^*$. Thus, EDMD could successfully reconstruct the phase and amplitude functions from time-series data in this example.

5 Summary

We have briefly reviewed the concepts of asymptotic phase and amplitudes of deterministic limit-cycle oscillators and demonstrated their natural generalization to stochastic oscillators from the viewpoint of Koopman operator theory. We then discussed the data-driven reconstruction of the asymptotic phase and amplitude functions from time-series data of stochastic oscillators, illustrating that the standard Extended Dynamic Mode Decomposition (EDMD) algorithm for estimating Koopman eigenvalues and eigenfunctions could successfully reconstruct those functions for an example of the stochastic FitzHugh-Nagumo model with reasonable accuracy. In this article, we focused solely on the definition and estimation of the phase and amplitude functions. In controlling stochastic oscillators based on the estimated functions, such as synchronizing them with periodic input signals, it is also crucial to characterize the effect of external perturbations on the phase and amplitudes of the stochastic oscillators. Further research will be required on this issue.

Acknowledgement. We acknowledge JSPS KAKENHI 22K14274, 22K11919, 22H00516 and JST CREST JP-MJCR1913 JST PRESTO **JPMJPR24K3** for financial support.

References

1. Winfree, A.T.: The Geometry of Biological Time. Springer, New York (2001)
2. Kuramoto, Y.: Chemical Oscillations, Waves, and Turbulence. Springer, Berlin (1984)
3. Pikovsky, A., Rosenblum, M., Kurths, J.: Synchronization: A Universal Concept in Nonlinear Science. Cambridge University Press, Cambridge (2001)
4. Ermentrout, G.B., Terman, D.H.: Mathematical Foundations of Neuroscience. Springer, New York (2010)
5. Shirasaka, S., Kurebayashi, W., Nakao, H.: Phase-amplitude reduction of transient dynamics far from attractors for limit-cycling systems. Chaos: Interdisc. J. Nonlinear Sci. **27**(2) (2017)
6. Monga, B., Wilson, D., Matchen, T., Moehlis, J.: Phase reduction and phase-based optimal control for biological systems: a tutorial. Biol. Cybern. **113**(1), 11–46 (2019)
7. Kuramoto, Y., Nakao, H.: On the concept of dynamical reduction: the case of coupled oscillators. Philos. Trans. R. Soc. A **377**(2160), 20190,041 (2019)
8. Shirasaka, S., Kurebayashi, W., Nakao, H.: Phase-amplitude reduction of limit cycling systems. The Koopman Operator in Systems and Control: Concepts, Methodologies, and Applications, pp. 383–417 (2020)
9. Nakao, H.: Phase and amplitude description of complex oscillatory patterns in reaction-diffusion systems. In: Physics of Biological Oscillators: New Insights into Non-Equilibrium and Non-Autonomous Systems, pp. 11–27. Springer (2021)
10. Mezić, I.: Analysis of fluid flows via spectral properties of the Koopman operator. Annu. Rev. Fluid Mech. **45**, 357–378 (2013)
11. Mauroy, A., Susuki, Y., Mezić, I.: The Koopman Operator in Systems and Control (2020)
12. Mauroy, A., Mezić, I., Moehlis, J.: Isostables, isochrons, and Koopman spectrum for the action-angle representation of stable fixed point dynamics. Physica D **261**, 19–30 (2013)
13. Thomas, P.J., Lindner, B.: Asymptotic phase for stochastic oscillators. Phys. Rev. Lett. **113**(25), 254,101 (2014)
14. Kato, Y., Zhu, J., Kurebayashi, W., Nakao, H.: Asymptotic phase and amplitude for classical and semiclassical stochastic oscillators via Koopman operator theory. Mathematics **9**(18), 2188 (2021)
15. Williams, M.O., Kevrekidis, I.G., Rowley, C.W.: A data-driven approximation of the Koopman operator: extending dynamic mode decomposition. J. Nonlinear Sci. **25**(6), 1307–1346 (2015)
16. Kutz, J.N., Brunton, S.L., Brunton, B.W., Proctor, J.L.: Dynamic mode decomposition: data-driven modeling of complex systems. SIAM (2016)
17. Klus, S., Nüske, F., Peitz, S., Niemann, J.H., Clementi, C., Schütte, C.: Data-driven approximation of the Koopman generator: model reduction, system identification, and control. Physica D: Nonlinear Phenomena **406**, 132,416 (2020)
18. Nakao, H.: Phase reduction approach to synchronisation of nonlinear oscillators. Contemp. Phys. **57**(2), 188–214 (2016)

19. Guckenheimer, J., Holmes, P.: Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields. Springer, New York (1983)
20. Winfree, A.T.: Biological rhythms and the behavior of populations of coupled oscillators. J. Theor. Biol. **16**(1), 15–42 (1967)
21. Guckenheimer, J.: Isochrons and phaseless sets. J. Math. Biol. **1**(3), 259–273 (1975)
22. Wilson, D., Moehlis, J.: Isostable reduction of periodic orbits. Phys. Rev. E **94**(5), 052,213 (2016)
23. Wilson, D., Ermentrout, B.: Greater accuracy and broadened applicability of phase reduction using isostable coordinates. J. Math. Biol. **76**(1), 37–66 (2018)
24. Mauroy, A., Mezić, I.: Global computation of phase-amplitude reduction for limit-cycle dynamics. Chaos: Interdisc. J. Nonlinear Sci. **28**(7), 073,108 (2018)
25. Takata, S., Kato, Y., Nakao, H.: Fast optimal entrainment of limit-cycle oscillators by strong periodic inputs via phase-amplitude reduction and Floquet theory. Chaos: Interdisc. J. Nonlinear Sci. **31**(9), 093,124 (2021)
26. Teramae, J., Nakao, H., Ermentrout, G.B.: Stochastic phase reduction for a general class of noisy limit cycle oscillators. Phys. Rev. Lett. **102**(19), 194,102 (2009)
27. Goldobin, D.S., Teramae, J., Nakao, H., Ermentrout, G.B.: Dynamics of limit-cycle oscillators subject to general noise. Phys. Rev. Lett. **105**(15), 154,101 (2010)
28. Wedgwood, K.C., Lin, K.K., Thul, R., Coombes, S.: Phase-amplitude descriptions of neural oscillator models. J. Math. Neurosci. **3**(1), 1–22 (2013)
29. Kotani, K., Ogawa, Y., Shirasaka, S., Akao, A., Jimbo, Y., Nakao, H.: Nonlinear phase-amplitude reduction of delay-induced oscillations. Phys. Rev. Res. **2**(3), 033,106 (2020)
30. Mauroy, A., Mezić, I.: Global stability analysis using the eigenfunctions of the Koopman operator. IEEE Trans. Autom. Control **61**(11), 3356–3369 (2016)
31. Gardiner, C.: Stochastic Methods. Springer, Cham (2009)
32. Pavliotis, G.A.: Stochastic Processes and Applications. Springer, Cham (2016)
33. Mezić, I.: Spectral properties of dynamical systems, model reduction and decompositions. Nonlinear Dyn. **41**(1–3), 309–325 (2005)
34. Schwabedal, J.T., Pikovsky, A.: Phase description of stochastic oscillations. Phys. Rev. Lett. **110**(20), 204,102 (2013)
35. Kato, Y., Nakao, H.: A definition of the asymptotic phase for quantum nonlinear oscillators from the Koopman operator viewpoint. Chaos: Interdisc. J. Nonlinear Sci. **32**(6) (2022)
36. Kato, Y., Nakao, H.: Quantum asymptotic phases reveal signatures of quantum synchronization. New J. Phys. **25**(2), 023,012 (2023)
37. Pérez-Cervera, A., Lindner, B., Thomas, P.J.: Isostables for stochastic oscillators. Phys. Rev. Lett. **127**(25), 254,101 (2021)
38. Pérez-Cervera, A., Gutkin, B., Thomas, P.J., Lindner, B.: A universal description of stochastic oscillators. Proc. Natl. Acad. Sci. **120**(29), e2303222,120 (2023)
39. Korda, M., Mezić, I.: On convergence of extended dynamic mode decomposition to the Koopman operator. J. Nonlinear Sci. **28**(2), 687–710 (2018)
40. Li, Q., Dietrich, F., Bollt, E.M., Kevrekidis, I.G.: Extended dynamic mode decomposition with dictionary learning: a data-driven adaptive spectral decomposition of the Koopman operator. Chaos: Interdisc. J. Nonlinear Sci. **27**(10) (2017)



Comparison Between Linear and Nonlinear System Identification Based on Response Probability Density Function of 1-DOF System Which is Subjected to White Noise Excitation

Soichiro Takata[1(✉)], Hiroharu Matsubara[1], Naoko Watanabe[2], Shuya Kubota[2], and Kaito Araki[3]

[1] National Institute of Technology, Tokyo College, 1220-2, Kunugida-machi, Hachioji, Tokyo, Japan
takata@tokyo-ct.ac.jp

[2] Advanced Course, National Institute of Technology, Tokyo College, 1220-2, Kunugida-machi, Hachioji, Tokyo, Japan

[3] Toyohashi University of Technology, 1-1 Hibarigaoka, Tenpaku, Toyohashi, Aichi, Japan

Abstract. Mechanical structural vibration is traditionally addressed using the system identification method, which uses a random vibration response. In this study, we propose an identification method based on a probability density function. The probability density function has a shape that is determined by system parameters, such as the spring constant (linear or nonlinear), damping constant, and diffusion coefficient of the input white noise. These characteristics are not directly related to the frequency spectrum. Therefore, a probability density function-based method is expected to be applicable to systems with low S/N ratios. An accuracy comparison was conducted in this study. As a result, when the estimation error rate exceeded 1%, the nonlinear model reduced the error rates of both the linear and nonlinear spring constants. On the contrary, the estimation results based on the linear model revealed a high-accuracy region in the low μ/k values. To obtain the guideline for PDF model selectivity, we introduced the average error rate in spring constant estimation results. The optimized ratio of μ/k was observed in the region between 0.5% and 1%. Therefore, in a weakly nonlinear system (i.e., a ratio of μ/k less than around 1%), the linear spring constant can be estimated using both linear and nonlinear models.

Keywords: Random vibration · Nonlinear vibration · System identification · Fokker-Planck equation · Maximum likelihood estimation

1 Introduction

The system identification method, which uses a random vibration response, has traditionally been used for mechanical structural vibration. The auto-regressive method (ARM) is the most widely used method [1–3] because it only outputs data. In field tests, obtaining input data is generally difficult, and ARM is often used for actual field data.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 154–161, 2025.
https://doi.org/10.1007/978-3-031-72794-8_11

However, because it is dependent on the output data, it has low precision in predicting error variance. The establishment of output-only system identification technology is expected to be applicable for various field vibration diagnoses [4, 5].

In recent years, the above issue has been addressed using the auto-regressive exogenous (ARX) model, which uses input–output data [6, 7]. Furthermore, operational modal analysis (OMA) was recently developed in the field of stochastic signal-processing technologies [8–12]. However, because it is based on the frequency response function method, OMA cannot be applied to large-damping structural and sampling fluctuating systems [10].

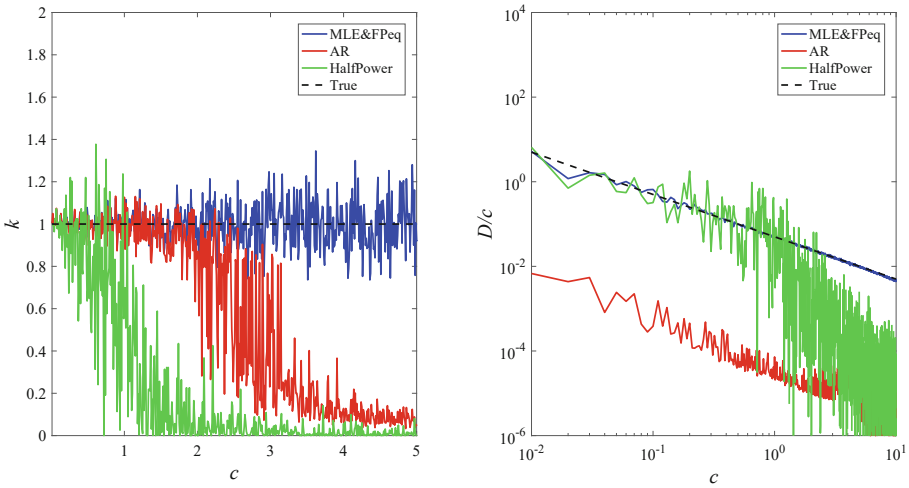
In this study, we proposed an identification method based on a probability density function (PDF). The shape of the PDF is influenced by system parameters, such as the spring constant (linear-nonlinear), damping constant, and diffusion coefficient of the input white noise [13–16]. The analytical solutions are obtained by solving the Fokker-Planck (FPK) equation. These characteristics of the PDF are not directly related to the frequency spectrum. Therefore, a method based on the PDF is expected to be applicable to a low S/N ratio system.

In our previous study, we proposed an identification algorithm based on maximum likelihood estimation (MLE) using the analytical solution of the FPK equation for the linear and nonlinear single-degree-of-freedom system which is subjected to white noise excitation [17, 18]. In the proposed method, the likelihood function was unprecedentedly composed using the analytical solution of FPK equation. Generally, the physical parameters (i.e., spring constant, damping constant, etc.) are not included in the likelihood function. Thus, the estimation formula of the explicit form of the physical parameters cannot be obtained. However, the analytical solution of FPK equation contains the physical parameters. Therefore, by optimizing the physical parameters, the proposed likelihood function can provide for their explicit estimation.

In a previous study, the benchmark test results were obtained in terms of damping coefficient dependency [19]. The identification results of a linear single-degree-of-freedom system, subjected to random white noise excitation, are shown in Fig. 1. Here, blue, red, and green solid lines show the results of the proposed, AR, and half-power methods, respectively; the black dotted line shows the true values. In this study, we focused on the estimation result of spring constant (case (a)). The proposed method agreed well with the true values throughout the range of the damping constant values. Conversely, HPM and ARM did not agree with the true values for a high damping constant. Furthermore, we focused on the estimation result of the ratio between the diffusion coefficient and damping constant (case (b)). The proposed method agreed well with the true values throughout the entire range of ratio between the diffusion coefficient and damping constant. HPM did not agree well with the true values for a high damping constant. In addition, ARM did not agree with the true values throughout the entire range of the D/c parameter.

The advantage of the method proposed in our proposed study was revealed in case of a linear system. However, the case of nonlinear system was limited to principle verification. Moreover, a comprehensive analysis comparing linear and nonlinear models has not been conducted.

In this study, we compared the accuracies of linear and nonlinear systems. First, the overview of the proposed identification formula is described. Furthermore, we also conducted numerical simulations.



(a) Linear spring constant estimation (b) D/c ratio estimation

Fig. 1. Benchmark test results in terms of damping coefficient dependency [19].

2 Maximum Likelihood Estimation Based on the Probability Density Function of Fokker-Planck Equation

We consider the Langevin equation in the case of a unit mass which is expressed in Eq. (1) as follows:

$$dy_1/dt = y_2, \qquad dy_2/dt = -cy_2 - ky_1 - \mu y_1^3 + W(t) \tag{1}$$

where k, c and μ represent the spring, damping, and nonlinear spring constants, respectively. In addition, y_1 and y_2 represent the displacement and velocity, respectively, and $W(t)$ represents white noise excitation. If the nonlinear spring constant is zero, the system can be considered as a linear system.

The stationary Fokker Plank equation becomes

$$-y_2\frac{\partial f}{\partial y_1} + \left(ky_1 + \mu y_1^3 + cy_2 + cf\right)\frac{\partial f}{\partial y_2} + D\frac{\partial^2 f}{\partial {y_2}^2} = 0, \tag{2}$$

where $f(y_1, y_2) = f$ represents the PDF of the random response and D represents the diffusion coefficient. These analytical solutions are expressed as follows:

$$f_{\text{Linear}}(y_1, y_2) = \frac{c\sqrt{k}}{2\pi D}\exp\left[-\frac{c}{2D}(ky_1^2 + y_2^2)\right], \tag{3}$$

$$f_{\text{Nonlinear}}(y_1, y_2) = \frac{\sqrt{\mu c/kD\pi}}{\exp\left[\frac{k^2 c}{8\mu D}\right] K_{\frac{1}{4}}\left(\frac{k^2 c}{8\mu D}\right)} \exp\left[-\frac{c}{D}\left(\frac{ky_1^2}{2} + \frac{\mu y_1^4}{4} + \frac{y_2^2}{2}\right)\right], \quad (4)$$

where, $y_{1\alpha}$ and $y_{2\alpha}$ represent the discrete variables of relative displacement and velocity, respectively, N denotes the number of samples, and α is 1,2, …, N. In addition, the simultaneous measurement of relative displacement and velocity is assumed.

If the sensor data $\mathbf{D} = \left\{(y_{1,1}, y_{2,1}), \ldots, (y_{1,\alpha}, y_{2,\alpha}) \ldots, (y_{1,N}, y_{2,N})\right\}$, the likelihood function is defined using the analytical solution of the stochastic differential equation given in Eq. (1) as follows:

$$L_{\text{Linear}} = \prod_{\alpha=1}^{N} f_{\text{Linear}}(y_{1\alpha}, y_{2\alpha}) = \prod_{\alpha=1}^{N} \frac{\sqrt{k}}{2\pi(D/c)} \exp\left[\frac{-ky_{1\alpha}^2}{2(D/c)}\right] \exp\left[\frac{-y_{2\alpha}^2}{2(D/c)}\right], \quad (5)$$

$$\begin{aligned} L_{\text{Nonlinear}} &= \prod_{\alpha=1}^{N} f_{\text{Nonlinear}}(y_{1\alpha}, y_{2\alpha}) \\ &= \prod_{\alpha=1}^{N} \frac{\sqrt{\frac{\mu}{k}}}{\sqrt{\pi}\left(\frac{D}{c}\right) K_{\frac{1}{4}}\left(\frac{k^2}{8\mu\left(\frac{D}{c}\right)}\right)} \exp\left[\frac{-k^2}{8\mu\left(\frac{D}{c}\right)}\right] \exp\left[\frac{-1}{2\left(\frac{D}{c}\right)}\left\{ky_{1\alpha}^2 + \frac{\mu y_{1\alpha}^4}{2}\right\}\right] \exp\left[\frac{-y_{2\alpha}^2}{2\left(\frac{D}{c}\right)}\right] \end{aligned} \quad (6)$$

where, Eq. (5) represents the case for the linear model and Eq. (6) represents a nonlinear model. $K_*(\cdot)$ in Eq. (6) represents the modified Bessel function of the second kind.

Identification algorithms were derived using the maximum likelihood estimation method. The estimated values of the unknown parameters were obtained by maximizing the likelihood function of Eqs. (5) and (6). The mathematical procedure was realized using partial differentiation along the unknown parameters.

The estimation formulae were obtained using Eqs. (7)–(9). The formulae were obtained in explicit form. y_{est} was obtained from the common estimation formula between the cases of the linear system model and nonlinear system model.

Case of linear system model:

$$k_{\text{est}} = \sum_{\alpha=1}^{N} y_{2\alpha}^2 \Big/ \sum_{\alpha=1}^{N} y_{1\alpha}^2, \qquad y_{\text{est}} = \frac{1}{N} \sum_{\alpha=1}^{N} y_{2\alpha}^2. \quad (7)$$

Case of nonlinear system model:

$$k_{\text{est}} = \frac{1}{2}\left(-X_{\text{est}}\left(\frac{\frac{1}{N}\sum_{\alpha=1}^{N} y_{1\alpha}^2}{\frac{1}{N}\sum_{\alpha=1}^{N} y_{1\alpha}^4}\right) + \sqrt{X_{\text{est}}^2\left(\frac{\frac{1}{N}\sum_{\alpha=1}^{N} y_{1\alpha}^2}{\frac{1}{N}\sum_{\alpha=1}^{N} y_{1\alpha}^4}\right)^2 + \frac{4y_{\text{est}}X_{est}}{\frac{1}{N}\sum_{\alpha=1}^{N} y_{1\alpha}^4}}\right), \quad (8)$$

$$\mu_{\text{est}} = \frac{k_{\text{est}}^2}{X_{\text{est}}}$$

$$G(X) = 2\frac{\left(\frac{1}{N}\sum_{\alpha=1}^{N} y_{1\alpha}^2\right)^2}{\left(\frac{1}{N}\sum_{\alpha=1}^{N} y_{1\alpha}^4\right)}X - \left(\frac{\left(-y_{\text{est}} - \frac{X}{2} + \frac{X}{4}\text{Bes}(X)\right)^2}{\left(y_{\text{est}} + \frac{X}{4} - \frac{X}{8}\text{Bes}(X)\right)}\right) = 0 \tag{9}$$

$$X_{\text{est}} = \min_X |G(X)|, \qquad \text{Bes}(X) = \left\{K_{-\frac{3}{4}}\left(\frac{X}{8y_{\text{est}}}\right) + K_{\frac{5}{4}}\left(\frac{X}{8y_{\text{est}}}\right)\right\} / K_{\frac{1}{4}}\left(\frac{X}{8y_{\text{est}}}\right)$$

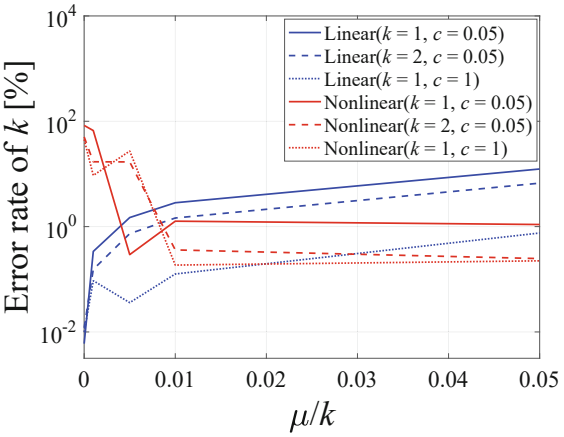
3 Numerical Simulation

The time series for the numerical identification was generated using the fourth order Runge-Kutta method. The numerical conditions were as follows: $k = 1,\ c = 0.05,\ 1.0,\ \mu = 0 - 0.05$, and $D = 0.05$. The variance of white noise was $\sigma_w^2 = 1$ the average was zero, the initial values were $y(0) = 0, y'(0) = 0$, and the sampling period was $\Delta t = 0.1$. Furthermore, 50,00,000 samples were used for identification. The identification results of the linear and nonlinear spring constants are shown in Fig. 2. Here, the horizontal axis shows the ratio between the nonlinear and linear spring constants in true values; on the other hand, the vertical axis shows the error rate. Figure 2 (a) represents the case of the linear spring constant estimation, and Fig. 2 (b) represents the case of the nonlinear spring constant estimation. Furthermore, the solid line represents the case of the condition $k = 1,\ c = 0.05$, the dotted line represents the case of the condition $k = 2,\ c = 0.05$ (that is, the high rigidity case), and the dashed line represents the case of condition $k = 1,\ c = 1$ (that is, large damping case).

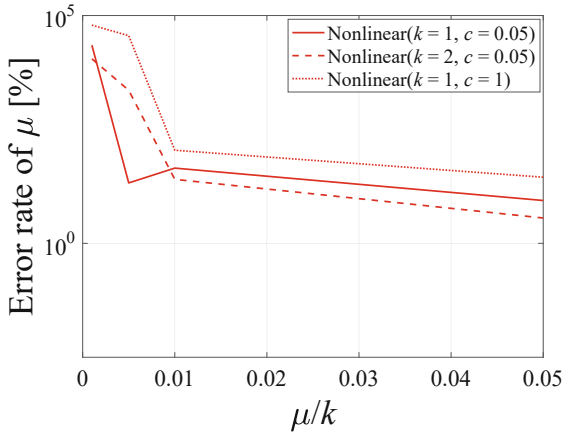
The error rate of linear spring constant estimation in the linear system model (the case of all blue line) increased by increasing μ/k; however, the error rate in the nonlinear system model (the case of all red line) decreased by increasing μ/k, as shown in Fig. 2 (a). The results show the resolution of the nonlinearity detection in the nonlinear model and the application limit in the linear model. In particular, the estimation error rate decreased in the case of more than 1% of μ/k. The linear model had a high accuracy region in the low μ/k values. This was considered because the response of the linear model was in agreement with the response of the true calculation model. However, the estimation accuracy was decreased because of the response of the linear model was misfitted with the response of the true calculation in the high μ/k values.

The error rate of the nonlinear spring constant in the case of the nonlinear system model (red) decreased by increasing the μ/k as shown in Fig. 2 (b). The results showed the resolution of the nonlinearity detection. The estimation error rate decreased in the case of more than 1% of μ/k. The findings of the investigation suggest that the relative contribution of the nonlinear restoring force to the entire restoring force was small in the region of response displacement.

To obtain the guidelines for PDF model selectivity, we introduced the average error rate in the estimation result of the spring constant. The calculation results are shown in Fig. 3. The average error rates were calculated by averaging the error rates obtained using linear and nonlinear models. Here, the horizontal axis shows the ratio between the nonlinear and linear spring constants in true values, on the other hand, the vertical axis shows the averaged error rate. Furthermore, the solid line represents the case of



(a) Linear spring constant estimation



(b) Nonlinear spring constant estimation

Fig. 2. Comparison of estimation accuracy (using error rate).

the condition $k = 1,\ c = 0.05$, the dotted line represents the case of the condition $k = 2,\ c = 0.05$ (that is, high rigidity case), and the dashed line represents the case of the condition $k = 1,\ c = 1$ (that is the large damping case). The averaged error rates showed the behavior of single-valley structures because there was a trade-off between the cases of linear and nonlinear models. The minimum value of the average error rate was obtained in the large damping case ($k = 1,\ c = 1$). Similar values were obtained in the low damping case ($k = 1,\ c = 0.05$ and $k = 2,\ c = 0.05$). The optimized ratio of μ/k was observed in the region between 0.5 and 1%. Therefore, in the weakly nonlinear system (i.e., ratio of μ/k less than approximately 1%), the linear spring constant could be estimated using linear and nonlinear models. However, the estimation of nonlinear spring constant was desired region of the ratio of μ/k more than approximately 1%.

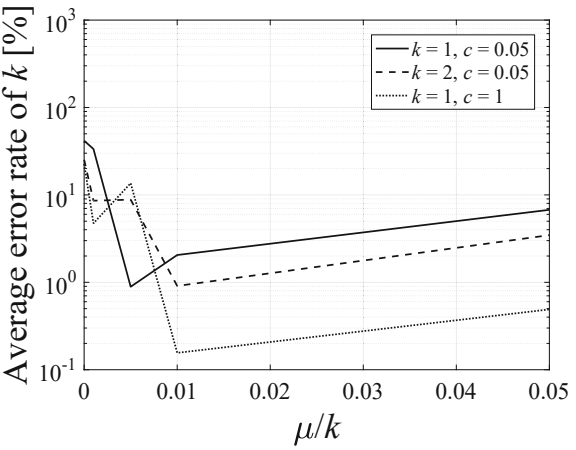


Fig. 3. Comparison of average error of linear spring constant.

4 Conclusion

In this study, we discuss an identification method based on the MLE that uses the analytical solution of the Fokker-Planck equation, and the following results were obtained:

(1) A comparison between the linear and nonlinear models was conducted. The formulations were easily reviewed, and the estimation error rate was considered.
(2) In the case of more than 1% of μ/k, the estimation error rate of the linear spring constant and the nonlinear spring constant using the nonlinear model was reduced.
(3) The estimation results based on the linear model showed a high accuracy region for the low μ/k values.
(4) To obtain the guideline for PDF model selectivity, we introduced the average error rate in spring constant estimation results. The optimized ratio of μ/k were observed in the region between 0.5 and 1%. Therefore, in a weakly nonlinear system (i.e., ratio of μ/k less than approximately 1%), the linear spring constant can be estimated using both linear and nonlinear models.

The proposed identification method for the nonlinear system was limited to the hardening nonlinear system and weak nonlinearity. We will expand the proposed method to the softening nonlinear system and strong nonlinearity. In the future work, we will focus on the experimental identification.

Acknowledgements. This work was supported by JST Grant Number JPMJPF2115.

References

1. Lu, Y., Gao, F.: A novel time-domain auto-regressive model for structural damage diagnosis. J. Sound Vibr. **283**, 1031–1049 (2005)

2. Sohn, H., Farrar, C.R.: Damage diagnosis using time series analysis of vibration signals. Smart Mater. Struct. **10**, 446–451 (2001)
3. Suzuki, K., Nakashima, A.: Damping estimation method using the spectral analysis technique based on the auto-regressive model fitting. Bull. JSME **26**(215), 832–838 (1983)
4. Kubota, S., Takata, S.: Automatically discriminations system of landslide disaster sign based on detection of resonance longitudinal vibration. In: Proceedings of 2022 JSME-IIP/ASME-ISPS Joint International Conference on Micromechatronics for Information and Precision Equipment (MIPE 2022), D3-3-05, Nagoya, Japan (2022)
5. Takata, S., Kubota, S., Watanabe, N.: Prototyping of vibration-sensing-actuation device to realization of future intelligent infrastructure. J. Adv. Mech. Des. Syst. Manuf. **17**(4) (2023)
6. Maurya, D., Tangirala, A.K., Narasimhan, S.: ARX model identification using generalized spectral decomposition. IFAC-PapersOnLine **54**(9), 690–695 (2021)
7. Diversi, R., Guidorzi, R., Soverini, U.: Identification of ARX and ARARX models in the presence of input and output noises. Eur. J. Control. **16**(3), 242–255 (2010)
8. Zhu, Y.C., Au, S.K.: Bayesian operational modal analysis with asynchronous data, part 1: most probable value. Mech. Syst. Sig. Process. **98**, 652–666 (2018)
9. Zhu, Y.C., Au, S.K.: Bayesian operational modal analysis with asynchronous data, part 2: posterior uncertainty. Mech. Syst. Sig. Process. **98**, 920–935 (2018)
10. Zhu, Y.C., Au, S.K., Brownjohn, J.M.W.: Bayesian operational modal analysis with buried modes. Mech. Syst. Sig. Process. **121**, 246–263 (2019)
11. Brincker,R., Anderson, P., Zhang, L.: Modal identification from ambient responses using frequency domain decomposition. In: Proceedings of the 18th International Modal Analysis Conference (IMAC), pp. 625–630 (2000)
12. Krishnan, N.K., Kiremidjian, A.S.: Time series based structural damage detection algorithm using Gaussian mixtures modeling. Trans. ASME J. Dyn. Syst. Meas. Control **129**, 285–293 (2007)
13. Lin, Y.K.: Probabilistic Structural Dynamics. McGraw-Hill, New York (1967)
14. Caughey, T.K.: Derivation and application of the Fokker-Planck equation to discrete nonlinear dynamic systems subjected to white random excitation. J. Acoust. Soc. Am. **35**(11), 1683–1692 (1963)
15. Hesam, S., Nazemi, A.R., Haghbin, A.: Analytical solution for the Fokker-Planck equation by differential transform method. Scientia Iranica B **19**(4), 1140–1145 (2012)
16. Suciu, N., Radu, F.A., Attinger, S., Schuler, L., Knabner, P.: A Fokker-Planck approach for probability distributions of species concentrations transported in heterogeneous media. J. Comput. Appl. Math. **289**, 241–252 (2015)
17. Takata, S.: Experimental identification of a lumped mass-spring system based on the maximum likelihood estimation using the steady response distribution of Fokker-Planck equation. Mech. Eng. J. **10**(1) (2023)
18. Araki, K., Takata, S.: System identification of nonlinear 1-DOF system which subjected to white noise excitation based on analytical solution of Fokker-Planck equation. In: Proceedings of 2022 JSME-IIP/ASME-ISPS Joint International Conference on Micromechatronics for Information and Precision Equipment (MIPE2022), A2-1-01, Nagoya, Japan (2022)
19. Takata, S.: Identification of linear 1-dof system based on the maximum likelihood estimation using the analytical solution of Fokker-Planck equation. Bull. JSME Trans. JSME **87**(893) (2021). (in Japanese)

Calculus and Computation

Autotuners and Fractional Calculus. Applications to the Control of Nonlinear Processes

Cristina Muresan[1], Isabela Birs[1,2,3(✉)], and Robin De Keyser[2]

[1] Department of Automation, Technical University of Cluj Napoca, Memorandumului no. 28, 400114 Cluj-Napoca, Romania
Cristina.Muresan@aut.utcluj.ro

[2] Department of Electromechanical, Systems and Metal Engineering Flanders Make EEDT Core Lab, Ghent University, Tech Lane Science Park 125 and 131, 9052 Gent, Belgium
{IsabelaRoxana.Birs,Robain.DeKeyser}@ugent.be

[3] Flanders Make, EEDT - Decision and Control Group, Tech Lane Science Park 131, 9052 Ghent, Belgium

Abstract. Fractional calculus has seen a growing interest in the control community during the past two decades. Most papers focus on the design and analysis of fractional order controllers, as generalization of the standard PID controller. Applications range from chemical plants, biomedical engineering, vibration control and mechanical systems, to name only a few. Most of the times, the tuning of such controllers is done based on a process model. Identifying a process model can be time consuming and difficult to determine accurately. In the case of nonlinear processes, usually very few operating points are of interest. A procedure to extract relevant process information in these operating points and use this information to tune fractional order controllers would reduce the design steps by eliminating process modeling. In this paper, an alternative approach to sine wave autotuners is considered based on a single sine test applied to the process. The methods extracts relevant process data that is further used to tune PID-type controllers. Experimental validations and comparisons to existing sine based autotuners are provided on two nonlinear processes: mass-spring-damper system and a vertical take off and landing platform.

Keywords: fractional order control · fractional order autotuner · industrial control

1 Introduction

It is undeniable that the PID controller holds a prominent position in industrial control. With escalating customer demands for high product quality, increasingly stringent safety regulations, and intensifying global competition, industrial processes often operate under conditions that emphasize inherent nonlinearities. For

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 165–183, 2025.
https://doi.org/10.1007/978-3-031-72794-8_12

large industrial plants characterized by highly coupled loop interactions, obtaining an accurate process model becomes a laborious task, justifying the utilization of auto-tuning methods to ascertain the controller's parameters [1–3].

Recent advances in controller auto-tuning have enhanced traditional PID controllers by integrating fractional-order differ-integral operators, first introduced by I. Podlubny [4]. Numerous studies confirm the superiority of fractional-order PID controllers, showing improvements in closed-loop system performance, stability, and robustness via arbitrary orders of integration and differentiation [5–9]. While various tuning methods exist [10], strategies for fractional-order auto-tuning are still limited and often face constraints [11].

One of the early endeavors in fractional-order auto-tuning was presented by Chen in 2004 [12], where the controller design was grounded in the shaping of a 'flat-phase' around the gain crossover frequency. This approach yielded a controller akin to the classical PID, albeit with the addition of a fractional component s^{λ}. Another method, as elucidated in [6], adopted the relay test, commonly employed in integer-order PID controller design, to derive a fractional-order PI controller in conjunction with a fractional-order PD controller accompanied by a filter. This method necessitates specifications linked to the gain crossover frequency, phase margin, and robustness, entailing significant effort in determining the controller's parameters.

The Ziegler-Nichols (ZN) method, widely recognized for its simplicity, determines integer-order PID parameters based on the process's critical gain and frequency. Despite its effectiveness in disturbance rejection, it often struggles with reference tracking [1]. Various extensions have been proposed to address its limitations. Yeroglu et al. [13] combine ZN for proportional and integral gains, Åström-Hägglund for initial derivative gain, and fine-tuning for optimal derivative gain. However, this method is complex and sensitive to initial values. Valerio et al. [14] extend ZN into the fractional-order domain but are limited to systems with S-shaped step responses. Other extensions exist [15–18], yet all face challenges such as extensive tuning efforts and requirements for specific process response shapes.

De Keyser et.al. propose alternative approaches in [19] and [20], introducing the 'Sine Test auto-tuner' and the 'FO-KC auto-tuner', respectively. A single sine test at the gain crossover frequency is used to measure output signal magnitude and phase, followed by innovative filtering techniques to determine phase slope. Nonlinear equations incorporating experimental data are then solved to establish parameters like gain crossover frequency and phase margin. The methodology utilizes phase margin and isodamping frequency domain specifications to create a forbidden region circle on the Nyquist diagram. A single sine test estimates process phase and slope, with the tuning process minimizing slope differences to achieve the optimal fractional-order PID controller.

Two approaches are presented in this chapter to estimate accurately the process frequency response and slope, even in the presence of disturbances and noise. Both of these methods require a simple short sinusoidal test to be performed on the process and thus become suitable for a large range of processes. Both meth-

ods require selection of a test frequency $\bar{\omega}$, that usually represents the desired gain crossover frequency of the loop frequency response or the critical frequency, i.e. the frequency where the process is at its limit of stability.

The first method, known as KCT, consists in a sine test performed on the process, followed by an offline filtering technique [21]. The frequency of the sine is the desired gain crossover frequency, which is also one of the imposed performance specifications [22]. The second method consists in applying a short sequence of sine signals. Based on the measured output signal and using filtering methods, the impulse response of the process is estimated, which is later used to compute the process frequency response and its slope at a specific frequency. The test frequency of the sine is the process critical frequency and each section of the test sequence lasts as long as the process settling time [23]. Both methods provide accurate approximations for the process frequency response and slope at the test frequency, despite possible disturbances or noisy signals. Once these have been estimated, the FOPID parameters can be determined based on the imposed performance criteria. Both methods have been tested on a vertical take-off and landing (VTOL) unit, as well as on a mass-spring-damper (MSD) system, to name just a few possible applications. Analytical analysis as well as experimental tests have shown that these processes exhibit nonlinear character.

The novelty of the present work lies in the alternative method and its validation on the nonlinear experimental units. Furthermore, the results are compared to the already existing sine based KCT autotuner.

2 Fractional Order Control

Commencing with an integer-order PID controller as expressed by the equation:

$$C_{PID} = k_p \left(1 + \frac{1}{T_i s} + T_d s\right) \tag{1}$$

and extending this concept into the realm of fractional orders yields the following expression:

$$C_{FOPID} = k_p \left(1 + k_i \frac{1}{s^\lambda} + k_d s^\mu\right), \tag{2}$$

In the above equations, where k_p, k_i, and k_d represent proportional gains, T_i and T_d denote the integral and derivative time constants, and λ and μ stand for the fractional orders of integration and differentiation, respectively.

The predominant approach for tuning fractional order controllers centers around the establishment of phase margin specifications connected to the gain crossover frequency ω_{gc}, phase margin φ_m. This method is widely documented in the literature [10,24–27]. Robustness is defined with respect to a property known as *isodamping*, which aims to maintain a flat phase response around the gain crossover frequency. This ensures that the system exhibits a certain level of resilience to variations in gain. For frequency changes in the interval encompassing the gain crossover frequency ω_{gc}, a straight phase plot gives the

same value when the gain changes, leading to a certain degree of robustness to gain uncertainties. The resulting set of nonlinear equations that determine the parameters of a fractional order controller can be expressed as follows:

$$\angle(C(j\omega_{gc})P(j\omega_{gc})) = -\pi + \varphi_m, \tag{3}$$

$$|C(j\omega_{gc})P(j\omega_{gc})| = 1, \tag{4}$$

$$\frac{d(\angle(C(j\omega)P(j\omega)))}{d\omega}\bigg|_{\omega=\omega_{gc}} = 0, \tag{5}$$

$$S(j\omega) = \frac{1}{1 + C(j\omega)P(j\omega)}, \tag{6}$$

$$T(j\omega) = \frac{C(j\omega)P(j\omega)}{1 + C(j\omega)P(j\omega)}. \tag{7}$$

Sensitivity, denoted as $S(j\omega)$, is a metric used to quantify the reaction of a closed-loop system in response to disturbances or alterations in the transfer function of the plant [28]. Complementary Sensitivity $T(j\omega)$ refers to the closed-loop system's responsiveness to the reference input or intended setpoint [29].

3 Autotuning Methods

3.1 Sine Wave Autotuner

The approach was originally proposed in [19]. An experimental validation for the sine wave auto-tuner can be found in [30], where a fractional order PD controller was implemented to mitigate undesired vibrations in a smart-beam platform. The fundamental concept involves stimulating the process denoted as P using a sine wave with an amplitude of A_i and the desired gain crossover frequency (ω_{gc}), represented as $u(t) = A_i \sin(\omega_{gc}t)$. As a result, the generated output y exhibits characteristics such as a different output amplitude A_o, the same ω_{gc} frequency, and a time shift denoted as $\tau = t_i - t_o$. The magnitude M and phase φ can be determined as follows:

$$M = |P(j\omega_{gc})| = \frac{A_o}{A_i}, \tag{8}$$

$$\varphi = \angle P(j\omega_{gc}) = \omega_{gc}\tau = \omega_{gc}(t_i - t_o), \tag{9}$$

where t represents time, and $j = \sqrt{-1}$.

The innovation in this approach centers around the challenge of calculating the phase derivative, as there is no simple, direct method to accomplish this task. The concept involves routing the output through a second-order filter to generate an additional signal: the output representing the process derivative, followed by the phase slope. Assuming an input signal is denoted as $v(t) = tu(t)$,

which results in an output $x(t)$, the following relationships apply in the Laplace domain

$$\mathscr{L}\{tu(t)\}P(s) = X(s), \tag{10}$$

which gives

$$-\frac{dU(s)}{ds}P(s) = X(s). \tag{11}$$

When the initial input signal is fed to the process derivative, the resulting output $\bar{y}$ is defined as

$$\frac{dP(s)}{ds}U(s) = \bar{Y}(s). \tag{12}$$

The derivative of the output signal is

$$\begin{aligned} \frac{dY(s)}{ds} &= \mathscr{L}\{-ty(t)\} \implies \\ \mathscr{L}^{-1}\left\{\frac{dY(s)}{ds}\right\} &= \mathscr{L}^{-1}\left\{\frac{d(P(s)U(s))}{ds}\right\}. \end{aligned} \tag{13}$$

Rewriting (13)

$$\frac{d(P(s)U(s))}{ds} = \bar{Y}(s) - X(s), \tag{14}$$

gives

$$-ty(t) = \mathscr{L}^{-1}\{\bar{Y}(s) - X(s)\}. \tag{15}$$

The Laplace transform of a sinusoidal input can be written as

$$U(s) = \frac{A_i\omega_{gc}}{s^2 + \omega_{gc}^2} \tag{16}$$

while it's derivative is obtained as

$$\frac{dU(s)}{ds} = -\frac{2A_i\omega_{gc}s}{(s^2 + \omega_{gc}^2)^2}. \tag{17}$$

This leads to

$$\begin{aligned} X(s) &= -\frac{dU(s)}{ds}P(s) = -\frac{2A_i\omega_{gc}s}{(s^2 + \omega_{gc}^2)^2}P(s) \\ &= \frac{A_i\omega_{gc}}{s^2 + \omega_{gc}^2}\frac{2\,s}{s^2 + \omega_{gc}^2}P(s). \end{aligned} \tag{18}$$

Considering (16) with respect to $U(s)P(s) = Y(s)$ allows the computation of $X(s)$ as

$$X(s) = \frac{2\,s}{s^2 + \omega_{gc}^2}Y(s). \tag{19}$$

$x(t)$ can be determined by incorporating a second-order filter, allowing the computation of the process derivative $\bar{y}(t)$ as the disparity between $x(t)$ and $ty(t)$. The procedure is illustrated in Fig. 1.

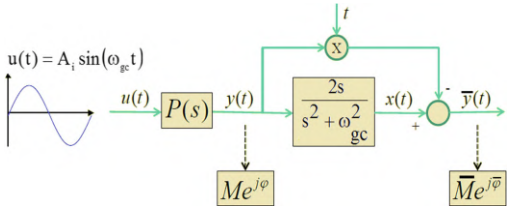


Fig. 1. Experimental block scheme for the sine wave auto-tuner

The elements observed at the lower section of Fig. 1 correspond to the complex representations of the signals $y(t)$ and $\bar{y}(t)$

$$\bar{M} = \frac{A_{\bar{y}}}{A_i} \tag{20}$$

$$\bar{\varphi} = \omega_{gc}\tau_{\bar{y}} = \omega_{gc}(t_i - t_{\bar{y}}) \tag{21}$$

with the slope of the phase evaluated at the gain crossover frequency

$$\left.\frac{d\angle P(j\omega)}{d\omega}\right|_{\omega=\omega_{gc}} = \left.\frac{d\varphi}{d\omega}\right|_{\omega=\omega_{gc}} = \frac{\bar{M}}{M} cos\left(\bar{\varphi} - \varphi\right). \tag{22}$$

The tuning procedure should be executed following the subsequent steps:

1. Determine the target gain crossover frequency ω_{gc} and phase margin φ_m based on the desired criteria for the closed-loop system's settling time and overshoot. Select ω_{gc} and φ_m based on practical insights into the behavior of the controlled process. The choice of frequency should align with the type of controller and the critical frequency of the process. For instance, when employing a fractional order PI controller that introduces negative phase, it's important to avoid selecting a frequency equal to the critical frequency, as it could render the closed-loop system unstable.
2. Stimulate the process using a sine wave test and record both the input and output signals, denoted as $u(t)$ and $y(t)$, respectively. If the experimental setup permits, it is advisable to use a sine wave input with an amplitude of 1 to simplify the calculations ($A_i = 1$).
3. Calculate $\bar{y}(t)$ according to (19). The implementation of the filtering process can occur in real-time within the process itself, or alternatively, the recorded signals from Step 2 can be utilised in a software environment such as Simulink to construct the phase derivative. In the case of excessive noise in the output signal, an alternative approach to filtering involves employing a pure sine signal possessing identical qualities for the purpose of calculating the phase derivative.
4. Using (8), (9), (20), (21) and (22), extract the modulus, phase and phase derivative of the process.
5. Resolve the set of nonlinear equations defined by (3), (4), and (5) in order to ascertain the controller's parameters. Employing a graphical method to solve

the system of nonlinear equations is a direct and uncomplicated approach, in contrast to employing intricate mathematical techniques. It is imperative to consistently verify the physical relevance of the acquired controller, ensuring that fractional orders fall within the interval of (0, 2) and that proportionate gains are greater than zero [31].

3.2 Alternative Approach

This approach represents an effective and practical methodology for extracting pertinent process information from datasets that are quite brief and contain minimum information. The methodology necessitates the implementation of a brief sine test that may be applied to nominal process inputs in either an open or closed loop configuration. This test allows for the estimation of the process impulse response, resulting in a frequency response that is both band limited and reasonably precise. The aforementioned data, in conjunction with performance standards, is subsequently utilized to adjust the parameters of a PID controller [23].

A test signal u(t) is applied representing the impulse response of a generating filter F(s). δ(t) is the Dirac impulse, where $\Delta(s) = \mathcal{L}\{\delta(t)\} = 1$. The filter denoted by

$$F(s) = \frac{\bar{\omega}}{s^2 + \bar{\omega}^2} \tag{23}$$

gives the test signal

$$u(t) = \sin(\bar{\omega}t) = \sin(\frac{2\pi}{T_p}t). \tag{24}$$

with T_p representing the oscillation period. F(s) can be defined as the Laplace transform of the deterministic signal. Figure 2 showcases a range of instances illustrating the generating filters. Additionally, it includes the deterministic test signals connected with these filters and the appropriate impulse responses denoted as F(s).

The step and sine waveforms are widely recognized as the most commonly utilized test signals in industrial applications. The proposed approach is built on a sine test, for illustrating the autotuning principles.

Feeding the process $P(s)$ with the sinusoidal signal from (24) gives the output $y(t)$ as shown by Fig. 3. $h(t)$ represents the impulse response of the unknown process $P(s)$. Then $Y(s) = F(s)H(S)$, where $H(s) = P(s)\Delta(s)$ is the Laplace transform of $h(t)$. Further filtering the output signal $y(t)$ using $\frac{1}{F(s)}$ leads to the estimated impulse response $\hat{h}(t)$ of the unknown process. Obviously, in lack of disturbances and noisy signals, $\hat{h}(t) = h(t)$.

In real-world scenarios, the presence of noise and disturbances has a significant impact on the output of a given process, denoted as $y(t)$. A more pragmatic approach must be taken into account in order to address the impact of random disturbances, as depicted in Fig. 4. The variable d(t) is used to denote the aggregate impact of all disturbances that influence the observed process output over the duration of the test. In the subsequent analysis, we shall regard d(t) as a

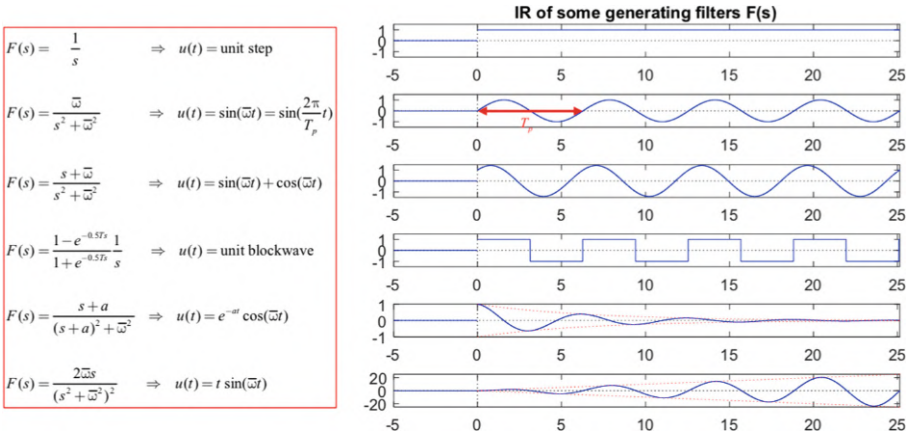


Fig. 2. Examples of impulse responses associated to different F(s) generating filters

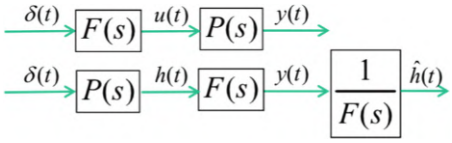


Fig. 3. Block diagram that produces the process impulse response

stochastic signal characterised by coloured noise and possessing a mean value of zero. Then, the following equations hold:

$$\bar{H}(s) = \frac{1}{F(s)}Y(s) = \frac{1}{F(s)}[P(s)U(s) + D(s)] \tag{25}$$

and

$$\bar{H}(s) = \frac{1}{F(s)}P(s)F(s)\Delta(s) + \frac{1}{F(s)}D(s) = H(s) + \frac{1}{F(s)}D(s). \tag{26}$$

Considering $\hat{h}(t) = h(t) + n(t)$, the error signal $n(t)$ is the process output disturbance $d(t)$ filtered by the inverse $\frac{1}{F(s)}$ of the generating filter.

The design of the generating filter should ensure that the disturbance signal d(t) is sufficiently attenuated at the frequency of interest $\bar{\omega}$, as shown by the inverse of the transfer function $\frac{1}{F(j\omega)}$. For instance, when the test signal u(t) is the unit step signal, the Laplace transform of the signal, denoted as F(s), is equal to $\frac{1}{s}$. In the above scenario, the inverse of the generating filter may be expressed as $F^{-1}(s) = s$. Furthermore, the magnitude of $F^{-1}(j\bar{\omega})$ can be denoted as $|F^{-1}(j\bar{\omega})| = \bar{\omega}$.

However, in the case where the test signal is the sinusoidal signal described in (24), the generating filter may be determined using (23). The inverse of the

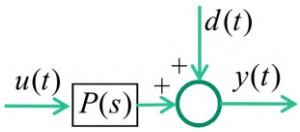


Fig. 4. Block diagram of a process affected by disturbances

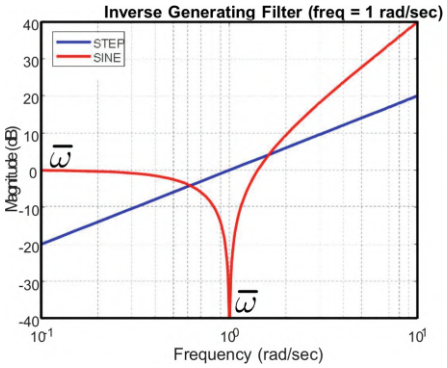


Fig. 5. Magnitude plot of the inverse generating filters corresponding to step and sine signals

generating filter is denoted as $F^{-1}(s) = \frac{s^2+\bar{\omega}^2}{\bar{\omega}}$. It is worth noting that the magnitude of $F^{-1}(j\bar{\omega})$ is equal to zero. In this particular scenario, the inverse sine filter effectively attenuates the disturbance occurring at the specific frequency of interest denoted as $\bar{\omega}$. The correct calculation of the process frequency response and its slope can be achieved even in the presence of disturbances. The filter also significantly lowers disturbances at frequencies in close proximity to $\bar{\omega}$, as evidenced by the magnitude plot of the two producing filters shown in Fig. 5.

Let us consider a stable process denoted as P(s), which exhibits a finite settling time denoted as ST=4. In practical applications, the IR (Inverse Response) signal h(t) exhibits a consistent duration, denoted as ST, during which it gradually diminishes towards zero. In practical terms, the duration of the impulse response (IR) can be equated to the settling time (ST). One can construct a test with a duration beyond ST by skillfully integrating multiple parts, each with a duration equivalent to ST, as depicted in Fig. 6. This approach enables the extraction of several IR estimations from a single experiment, as will be shown further. The act of calculating the average of the IR values will subsequently enhance the ratio of the signal to the noise.

After the signal $u(t)$ is applied at the input of the process and the measured output signal $y(t)$ is obtained, the IR-extractor can be computed [23]

$$h = \frac{\ddot{y}(t) + \bar{\omega}^2 y(t)}{\bar{\omega}} \tag{27}$$

and averaged to obtain the process impulse response $h(t)$. Considering

$$P(s) = \mathcal{L}\{h(t)\} = \int_0^\infty h(t)e^{-st}dt, \tag{28}$$

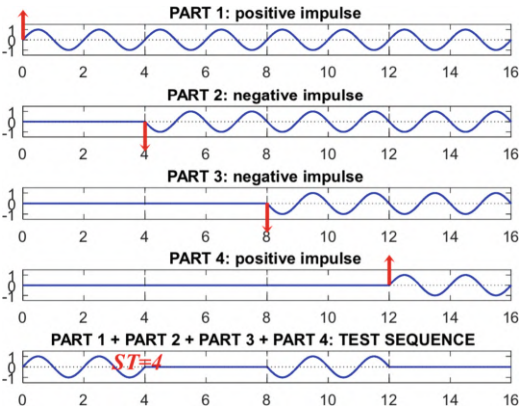


Fig. 6. Test signal combination to obtain multiple impulse response estimations

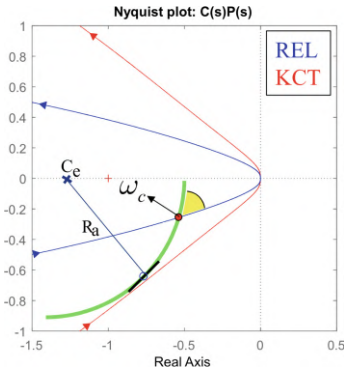


Fig. 7. Concept of the KC autotuner (KCT- Kiss Circle Tuner)

gives the corresponding frequency

$$P(j\omega) = \int_0^\infty h(t)e^{-j\omega t}dt \tag{29}$$

and its slope

$$\frac{dP(j\omega)}{d\omega} = -j\int_0^\infty th(t)e^{-j\omega t}dt. \tag{30}$$

In practice, for the given test frequency $\bar{\omega}$, the Discrete Fourier Transform is used which can be approximated to:

$$P(j\bar{\omega}) \cong T_s \sum_{k=0}^{N_s-1} h_k e^{-j\Omega k}, \tag{31}$$

with $t = kT_s$, T_s sampling periods, N_s number of measured samples, $h_k = h(kT_s)$ and $\Omega = \bar{\omega}T_s$ the constant discrete frequency. A similar procedure is used to evaluate the slope, based on the IR-extractor. The frequency response value $P(j\bar{\omega})$ and the slope $\frac{dP(j\omega)}{d\omega}|_{\omega=\bar{\omega}}$ will be calculated accurately even in the presence of disturbances.

The IR approach is employed to compute the frequency response of a process and its slope at a certain frequency. This information can subsequently be utilised to tune PID controllers by employing the Kiss Circle tuner (KCT) as proposed by [21,32]. The operational concept of the KC tuner is depicted in Fig. 7. It involves the adjustment of the red point's position along the green circle, commonly referred to as the Kiss Circle (KC), with the objective of achieving a yellow angle of zero. The frequency response of the loop subsequently intersects with the circle. Any circle that effectively isolates the point -1 can be selected. The circle utilised in this study is characterised by a centre $C_e = -1.25$ and a radius

$R_a = 0.75$. It serves to delineate the "forbidden region" that the loop frequency response must steer clear of.

The following optimization problem has to be solved to determine the parameters of the PID controller

$$\min_{\alpha} \left[\left. \frac{dIm}{dRe} \right|_{\alpha} - \left. \frac{dI_L}{dR_L} \right|_{\omega} \right], \tag{32}$$

where $\left. \frac{dIm}{dRe} \right|_{\alpha} = \frac{cos(\alpha)}{sin(\alpha)}$ represents the slope of the forbidden region with respect to $\alpha \in [0, \alpha_{max}]$. The loop frequency response is defined as $L(j\omega) = C(j\omega)P(j\omega)$ or $L(j\omega) = R_L(j\omega) + jI_L(j\omega)$. Its slope, denoted as $\left. \frac{dI_L}{dR_L} \right|_{\bar{\omega}}$, can be computed using the derivative of the loop frequency response:

$$\left. \frac{dL(j\omega)}{d\omega} \right|_{\bar{\omega}} = P(j\bar{\omega}) \left. \frac{C(j\omega)}{d\omega} \right|_{\bar{\omega}} + C(j\bar{\omega}) \left. \frac{P(j\omega)}{d\omega} \right|_{\bar{\omega}} \tag{33}$$

or $\left. \frac{dL(j\omega)}{d\omega} \right|_{\bar{\omega}} = \left. \frac{R_L(j\omega)}{d\omega} \right|_{\bar{\omega}} + j \left. \frac{I_L(j\omega)}{d\omega} \right|_{\bar{\omega}}$. The design procedure based on the IR method can be summarized in a step-by-step approach:

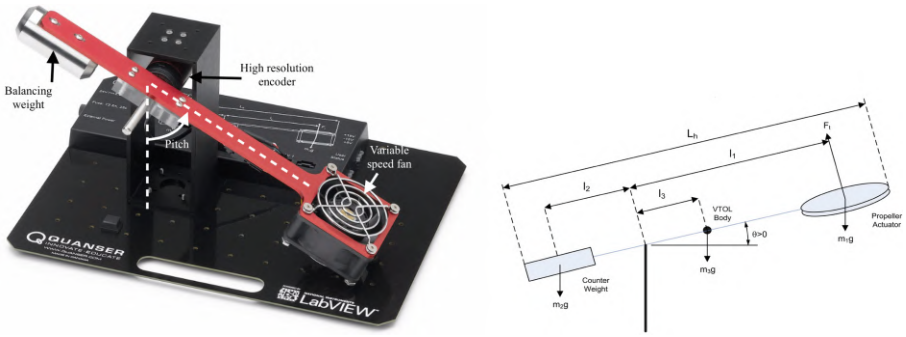
1. Select a test frequency denoted by $\bar{\omega}$. Typically, the aforementioned process is commonly referred to as the critical frequency process denoted as ω_c, at which the phase angle of $P(j\omega_c)$ equals $-180°$.
2. Construct a sinusoidal test input u(t), the IR-explorer signal, similarly to Fig. 6.
3. Apply u(t) as the input of the process, and subsequently measure the resultant output.
4. To acquire the process impulse response h(t), the signal h, as specified in (27), needs to be calculated using the IR-extractor. Subsequently, the signal h should be averaged.
5. Determine the values of the two complex numbers P=$P(j\bar{\omega})$ and dP=$\frac{dP(j\omega)}{d\omega}|_{\bar{\omega}}$, as well as the frequency response and its derivative at the test frequency $\bar{\omega}$.
6. The PID controller can be tuned using the KCT approach by utilising the parameters P and dP. Equation (32) is solved in order to adjust the red point's position along the KC circle with the main objective of achieving a yellow angle of zero, according to the tuning methodology from Fig. 7.

4 Experimental Validation

4.1 Vertical Take-Off and Landing

The autotuners discussed in Sect. 3 will undergo testing using a Vertical Take-Off and Landing (VTOL) Platform provided by Quanser. Figure 8a provides a visual representation of the experimental setup, while Fig. 8b illustrates a schematic model of the revolving beam.

The arm in Fig. 8a, indicated by the colour red, is outfitted with a counterweight on the left side and a fan with adjustable speed on the right side. Applying a voltage within the range of 0 to 10 V to the ventilator results in the rotational motion of the arm. This motion occurs about a stationary point located at one-third of the arm's overall length. The output signal corresponds to the angular displacement of the beam, which is determined by a high-resolution encoder as the angle formed between a vertical reference line and the moving arm. The output values range from -26 to 60°, with the reference point of 0° being the horizontal position of the beam. The process is highly nonlinear due to the angular movement of the fan/weight. The initial position of the VTOL platform in the lack of any input signal is at $-26°$. Figure 8 shows an experimental test performed on the VTOL unit where inputs from 3 to 9 V are given with 1 V increments. The nonlinear behavior of the platform can be observed in the pitch, since for equal step variations the response of the rotating beam varies in terms of amplitude, oscillations and settling time.

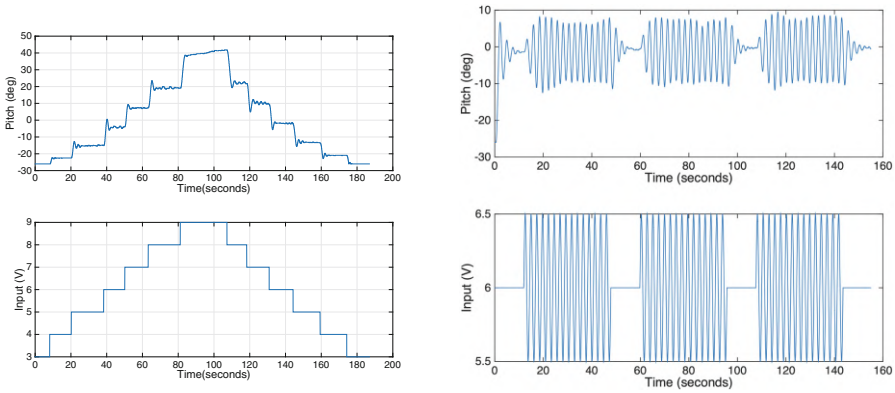


(a) The VTOL process (b) VTOL schematic

Fig. 8. Experimental tests on the VTOL unit

The process exhibits a significant degree of nonlinearity as a result of its design, as depicted in Fig. 8b. Empirical evidence demonstrates that the 0° position can be achieved by adding a voltage of 6.3 V to the input. Figure 9a illustrates the reaction of the Vertical Takeoff and Landing (VTOL) unit to step inputs, effectively demonstrating its nonlinearity. The response of the experimental unit when stimulated with a sine wave of 1 V amplitude and a critical frequency of 2.54 rad/s (equal to 0.4 Hz), with an amplitude of 6.3 V, is illustrated in Fig. 9b. Initially, the platform attains a horizontal position, which is then succeeded by three sinusoidal inputs of equal magnitude and frequency, commencing at $t = 14$s.

The critical frequency of the VTOL system has been determined through experimental investigation to be 2.54 rad/s. At the critical frequency, the system exhibits limited stability, as shown by a phase angle of $-180°$. The phase introduced by a fractional order proportional-integral (PI) controller ranges



(a) Step response

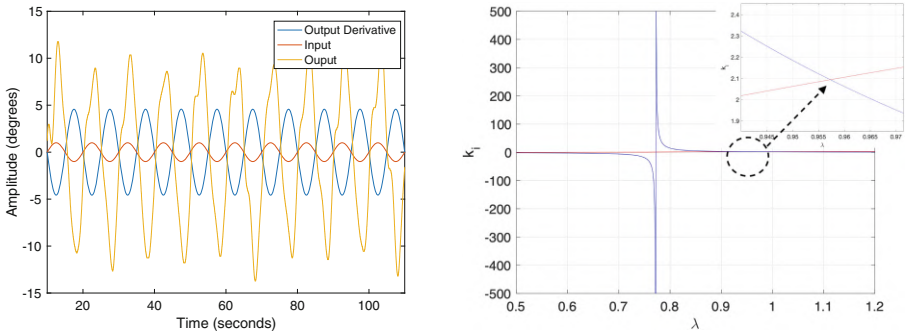
(b) Sine response
$A_i = 1V, \omega = \omega_c = 2.54 rad/s$

Fig. 9. The experimental unit

from -90λ to 0 degrees. Hence, it is advisable to select ω_{gc} to be significantly less than the critical frequency, thereby making a trade-off between settling time and stability. The selection of the gain crossover frequency, denoted as $\omega_{gc} = 0.1$ Hz $= 0.62831$ rad/s, has been made without any specific rationale. Additionally, a phase margin of $\varphi_m = 95°$ has been incorporated to guarantee the stability of the closed-loop system.

The subsequent procedure involves the application of a sine signal to the process and recording of the resulting output data. As previously, the Vertical Takeoff and Landing (VTOL) procedure requires approximately 15 s to attain the 0° orientation. Therefore, an input signal with a magnitude of 6.3 V is initially applied for a duration of 15 s. Subsequently, a sinusoidal waveform with an amplitude of 1 V and a frequency of 0.62831 rad/s is introduced. Figure 10a illustrates the normalised input and output waveforms. As can be observed, the output of the procedure deviates significantly from an ideal sinusoidal waveform. Therefore, the sine wave output is regenerated and subsequently supplied to the filter described in Eq. (19). The resulting output corresponds to the derivative of the phase, denoted as $\bar{y}$, which is also depicted in Fig. 10a. The computational framework employed in this stage consists of Matlab/Simulink. One potential approach involves the direct computation of the phase derivative within the LabVIEW environment, accompanied by real-time filtering applied to the output of the VTOL system. If the latter method is selected, it is imperative to ensure that the real-time denoiser does not cause any delays in the signal that is being supplied to the second order filter.

The process magnitude is obtained as $M(j\omega_{gc}) = 10.757$, while the phase $\varphi(j\omega_{gc}) = -15.5043°$ is determined by multiplying the frequency with the time difference between two consecutive zero crossings of the input and out-



(a) Normalized input, output and output derivative signals

(b) Graphical solution for k_i

Fig. 10. Controller tuning using the sine test auto-tuner

put signals, respectively. $\bar{M}(j\omega_{gc}) = 4.5764$ and $\bar{\varphi}(j\omega_{gc}) = -3.1418$ rad lead to $\frac{d\angle P(j\omega)}{d\omega}\Big|_{\omega=\omega_{gc}} = -0.4099$ rad.

Solving the system of nonlinear Eqs. (3)–(5) considering a fractional order PI controller ($\mu = 0$ in (2))

$$C_{FOPI}(s) = k_p\left(1 + \frac{k_i}{s^{\lambda}}\right) \tag{34}$$

is realized using a graphical approach shown in Fig. 10. The integral gain $k_i = 2.0942$ is at the intersection of the two plots, for $\lambda = 0.9572$, with $k_p = 0.0267$ obtained by replacing k_i and λ in the magnitude equation from (3).

This leads to the following fractional order controller

$$C_{Sine-Test}(s) = 0.0267\left(1 + \frac{2.0942}{s^{0.9572}}\right). \tag{35}$$

Experimental validations of the controller from (35) are performed in Fig. 11 for a sequence of step reference inputs. As can be seen, the platform is stable and reaches the desired setpoint value with a settling time of approx. 7 s and zero overshoot.

4.2 Mass Spring Damper

Let us examine the mass-spring-damper system depicted in Fig. 12. Mass-spring-damper systems are commonly employed for the investigation of yarn behavior in weaving machines, active suspension systems, trains or car chains on highways, tall buildings and aeroplane wings, among other applications. The output signal

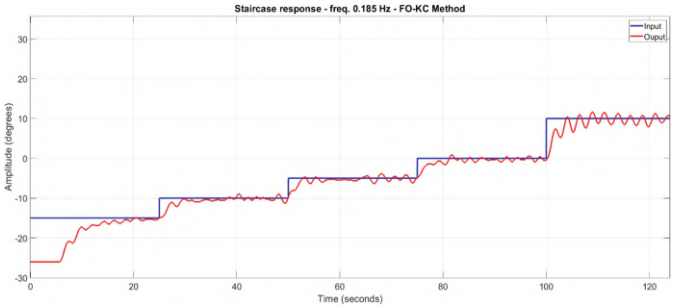


Fig. 11. Experimental validation of the Sine-Test autotuner on the VTOL process

and its derivative are generated by applying a sine-test signal with a frequency of 33.4 rad/s, as shown in Fig. 13. The measured output signal is evidently affected by stochastic disturbances.

The experimental apparatus under consideration is a mass-spring-damper system. The primary focus of measurement lies in determining the position of the second mass, which is situated at the midpoint of the system. The linear motor positioned on the leftmost side of the configuration serves as the input. The damper located on the leftmost side of the setup is intentionally left unconnected, hence resulting in inadequate damping within the system's dynamics. The mass spring damper has a nonlinear relationship between force and displacement of the masses.

In this subsection, the proposed technique is assessed using an integrative poorly damped system known as the mass-spring-damper (MSD) process, as depicted in Fig. 12. The plant under consideration consists of a single input, namely the voltage v(t) applied to the motor. Additionally, it comprises two outputs, which correspond to the locations of the masses M1 and M2. These positions are measured in centimetres and are obtained through the use of encoders. The actuator is the force applied by the motor, which is directly proportional to the input voltage v(t). The stimulus is exerting a direct effect on the M1 receptor. The primary objective of this study is to regulate the variable M1. The disconnection of the first and third springs in this configuration results in the transformation of the system into an integrating system.

The critical frequency of the process has been estimated as $\bar{\omega} = \omega_c = 5.5$ rad/s. Using the IR extractor method, the signal u(t) is applied to the process in Fig. 13. The real unit impulse response obtained using a linearised model of the process is also given in Fig. 13. In the presence of disturbances, the extracted and real unit impulse responses are different. However, the IR-extractor signal can be efficiently used to accurately estimate the process frequency response P and the slope dP, at the critical frequency: P = −1.92-0.06j and dP=0,63+17j. These estimated values are used to compute the FOPID controller using the KCT method from Sect. 3.

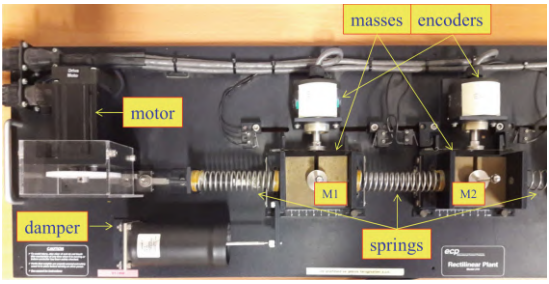


Fig. 12. The mass-spring-damper system used in this case study

Table 1. PID parameters for the mass-spring-damper system

Method	K_p	T_i	T_d
KCT	0.59	0.495	0.177
REL	0.31	0.566	0.142

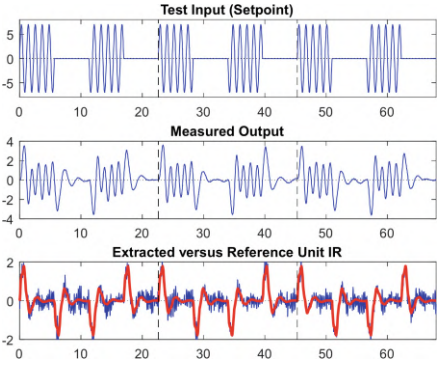


Fig. 13. The IR-explorer and IR-extractor for the mass-spring-damper process

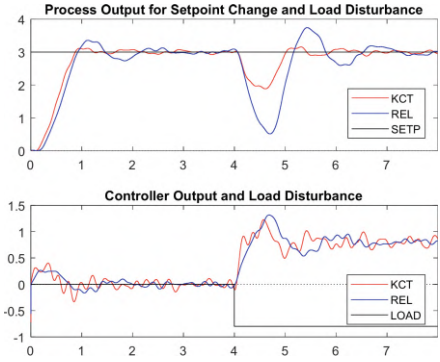


Fig. 14. Closed loop results for the mass-spring-damper system

For comparison purposes, the popular Relay tuner (REL) from [33] is also used to tune a PID controller. For this approach the critical gain is $K_c = -\frac{1}{real(P)} = 0.52$ with an oscillation period $T_c = 1.13$. For the REL approach, the resulting critical gain and period of oscillations are: $K_c = -\frac{1}{real(P)} = 0.52$ and $T_c = 1.13$. The FOPID controller parameters for the REL and KCT methods are indicated in Table 1. Figure 14 shows the closed loop results of the Mass-Spring-Damper system with the two KCT and REL autotuner methods. Note that for both controllers, the fractional order of differentiation and integration are obtained as 1, making these controllers classical, integer order, PIDs. The superiority of the KCT design methodology is highlighted through a smaller overshoot and improved settling time when compared to the REL autotuning methods. In addition, the KCT controller outperforms the REL controller from the disturbance rejection point of view.

5 Conclusion

Efficient tuning of controllers does not necesarily imply the use of a process model, even in the case of nonlinear processes. Most of the times, these need to be controlled in certain operating points. Knowledge about the process in this operating point is enough to design reliable and efficient fractional order PIDs. In this manuscript, two methods for extracting relevant process information are presented. Both are simple and efficient and can be applied to various types of stable processes. The methods lead to process frequency response and its slope which can be then used to design FOPIDs. The proposed methods are exemplified on two nonlinear poorly damped processes, a VTOL unit and a MSD system. The reference tracking results show the efficiency of the proposed methods.

This method is suitable for nonlinear systems because the input sine signal is applied in one operating point of the nonlinear system, where the relationship between the input and output signals can be considered linear.

Acknowledgment. This work was financed by grants of the Romanian Ministry of Research, Innovation and Digitization, PNRR-III-C9-2022-I8, grant number 760068/23.05.2023 and PNRR-III-C9-2022-I9, grant number 760018/27.01.2023. This work was partially supported by the grant of the Romanian Ministry of Education and Research, CNCS-UEFISCDI, project number PN-III-P1-1.1-PD-2021-0204.

References

1. Hornsey, S.: A review of relay auto-tuning methods for the tuning of PID-type controllers. Reinvention: Int. J. Undergraduate Res. **5** (2012)
2. Wang, Q.-G.: Autotuning of PID controllers. J. Process Control **11** (2001)
3. Zhuang, M., Atherton, D.P.: Automatic tuning of optimum PID controllers. IEE Proc. D: Control Theory Appl. **140** (1993)
4. Podlubny, I.: Fractional-order systems and piλdμ-controllers. IEEE Trans. Autom. Control **44**(1), 208–214 (1999)
5. Chen, Y.Q., Petráš, I., Xue, D.: Fractional order control - a tutorial (2009)
6. Monje, C.A., Vinagre, B.M., Feliu, V., Chen, Y.Q.: Tuning and auto-tuning of fractional order controllers for industry applications. Control Eng. Pract. **16**, 798–812 (2008)
7. Sondhi, S., Hote, Y.V.: Fractional order controller and its applications: a review. In: Proceedings of AsiaMIC (2012)
8. Wang, Z., Wang, C., Ding, L., Wang, Z., Liang, S.: Parameter identification of fractional-order time delay system based on legendre wavelet. Mech. Syst. Signal Process. **163**, 108141 (2022)
9. Birs, I., Copot, D., Muresan, C.I., De Keyser, R., Ionescu, C.M.: Robust fractional order pi control for cardiac output stabilisation, vol. 52 (2019)
10. Birs, I., Muresan, C., Nascu, I., Ionescu, C.: A survey of recent advances in fractional order control for time delay systems. IEEE Access **7**, 30951–30965 (2019)
11. Cristina I. Muresan, Isabela Birs, Clara Ionescu, Eva H. Dulf, and Robin De Keyser. A review of recent developments in autotuning methods for fractional-order controllers. *Fractal and Fractional*, 6(1), 2022

12. Chen, Y.Q., Moore, K., Vinagre, B., Podlubny, I.: Robust PID controller autotuning with a phase shaper (2004)
13. Yeroglu, C., Onat, C., Tan, N.: A new tuning method for piλdμ controller, p. II–312 (2009)
14. Valerio, D., da Costa, J.S.: Tuning of fractional PID controllers with Ziegler-Nichols-type rules. Signal Process. **86**(10), 2771–2784 (2006). Special Section: Fractional Calculus Applications in Signals and Systems
15. Tajjudin, M., Rahiman, M.H.F., Arshad, N.M., Adnan, R.: Robust fractional-order pi controller with Ziegler-Nichols rules. Int. J. Electr. Comput. Eng **7** (2013)
16. Gude, J.J., Kahoraho, E.: Modified Ziegler-Nichols method for fractional pi controllers. In: 2010 IEEE 15th Conference on Emerging Technologies & Factory Automation (ETFA 2010), pp. 1–5 (2010)
17. Mughees, A., Mohsin, S.A.: Design and control of magnetic levitation system by optimizing fractional order PID controller using ant colony optimization algorithm. IEEE Access **8** (2020)
18. Chen, P., Luo, Y., Peng, Y., Chen, Y.Q.: Optimal robust fractional order piλd controller synthesis for first order plus time delay systems. ISA Trans. **114** (2021)
19. De Keyser, R., Muresan, C.I., Ionescu, C.M.: A novel auto-tuning method for fractional order PI/PD controllers. ISA Trans. **62**, 268–275 (2016)
20. De Keyser, R., Muresan, C.I., Ionescu, C.M.: Autotuning of a robust fractional order PID controller, vol. 51, pp. 466–471. Elsevier B.V. (2018)
21. De Keyser, R., Muresan, C.I., Ionescu, C.M.: Universal direct tuner for loop control in industry. IEEE Access **7**, 81308–81320 (2019)
22. Birs, I., Muresan, C., Mihai, M., Dulf, E., De Keyser, R.: Tuning guidelines and experimental comparisons of sine based auto-tuning methods for fractional order controllers. IEEE Access **10**, 86671–86683 (2022)
23. De Keyser, R., Birs, I., Muresan, C., Ionescu, C.: Process information extraction for controller tuning from minimal datasets with low excitatory content. ISA Trans. (under review)
24. Li, Z., Liu, L., Dehghan, S., Chen, Y.Q., Xue, D.: A review and evaluation of numerical tools for fractional calculus and fractional order controls. Int. J. Control **90**, 1165–1181 (2015)
25. Cajo, R., Mac, T.T., Plaza, D., Copot, C., De Keyser, R., Ionescu, C.: A survey on fractional order control techniques for unmanned aerial and ground vehicles. IEEE Access **7** (2019)
26. Tepljakov, A., et al.: Towards industrialization of fopid controllers: a survey on milestones of fractional-order control and pathways for future developments. IEEE Access **9** (2021)
27. Kang, S., Wu, H., Yang, X., Li, Y., Pan, L., Chen, B.: Fractional robust adaptive decoupled control for attenuating creep, hysteresis and cross coupling in a parallel piezostage. Mech. Syst. Signal Process. **159**, 107764 (2021)
28. Podlubny, I.: Fractional-order systems and piλdμ controllers. IEEE Trans. Autom. Control **44**(1), 208–214 (1999)
29. Petras, I., Tenreiro Machado, J.A., Baleanu, D.: Fractional Order Systems: Modeling and Control Applications. Springer (2018)
30. Muresan, C.I., De Keyser, R., Birs, I.R., Folea, S., Prodan, O.: An autotuning method for a fractional order PD controller for vibration suppression, pp. 245–256. Springer, Cham (2019)
31. Monje, C., Chen, Y.Q., Vinagre, B., Xue, D., Feliu, V.: Fractional Order Systems and Control - Fundamentals and Applications (2010)

32. De Keyser, R., Ionescu, C.M., Muresan, C.I.: Comparative evaluation of a novel principle for PID autotuning. In: 11th Asian Control Conference (ASCC), Gold Coast, Australia, 17–20 December 2017, pp. 1164–1169 (2017)
33. Åström, K.J., Hägglund, T.: Revisiting the Ziegler-Nichols step response method for PID control. J. Process Control **14**(6), 635–650 (2004)



Fractional-Order Servo Linear Quadratic Regulator Control for a Magnetic Levitation System

Yuki Moriguchi[1], Masaharu Kuroda[2(✉)], and Natsuki Kawaguchi[2]

[1] Mazda Motor Corporation, Fuchu-cho, Aki-gun, Hiroshima 730-8670, Japan
[2] University of Hyogo, 2167 Shosha, Himeji, Hyogo 671-2280, Japan
m-kuroda@eng.u-hyogo.ac.jp

Abstract. Under uncontrolled conditions, voltage-controlled suspension-type magnetic levitation systems are unstable systems in which a levitated object either falls or is attracted to the electromagnet. Furthermore, the magnetic force and the electrical characteristics of the circuit have strong nonlinearities. This study applies fractional-order linear quadratic regulator (LQR) control, which combines fractional calculus and modern control theory, to a benchmark problem: a voltage-controlled suspension-type magnetic levitation system. Furthermore, a fractional-order servo LQR control method is developed for a case in which a levitated object is controlled to a target position that deviates from the equilibrium point. To perform state-feedback control such as fractional-order servo LQR control, which requires information of all the states, a fractional-order state-observer is designed to estimate fractional-order states. Simulation results demonstrate that fractional-order servo LQR control enables us not only to achieve stabilization of the equilibrium point; it also enables target-value tracking. Experimental results confirmed that the fractional-order servo LQR control provides higher control performance than that of conventional integer-order servo LQR control.

Keywords: Fractional Calculus · Linear Quadratic Regulator (LQR) · Magnetic Levitation System · Servo Control

1 Introduction

Fractional calculus refers to calculus operations that extend the order of differentiation and integration to non-integers. The application of fractional calculus to engineering has enabled the concise description of dynamic properties of viscoelastic materials such as silicone gel [1]. In the field of control engineering, fractional-order proportional-integral-differential (PID) control, in which fractional calculus is applied to PID control in classical control theory, already exists [2]. Reportedly, it is robust against nonlinearities [3]. Nevertheless, the application of fractional calculus to modern control theory has long stagnated because the stability, controllability, and observability of systems described using a state-space representation with fractional-order derivatives have remained unclear. However, their clarification in the late 1990s by Matignon et al. [4, 5] made it possible to apply fractional calculus to modern control theory.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 184–198, 2025.
https://doi.org/10.1007/978-3-031-72794-8_13

A magnetic levitation system can levitate a magnetic object in the air by magnetic force, notably without contact. The several categories of levitating an object by magnetic force include repulsion-type magnetic levitation, which uses the repulsive force acting between permanent magnets, and attraction-type magnetic levitation, which controls the attractive force acting between an electromagnet and a magnetic object. For example, active magnetic bearings employ attraction-type (suspension-type) magnetic levitation [6]. Suspension-type magnetic levitation systems are unstable because, without control, the levitated object is either attracted to the electromagnet, or it falls. Moreover, the magnetic force and the electrical characteristics of the circuit have strong nonlinearities. Additionally, a voltage-controlled system is a system in which the input to the system is a voltage, with two degrees of freedom: the gap length between the levitated object and the electromagnet, and the electrical current flowing in the circuit. Furthermore, the input voltage acts indirectly on the gap length through the electrical current. Consequently, it is also an underactuated system with fewer inputs than outputs.

Therefore, taking up the achievement of stable servo control of a magnetic levitation system as a benchmark problem, this study was conducted to combine a linear quadratic regulator (LQR), which is a modern control theory, with fractional calculus, which is regarded as robust against nonlinearity, to stabilize an unstable nonlinear system.

2 Preliminary Knowledge Related to Fractional Calculus

For order α, which is $0 < \alpha < 1$, the fractional-order derivative and integral are defined as [7]

$$ {}^{RL}D_t^{\alpha} f(t) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_0^t (t-\tau)^{-\alpha} f(\tau) d\tau, \tag{1} $$

$$ {}^{RL}D_t^{-\alpha} f(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-\tau)^{\alpha-1} f(\tau) d\tau, \tag{2} $$

$$ {}^{C}D_t^{\alpha} f(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-\tau)^{-\alpha} \frac{df(\tau)}{d\tau} d\tau, \tag{3} $$

where ${}^{RL}D_t^{\alpha}$ and ${}^{RL}D_t^{-\alpha}$ respectively represent the Riemann–Liouville fractional-order derivative and integral operators, ${}^{C}D_t^{\alpha}$ denotes the Caputo fractional-order differential operator, and Γ represents the gamma function. The Caputo fractional-order derivative has the following commutativity and additivity [8]:

$$ {}^{C}D_t^{\beta} \left\{ {}^{C}D_t^{\alpha} f(t) \right\} = {}^{C}D_t^{\alpha} \left\{ {}^{C}D_t^{\beta} f(t) \right\} = {}^{C}D_t^{\alpha+\beta} f(t). \tag{4} $$

The relation between the Riemann–Liouville definition and the Caputo definition can be described as shown. If the initial condition $f(0)$ is equal to zero, then the Riemann–Liouville derivative and the Caputo derivative are equivalent.

$$ {}^{RL}D_t^{\alpha} f(t) = {}^{C}D_t^{\alpha} f(t) + \frac{t^{-\alpha}}{\Gamma(1-\alpha)} f(0) = {}^{C}D_t^{\alpha} f(t) \tag{5} $$

Therefore, when the initial condition $f(0)$ is set to zero, Caputo differentiation after Riemann–Liouville integration is equivalent to Riemann–Liouville differentiation after Riemann–Liouville integration. Riemann–Liouville integration followed by Riemann–Liouville differentiation returns the function to its original form below.

$$ {}^{C}D_t^{\alpha}\left\{{}^{RL}D_t^{-\alpha}f(t)\right\} = {}^{RL}D_t^{\alpha}\left\{{}^{RL}D_t^{-\alpha}f(t)\right\} = {}^{RL}D_t^{\alpha-\alpha}f(t) = f(t) \tag{6} $$

3 Suspension-Type Magnetic Levitation System

3.1 Simultaneous Nonlinear Differential Equations for Magnetically Levitated System

In a suspension-type magnetic levitation system, a steel ball is levitated in the air by the magnetic force generated by an electromagnet. Let $x(t)$ be the gap length between the electromagnet and the steel ball. The steel ball and the coil of the electromagnet produce inductance $L(x)$. This inductance can be modeled as

$$ L(x) = \frac{Q}{x(t) + X_{\infty}} + L_{\infty}, \tag{7} $$

where Q, X_{∞}, and L_{∞} respectively denote physical parameters determined by the relation between the steel ball and the electromagnet.

First, the integer-order differential equations for the magnetic levitation system are expressed as [9, 10]

$$ \begin{cases} \dot{x}_1(t) = x_2(t) \\ \dot{x}_2(t) = g - \frac{Qx_3^2(t)}{2M\{x_1(t)+X_{\infty}\}^2} \\ \dot{x}_3(t) = \frac{x_3(t)\left[Qx_2(t) - R\{x_1(t)+X_{\infty}\}^2\right]}{Q\{x_1(t)+X_{\infty}\}+L_{\infty}\{x_1(t)+X_{\infty}\}^2} + \frac{x_1(t)+X_{\infty}}{Q+L_{\infty}\{x_1(t)+X_{\infty}\}}u(t) \end{cases}, \tag{8} $$

where M stands for the mass of the levitated steel ball, R expresses the electric resistance of the entire circuit, and g signifies the acceleration of gravity. State variable $x_1(t)$ represents the gap length between the electromagnet and the steel ball, $x_2(t)$ denotes the steel ball velocity, and $x_3(t)$ stands for the electric current flowing through the circuit. Additionally, $u(t)$ is the voltage applied to the electromagnet as a control input.

3.2 Linearization Around an Equilibrium Point

Deviations from the equilibrium points in the state and in the input are introduced as

$$ \boldsymbol{x_d}(t) = \boldsymbol{x}(t) - \boldsymbol{x_{st}}, u_d(t) = u(t) - u_{st}. \tag{9} $$

With these, Eq. (8) is linearized around the equilibrium point $\boldsymbol{x_{st}} = \begin{bmatrix} x_{1st} & x_{2st} & x_{3st} \end{bmatrix}^T$.

$$ \frac{d}{dt}\boldsymbol{x_d}(t) = \boldsymbol{A}\boldsymbol{x_d}(t) + \boldsymbol{B}u_d(t) \tag{10} $$

Therein, $\boldsymbol{A}$and $\boldsymbol{B}$ are the matrices shown as

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 \\ k_a & 0 & k_b \\ 0 & k_c & k_d \end{bmatrix}, \ \boldsymbol{B} = \begin{bmatrix} 0 & 0 & k_e \end{bmatrix}^T. \tag{11}$$

In the matrices described above, k_a, k_b, k_c, k_d, and k_e are expressed respectively as

$$\begin{aligned} &k_a = \frac{Qx_{3st}^2}{M\{x_{1st} + X_\infty\}^3}, k_b = -\frac{Qx_{3st}}{M\{x_{1st} + X_\infty\}^2}, k_c = \frac{Qx_{3st}}{L_{st}(x_{1st} + X_\infty)^2}, \\ &k_d = -\frac{R}{L_{st}}, \ and \ k_e = \frac{1}{L_{st}}, \end{aligned} \tag{12}$$

where L_{st} denotes the inductance at the equilibrium point.

3.3 Extension to Fractional-Order State Equation

A voltage-controlled suspension-type magnetic levitation system usually has only integer-order states. Therefore, it must be rewritten as a 0.5-order differential equation model based on the additivity of the Caputo derivative presented in Eq. (4).

First, $\boldsymbol{x_{df}}(t)$ represents an extended new state vector consisting of integer-order derivative states and fractional-order derivative states, as shown below.

$$\begin{aligned} &\boldsymbol{x_{df}}(t) = \\ &\begin{bmatrix} x_{df,1}(t) & x_{df,2}(t) & x_{df,3}(t) & x_{df,4}(t) & x_{df,5}(t) & x_{df,6}(t) \end{bmatrix}^T \\ &= \begin{bmatrix} x_{d,1}(t) & {}^C D_t^{0.5} x_{d,1}(t) & {}^C D_t^{1.0} x_{d,1}(t) \\ {}^C D_t^{1.5} x_{d,1}(t) & x_{d,3}(t) & {}^C D_t^{0.5} x_{d,3}(t) \end{bmatrix}^T \end{aligned} \tag{13}$$

Using this new state vector $\boldsymbol{x_{df}}$, the state-space model extended to a 0.5-order differential equation is obtainable as.

$${}^C D_t^{0.5} \boldsymbol{x_{df}}(t) = \boldsymbol{A_f} \boldsymbol{x_{df}}(t) + \boldsymbol{B_f} u_d(t), y_d(t) = \boldsymbol{C_f} \boldsymbol{x_{df}}(t), \tag{14}$$

where $\boldsymbol{A_f}$, $\boldsymbol{B_f}$, and $\boldsymbol{C_f}$ are the matrices shown as the following.

$$\boldsymbol{A_f} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ k_a & 0 & 0 & 0 & k_b & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & k_c & 0 & k_d & 0 \end{bmatrix}, \boldsymbol{B_f} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ k_e \end{bmatrix}, \boldsymbol{C_f} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}^T \tag{15}$$

4 Fractional-Order Servo LQR Control

4.1 Extension to Fractional-Order Servo LQR Control

In actuality, LQR control sets all states to zero. However, for the magnetic levitation system considered here, the levitated object must be tracked to a target position that is distant from the equilibrium point. Therefore, the control system must be extended to a servo control system that includes an integrator.

First, a new state $z_1(t)$ is configured as the integral of the error between the target value r and the output $y_d(\tau)$ as

$$z_1(t) = \int_0^t (y_d(\tau) - r)d\tau = {}^{RL}D_t^{-1.0}(y_d(t) - r). \tag{16}$$

Second, the initial condition is presumed to be zero; then the 0.5-order derivative of $z_1(t)$ is derived as

$$\begin{aligned} {}^{C}D_t^{0.5}\{z_1(t)\} &= {}^{C}D_t^{0.5}\left\{{}^{RL}D_t^{-1.0}(y_d(t) - r)\right\} \\ ={}^{RL}D_t^{0.5}\left\{{}^{RL}D_t^{-1.0}(y_d(t) - r)\right\} &= {}^{RL}D_t^{-0.5}(y_d(t) - r) = z_2(t). \end{aligned} \tag{17}$$

Third, the 0.5-order derivative of $z_2(t)$ yields the next equation based on Eq. (6):

$$\begin{aligned} {}^{C}D_t^{0.5}\{z_2(t)\} &= {}^{C}D_t^{0.5}\left\{{}^{RL}D_t^{-0.5}(y_d(t) - r)\right\} \\ ={}^{RL}D_t^{0.5}\left\{{}^{RL}D_t^{-0.5}(y_d(t) - r)\right\} &= y_d(t) - r = \boldsymbol{C_f x_{df}}(t) - r. \end{aligned} \tag{18}$$

Consequently, to provide trackability to target value r, the extended state equation, which is a combination of the fractional-order state equation (14), and Eqs. (17) and (18), can be composed as

$$\begin{aligned} {}^{C}D_t^{0.5}\boldsymbol{x_w}(t) &= \begin{bmatrix} 0 & 1 & \boldsymbol{0} \\ 0 & 0 & \boldsymbol{C_f} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A_f} \end{bmatrix}\boldsymbol{x_w}(t) + \begin{bmatrix} 0 \\ 0 \\ \boldsymbol{B_f} \end{bmatrix}u_d(t) + \begin{bmatrix} 0 \\ -r \\ \boldsymbol{0} \end{bmatrix} \\ &= \boldsymbol{A_w x_w}(t) + \boldsymbol{B_w}u_d(t) + \boldsymbol{E_w}, \end{aligned} \tag{19}$$

where $\boldsymbol{x_w} = \begin{bmatrix} z_1 & z_2 & \boldsymbol{x_{df}} \end{bmatrix}^T$.

For this extended state equation, the fractional-order LQR controller can be derived by finding the feedback gains which minimize the following evaluation function J as

$$J = \int_0^\infty \{\boldsymbol{x_w}^T(t)\boldsymbol{Q}\boldsymbol{x_w}(t) + Ru_d^2(t)\}dt, \tag{20}$$

where $\boldsymbol{Q}$ and R are the respective weights assigned for the states and the input. This evaluation function is minimized by numerical calculation because the minimum value cannot be obtained analytically [11].

Based on the extended system, the optimal feedback gains are calculated numerically using the gradient method [12], in which the feedback gain vector $\boldsymbol{F}$ obtained from solution $\boldsymbol{P}$ of the algebraic Riccati equation is used for the initial values [13].

Algebraic Riccati Equation:

$$\boldsymbol{P}\boldsymbol{A}_{\boldsymbol{w}} + \boldsymbol{A}_{\boldsymbol{w}}^{T}\boldsymbol{P} - \boldsymbol{P}\boldsymbol{B}_{\boldsymbol{w}}R^{-1}\boldsymbol{B}_{\boldsymbol{w}}^{T}\boldsymbol{P} + \boldsymbol{Q} = \boldsymbol{0} \tag{21}$$

Optimal Feedback Gain Vector:

$$\boldsymbol{F} = R^{-1}\boldsymbol{B}_{\boldsymbol{w}}^{T}\boldsymbol{P} \tag{22}$$

It is noteworthy that the third term on the right-hand side of Eq. (19) is a constant term. It does not affect the design of the optimal feedback gain.

4.2 Configuration of Weight Matrices Q and R

For integer-order systems, the states are integer-order. Consequently, the weight matrix $\boldsymbol{Q}$ in the evaluation function for integer-order LQR control has components for the first-order integration, displacement, velocity, and current terms. By contrast, for fractional-order systems, fractional-order states exist in addition to integer-order states.

For a fair comparison between integer-order LQR control and fractional-order LQR control, the weight matrix $\boldsymbol{Q}$ for fractional-order LQR control should still have components only for the first-order integration, displacement, velocity, and current terms. The weight components corresponding to the fractional-order states are therefore set to zero as presented below.

$$\boldsymbol{Q} = diag\left[\,10^{6}\ 0\ 10^{3}\ 0\ 10^{6}\ 0\ 1\ 0\,\right], R = 1. \tag{23}$$

4.3 Fractional-Order State Observer

Fractional-order servo LQR control is state feedback that requires all states. The output signal here is the displacement of the levitated object. Therefore, the state observer must be configured to estimate other states. However, estimating fractional-order states is impossible with conventional state observers. A fractional-order state observer that can estimate not only velocity and current is needed, but it must include fractional-order states, which cannot be measured using a displacement sensor. To do so, a fractional-order state-observer is used [14]. The state equation for the fractional-order state-observer is given as

$${}^{C}D_{t}^{0.5}\hat{\boldsymbol{x}}_{\boldsymbol{df}}(t) = \boldsymbol{A}_{\boldsymbol{f}}\hat{\boldsymbol{x}}_{\boldsymbol{df}}(t) + \boldsymbol{B}_{\boldsymbol{f}}u_{d}(t) + \boldsymbol{H}\big(y_{d}(t) - \boldsymbol{C}_{\boldsymbol{f}}\hat{\boldsymbol{x}}_{\boldsymbol{df}}(t)\big), \tag{24}$$

where $\hat{\boldsymbol{x}}_{\boldsymbol{df}}(t)$ denotes the estimated states and where $\boldsymbol{H}$ is the observer gain obtained using the pole-assignment method. A block diagram of a fractional-order state observer is shown here. The 1/2-order integral element is approximately realized using an Oustaloup filter [7].

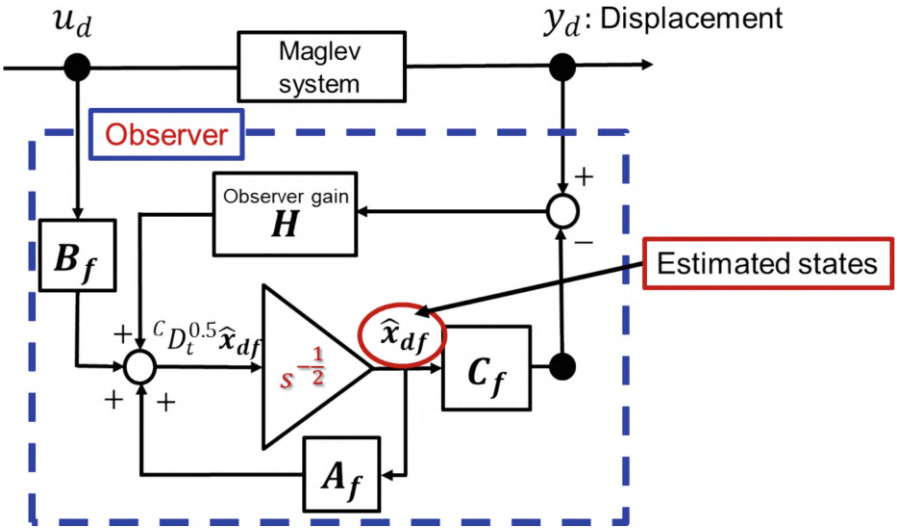


Fig. 1. Block diagram of fractional-order state observer.

Finally, the input to minimize the evaluation function for the extended system is expressed as

$$u(t) = u_{st} - F_{z_1} \int_0^t (y_d(\tau) - r)d\tau - F_{z_2}{}^{RL}D_t^{-0.5}(y_d(t) - r) - \boldsymbol{F_x}\hat{\boldsymbol{x}}_{\boldsymbol{df}}(t), \tag{25}$$

where F_{z_1}, F_{z_2} and $\boldsymbol{F_x}$ denote the optimal feedback gains.

4.4 Simulation Results

Figure 2 presents a control result obtained using the control input expressed by Eq. (25) and the fractional-order state observer shown in Fig. 1. The magnetic levitation model is linearized at the stage of designing the fractional-order servo LQR controller, but the obtained fractional-order servo LQR control is applied, of course, to the original nonlinear differential equation model of the magnetic levitation system in this simulation. The equilibrium point x_{1st} is set as 5 mm. Figure 2 confirms that the fractional-order servo LQR control can achieve not only stabilization of the equilibrium point but also target-value tracking while the magnetic levitation system is stabilized.

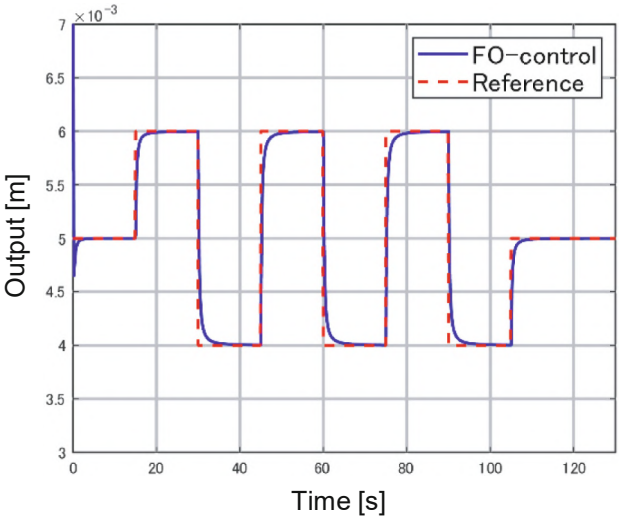


Fig. 2. Fractional-order (FO) servo LQR control result (output).

An impulse-like response at the initial time of the simulation in Fig. 2 is presumably caused by the fact that the estimation of states has not yet converged near the initial time, resulting in a large control input. However, no such impulse-like response occurred near the initial time in the experiments. Therefore, the cause of the impulse-like response at the initial time is left as an investigation for future work.

5 Experimental Results

5.1 Experimental Setup Including the Control Loop

Figure 3 portrays a schematic diagram of the entire system of a voltage-controlled attractive-force-type magnetic levitation system. An electromagnet is installed at the top. Also, a steel ball, which is the levitated object, is placed below the electromagnet. A laser displacement sensor is installed below the steel ball to measure its position.

First, the voltage signal from the laser displacement sensor is sent to the digital signal processor (DSP) via the A/D board. Based on the displacement signal, the DSP estimates each state and generates the control input. The control signal voltage from the DSP is output through the D/A board and is amplified by the bipolar power supply. It is then input to the electromagnet coil. Here, the sampling period is 1 ms. The input voltage is limited to 0.01–35 V.

The physical parameters obtained from the identification experiments are shown in Table 1. The displacement of the equilibrium point is set as 5 mm.

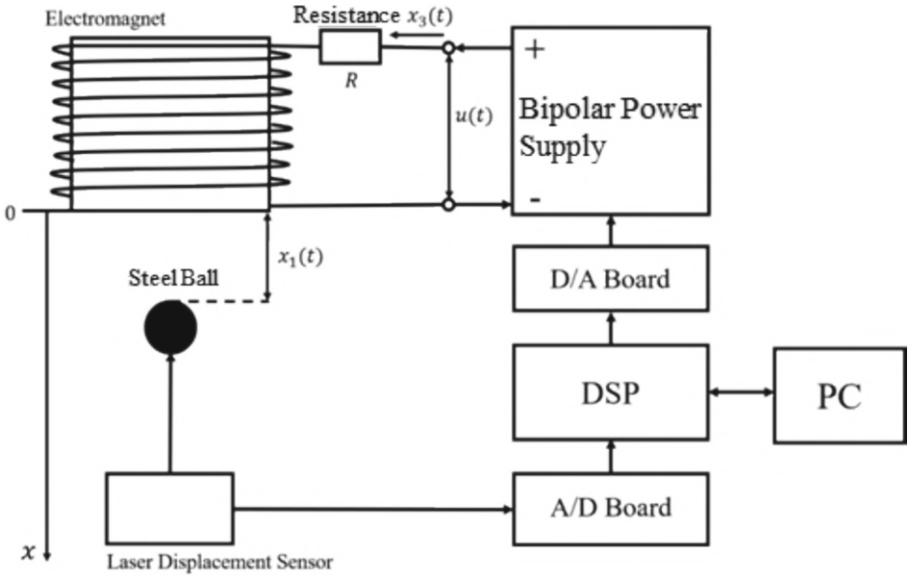


Fig. 3. Whole image of a voltage-controlled suspension-type magnetic levitation system.

Table 1. Physical parameters of the magnetic levitation system

Symbol	Description	Value
M	Mass of the steel ball	63.669×10^{-3} [kg]
Q	Coefficient of the coil	2.22×10^{-4} [Hm]
L_∞	Inductance determined by leakage flux	0.446 [H]
X_∞	Distance determined by relative permeability	6.1×10^{-3} [m]
R	Resistance of the whole circuit	31.12 [Ω]
g	Acceleration of gravity	9.81 $\left[\mathrm{m/s^2}\right]$

5.2 Servo LQR Feedback Gains and State Observer Gains

Numerical calculations were performed using the gradient method under the condition of tracking control from the initial displacement of 7 mm to the target value of 5 mm. The obtained fractional-order servo LQR feedback gains $\boldsymbol{F_{fo}}$ are shown below.

$$\begin{aligned}\boldsymbol{F_{fo}} &= \left[F_{z1}\ F_{z2}\ \boldsymbol{F_x} \right] \\ &= \left[-4995.3\ -11140\ -11149\ -5460.6 \right. \\ &\quad \left. -853.8133\ -47.9512\ 80.5297\ 4.5787 \right]\end{aligned} \tag{26}$$

The integer-order servo LQR feedback gains $\boldsymbol{F_{io}}$ are also shown for comparison.

$$\boldsymbol{F_{io}} = \left[F_{ioz}\ \boldsymbol{F_{iox}} \right] = \left[-1000.0\ -14392\ -1054.9\ 124.2715 \right] \tag{27}$$

The fractional-order state observer gains $\boldsymbol{H_{fo}}$ obtained using the pole assignment method are

$$\begin{aligned} \boldsymbol{H_{fo}} = [&-3.1732 \times 10^{-11}\ 263.2294\ -5.2710 \times 10^{-9} \\ &20316\ 3.0418 \times 10^{-9}\ -2611.5]. \end{aligned} \tag{28}$$

The following method is used for determining the poles used in the pole assignment method. First, a necessary and sufficient condition for the dynamic characteristics of an α-order fractional system to be asymptotically stable is that the argument of its pole p_i satisfies the following equation [15].

$$|\arg(p_i)| > \frac{\alpha\pi}{2}, i = 1, 2, \cdots, n \tag{29}$$

In the case examined for this study, $\alpha = 0.5$. Moreover, the eigenvalues of the 0.5-order fractional system are the 0.5th powers of the eigenvalues of the original integer order system. Therefore, the 0.5th powers of the poles used to obtain the gains of the original integer-order state observer are used to obtain the gains of the fractional-order state observer. As a result, the poles used for obtaining the fractional-order state observer gains are

$$\left[\pm 10\mathrm{i}\ \pm\sqrt{110}i\ \pm\sqrt{120}i \right]. \tag{30}$$

The integer-order state observer gains $\boldsymbol{H_{io}}$ are also shown for comparison.

$$\boldsymbol{H_{io}} = \left[263.2294\ 20316\ -2611.5 \right]. \tag{31}$$

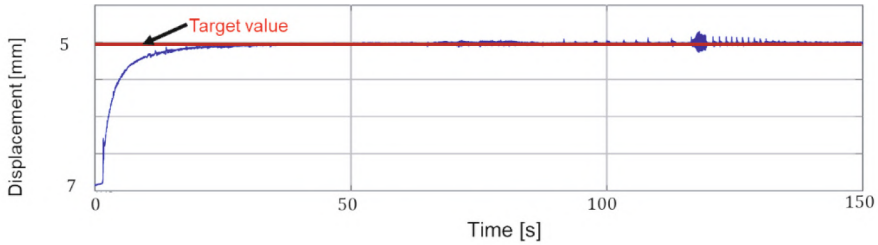
The poles used for obtaining the integer-order state observer gains are shown below.

$$\left[-100\ -110\ -120 \right] \tag{32}$$

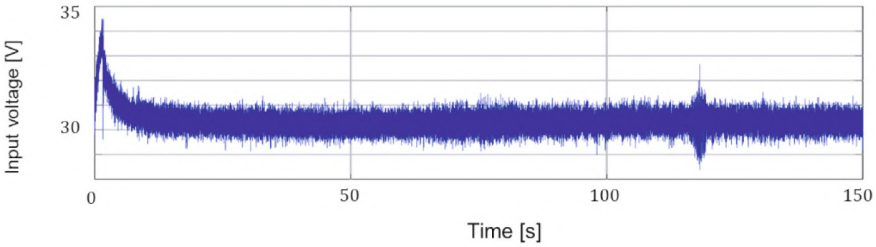
These gains were used in actual experiments described in the following sections.

5.3 Tracking from 7 to 5 mm by Fractional-Order Servo LQR

The results obtained for tracking control from 7 to 5 mm by fractional-order servo LQR control are presented in Fig. 4. The upper graph shows the steel ball position. The lower graph shows the control input.

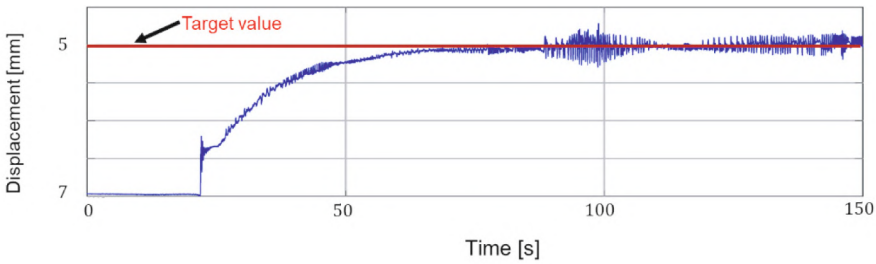


(a) Output measured using laser displacement sensor (equilibrium point: 5 mm)

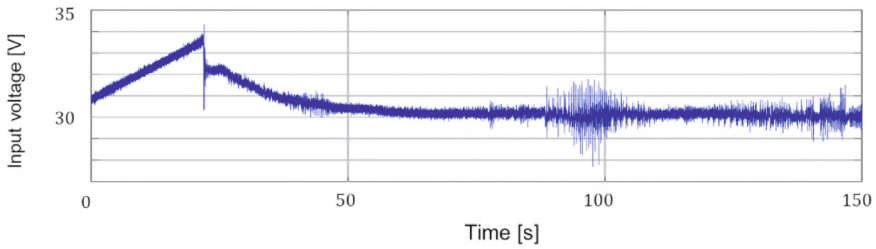


(b) Fractional-order servo LQR control input

Fig. 4. Tracking control by fractional-order servo LQR.



(a) Output measured using laser displacement sensor (equilibrium point: 5 mm)



(b) Integer-order servo LQR control input

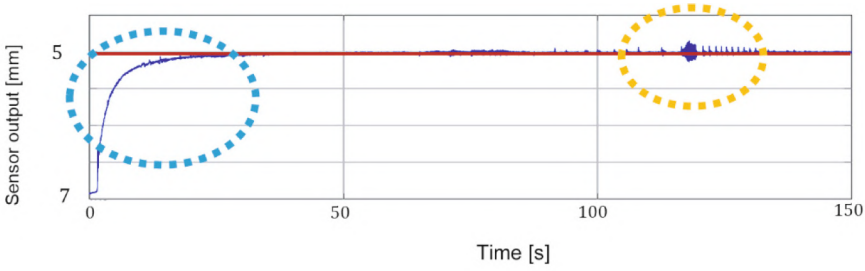
Fig. 5. Tracking control by integer-order servo LQR.

5.4 Tracking from 7 to 5 mm by Integer-Order Servo LQR

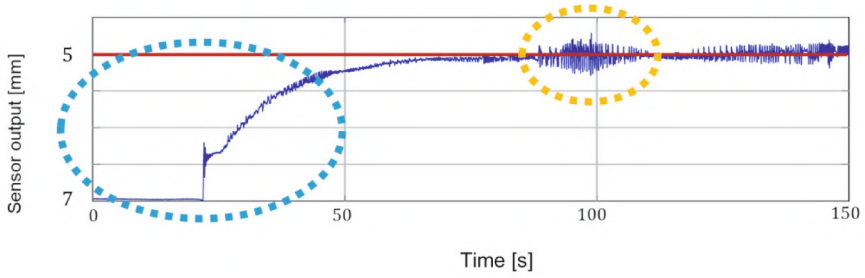
For comparison, the results obtained for the same tracking control by integer-order servo LQR control are presented in Fig. 5.

5.5 Comparison Between Fractional-Order and Integer-Order Servo LQRs

As displayed in Fig. 6, a comparison of the control results for fractional-order servo LQR with those for integer-order servo LQR indicates that the conventional method has a slow start-up, whereas the proposed method quickly reaches the target value (blue dotted circle). Convergence is also improved by fractional-order servo LQR control. A small disturbance induces oscillations (yellow dotted circle). These oscillations are large for the integer-order servo LQR but they are easily suppressed for the fractional-order servo LQR. These results demonstrate the benefit of fractional-order state feedback.



(a) Fractional-order servo LQR control (proposed method)



(b) Integer-order servo LQR control (conventional method)

Fig. 6. Comparison between fractional-order and integer-order servo LQRs.

6 Analysis of Feedback Signals

6.1 Influence Ratio for State Feedback Signals

The effectiveness of fractional-order LQR control is discussed in this section. As shown in Eq. (25), the control input to the magnetic levitation system is the sum of the equilibrium input, the integral inputs of the deviation between the output and the target value,

and the state feedback input. The first three terms become constant after the levitated ball reaches the target position. Therefore, the fourth term, the state feedback input, is dominant in the steady state.

Consequently, the effect of steady-state feedback on disturbances is considered by determining the influence ratio of the steady-state feedback signals. The influence ratio $p_i(t)$ is defined as the absolute value of each feedback signal divided by the sum of the absolute values of all feedback signals as presented below.

$$p_i(t) = \frac{\left| -F_{x,i} \times \hat{x}_{df,i}(t) \right|}{\sum_{j=1}^{6} \left| -F_{x,j} \times \hat{x}_{df,j}(t) \right|} \times 100, (i = 1, 2, \cdots, 6) \tag{33}$$

The following graph presents results of influence ratio calculations based on data from 117 s to 118 s in Fig. 6(a), for which oscillations attributable to external disturbances are observed.

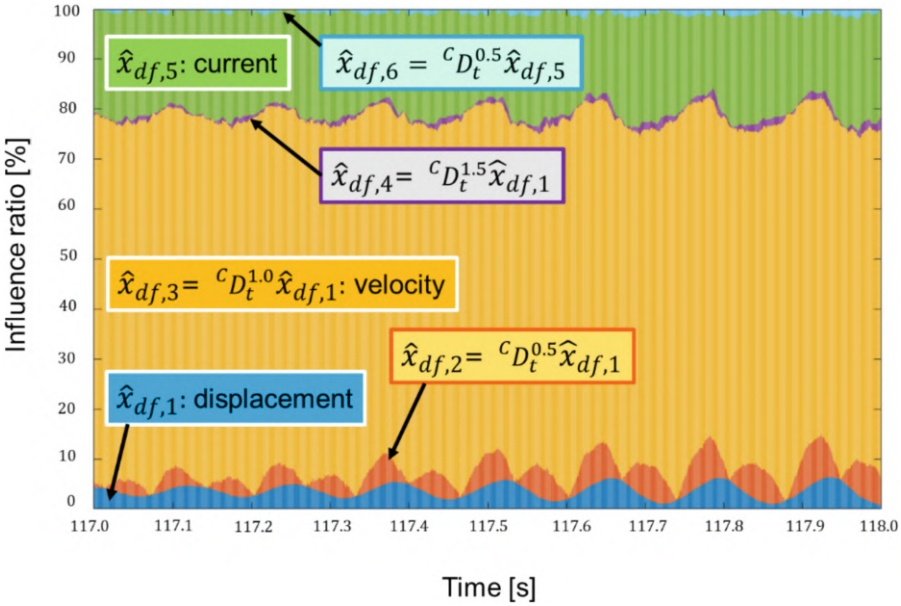


Fig. 7. Influence ratio for each state feedback signal.

6.2 Discussion of the Influence of State Feedback Signals

In Fig. 7, the horizontal axis represents time. The vertical axis is the influence ratio. It is apparent that inputs from the velocity and current feedback account for a large proportion of all control inputs.

Table 2 presents the maximum influence ratio for the feedback signal of each state. Although the influence of feedback from the integer-order states is considerable, the

influence of the fractional-order feedback is also non-negligible. It can be inferred that the fractional-order state feedback control, which supplements the integer-order state feedback control as the main control, functions to suppress vibration.

Table 2. Maximum influence ratio for the feedback signal of each state

Maximum influence ratio for each feedback signal ($117 \leq t \leq 118$)	
$\hat{x}_{df,1}$(displacement)	6.37%
$\hat{x}_{df,2} = {}^{C}D_t^{0.5}\hat{x}_{df,1}$	9.09%
$\hat{x}_{df,3}={}^{C}D_t^{1.0}\hat{x}_{df,1}$(velocity)	77.06%
$\hat{x}_{df,4}={}^{C}D_t^{1.5}\hat{x}_{df,1}$	2.42%
$\hat{x}_{df,5}$(current)	22.40%
$\hat{x}_{df,6}={}^{C}D_t^{0.5}\hat{x}_{df,5}$	2.08%

Furthermore, it has been confirmed through experimentation that the influence of the fractional-order state feedback terms is present on the order of several percent, even before the steady state is reached. Therefore, it can be inferred that the contribution of fractional-order state feedback to the improvement of rise time, as pointed out in Sect. 5.5, is also considerable. However, because the effect of the changes in the integral terms of the deviation must also be considered before the steady state in these experiments, this possibility can only be noted for now.

7 Conclusions

Fractional calculus was applied to LQR control for the servo control of an attractive-force-type magnetically levitated object. A setup for experimentation was constructed to evaluate a magnetic levitation system. The control system was able to levitate a steel ball stably with fractional-order servo LQR control using a fractional-order state observer. The application of linear control to a strongly nonlinear system performed better than the conventional control method, confirming the feasibility of fractional-order state feedback. The influence of each state feedback on the control input was investigated. The findings confirmed fractional-order state feedback as suitable for vibration control.

Future work shall include clarification of the reason for the impulse-like response which appears near the initial time in the simulation, and examination of the effect of the fractional-order state feedback terms on shortening of the rise time before the steady state is reached.

Acknowledgments. This work was supported by JSPS KAKENHI Grant Number JP23K03749. The authors appreciate constructive comments offered by the reviewers, which helped to improve the quality of this paper.

References

1. Shimizu, N.: Dynamic characteristics of a viscoelastic oscillator. Trans. Jpn. Soc. Mech. Eng. Ser. C **61**(583), 902–906 (1995)
2. Petráš, I.: Tuning of the non-linear fractional-order controller. In: Proceedings of the 2019 20th International Carpathian Control Conference (ICCC), Krakow-Wieliczka, Poland, pp. 1–4 (2019)
3. Odai, M., Hori, Y.: Controller design robust to nonlinear elements based on fractional order control system. IEEJ Trans. Ind. Appl. **120**(1), 11–18 (2000)
4. Matignon, D.: Stability results for fractional differential equations with applications to control processing. Comput. Eng. Syst. Appl. **2**, 963–968 (1996)
5. Matignon, D., d'Andrea-Novel, B.: Some results on controllability and observability of finite-dimensional fractional differential systems. Comput. Eng. Syst. Appl. **2**, 952–956 (1996)
6. Ji, J.C., Hansen, C.H., Zander, A.C.: Nonlinear dynamics of magnetic bearing systems. J. Intell. Mater. Syst. Struct. **19**(12), 1471–1491 (2008)
7. Xue, D.: Fractional-Order Control Systems: Fundamentals and Numerical Implementations. De Gruyter, Berlin, Boston (2017)
8. Podlubny, I.: Fractional Differential Equations. Academic Press, San Diego (1999)
9. Yang, Z.-J., Miyazaki, K., Jin, C.-Z., Wada, K.: Physical parameter identification of a magnetic levitation system under a robust nonlinear controller. In: Proceedings of the 15th Triennial World Congress of the International Federation of Automatic Control, p. 6 (2002)
10. Yang, Z.-J., Tsubakihara, H., Kanae, S., Wada, K.: Robust nonlinear control of a voltage-controlled magnetic levitation system using disturbance observer. IEEJ Trans. Electron. Inf. Syst. **127**(12), 2118–2125 (2007)
11. Ikeda, F., Kawata, S., Watanabe, A.: An optimal regulator design of fractional differential systems. Trans. Soc. Instrum. Control Eng. **37**(9), 856–861 (2001)
12. Arabi, S.H., Merrikh-Bayat, F.: A practical method for designing linear quadratic regulator for commensurate fractional-order systems. J. Optim. Theory Appl. **174**, 550–566 (2017)
13. Sierociuk, D., Vinagre, B.M.: Infinite horizon state-feedback LQR controller for fractional systems. In: Proceedings of the 49th IEEE Conference on Decision and Control (CDC), Atlanta, GA, USA, pp. 6674–6679 (2010)
14. Dadras, S., Momeni, H.R.: A new fractional order observer design for fractional order nonlinear systems. In: Proceedings of the ASME 2011 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Volume 3: 2011 ASME/IEEE International Conference on Mechatronic and Embedded Systems and Applications, Parts A and B, pp. 403–408. ASME, Washington, DC, USA (2011)
15. Hartley, T.T., Lorenzo, C.F.: Dynamics and control of initialized fractional-order systems. Nonlinear Dyn. **29**, 201–233 (2002)



Rigorous Numerical Computations with the Chebyshev Interpolation Using the Julia Language

Shinsuke Kondo[1] and Akitoshi Takayasu[2(✉)]

[1] Graduate School of Science and Technology, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
s2220530@s.tsukuba.ac.jp

[2] Institute of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
takitoshi@risk.tsukuba.ac.jp

Abstract. Chebyshev interpolation is a method for approximating a function that is numerically stable and can provide accurate approximations even for highly oscillatory functions. In this study, we focus on numerical implementations of Chebyshev interpolation. In particular, we introduce a method for numerically computing the derivative, rootfinding, computing maxima and minima of a given Chebyshev interpolation function using the Julia language. The main contribution of this study is that our implementation can be rigorously computed based on interval arithmetic.

Keywords: Chebyshev interpolation · Rigorous numerical computations · Julia language

1 Introduction

Numerical computations with computers are subject to errors such as rounding errors, truncation errors, and discretization errors. In most cases, these errors are small enough not to worry about them all the time and do not significantly affect the computational results. However, in some cases, they can have a large impact on computational results, which can be considered a risk for research using numerical computation in various fields. Therefore, it is very important to understand the errors in numerical computations and to find them properly. To address the problem of errors in numerical computations, *interval arithmetic* is used as a powerful tool [1]. This approach involves representing numbers as intervals, thereby capturing the uncertainty and potential error range inherent in numerical computations. Interval arithmetic then provides a clearer picture of the error bounds that arise in numerical computations. For example, instead of representing a number as a single value, it is expressed as an interval, such as $[a, b]$, indicating that the exact value lies somewhere between a and b.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 199–209, 2025.
https://doi.org/10.1007/978-3-031-72794-8_14

The Chebyshev interpolation is introduced in a number of textbooks on numerical analysis, e.g. [3]. It is well known that the approximation of a function by the Chebyshev polynomial is the best possible foundation for work on many topics in numerical analysis [4]. Recently, from the point of view that approximation theory is extremely close to computation, L. N. Trefethen and his collaborators have introduced a paradigm of numerical computation with Chebyshev interpolations. It is called *Chebfun* [5], which is also the name of their MATLAB toolbox. In their works with the Chebfun, classical computations by numbers are beautifully replaced by the Chebyshev interpolations, and many computations in numerical analysis are performed by function-based computations. This is our main motivation in this study. On the other hand, the Chebfun paradigm still ignores the errors in numerical computations. No matter how accurate a function approximation is, it is not possible to get out of the classical approximation framework. In this study, we implement a *rigorous* numerical computing framework with Chebyshev interpolation using interval arithmetic and try to free ourselves from "approximation".

Our implementation is based on the Julia language, which is a high-level dynamic programming language widely used in scientific and numerical computing. Like other scientific computing languages such as MATLAB and Python, Julia allows for fast computations with simple syntax. As an preceding work, there is a numerical analysis package called *ApproxFun.jl* [6] in the Julia language. This package is in a similar vein to the Chebfun, which provides fast interpolation by approximating functions and implements many functions, including differentiation, integration, and calculation of maxima and minima of the Chebyshev interpolation. In this context, our main contribution is to develop a computational tool for performing mathematically rigorous computations using the Chebyshev interpolation. In practical implementation, the Julia language provides an interval arithmetic package called *IntervalArithmetic.jl* [2]. This package facilitates the use of interval arithmetic in computational tasks. It provides a comprehensive set of tools and functions that allow us to easily implement interval arithmetic in our numerical computations. By using this package, one can not only perform interval arithmetic operations, but also gain insight into the error margins of their computations, thus increasing the reliability of their numerical results.

The paper is organized as follows: In Sect. 2, we present the definition of Chebyshev interpolation and show how we implement the computation of the coefficients of Chebyshev interpolation (Chebyshev coefficients). In Sect. 3, we introduce a procedure for obtaining the coefficients in the derivative of the Chebyshev interpolation. Section 4 describes how to rigorously compute the eigenvalues of a matrix, called the Colleague matrix, to find the roots of the Chebyshev interpolation. In Sect. 5, we show numerical results of rigorous computation of the maximum and minimum values of the Chebyshev interpolation using the above computational techniques. In Sect. 6, we conclude the paper and describe future works in this study.

2 Chebyshev Interpolation

Let $T_n(x) := \cos(n\theta)$ ($\theta = \arccos(x)$, $x \in [-1, 1]$) be the n th order *Chebyshev polynomial* of the first kind. The Chebyshev interpolation is given by a linear combination of the Chebyshev polynomials, that is

$$p(x) = a_0 + 2\sum_{n=1}^{M-1} a_n T_n(x), \tag{1}$$

where a_n is called the *Chebyshev coefficients*. Since the Chebyshev series is the cosine Fourier series with the transformation $x = \cos\theta$, the Chebyshev coefficients are numerically given via the FFT [4] by

$$a_n = \frac{1}{2M-2}\sum_{j=0}^{2M-3} f(\cos\theta_j)e^{-\pi i\frac{nj}{M-1}}, \quad \theta_j = \frac{\pi j}{M-1}, \quad h = \frac{\pi}{M-1}, \quad 0 \le n < M.$$

The Chebyshev interpolation is used to approximate a given function $f(x)$ by interpolation at the *Chebyshev points* $x_j = \cos\theta_j$ (of the second kind), which are the points most often used in practical computation. Using polynomial approximation in general, it is possible to approximate numerical solutions of complex equations or differential equations. The Chebyshev interpolation is used to handle data obtained through polynomial approximation more efficiently. This can be applied to the solutions of equations modeling phenomena in engineering and other fields. In order to perform the Chebyshev interpolation, it is important to know the value of the Chebyshev coefficients a_n. In actual implementation, the number of bases M is systematically determined because the number of bases is closely related to the accuracy of the approximation. The following is an algorithm for obtaining the Chebyshev coefficients with the appropriate number of terms:

Step1. Compute $2M-2$ Fourier coefficients by performing Fast Fourier Transform (FFT) based on the function values at $2M-2$ (starting from $M = 5$) Chebyshev points.

Step2. Extract M Chebyshev coefficients from the relationship between the Chebyshev coefficients with the Fourier coefficients. Step 1 & 2 are done with `chebcoeffs` function.

Step3. Increase M to $2^i + 1 (i = 2, 3, ...)$ and compute the same procedure in Steps 1 &2.

Step4. Determine the appropriate number M for which the last term a_{M-1} of the Chebyshev coefficients is less than a given tolerance (set 5×10^{-15} in this case), and finish.

Conducting with the above procedure, we implement the following Julia code to compute the Chebyshev coefficients within the tolerance. The code returns the Chebyshev coefficients with the appropriate number of terms for interpolating the given function.

```
function cheb(f,I=[-1;1];tol = 5e-15,Nmax = 10000)
    a = I[1]; b = I[2]; m = 0.5*(a+b); r = 0.5*(b-a); x = rand(5)
    x1 = m .+ x*r; x2 = m .- x*r
    if f.(x1) = f.(x2)
        odd_even = 1 # even function: 1
    elseif f.(x1) = -f.(x2)
        odd_even = -1 #  odd function: -1
    else
        odd_even = 0 # otherwise: 0
    end
    i = 3
    schbc = 0 # sampling chebyshev coefficients
    while true
        schbc = chebcoeffs(f,2^i+1,I)
        if all(abs.(schbc[end-2:end]) .< tol) || (2^i+1 > Nmax)
            break
        end
        i += 1
    end
    M = findlast(abs.(schbc) .> tol)
    cc = schbc[1:M]
    if odd_even == 1 # even function
        cc[2:2:end] .= 0
    elseif odd_even == -1 # odd function
        cc[1:2:end] .= 0
    end
    return cc # return Two-sided Chebyshev
end
```

2.1 Rigrous Error Bounds for Chebyshev Interpolation

Let $\mathcal{E}_\rho$ denote the *Bernstein* ellipse by

$$\mathcal{E}_\rho := \left\{ \frac{\rho z + (\rho z)^{-1}}{2} : z = \exp(2i\pi x),\ x \in [-1,1] \right\}, \quad \rho > 0.$$

Letting f be analytic in $\mathcal{E}_\rho$, the interpolation error between the original function $f(x)$ and its Chebyshev interpolation, say $p(x)$, is bounded by the following formula [4, Theorem 8.2].

$$\sup_{x\in[-1,1]} |f(x) - p(x)| \le \frac{4\rho^{-(M-1)}}{\rho - 1} \sup_{z\in\mathcal{E}_\rho} |f(z)|. \tag{2}$$

The term $\sup_{z\in\mathcal{E}_\rho} |f(z)|$ in the right hand side is the supremum of absolute value of f in the Bernstein ellipse. Since f is assumed to be analytic in the ellipse, one can obtain the rigorous truncation error bound based on this formula. For

example, let $f(x) = \exp(1-x^2)$, which is obviously analytic in $\mathcal{E}_\rho$. The truncation error for the Chebyshev interpolation can be rigorously estimated via interval arithmetic by

$$\frac{4\rho^{-(M-1)}}{\rho - 1} \sup_{z \in \mathcal{E}_\rho} |f(z)| \leq 1.22 \cdot 10^{-13},$$

where ρ is the sum of its semimajor plus semiminor axis lengths for the ellipse and is taken as $\rho = 4.79$.

3 Derivative of Chebyshev Interpolation

Let the derivative of the Chebyshev interpolation (1) be denoted by

$$p'(x) = b_0 + 2 \sum_{n=1}^{M-2} b_n T_n(x).$$

The relationship between a_n and b_n is obtained in [3] by reversing a process in the integration of the Chebyshev polynomials. Resulting relation is given by the form $b_{n-1} = b_{n+1} + 2na_n$ with $b_{M+1} = b_M = 0$ for $n = M-1, M-2, \ldots, 1$. Therefore, the coefficients b_n are derived from a_n. More precisely, we have

$$b_n = \sum_{\substack{k=n+1 \\ k-n:\text{odd}}}^{M-1} 2ka_k. \tag{3}$$

We implement the above procedure. This Julia function maps the Chebyshev coefficients a_n to its derivative b_n in (3). It is worth noting that this operation can be computed rigorously based on the interval arithmetic.

```
function chebdiff(ia::Vector{Interval{T}}; I=[-1,1]) where T
    M = length(a)
    ib = zeros(Interval{T},M+1)
    for r = M-1:-1:1
        ib[r] = ib[r+2] + 2*r*ia[r+1]
    end
    ib[1] /= 2.0
    return ib[1:end-2]*(2/(interval(I[2])-I[1]))
end
```

4 Rootfinding Procedure

It is well known that the roots of a linear combination of monomials x^k are equal to the eigenvalues of a certain companion matrix formed from its coefficients. Similarly, the real roots of the Chebyshev interpolation of the form

$$p(x) = \sum_{n=0}^{M-1} c_n T_n(x), \quad c_{M-1} \neq 0, \quad c_n = \begin{cases} a_0 & (n=0) \\ 2a_n & (n \geq 1) \end{cases}$$

are the real eigenvalues of the matrix

$$C = \begin{pmatrix} 0 & 1 & & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & & \\ & \frac{1}{2} & 0 & \frac{1}{2} & & \\ & & \ddots & \ddots & \ddots & \\ & & & & & \frac{1}{2} \\ & & & & \frac{1}{2} & 0 \end{pmatrix} - \frac{1}{2c_{M-1}} \begin{pmatrix} & & & & \\ & & & & \\ & & & & \\ & & & & \\ c_0 & c_1 & c_2 & \cdots & c_{M-2} & c_{M-1} \end{pmatrix},$$

where entries not displayed are all zero, see, e.g., [4, Theorem 18.1,p58]. This matrix C is called the *Colleague matrix*.

Rigorous inclusion of each eigenvalue is obtained by using the following theorem:

Theorem 1 ([1, p. 86, **Theorem 3.16**]). *Given matrices A and $B \in \mathbb{C}^{n\times n}$, define the mapping $F : \mathbb{C}^{n+1} \to \mathbb{C}^{n+1}$ as*

$$F(\lambda, x) := \begin{bmatrix} x^H x - 1 \\ Ax - \lambda Bx \end{bmatrix}.$$

Let $\bar{x} \in \mathbb{C}^n$ and $\bar{\lambda} \in \mathbb{C}$ be an approximate solution satisfying $F(\bar{\lambda}, \bar{x}) \approx 0$, and let $R \in \mathbb{C}^{n+1\times n+1}$ be an approximate inverse of the Jacobian matrix ($R \approx DF(\bar{\lambda}, \bar{x})$) of the map F at $(\bar{\lambda}, \bar{x})$. Let $\boldsymbol{y} \in \mathbb{IC}^n$ and $\boldsymbol{d} \in \mathbb{IC}$ be an interval vector and an interval, respectively. Let us also define $\boldsymbol{w} := (\boldsymbol{d}, \boldsymbol{y}^T)^T \in \mathbb{IC}^{n+1}$, where ^{T} denotes the transpose of the vector. Moreover, let us define

$$g(\boldsymbol{w}) := z + \left(I - R \cdot DF\left(\bar{\lambda} + \boldsymbol{d}, \bar{x} + \boldsymbol{y}\right)\right) \boldsymbol{w},$$

$$DF(\lambda, x) := \begin{bmatrix} 0 & 2x^T \\ -B\boldsymbol{x} & A - \boldsymbol{\lambda} B \end{bmatrix},$$

$$z := -R \begin{bmatrix} \bar{x}_1^2 + \bar{x}_2^2 + \cdots + \bar{x}_n^2 - 1 \\ A\bar{x} - \bar{\lambda} B\bar{x} \end{bmatrix}.$$

If $g(\boldsymbol{w}) \subset \mathrm{int}(\boldsymbol{w})$ holds, then the exact eigenvalue λ of the generalized eigenvalue problem $Ax = \lambda Bx$ exists uniquely within the interval $\bar{\lambda} + \boldsymbol{d}$.

The following code rigorously includes eigenpair that solves the generalized eigenvalue problem based on Theorem 1.

```
function verifyeig(A,lam,x,B = I)
    x = x./sqrt(x'*x)
    ysize = length(x)
    ix = map(Interval,x)
    DF(w) = [0 transpose(2*(x+w[2:end])) ; -B*(x+w[2:end]) A-(lam.+w[1])*B]
    zero =zeros(ysize+1)
    R = inv(mid.(DF(zero)))
    z = -R*[dot(ix,ix)-1;A*ix-lam*B*ix]
    #Amount of modification of Newton's method
    eps = 2*maximum(sup,z)
    if isreal(lam) && isreal(x)
```

```
        lam = real(lam)
        x = real(x)
        id = eps*interval.(-1,1)
        iy =  eps*interval.(-ones(ysize),ones(ysize))
    else
        id =Complex(interval(- eps,eps),interval(-eps,eps))
        iy = Complex.(interval(-eps,eps)*ones(ysize), interval(-eps,eps)*ones(ysize))
    end
    iw = [id;iy]
    g(w) = z+(I-R*DF(w))*w
    gw = g(iw)
    if all(gw . iw)
        return lam .+ gw[1]
    else
        return NaN
    end
end
```

5 Numerical Results: Getting Maximum and Minimum Values of Analytic Functions

The actual procedure of rigorously computing the maximum and minimum values of an analytic function is as follows:

Step1. Compute the Chebyshev coefficients a_n to obtain the Chebyshev interpolation $p(x)$ of $f(x)$ using the method introduced in Sect. 2. The truncate error of Chebyshev interpolation is obtained by (2) via interval arithmetic, which is done by Julia's package IntervalArithmetic.jl.

Step2. Compute the coefficients b_n of the derivative $p'(x)$ using the formula (3) introduced in Sect. 3, where b_n is computed as the intervals.

Step3. The roots of $p'(x)$ (the values of x such that $p'(x) = 0$) are obtained by rigorously computing the eigenvalues of the Colleague matrix with respect to b_n. In this case, Theorem 1 is used to rigorously include the eigenvalues, and the exact roots, which denote extrema of $p(x)$, are computed via interval arithmetic.

Step4. The maximum and minimum values are determined by comparing the values at the endpoints with the extreme values calculated from the following equation.

Example. We perform an example of rigorously computing maximum and minimum values of the function

$$f(x) = \exp(\mathrm{erf}((x^2) + x^5)) \sin(5\pi x) + x.$$

Table 1. Chebyshev coefficients of $p(x)$.

n	Chebyshev coefficients a_n
1	0.28654
2	1.82193
⋮	⋮
51	$1.4160 \cdot 10^{-14}$
52	$-2.0906 \cdot 10^{-14}$
53	$8.5402 \cdot 10^{-15}$

Table 1 shows the results of computing the Chebyshev coefficients a_n using the method introduced in Step 1. Furthermore, the truncate error given by (2) is given by $2.63 \cdot 10^{11}$.

The results of computing the Chebyshev coefficients b_n for the derivative $p'(x)$ using Step 2 are shown in Table 2. Since interval arithmetic is performed, the result is displayed by intervals. We then compute the eigenvalues of the Colleague matrix for b_n. The root of $p'(x)$ is obtained via applying Theorem 1. Computing the function value at each root, we have extrema of the function shown in Table 3. In addition, the function values of the Chebyshev interpolation at $x = -1, 1$ are obtained by the form

Table 2. Interval inclusion of Chebyshev coefficients of $p'(x)$.

n	Chebyshev coefficients b_n
1	$[-5.98664, -5.98663]$
2	$[-9.97652, -9.97651]$
⋮	⋮
50	$[2.30413 \cdot 10^{-12}, 2.30414 \cdot 10^{-12}]$
51	$[-2.13245 \cdot 10^{-12}, -2.13244 \cdot 10^{-12}]$
52	$[8.88178 \cdot 10^{-13}, 8.88179 \cdot 10^{-13}]$

$$p(-1) = \sum_{n=0}^{M} (-1)^{n-1} a_n$$

$$p(1) = \sum_{n=0}^{M} a_n$$

Comparing the values at the endpoints with extrema, one can determine the maximum and minimum value of the function $f(x)$. The result is shown in

Table 3. Interval inclusion of the roots of $p'(x)$ and extrema of $p(x)$.

x	$p(x)$
$[-0.894769, -0.894768]$	$[-2.07482, -2.07481]$
$[-0.697367, -0.697366]$	$[0.71125, 0.711251]$
$[-0.506159, -0.506158]$	$[-1.78375, -1.78374]$
$[-0.298876, -0.298875]$	$[0.804071, 0.804072]$
$[-0.104963, -0.104962]$	$[-1.11439, -1.11438]$
$[0.104967, 0.104968]$	$[1.1144, 1.11441]$
$[0.299221, 0.299222]$	$[-0.809362, -0.809361]$
$[0.508625, 0.508626]$	$[1.87254, 1.87255]$
$[0.708057, 0.708058]$	$[-1.28889, -1.28888]$
$[0.919009, 0.91901]$	$[4.88583, 4.88584]$

Table 4. Figure 1 shows the profile of $f(x)$ in $[-1, 1]$, which matches our rigorous computing result in Table 4.

Table 4. Interval inclusion of extrema and function values at end points: bold values are the maximum and minimum value of $f(x)$ in $x \in [-1, 1]$.

x	$p(x)$
-1	$[-1.00001, -0.999999]$
$\mathbf{[-0.894769, -0.894768]}$	$\mathbf{[-2.07482, -2.07481]}$
$[-0.697367, -0.697366]$	$[0.71125, 0.711251]$
$[-0.506159, -0.506158]$	$[-1.78375, -1.78374]$
$[-0.298876, -0.298875]$	$[0.804071, 0.804072]$
$[-0.104963, -0.104962]$	$[-1.11439, -1.11438]$
$[0.104967, 0.104968]$	$[1.1144, 1.11441]$
$[0.299221, 0.299222]$	$[-0.809362, -0.809361]$
$[0.508625, 0.508626]$	$[1.87254, 1.87255]$
$[0.708057, 0.708058]$	$[-1.28889, -1.28888]$
$\mathbf{[0.919009, 0.91901]}$	$\mathbf{[4.88583, 4.88584]}$
1	$[0.999999, 1.00001]$

In this example, we have presented a result of the rigorous numerical computation for obtaining the maximum and minimum values of $f(x)$. The above procedure does not produce similar results for all functions, but it is guaranteed that the result is rigorous if the procedure succeeds. For example, the following result is returned for $g(x) = \cos(3\pi x)$ because the procedure fail to rigorously

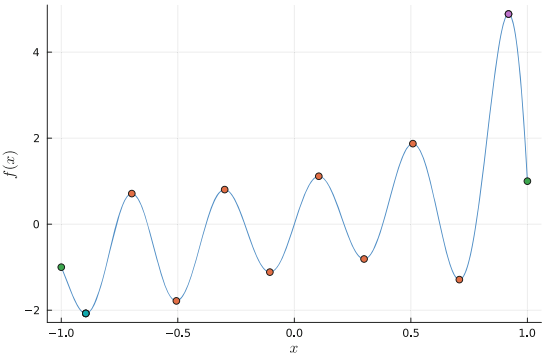


Fig. 1. The profile of $f(x)$: extrema and function values at end points are plotted by orange points.

obtain the eigenvalue of the Colleague matrix, which is Step 3. In our implementation, NaN is returned as shown in Table 5 if the procedure fails.

Table 5. A result of failure for $g(x)$: obtaining the rigorous inclusion of eigenvalues of the Colleague matrix (Step 3) is failed. Some values are succeeded actually but all eigenvalues are not obtained.

Interval inclusion eigenvalues
[NaN, NaN]
$[-0.666668, -0.666666]$
[NaN, NaN]
$[-4.66313 \cdot 10^{-9}, 4.66313 \cdot 10^{-9}]$
[NaN, NaN]
$[0.666665, 0.666668]$
[NaN, NaN]

6 Conclusion and Future Works

In this study, we developed a computational tool for Chebyshev interpolation and rigorous numerical computations using the Julia language. Our procedure is based on the Chebyshev interpolation with interval arithmetic. The function values at the Chebyshev point are transformed into the Chebyshev coefficients. The derivative operation is performed by the relational formula (3). The rigorous computation of the roots of the Chebyshev interpolation is based on the rigorous inclusion of the eigenvalues of the Colleague matrix. If we obtain rigorous inclusion of such eigenvalues, then extrema of the target function are computed

via interval arithmetic. Then, by comparing the extrema with the function values at the endpoints, our procedure provides rigorous inclusion of the maximum and minimum values of the target function. Our future work is to apply this computational methodology to the rigorous computation of numerical solutions to differential equations used as mathematical models for real-world problems.

Acknowledgement. AT is partially supported by JSPS KAKENHI Grant Numbers JP22K03411, JP21H01001, and JP20H01820.

References

1. Oishi, S., et al.: Principle of Verified Numerical Computations. Corona Publishers, Tokyo, Japan (2018). (in Japanese)
2. Julia, IntervalArithmetic.jl (2017). https://juliapackages.com/p/intervalarithmetic
3. Mason, J.C., Handscomb, D.C.: Chebyshev Polynomials, 1st edn. Chapman and Hall/CRC (2002)
4. Trefethen, L.N.: Approximation Theory and Approximation Practice. Extended Edition. SIAM (2019)
5. Driscoll, T.A., Hale, N., Trefethen, L.N. (eds.): Chebfun Guide. Pafnuty Publications, Oxford (2014)
6. Olver, S., Townsend, A.: A practical framework for infinite-dimensional linear algebra. In: Proceedings of the 1st First Workshop for High Performance Technical Computing in Dynamic Languages, pp. 57–62 (2014)



Parallel Implementation of a Verified Contour Integral Based Eigensolver and Its Performance Evaluation Using the Julia Language

Shota Seto[1(✉)] and Akitoshi Takayasu[2]

[1] Graduate School of Science and Technology, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
s2320544@u.tsukuba.ac.jp

[2] Institute of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan

Abstract. In this paper, we introduce a parallel implementation of a numerical verification method for generalized Hermitian eigenvalue problems to compute rigorous error bounds for eigenvectors and eigenvalues in a prescribed domain. The verification method is based on the Sakurai-Sugiura method that has hierarchical parallelism and fully utilizes the performance by parallel implementation. We implement a rigorous version of the Sakurai-Sugiura method using the Julia language and evaluate its performance. This paper provides the development in the fields of mechanics and electromagnetics etc. because it reduces the computation time, while computes error bounds rigorously for eigenvectors and eigenvalues.

1 Introduction

In various fields of science and engineering, such as meteorological observations, structural design, and fluid analysis, the technology for numerically solving various equations using computers has advanced. In addition, recent developments in multi-core processors and general-purpose graphics processing units (GPUs) have made it possible to perform large and complex numerical simulations using extensive computing resources. In many numerical simulations, the complex nonlinear phenomena to be solved are modeled as nonlinear dynamical systems. The modeled mathematical problem, mainly described by a system of nonlinear differential equations, is then reduced to a model for handling by a computer. That is, linear problems such as linear systems of equations and eigenvalue problems are commonly handled numerically. However, as numerical simulations have become larger and more complex, the size of the linear problems to be solved has also increased dramatically. In recent year, there has been a growing need to efficiently solve numerical problems with dimensions in the tens of thousands or even billions.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 210–220, 2025.
https://doi.org/10.1007/978-3-031-72794-8_15

In this paper, we focus on the eigenvalue problem, which is closely related to nonlinearities and dynamics. In general, nonlinear systems exhibit more complex behavior than linear systems, and eigenvalues are essential to understanding such dynamics. For example, in nonlinear vibration and stability analysis, eigenvalues determine the stability and resonance properties of a target system. In addition, in chaos theory in nonlinear dynamics, eigenvalues are important for understanding the long-term behavior of the system. Numerical methods are necessary to compute these eigenvalues and are indispensable tools in nonlinear system analysis. We implement parallel computations of the contour integral based eigensolver [1], one of the solution methods for eigenvalue problems, using verified numerical computations. We aim to enhance computational speed through the parallel implementation of this verified numerical computation. Here, the verification of eigenvalues and eigenvectors is to produce rigorous error bounds of numerically computed eigenvalues and eigenvectors, taking into account all possible errors, in particular rounding errors [2,3].

Previous studies of verified eigenvalue and eigenvector computations are categorized into two types: one is for the verification of specific eigenpairs, and the other is for the verification of all the eigenpairs at once. This study focuses on the former one. There are several approaches to this type of verification methods, but in previous research, a verification method using complex moments was proposed [4]. This method derives suitable formulas for interval arithmetic using the Hankel matrix approach [1] to evaluate truncation errors. It also employs the Rayleigh-Ritz procedure [5] to improve truncation errors and reduces the number of integration points required by half to meet the specified integration error. This contour integral based approach can be applied to various areas, including finite element solutions [6,7], eigenvalues of elliptic operators [8], and electronic structure calculations [9].

Our parallel implementation is performed using the high-level Julia language. The Julia language is a scripting language capable of advanced computation and numerical analysis. It offers high computational performance and facilitates seamless integration with external libraries, making complex data analysis tasks feasible. The main aim of this paper is to show that a parallel implementation of the verified contour integral based method with hierarchical parallelism can reduce the computational time. It is also shown that such a parallel implementation is easily realized by using the Julia language.

The paper is organized as follows: In Sect. 2, we explain a generalized eigenvalue solution method based on contour integral based eigensolver. In Sect. 3, we discuss verified version of the contour integral based eigensolver. In Sect. 4, we address its parallel implementation, and in Sect. 5, we preform the implementation of parallel computing that yields rigorous inclusion of the partial eigenvalues. Finally, we present concluding remarks in Sect. 6.

2 Coutour Integral Based Eigen Solvers

In this paper, we focus on the Sakurai-Sugiura method [1], which is one of various methods for solving generalized eigenvalue problems. In this section, we will provide a detailed explanation of Sakurai-Sugiura method's specific techniques.

The Sakurai-Sugiura method is presented in [1] for solving generalized eigenvalue problems to find $(\lambda, x) \in \Omega \times \mathbb{C}^n \setminus \{0\}$ such that

$$Ax = \lambda Bx, \quad A, B \in \mathbb{C}^{n \times n}, \tag{1}$$

where Ω is a given prescribed domain. The Sakurai-Sugiura method is considered as a projection method that generates a subspace containing only eigenvectors corresponding to eigenvalues contained in the domain Ω by integrating around the domain, and extracts eigenpairs from the subspace. Let Γ be a closed curve along Ω unti-clockwisely. Let L be a block size and suppose that L is greater than or equal to the largest degree of multiplicity of eigenvalues contained in Ω. Let M be a moment order and LM greater than or equal to the number m of eigenvalues in Ω ($m \leq LM$). For a given matrix $V \in \mathbb{C}^{n \times L}$, we define a matrix S using integration around Γ as follows:

$$S := [S_0, S_1, ..., S_{M-1}] \in \mathbb{C}^{n \times LM}$$

$$S_k := \frac{1}{2\pi \mathrm{i}} \oint_{\Gamma} z^k (zB - A)^{-1} BV dz, \ \ k = 0, 1, ..., M-1, \tag{2}$$

where $\mathrm{i} = \sqrt{-1}$ is an imaginary unit. The column vector of S is represented by a linear combination of eigenvectors corresponding to the eigenvalues in Ω. That is,

$$\mathrm{span}\, S = \mathrm{span}\{x_1, ..., x_m\}$$

holds.

As a method for extracting eigenvalues and corresponding eigenvectors from a matrix S, we use the Rayleigh-Ritz procedure introduced in [5].

The singular value decomposition of S is given as

$$S = U \Sigma W^H.$$

Here, $\Sigma = \mathrm{diag}(\sigma_1, ..., \sigma_{LM}) \in \mathbb{R}^{LM \times LM}$, $U = [u_1, ..., u_{LM}] \in \mathbb{C}^{n \times LM}$, $W \in \mathbb{C}^{LM \times LM}$. Assume that $\mathrm{rank} S = m$, the singular values of S satisfy the following relationship:

$$\sigma_1 \geq ... \geq \sigma_m \geq \sigma_{m+1} = ... = \sigma_{LM} = 0. \tag{3}$$

We define a matrix composed of the left singular vectors corresponding to the non-zero singular values as

$$U_m = [u_1, ..., u_m] \in \mathbb{C}^{n \times m}.$$

Using this matrix, we can solve a reduced-size generalized eigenvalue problem with the dimension m

$$U_m^H A U_m r = \omega U_m^H B U_m r.$$

Solving this equation yields $\lambda = \omega$ and $x = U_m r$, allowing us to obtain the eigenvalues and corresponding eigenvectors of the original eigenvalue problem.

The Sakurai-Sugiura method is designed on the basis of contour integrals around the domain Ω, but since it is difficult to compute the contour integrals rigorously on a computer, numerical quadrature is used to approximate the contour integral. Let us assume that Γ is an ellipse centered at γ with radius along the real axis ρ and the ratio of the radii along the imaginary and real axes α. The numerical quadrature is performed by the N-point trapezoidal rule on Γ. Therefore, (2) is approximated by the formula

$$S_k \approx \hat{S}_k = \frac{\rho}{N} \sum_{j=1}^{N} \omega_j \rho^k \zeta_j^k (z_j B - A)^{-1} BV, \tag{4}$$

where $z_j = \rho(\cos(\theta_j) + \mathrm{i}\alpha \sin(\theta_j))$, $\omega_j = \alpha \cos(\theta_j) + \mathrm{i} \sin(\theta_j)$, $\zeta_j = \cos(\theta_j) + \mathrm{i}\alpha \sin(\theta_j)$, and $\theta_j = \frac{2\pi}{N}\left(j - \frac{1}{2}\right)$, $(j = 1, \ldots, N)$. In the computation of numerical quadrature, it is necessary to solve large scale linear equations, and a large part of the computation time of the Sakurai-Sugiura is spent in solving the linear equations at each grid point.

Performing the Sakurai-Sugiura method, the grid points on the unit circle are taken as $z_j = \cos\theta_j + \mathrm{i}\sin\theta_j = \cos\left(\frac{\pi}{N}j\right) + \mathrm{i}\sin\left(\frac{\pi}{N}j\right)$, where $\theta_j = \frac{\pi}{N}j (j = 0, 1, \ldots, 2N)$. This yields a symmetry for the conjugate of grid points, that is $\overline{z_j} = \cos\left(\frac{\pi}{N}j\right) - \mathrm{i}\sin\left(\frac{\pi}{N}j\right) = \cos\left(2\pi - \frac{\pi}{N}j\right) + \mathrm{i}\sin\left(2\pi - \frac{\pi}{N}j\right)$. We note that we take $z_j = \overline{z_j}$, when $j = 0$ or $j = N$. Therefore, the number of actual grid points to compute (4) becomes half because of this symmetry. If A, B, and V are real matrices, we have the formula taking advantage of the symmetry in the arrangement of integration points

$$\begin{aligned} S_k &\approx \frac{\rho}{N} \sum_{j=1}^{N/2} \omega_j \rho^k \zeta_j^k (z_j B - A)^{-1} BV + \frac{\rho}{N} \sum_{j=N/2+1}^{N} \omega_j \rho^k \zeta_j^k (z_j B - A)^{-1} BV \\ &= \frac{\rho}{N} \sum_{j=1}^{N/2} \omega_j \rho^k \zeta_j^k (z_j B - A)^{-1} BV + \frac{\rho}{N} \sum_{j=1}^{N/2} \overline{\omega_j} \rho^k \overline{\zeta_j}^k (\overline{z_j} B - A)^{-1} BV \\ &= 2\mathrm{Re}\left(\frac{\rho}{N} \sum_{j=1}^{N/2} \omega_j \rho^k \zeta_j^k (z_j B - A)^{-1} BV \right). \end{aligned} \tag{5}$$

The Sakurai-Sugiura method is a technique for determining eigenvalues within a given domain by contour integration. On a computer, contour integration can be calculated by placing grid points, and by adjusting the position of the grid points, the computational effort can be reduced by half. Next, we are considering the application of verified numerical calculations to this Sakurai-Sugiura method.

3 Verified Computations for Sakurai-Sugiura Method

Verified computations for the Sakurai-Sugiura method was proposed in [4]. The verification method evaluates all the errors in the computations of the contour integral defined in (2), which reduce the size of eigenvalue problem to LM and rigorously includes the eigenvalues in Ω with corresponding eigenvectors. The errors in computing the complex moment are split into truncation errors of numerical quadrature and rounding errors in numerical computations. An interval arithmetic based formula is derived to evaluate the truncation error. This idea enables the verification of eigenvectors and eigenvalues in Ω.

Let us define M_k as follows:

$$M_k := \frac{1}{2\pi \mathrm{i}} \oint_{\Gamma} z^k (zB - A)^{-1} dz, \quad k = 0, 1, ..., M-1. \tag{6}$$

Numerical approximation of the complex moment M_k is given by

$$M_k \approx M_k^{(N)} = \sum_{i=1}^{r} (\lambda_i - \gamma)^k d_i^{(N)} x_i x_i^H, \tag{7}$$

where

$$d_i^{(N)} = \begin{cases} \dfrac{1}{1 - \left(\frac{\lambda_i - \gamma}{\rho}\right)^N}, & i = 1, 2, \ldots, t, \\ \dfrac{-\left(\frac{\rho}{\lambda_i - \gamma}\right)^N}{1 - \left(\frac{\rho}{\lambda_i - \gamma}\right)^N}, & i = t+1, t+2, \ldots, r. \end{cases} \tag{8}$$

To see the effect of the eigenvalues inside and outside the interval Ω on the quadrature errors and for notational convenience, we split the complex moment into two parts

$$M_k^{(N)} = M_{k,in}^{(N)} + M_{k,out}^{(N)}, \tag{9}$$

where we denote

$$\begin{aligned} M_{k,in}^{(N)} &= X_{\Omega} (\Lambda_{\Omega} - \gamma I_t)^k D_{\Omega}^{(N)} X_{\Omega}^H, \\ M_{k,out}^{(N)} &= X_{\Omega^C} (\Lambda_{\Omega^C} - \gamma I_t)^k D_{\Omega^C}^{(N)} X_{\Omega^C}^H, \\ D_{\Omega}^{(N)} &= diag(d_1^{(N)}, d_2^{(N)}, \ldots, d_t^{(N)}), \\ D_{\Omega^C}^{(N)} &= diag(d_{t+1}^{(N)}, d_{t+2}^{(N)}, \ldots, d_r^{(N)}), \\ X_{\Omega^C} &= [x_{t+1}, x_{t+2}, \ldots, x_r], \\ \Lambda_{\Omega^C} &= diag(\lambda_{t+1}, \lambda_{t+2}, \ldots, \lambda_r). \end{aligned}$$

Similarly, S_k in (2) can also be expressed

$$\begin{aligned} S_k^{(N)} &= S_{k,in}^{(N)} + S_{k,out}^{(N)} \\ S_{k,in}^{(N)} &= M_{k,in}^{(N)} BV, \\ S_{k,out}^{(N)} &= M_{k,out}^{(N)} BV \end{aligned} \tag{10}$$

Here, we denote

$$S_{in}^{(N)} = [S_{0,in}^{(N)}, S_{1,in}^{(N)}, ..., S_{M-1,in}^{(N)}] \tag{11}$$

and assume that $rankS = t$, then $S^H(A-zB)S$ and $(S_{in}^{(N)})^H(A-zB)S_{in}^{(N)}$ have the same eigenvalues. Denote the quantity obtained by numerically computing $M_k^{(N)}$ by $\bar{M}_k^{(N)}$. Defining the interval matrix with radius $R \in \mathbb{R}^{n\times n}$ and center $C \in \mathbb{R}^{n\times n}$ as $\langle C, R\rangle$, $M_{k,in}^{(N)}$ is given by:

$$\begin{aligned} M_{k,in}^{(N)} &\in \langle M_k^{(N)}, |M_{k,out}^{(N)}|\rangle \\ &\subset \langle \bar{M}_k^{(N)}, |M_{k,out}^{(N)}| + |\bar{M}_k^{(N)} - M_k^{(N)}|\rangle. \end{aligned} \tag{12}$$

An enclosure of $|M_{k,out}^{(N)}|$ is obtained as follows. Let B be a Hermite positive semidefinite matrix. Assume $2M-1 < N$ and that $\bar{\lambda} \in \mathbb{R}$ satisfies $|\bar{\lambda}-\gamma| = \min_{i=t+1,t+2,...,r} |\lambda_i - \gamma|$. Then, it follows from [4] that

$$|M_{k,out}^{(N)}| \leq (r-t)|\bar{\lambda}-\gamma| \left(\frac{\left(\frac{\rho}{|\bar{\lambda}-\gamma|}\right)^{2N}}{1-\left(\frac{\rho}{|\bar{\lambda}-\gamma|}\right)^{2N}} \right) \parallel V^H B V \parallel_F \tag{13}$$

for $k = 0, 1, \ldots, 2m-1$.

Since the eigenvalues of S and $S_{in}^{(N)}$ given in (2) and (11) are the same, it follows that

$$S_{out}^{(N)} = \left[S_{0,out}^{(N)}, S_{1,out}^{(N)}, ..., S_{M-1,out}^{(N)} \right]. \tag{14}$$

Then, rigorously enclose the approximate transformation matrix given by

$$\begin{aligned} S_{in}^{(N)} &\in \langle S^{(N)}, |S_{out}^{(N)}|\rangle \\ &\subset \langle \bar{S}^{(N)}, |S_{out}^{(N)}| + |\bar{S}^{(N)} - S^{(N)}|\rangle. \end{aligned} \tag{15}$$

Assume that B is a Hermitian and positive definite matrix. Assume also $2M-1 < N$ and that $\bar{\lambda} \in \mathbb{R}$ satisfies $|\bar{\lambda}-\gamma| = \min_{i=t+1,t+2,...,r} |\lambda_i - \gamma|$. Then, $S_{k,out}^{(N)}$ is bounded [4] by the form

$$|S_{k,out}^{(N)}| \leq (n-t)|\bar{\lambda}-\gamma|^k \left(\frac{\left(\frac{\rho}{|\bar{\lambda}-\gamma|}\right)^{N}}{1-\left(\frac{\rho}{|\bar{\lambda}-\gamma|}\right)^{N}} \right) (\parallel V^H B V \parallel_F)^{1/2}. \tag{16}$$

4 Parallel Implementation

It is known [10] that the Sakurai-Sugiura method has hierarchical parallelism, as shown in Fig. 1, which effectively uses parallel implementation to improve performance. This parallel implementation consists of three different layers.

At the top layer, the Sakurai-Sugiura method demonstrates top-level parallelism by allowing the placement of multiple domains. Within each domain, all computations can be executed independently, optimising parallel processing. This layer is referred to as top-level parallelism.

Moving to the middle layer, grid points are used to approximate the integrals around the contour within each domain. The calculations at these grid points can be performed simultaneously, exploiting the independence of the coefficient matrices for each linear system. This layer is referred to as the middle layer parallelism.

Finally, at the bottom layer, the Sakurai-Sugiura method uses parallel linear equation solvers such as BLAS/LAPACK to compute the solution for each linear system. This layer is called bottom layer parallelism.

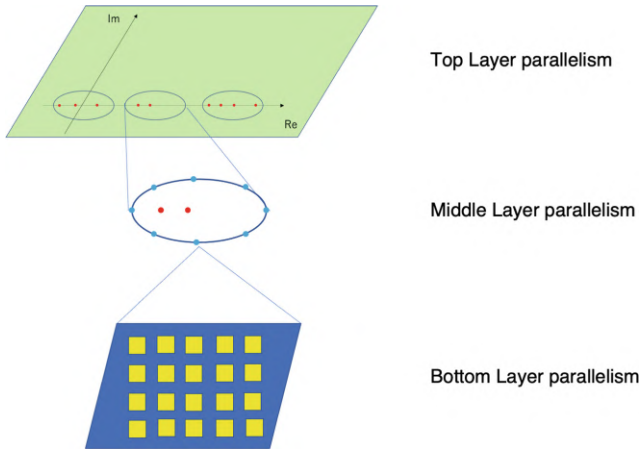


Fig. 1. Outline of hierarchical parallelism diagram.

The computation of each quadrature point and the computation of solving the linear equations are independent. Therefore, by concurrently processing multiple processes or threads, efficient parallel processing can be achieved. This parallel implementation allows maximizing the concurrency of computations, and it is anticipated to reduce the overall execution time.

A parallel implementation of the Sakurai-Sugiura method was proposed in [10] for efficiently exploiting the performance of massively parallel computational environments. It is also considered to apply a contour-integral based method to a large dense problem in conjunction with a block Krylov subspace method as the linear solver. Consequently, the performance of the Sakurai-Sugiura method on a massively parallel computational environments is shown.

We used process parallelism and thread parallelism in the Julia language for the parallel implementation in the computation of each quadrature point and in the computation of solving the linear equations, respectively. Prior to implementation of the verified version of the Sakurai-Sugiura method, we implement

a non-verified version as a test case. Table 1 and Table 2 show the computation time of the contour integral when the number of processes is varied from 1 to 6. The implementation environment is shown in Table 3.

Table 1. Computational time ($n = 400,\ N = 2000$).

process	1	2	3	4	5	6
time (sec)	63.493248	34.683393	28.590250	22.583462	20.102951	18.277447

Table 2. Computational time ($n = 400,\ N = 5000$).

process	1	2	3	4	5	6
time (sec)	142.651353	72.369486	54.419968	46.156045	33.552329	27.600419

Table 3. Implementation environment.

CPU	Intel Xeon (R) Gold 6128 3.40 GHz
Memory	376 GB
OS	Ubuntu 16.04.7 LTS
# of cores	6

As shown in Table 1 and Table 2, an appropriate control of process parallelism improves the computation speed. This trend is also valid for the verified version, and it is expected to improve the computation speed in parallel implementation. It is expected that saturation will occur, but thanks to the high parallel performance, a higher number of processes would be required to reach the saturation point. The bottleneck of this implementation is when the communication time to collect results etc. is more than the computation time for each linear solver, which depends on the size of the computing environment. Our implementation does not have enough computing power to cause such a bottleneck.

5 Parallel Implementation of a Verified Contour Integral Based Eigensolver

As a next step, we compute rigorous error bounds for eigenpairs using this parallel implementation of the Sakurai-Sugiura methods. We implement a verified version of the Sakurai-Sugiura method using the Julia language and evaluate its performance.

The performance of the computer used for the experiment is the same as the previous one shown in Table 3. Additionally, Julia version 1.7.3 was used. The calculation times for $n = 256$ and $N = 82$ will be shown in Table 4 and Fig. 2. In Fig. 2, the vertical axis represents the computation time, and the horizontal axis is defined as 'procs', indicating the number of threads used in parallel computing.

Table 4. Computational time of the contour integral.

process	1	2	3	4	5	6
time (sec)	85.010291	43.796222	31.584319	23.323432	19.532687	15.298769

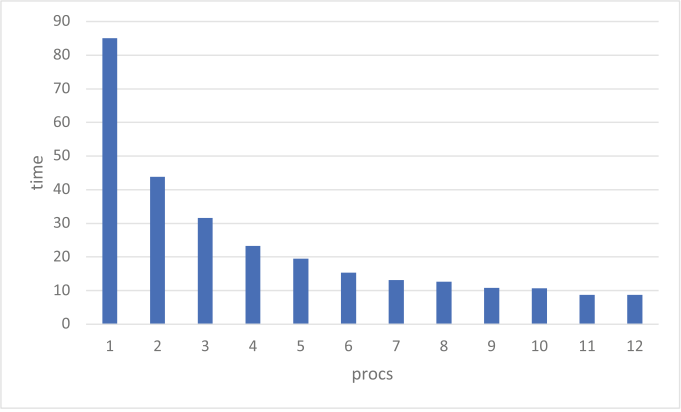


Fig. 2. Computation time of the contour integral.

From these results, it is shown that increasing the number of processes reduces the computation time. In other words, parallelization, achieved by increasing the number of processes, leads to efficient reduction in computation time. It is worth noting that the parallel implementation of the verified version of the Sakurai-Sugiura method is firstly achieved in this work. This result is considered as an pioneering result for the parallel implementation for the verified version of the Sakurai-Sugiura method.

Figure 3 illustrates the proportion of computation time occupied by each computational step in the Sakurai-Sugiura method. In Fig. 3, similarly to Fig. 2, with $n = 256$ and $N = 82$, the vertical axis represents the ratio of each process to the total computation time, and the horizontal axis indicates the number of threads used in parallel computation. The blue, orange, and grey bars represent the proportion of execution time spent on adding processes, performing contour integrations, and applying the Rayleigh-Ritz procedure, respectively. The orange bar in Fig. 3 shows the time taken for the counter integration, and with a procs number of 6, almost all of this computation time, or 15.13 s, is for the linear

equation solver. The remaining time includes the creation of the empty vector and other calculations. From this figure, it can be seen that the proportion of computation time devoted to numerical quadrature of the contour integral has decreased in the overall computation time.

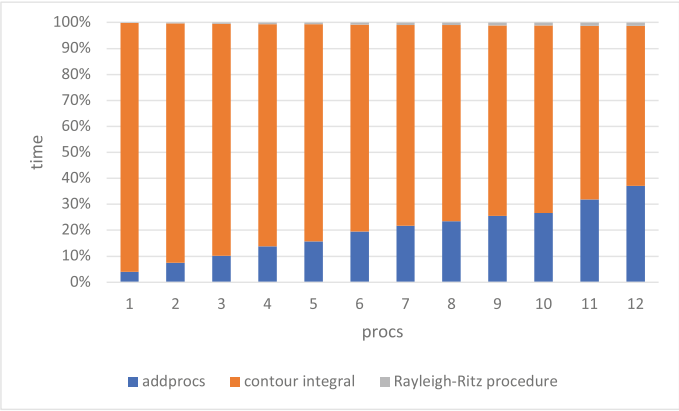


Fig. 3. The proportion of computational time of the Sakurai-Sugiura method.

6 Concluding Remarks

In this paper, we performed a parallel implementation of the Sakurai-Sugiura method, a representative contour integral based eigensolver, applied to the generalized eigenvalue problem using the Julia language's process controls for parallel computing. We conducted verified numerical computations in parallel computing, resulting in a reduction in computational time. Namely, we implemented a verified version of the Sakurai-Sugiura method introduced in [4] using the Julia language. The verified numerical computation is conducted with Julia's interval arithmetic package IntervalArithmetic.jl. Combining the implementation of verified version of the contour integral based eigensolver with parallel computing, we demonstrated the parallel performance of the implementation, which is the greatest feature of the SS method. The parallel implementation actually reduced the computation time, confirming the increase in computational efficiency thanks to parallelization.

This parallel implementation is intended for distributed parallel computing, and its usefulness will be demonstrated in the future in the rigorous computation of some eigenvalues of more larger matrices. Future challenges of this work include increasing the scale of computations, exploring methods to reduce computation time other than parallel computing. In particular, we explore a direction of accelerating parallel computing using general-purpose GPUs.

Acknowledgement. AT is partially supported by JSPS KAKENHI Grant Numbers JP22K03411 and JP21H03451.

References

1. Sakurai, T., Sugiura, H.: A projection method for generalized eigenvalue problems using numerical integration. J. Comput. Appl. Math. **159**(1), 119–128 (2003)
2. Nagatou, K.: Eigenvalue problem: numerical approximation and verified numerical computation. Bull. Jpn. Soc. Ind. Appl. Math. **13**(3), 244–257 (2003)
3. Rump, S.M.: Verification methods: rigorous results using floating-point arithmetic. Acta Numer. **19**, 287–449 (2010)
4. Imakura, A., Morikuni, K., Takayasu, A.: Verified eigenvalue and eigenvector computations using complex moments and the Rayleigh-Ritz procedure for generalized hermitian eigenvalue problems. J. Comput. Appl. Math. **424**, 114994 (2023)
5. Sakurai, T., Tadano, H.: CIRR: a Rayleigh-Ritz type method with contour integral for generalized eigenvalue problems. Hokkaido Math. J. **36**(4), 745–757 (2007)
6. Yamamoto, N., Nakao, M.T.: Numerical verifications of solutions for elliptic equations in nonconvex polygonal domains. Numer. Math. **65**(1), 503–521 (1993)
7. Watanabe, Y., Yamamoto, N., Nakao, M.T.: Verification methods of generalized eigenvalue problems and its applications. Trans. Jpn. Soc. Ind. Appl. Math. **9**(3), 137–150 (1999)
8. Toyonaga, K., Nakao, M., Watanabe, Y.: Verified numerical computations for multiple and nearly multiple eigenvalues of elliptic operators. J. Comput. Appl. Math. **147**(1), 175–190 (2002)
9. Hoshi, T., Ogita, T., Ozaki, K., Terao, T.: An a posteriori verification method for generalized real-symmetric eigenvalue problems in large-scale electronic state calculations. J. Comput. Appl. Math. **376**, 112830 (2020)
10. Yano, T., Futamura, Y., Imakura, A., Sakurai, T.: Performance evaluation of the Sakurai-Sugiura method with a block Krylov subspace linear solver for large dense Hermitian-definite generalized eigenvalue problems. JSIAM Lett. **10**, 77–80 (2018)
11. Ikegami, T., Sakurai, T.: Contour integral eigensolver for non-Hermitian systems: a Rayleigh-Ritz-type approach. Taiwanese J. Math. **14**, 825–837 (2010)

Analysis of Nonlinear Acoustic Waves Using Physics-Informed Neural Networks

Kazuya Yokota(✉), Takahiko Kurahashi, and Masajiro Abe

Nagaoka University of Technology, Kamitomioka-machi 1603-1, Nagaoka, Niigata, Japan
yokokazu@vos.nagaokaut.ac.jp

Abstract. In the acoustic analysis of mechanical systems that produce high sound pressures, the nonlinearity of sound waves must be considered. Under the conditions of nonlinearity, a change in the speed of sound distorts the waveform. This waveform distortion is important in the design of acoustic equipment. Numerical analysis is commonly used to analyze nonlinear sound waves. Recently, there have been many reports on the use of machine learning for numerical analysis. However, in data-driven approaches, the governing equations may not be satisfied by simulation results. To address this problem, we present a new approach for the analysis of nonlinear sound waves using Physics-Informed Neural Networks (PINNs). In this study, a PINN algorithm for nonlinear acoustic equations was presented and a one-dimensional traveling wave was analyzed. The results obtained were consistent with the nonlinearity of sound waves. Considering that PINN is a mesh-free numerical analysis method, and that trained PINN performed acoustic analysis faster than FDM, PINN is expected to be a new option for nonlinear acoustic wave analysis.

Keywords: nonlinear acoustics · physics-informed neural networks · numerical analysis

1 Introduction

In the design of loud mechanical systems, such as musical instruments, the nonlinearity of sound waves must be considered. Under the conditions of nonlinearity, the speed of sound changes depending on the particle velocity, leading to the distortion of the sound waveform [1]. For example, the brassy timbre of brass instruments is caused by the nonlinearity of sound waves [2]. Therefore, the nonlinearity of sound waves cannot be ignored in the design of loud mechanical systems.

Because most nonlinear partial differential equations cannot be solved analytically, numerical analysis has been used for simulation. In recent years, many machine learning methods have been proposed to accelerate numerical analysis and advance inverse analysis [3]. However, in data-driven approaches, simulation results may not satisfy the governing equations.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 221–228, 2025.
https://doi.org/10.1007/978-3-031-72794-8_16

To address this problem, Raissi et al. proposed Physics-Informed Neural Networks (PINNs) [4]. PINNs are methods for numerical analysis that introduce constraints on the governing partial differential equations (PDEs) in the loss functions of a neural network. The basic idea of PINNs is to define a loss function based on the PDE residuals evaluated by automatic differentiation (AD) [5] in addition to the loss functions of the initial and boundary conditions and then minimize these loss functions to obtain the simulation results. PINNs can be used for inverse problems to identify the uncertain parameters of the governing PDEs, and, in recent years, they have been successfully employed in thermodynamics and fluid mechanics, including parameter identification for the equation of state [6], application to inverse problems in supersonic flows [7], and identification of the thermal diffusivity of materials [8].

Despite the aforementioned advances in PINNs, their applications in the acoustic field are limited. The governing equations for acoustic phenomena, influenced by reflection, interference, and diffraction, yield solutions that span a wide range of amplitudes and frequencies. This diversity makes setting weight parameters for loss functions a challenging task. Consequently, it is difficult to analyze sound waves using PINNs [9], and there are few reports on the analysis of nonlinear sound waves using PINNs.

In this study, we present a method for the analysis of nonlinear sound waves using PINNs for application in the design of loud mechanical systems, such as musical instruments. In a fundamental study, a one-dimensional nonlinear traveling wave was analyzed to confirm the validity of the analytical model.

2 Methods

This section describes the wave equation analyzed in this study and the acoustic analysis method using PINNs.

2.1 Equation for the One-Dimensional Nonlinear Traveling Wave

In this section, the governing equation for the one-dimensional traveling wave analyzed herein is explained. From the equations of continuity and motion, the following equations are obtained for the particle velocity u and density ρ [1]:

$$\frac{\partial \rho}{\partial t} + u\frac{\partial \rho}{\partial x} = -\rho\frac{\partial u}{\partial x}, \tag{1}$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = -\frac{1}{\rho}\frac{\partial P}{\partial x}, \tag{2}$$

where

$$P = P_0 + p, \tag{3}$$

x is the spatial coordinate, t is time, ρ is the air density, P_0 is the equilibrium pressure, and p is the sound pressure. We assume that u is a function of ρ only; that is, $u = u(\rho)$ and $\rho = \rho(u)$. This assumption holds until a shock wave

is formed in the traveling wave [10]. Under this assumption, Eqs. (1) and (2) become:

$$\frac{d\rho}{du}\left(\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}\right)=-\rho\frac{\partial u}{\partial x}, \tag{4}$$

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}=-\frac{c^2}{\rho}\frac{d\rho}{du}\frac{\partial u}{\partial x}. \tag{5}$$

Here, the equation for the speed of sound $c=\sqrt{dP/d\rho}$ is used. Substituting the right-hand side of Eq. (5) into Eq. (4), we obtain:

$$\frac{d\rho}{du}=\pm\frac{\rho}{c}. \tag{6}$$

Substituting Eq. (6) into Eq. (5), we obtain the equation for the one-dimensional traveling wave analyzed in this study.

$$\frac{\partial u}{\partial t}+(u\pm c)\frac{\partial u}{\partial x}=0. \tag{7}$$

The speed of sound c is expressed as:

$$c=c_0\pm\frac{\gamma-1}{2}u, \tag{8}$$

where γ is the specific heat ratio, and c_0 is the speed of sound when $u=0$. Detailed derivations of Eqs. (8) is explained in Ref. [1].

As shown in Eqs. (7) and (8), when the particle velocity u is sufficiently low ($u \ll c_0$), Eq. (7) becomes linear. However, if the particle velocity u is not negligible with respect to c_0, the speed of sound c varies with the particle velocity, resulting in waveform distortion. The case where u is non-negligible with respect to c_0 corresponds to the case in which a large sound pressure is generated, and nonlinearity cannot be ignored during the acoustic analysis of equipment that generates large sound volumes, such as brass instruments [2].

2.2 Analytical Model

The structure of the analytical model used in this study is shown in Fig. 1. The inputs of the neural network are x_i and t_i (where i is the sample number), and the output is the predicted value of the particle velocity $\hat{u}_i$. This input-output relationship is denoted by the following equation:

$$\hat{u}_i=F\left(x_i,t_i;\Theta\right), \tag{9}$$

where F is the neural network operator, and Θ is a set of trainable parameters of the neural network. In this study, we used a fully connected neural network with six hidden layers and 1200 nodes; sine was used as the activation function.

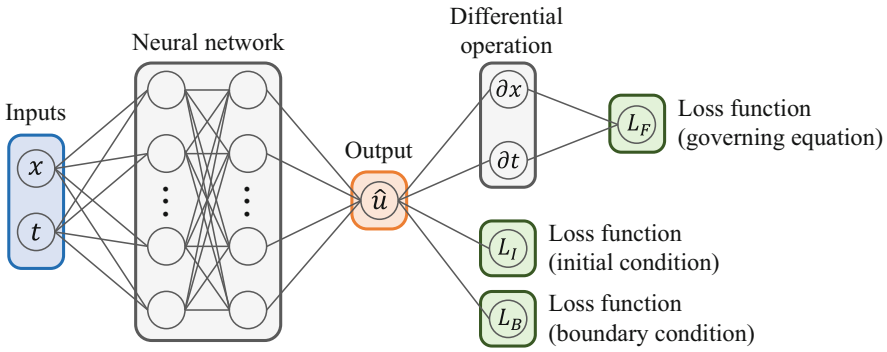


Fig. 1. Structure of the analytical model.

To force $\hat{u}$ to follow Eq. (7), the loss function L_F with respect to the governing equation is defined as follows:

$$L_F = \frac{1}{N_F} \sum_{i=1}^{N_F} \left\{ \frac{\partial \hat{u}_i}{\partial t_i} + (\hat{u}_i + c_i) \frac{\partial \hat{u}_i}{\partial x_i} \right\}^2, \tag{10}$$

where N_F is the number of collocation points for the governing equation and c_i is expressed as

$$c_i = c_0 + \frac{\gamma - 1}{2} \hat{u}_i. \tag{11}$$

The derivative of $\hat{u}$ that is required to calculate Eq. (10) is obtained by the automatic differentiation [5] of the neural network.

The loss function L_I for the initial condition is denoted by

$$L_I = \frac{1}{N_I} \sum_{i=1}^{N_I} \left(\hat{u}_i - \bar{u}_{i,I} \right)^2, \tag{12}$$

where N_I is the number of collocation points for the initial condition, and $\bar{u}_{i,I}$ is the particle velocity data given as the initial and boundary conditions. Similarly, the loss function L_B for the boundary condition is written as

$$L_B = \frac{1}{N_B} \sum_{i=1}^{N_B} \left(\hat{u}_i - \bar{u}_{i,B} \right)^2, \tag{13}$$

where N_B is the number of collocation points for the boundary condition, and $\bar{u}_{i,B}$ is the particle velocity data given as the boundary condition.

The loss function L for the entire network is calculated as follows:

$$L = \lambda_F L_F + \lambda_I L_I + \lambda_B L_B, \tag{14}$$

where λ_F, λ_I, and λ_B are the weight parameters of their respective loss functions. The optimization problem of the analytical model is formulated as

$$\min_{\Theta} L(\Theta). \tag{15}$$

By minimizing the loss function L, the trainable parameter Θ of the neural network was optimized to follow the governing equation and the initial and boundary conditions. Therefore, we used the Adam optimizer [11] to determine the optimal value of Θ through iterative calculations.

2.3 Implementation

We implemented the analytical model using the Deep Learning Toolbox in MATLAB (MathWorks, USA) and used the "dlfeval" function to code a custom training loop. We used the "sobolset" function from the Statistics and Machine Learning Toolbox to create datasets for x_i and t_i. The neural network was trained via GPU-assisted computation using a Parallel Computing Toolbox.

The neural network training and prediction were performed on a computer equipped with a Core i9-13900KS CPU (Intel, USA) and GeForce RTX 4090 GPU (NVIDIA, USA) with 128 GB of main memory and 24 GB of video memory.

3 Analysis of One-Dimensional Traveling Wave

In this section, we investigated the performance of the proposed analytical model by performing forward analysis.

3.1 Analysis Conditions

1D traveling wave analysis shown in Fig. 2 was performed to validate the analytical model. If nonlinearity is not considered, the waveform of the traveling wave does not change. However, when the nonlinearity of the sound waves is considered, the waveform changes as the sound wave propagates. In this section, we describe the waveform changes.

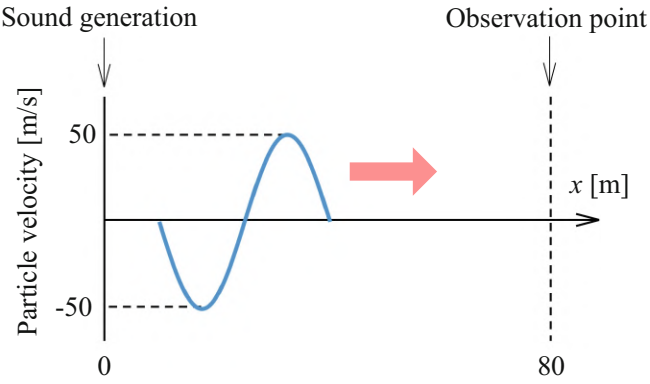


Fig. 2. Acoustic field to be analyzed.

We analyzed a sine wave generated at $x = 0$ m in the 1D space shown in Fig. 2. The particle velocity at $x = 0$ is given by

$$u_0 = A \sin\left(\frac{2\pi}{T} t_0\right), \tag{16}$$

where the amplitude is $A = 50$ m/s , the period is $T = 0.5$ s, and t_0 is the time when the sound wave is generated. The sine wave propagates in the $+x$ direction. We set c_0 to 340 m/s and γ to 1.4. For training, datasets of (x_i, t_i) were created using the MATLAB function "sobolset," which generates quasi-random numbers. Five thousand pairs for L_F, 1000 pairs for L_I, and 1000 pairs for L_B were created for the (x_i, t_i). That is, $N_F = 5000$, $N_I = 1000$, and $N_B = 1000$.

3.2 Weight Parameters of Loss Functions

As explained in Sect. 1, sound waves exhibit a broad range of amplitudes and frequencies, making the task of setting weight parameters in Eq. (14) challenging. In this study, the weight parameters were determined based on the amplitudes of u_0 in Eq. (16) and $\partial u_0/\partial t_0$, aiming to evaluate the three terms in Eq. (14) equitably. The derivative $\partial u_0/\partial t_0$ is given by

$$\frac{\partial u_0}{\partial t_0} = \frac{2\pi A}{T} \cos\left(\frac{2\pi}{T} t_0\right). \tag{17}$$

Based on the amplitudes of Eqs. (16) and (17), the weight parameters were calculated as follows:

$$\lambda_F = \left(\frac{T}{2\pi A}\right)^2, \tag{18}$$

$$\lambda_I, \lambda_B = \frac{1}{A^2}. \tag{19}$$

Without appropriately set weight parameters, accurate results cannot be achieved with PINNs. The effectiveness of Eqs. (18) and (19) in achieving balanced evaluation of the terms in Eq. (14) will be demonstrated in Sect. 3.4 through simulations.

3.3 Exact Solution

In this section, the exact solution is presented. Based on Eqs. (7) and (8), the travel distance of the sound wave x_l is expressed as

$$x_l = \left(u + c_0 + \frac{\gamma - 1}{2} u\right) t_l, \tag{20}$$

where t_l is the time taken for the sound wave generated at $x = 0$ to reach $x = x_l$. At this time, the observation point is at $x = 80$ m, as shown in Fig. 2.

By substituting $x_l = 80$ m and $u = u_0$, we obtain the exact solution for the arrival time of the sound wave at the observation point.

$$t_l = \frac{160}{2c_0 + A(\gamma+1)\sin\left(\frac{2\pi}{T}t_0\right)}. \tag{21}$$

Equation (21) is used to plot the exact solution in the next section.

3.4 Result

Figure 3 shows particle velocity waveforms obtained from the analytical model. The time 0 s on the horizontal axis is the time at which the sine wave arrives. Figure 3 shows that the analytical results from PINN are consistent with the waveform distortion observed in the exact solution. This result confirms that by performing the analysis using PINN, the exact solution of nonlinear acoustic waves can be analyzed with sufficient accuracy.

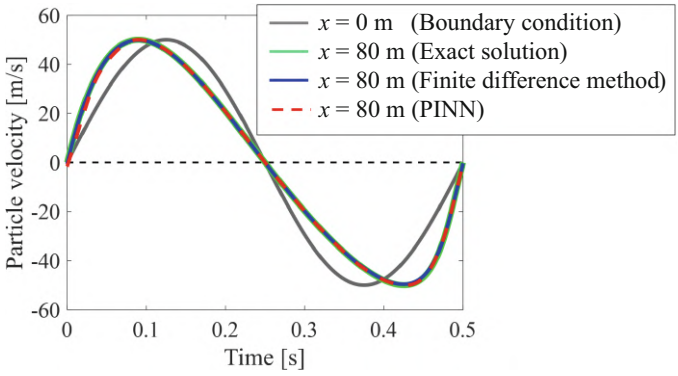


Fig. 3. Analyzed particle velocity waveform.

The PINN training time was 18 s, which was approximately 13 times longer than 1.4 s of the finite difference method (Lax-Wendroff method, time step = 5×10^{-5} s). However, after training, PINN could run the simulation in 0.044 s, which was approximately four times faster than the finite-difference method. Therefore, it is expected that PINNs can be used in applications where real-time performance is required for nonlinear acoustic wave analysis because PINNs can learn various boundary conditions and physical property parameters in advance. In addition, PINN is a mesh-free simulation method with no discretization errors. Because of these advantages, PINNs are expected to be a new option for nonlinear acoustic wave analysis.

Future research will focus on whether PINN can be used for analysis at higher waveform frequencies and whether the computational cost of training can be reduced.

4 Conclusion

In this study, we presented an algorithm for analyzing nonlinear sound waves using PINNs, aiming for its application in the design of mechanical systems that produce high sound pressure. The results obtained were consistent with the exact solution, and the validity of the analytical model was confirmed.

In recent years, PINNs have gained considerable attention for inverse analysis, such as physical property identification in thermodynamics and fluid mechanics. Further research into PINNs for nonlinear sound waves will significantly help with inverse problems in the design of loud mechanical systems such as musical instruments. In the future, we aim to perform an inverse analysis using the proposed PINN, including physical property identification and design optimization, as well as the analysis of more complex acoustic fields.

References

1. Hamilton, M.F., Blackstock, D.T.: Nonlinear Acoustics. Academic Press, Cambridge (1998)
2. Myers, A., Pyle, RW Jr, Gilbert, J., Campbell, DM., Chick, JP., Logie, S.: Effects of nonlinear sound propagation on the characteristic timbres of brass instruments. J. Acoust. Soc. Am. **131**(1) (2012)
3. Vinuesa, R., Brunton, S.L.: The potential of machine learning to enhance computational fluid dynamics, arXiv:2110.02085v1 (2021)
4. Raissi, M., Perdikaris, P., Karniadakis, G.E.: Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. **378** (2019)
5. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Netw. **61** (2015)
6. Mao, Z., Jagtap, A.D., Karniadakis, G.E.: Physics-informed neural networks for high-speed flows. Comput. Methods Appl. Mech. Eng. **360**, 112789 (2020)
7. Jagtap, A.D., Mao, Z., Adams, N., Karniadakis, G.E.: Physics-informed neural networks for inverse problems in supersonic flows. J. Comput. Phys. **466**, 111402 (2022)
8. Cai, S., Wang, Z., Wang, S., Perdikaris, P., Karniadakis, G.E.: Physics-informed neural networks for heat transfer problems. J. Heat Transf. **143**(6), 060801 (2021)
9. Moseley, B., Markham, A., Nissen-Meyer, T.: Solving the wave equation with physics-informed deep learning, arXiv:2006.11894 (2020)
10. Landau, L.D., Lifschitz, E.M.: Fluid Mechanics, vol. 6, 1st edn. Pergamon Press, Oxford (1959)
11. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization, arXiv:1412.6980 (2014)

Synchronization

Exciting of Slowly Modulated Vibrations with Unbalanced Rotor Exciters Using Nonlinear Resonance

Alexander Fidlin(✉) and Tunc Yüzbasioglu

Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
alexander.fidlin@kit.edu
http://www.itm.kit.edu/dynamik.php

Abstract. Vibratory exciters are essential in technical systems for their vast possibilities of applications. A new exciter with a pair of inertially coupled unbalanced rotors with an additional viscous coupling between them is proposed in the present work. It is capable of producing slowly modulated vibrations with large amplitudes in resonant regime. An asymptotic analysis is conducted to investigate the slow dynamics of the system using an averaging method for partially strongly damped systems. Different types of solutions are shown. The dependency of modulated vibrations on parameters is studied.

Keywords: modulated vibrations · self-synchronization · nonlinear resonance · averaging · unbalanced exciter

1 Introduction

Vibratory systems find many applications ranging from automotive or aerospace to civil engineering and manufacturing. The study of vibrational phenomena is crucial for a better understanding of natural occurrences and for developing and enhancing the performance of machines.

One fundamental aspect of vibration control and generation is the use of exciters. Unbalanced exciters have emerged as a powerful tool for generating controlled vibrations in various applications. They rely on the principle of mass imbalance to generate oscillatory motion. These devices consist of an eccentric mass, typically mounted on a rotating shaft, which creates a centrifugal force when the shaft is set in motion. The resulting unbalanced force induces vibrations in the attached structure. They are used widely in applications due to their simplicity, robustness, and versatility in generating controlled vibrations.

They can be used in conveyors, screens, feeders, and vibrating tables to transport and separate materials efficiently. In civil engineering and construction, they can be applied to compact concrete and reduce air voids to improve the stability and durability of structures. They can be used in test rigs to simulate real-world vibrations to assess the performance of products and structures. Depending on

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 231–244, 2025.
https://doi.org/10.1007/978-3-031-72794-8_17

the field of application, different kinds of exciters are used, taking advantage of different dynamical features. One significant use case is creating slowly modulated vibrations with varying amplitudes over a long time period. This kind of excitation is particularly relevant for test rigs to simulate different load levels over time, which might significantly affect the fatigue life.

One of the most important phenomena for exciters is capture into resonance (Sommerfeld effect), which was first observed in [21] and later explained by [3]. It describes for a single rotor exciter that the amplitudes of the carrying structure will increase instead of the rotation speed of the motor while accelerating if the right conditions are met. Depending on the application, this should be avoided to obviate damage to the machine because of the high vibration amplitudes, but it can also be used intentionally. The system can be operated at resonance to achieve high excitation amplitudes. Since its discovery, the Sommerfeld effect has been researched extensively; see [16–18]. The transient behavior of a strongly damped Sommerfeld system was investigated in [8].

The research was extended to systems with two coaxial unbalances to describe better the dynamics of many mechanical systems like automotive drive train elements with coaxial shafts [15] or some types of vibration exciters [4]. It was shown in [6] that a centrifugal exciter with two coaxial unbalances can create slowly modulated strong vibrations and slowly oscillating rotation speed if a non-small relative damping between the unbalanced rotors is taken into account. A further generalization concerns the self-balancing devices, which also have been known for a long time [22]. They are also extensively investigated in the literature; see [10,11,19,20].

Another possibility is to use multiple rotors with different energy sources. This can be helpful to distribute the load more evenly along the machine if several low-power exciters are used instead of one powerful exciter. More importantly, it allows coordinating rotor dynamics without kinematic connections between them. This phenomenon is called self-synchronization and was first observed by Huygens and later explained by Blekhman [3,4]. Since then, many different investigations have been conducted [1]. Different number of rotors [2], different carriers and configurations [25,26], self- and controlled synchronization [5,14] are considered, numerically [12] and experimentally [13,23] studied. Even though such systems offer broad possibilities as exciters, to the best knowledge of authors, they can not produce slowly modulated vibrations but only beating-like amplitudes.

This work aims to propose a new exciter that can generate slowly modulated vibrations, analyze it analytically, and show its capabilities. The analysis is done with the help of averaging, and it is shown that the system in consideration by proper choice of parameters can have the following features: Sommerfeld effect, synchronization, and slowly modulated vibrations both in overcritical and resonant regimes.

The paper is structured as follows. Section 2 presents the investigated model of an exciter with a pair of unbalanced rotors and relative viscous damping. In Sect. 3, the asymptotic analysis of the system is conducted using the averaging

method for partially strongly damped systems. The averaged system is investigated in Sect. 4. Different types of solutions and their existence are presented in phase space and time domain for two different sets of parameters. Section 5 conducts a small parameter study on slowly modulated vibrations. A summary of the most important results is given in Sect. 6.

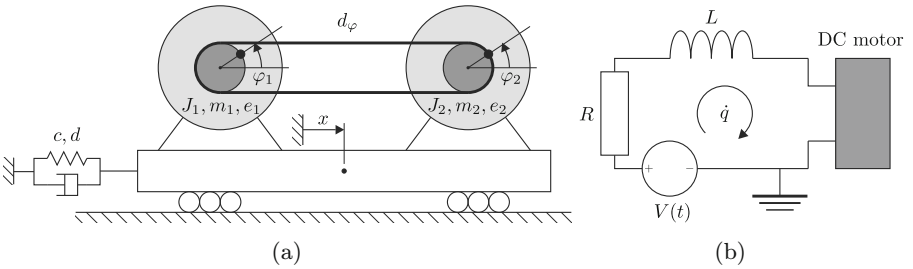


Fig. 1. (a) Investigated model and (b) electrical circuit of the DC motor

2 Pair of Unbalanced Rotors with Viscous Coupling

First, the investigated system is described, shown in Fig. 1a. Two unbalanced rotors of mass m_i, mass moment of inertia J_i, and eccentricity e_i, where $i = 1, 2$ corresponds to the index of rotors, are mounted on a carrier of mass M, which is elastically suspended with a linear spring-damper element of stiffness c and damping d in the horizontal direction. The rotors are driven in the same direction by DC motors. Furthermore, the two rotors have a viscous coupling, which acts on their relative angular velocity with damping coefficient d_φ.

The equations of motion of this system read

$$\begin{aligned}
&(M + m_1 + m_2)\ddot{x} + d\dot{x} + cx = \sum_{i=1}^{2} m_i e_i(\ddot{\varphi}_i \sin\varphi_i + \dot{\varphi}_i^2 \cos\varphi_i),\\
&(J_1 + m_1 e_1^2)\ddot{\varphi}_1 + d_\varphi(\dot{\varphi}_1 - \dot{\varphi}_2) = m_1 e_1 \ddot{x} \sin\varphi_1 + \tilde{k}_1 \dot{q}_1,\\
&(J_2 + m_2 e_2^2)\ddot{\varphi}_2 + d_\varphi(\dot{\varphi}_2 - \dot{\varphi}_1) = m_2 e_2 \ddot{x} \sin\varphi_2 + \tilde{k}_2 \dot{q}_2,\\
&L_i \ddot{q}_i + R_i \dot{q}_i + \tilde{k}_i \dot{\varphi}_i = V_i, \qquad i = 1, 2.
\end{aligned} \tag{1}$$

The last equation describes the electric circuit of the DC motors, where L_i is the inductance, R_i is the resistance, V_i is the voltage source and q_i is the current for each motor; see Fig. 1b. The electric subsystem can be reduced by assuming large electric resistance. This means that the characteristic electric time $t_e = L/R$ is much smaller than the characteristic mechanic time $t_m = \min\{1/k, 1/\sigma\}$, where k is the resonance frequency and σ is the damping ratio of

the mechanical system. In this case, the current can be replaced by its stationary solution

$$\dot{q}_i = \frac{V_i}{R_i} - \frac{\tilde{k}_i}{R_i}\dot{\varphi}_i. \tag{2}$$

With this assumption, the electric current is interpreted as a singularly perturbed variable and is replaced by its stationary solution, which results in a linearized motor characteristic. The torque acting on the rotors can then be written as

$$\tilde{k}_i\dot{q}_i = \frac{V_i\tilde{k}_i}{R_i} - \frac{\tilde{k}_i^2}{R_i}\dot{\varphi}_i = U_i(\omega_i^* - \dot{\varphi}_i), \tag{3}$$

where U_i describes the slope of the motor characteristic and ω_i^* the nominal rotation speed. Introducing the non-dimensional parameters

$$\begin{aligned}
&\mu_i = \frac{m_i}{M^*}, \quad \nu_i = \frac{e_i}{e^*}, \quad \varepsilon s_i = \frac{1}{1 + J_i/m_i e_i^2}, \quad u_i = \frac{U_i}{k m_i e_i^2},\\
&\lambda_i = \frac{\omega_i^*}{k}, \quad b_i = \frac{d_\varphi \varepsilon s_i}{k m_i e_i^2}, \quad 2\sigma = \frac{d}{kM^*}, \quad k^2 = \frac{c}{M^*},\\
&\xi = \frac{x}{e^*}, \quad e^* = (e_1 + e_2)/2, \quad M^* = M + m_1 + m_2, \quad \tau = kt,
\end{aligned} \tag{4}$$

the equations of motion for the non-dimensional displacement ξ and the rotor angles φ_1, φ_2 can be written as

$$\xi'' + 2\sigma\xi' + \xi = \sum_{i=1}^{2} \mu_i\nu_i(\varphi_i'' \sin\varphi_i + \varphi_i'^2 \cos\varphi_i), \tag{5}$$

$$\varphi_1'' + b_1(\varphi_1' - \varphi_2') = \varepsilon s_1 \left(\frac{1}{\nu_1}\xi'' \sin\varphi_1 + u_1(\lambda_1 - \varphi_1')\right) = \varepsilon f_1, \tag{6}$$

$$\varphi_2'' + b_2(\varphi_2' - \varphi_1') = \varepsilon s_2 \left(\frac{1}{\nu_2}\xi'' \sin\varphi_2 + u_2(\lambda_2 - \varphi_2')\right) = \varepsilon f_2. \tag{7}$$

The damping parameters σ and b_i are not small where the parameter ε is assumed to be small. This consideration is important to distinguish the slow and fast motions of the system.

3 Asymptotic Analysis

The averaging method for partially damped systems is used for the asymptotic analysis of the system, which was introduced in [9]. The method's goal is to separate slow and strongly damped variables to consider only the slow dynamics of the system. Similarly to [7], this system can be transformed into standard form to apply averaging.

Introducing the new variables, which correspond to the mean phase and the phase difference

$$\psi = \frac{\varphi_1 + \varphi_2}{2}, \tag{8}$$

$$\delta = \varphi_2 - \varphi_1, \tag{9}$$

$$\omega = \psi', \tag{10}$$

the equations can be written in the form

$$\psi' = \omega, \tag{11}$$

$$\omega' = -\frac{b_2 - b_1}{2}\delta' + \frac{1}{2}\varepsilon(f_1 + f_2), \tag{12}$$

$$\delta'' = -(b_1 + b_2)\delta' + \varepsilon(f_2 - f_1). \tag{13}$$

Splitting the phase difference δ in the following way

$$\delta = \tilde{\delta} + \beta, \tag{14}$$

$$\tilde{\delta} = \frac{\delta' + (b_1 + b_2)\delta}{b_1 + b_2}, \tag{15}$$

yields the equations

$$\omega' = \frac{1}{2}\varepsilon\left(f_1 + f_2 - \frac{b_2 - b_1}{b_1 + b_2}(f_2 - f_1)\right), \tag{16}$$

$$\tilde{\delta}' = \frac{\varepsilon}{b_1 + b_2}(f_2 - f_1), \tag{17}$$

$$\beta' = -(b_1 + b_2)\beta - \varepsilon\frac{f_2 - f_1}{b_1 + b_2}. \tag{18}$$

The variable β is strongly damped and, in the first-order approximation can be replaced by its stationary solution

$$\beta_0 = -\varepsilon\frac{f_2 - f_1}{(b_1 + b_2)^2}. \tag{19}$$

So, in the first-order approximation it holds

$$\delta = \tilde{\delta} + \mathcal{O}(\varepsilon). \tag{20}$$

With the introduced variable transformations, Eq. (5) can be rewritten as

$$\xi'' + 2\sigma\xi' + \xi = \omega^2\left(\mu_1\nu_1\cos\left(\psi - \frac{1}{2}\tilde{\delta}\right) + \mu_2\nu_2\cos\left(\psi + \frac{1}{2}\tilde{\delta}\right)\right) + \mathcal{O}(\varepsilon). \tag{21}$$

For the first approximation, this equation has the solution

$$\begin{aligned}\xi_p = \hat{\xi}_p\Bigg(&\mu_1\nu_1\left(\left(1 - \omega^2\right)\cos\left(\psi - \frac{1}{2}\tilde{\delta}\right) + 2\sigma\omega\sin\left(\psi - \frac{1}{2}\tilde{\delta}\right)\right)\\ &+\mu_2\nu_2\left(\left(1 - \omega^2\right)\cos\left(\psi + \frac{1}{2}\tilde{\delta}\right) + 2\sigma\omega\sin\left(\psi + \frac{1}{2}\tilde{\delta}\right)\right)\Bigg),\end{aligned} \tag{22}$$

$$\hat{\xi}_p = \frac{\omega^2}{(1 - \omega^2)^2 + 4\sigma^2\omega^2},$$

which can be used for the variable transformation

$$\xi = \xi_p + \rho_1 \sin(\kappa\tau) + \rho_2 \cos(\kappa\tau), \tag{23}$$

$$\xi' = \xi_p' + \rho_1\kappa \cos(\kappa\tau) - \rho_2\kappa \sin(\kappa\tau) - \sigma(\rho_1 \sin(\kappa\tau) + \rho_2 \cos(\kappa\tau)), \tag{24}$$

$$\kappa = \sqrt{1-\sigma^2}. \tag{25}$$

This transformation yields differential equations

$$\rho_1' = -\sigma\rho_1 + \mathcal{O}(\varepsilon), \tag{26}$$

$$\rho_2' = -\sigma\rho_2 + \mathcal{O}(\varepsilon), \tag{27}$$

for the strongly damped variables ρ_1 and ρ_2. Since their solutions are of the magnitude of order ε, they can also be neglected for the first-order approximation. Considering ψ as the independent variable, the system can finally be written in standard form

$$\frac{\mathrm{d}\omega}{\mathrm{d}\psi} = \frac{\varepsilon}{2\omega}\left(f_1 + f_2 - \frac{b_2 - b_1}{b_1 + b_2}(f_2 - f_1)\right), \tag{28}$$

$$\frac{\mathrm{d}\tilde{\delta}}{\mathrm{d}\psi} = \frac{\varepsilon}{(b_1 + b_2)\omega}(f_2 - f_1) \tag{29}$$

Analogously, the averaged equations of the first-order approximation are

$$\bar{\omega}' = \frac{\varepsilon}{2\bar{\omega}}\left(\bar{f}_1 + \bar{f}_2 - \frac{b_2 - b_1}{b_1 + b_2}(\bar{f}_2 - \bar{f}_1)\right), \tag{30}$$

$$\bar{\delta}' = \frac{\varepsilon}{(b_1 + b_2)\bar{\omega}}(\bar{f}_2 - \bar{f}_1), \tag{31}$$

where

$$\begin{aligned}\bar{f}_1 = {} & s_1 u_1 (\lambda_1 - \bar{\omega}) + \frac{s_1\bar{\omega}^4}{2\nu_1\left((1-\bar{\omega}^2)^2 + 4\sigma^2\bar{\omega}^2\right)}\left(\mu_2\nu_2\left(1-\bar{\omega}^2\right)\sin\bar{\delta}\right. \\ & \left. -2\sigma\bar{\omega}\left(\mu_1\nu_1 + \mu_2\nu_2\cos\bar{\delta}\right)\right),\end{aligned} \tag{32}$$

$$\begin{aligned}\bar{f}_2 = {} & s_2 u_2 (\lambda_2 - \bar{\omega}) + \frac{s_2\bar{\omega}^4}{2\nu_2\left((1-\bar{\omega}^2)^2 + 4\sigma^2\bar{\omega}^2\right)}\left(\mu_1\nu_1\left(\bar{\omega}^2-1\right)\sin\bar{\delta}\right. \\ & \left. -2\sigma\bar{\omega}\left(\mu_2\nu_2 + \mu_1\nu_1\cos\bar{\delta}\right)\right),\end{aligned} \tag{33}$$

$\bar{\omega}$ is the averaged mean rotor velocity and $\bar{\delta}$ averaged phase difference.

4 Analysis of the System Behavior

In this section, the stationary and periodic solutions of the averaged system are investigated. Where necessary, the basin of attraction of certain solutions is shown in phase space, and the dependency on initial values is explained. First, a system with almost identical rotors is investigated, where rotors have identical parameter values except for the nominal speed, and then a system with rotors of different masses.

4.1 System with Almost Identical Rotors

For the first case investigated, the following parameter values are considered:

$$\begin{aligned} &\varepsilon = 0.001, \quad \sigma = 0.1, \quad b_1 = 0.1, \quad s_1 = s_2 = 1, \quad \nu_1 = \nu_2 = 1, \\ &\mu_1 = \mu_2 = 0.2, \quad \lambda_1 = 1.5, \quad \lambda_2 = 1.7, \quad u_1 = u_2 = u \end{aligned} \tag{34}$$

Stationary solutions are derived by setting the right-hand side of Eqs. (30)–(31) to zero. It is not possible to solve the equations for the averaged variables explicitly. However, it is possible to eliminate the averaged phase difference and get an equation for stationary rotor velocity as a function of system parameters. Figure 2 depicts the dependence of stationary solutions on the parameter u, which corresponds to the motor power. This system possesses two stable stationary solutions. The undercritical stable solution, which lies below resonance frequency ($\bar{\omega} = 1$), is characterized as capture into resonance. The overcritical stable solution corresponds to the self-synchronization of the rotors. A system without relative damping also possesses the same curve, even though in such a system, other solutions coexist as well. This case is discussed in [24]. By inspecting the structure of the averaged equations, it is also possible to identify that the stationary solutions do not depend on relative damping between the rotors. If the right-hand side of (31) should be equal to zero, it means $\bar{f}_2 - \bar{f}_1 = 0$. Applying this to the right-hand side of (30) yields $\bar{f}_1 + \bar{f}_2 = 0$. Since $\bar{f}_1$ and $\bar{f}_2$ do not depend on relative damping, stationary solutions do not depend as well. Since multiple stable solutions can coexist in the system, the phase space will be evaluated for different values of parameter u to identify the basin of attraction of different stable solutions and the evolution of critical points. The phase space of the averaged system has the form of a cylinder surface because of the periodic variable $\bar{\delta}$.

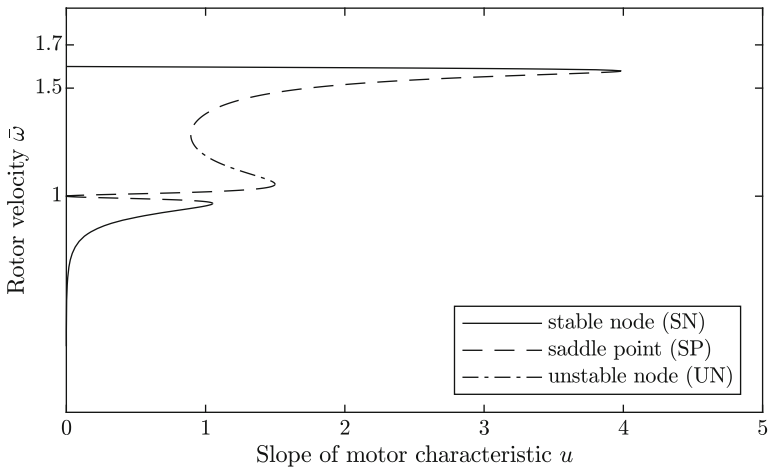


Fig. 2. Stationary rotor velocity $\bar{\omega}$ as a function of bifurcation parameter u for identical rotors

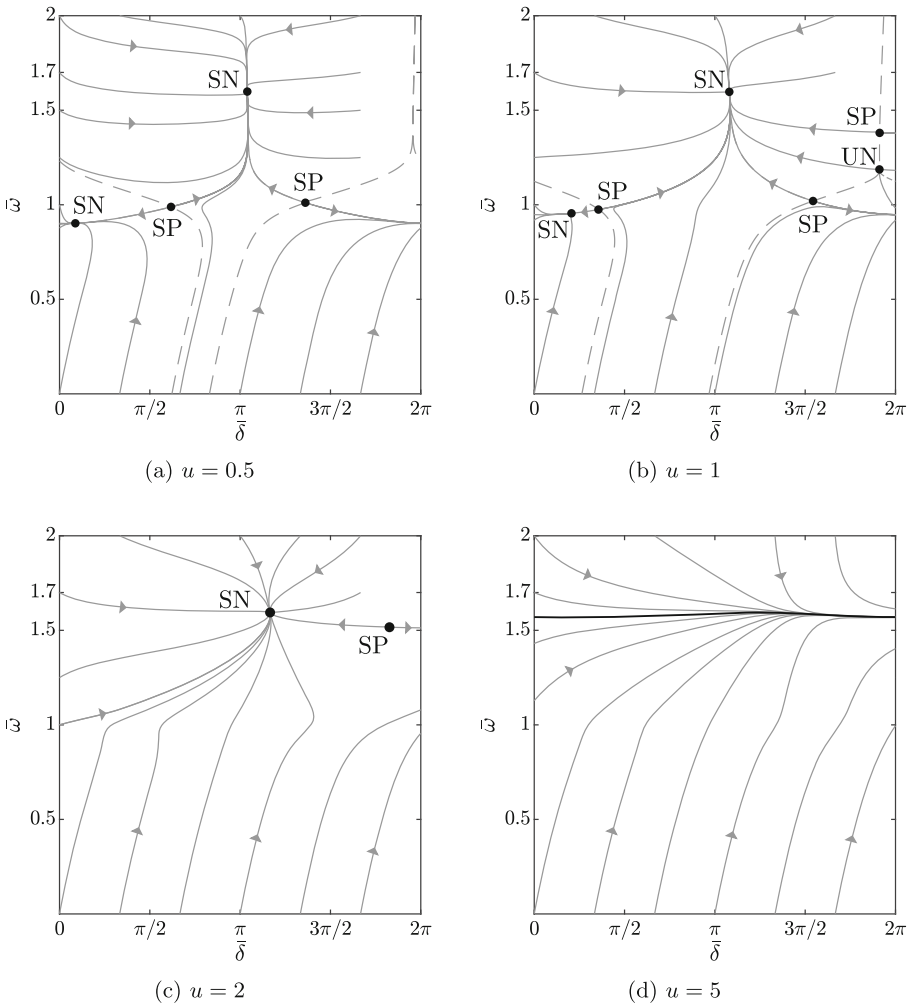
If the system is excited with relatively weak motors ($u = 0.5$), both capture into resonance and self-synchronization are possible. It depends on the initial conditions under which a stationary solution is reached. The phase space for this case can be seen in Fig. 3a. The stationary solutions are depicted as points and labeled depending on their types as stable nodes (SN), saddle points (SP), or unstable nodes (UN). The separatrixes that form the boundary between the basin of attraction of two different stable solutions are depicted as dashed curves. It can be seen that the system can accelerate from zero to self-synchronization without capturing into resonance if the initial phase difference is chosen accordingly. At the same time, it is also possible for the rotors to decelerate and be captured into resonance even if they start the motion above the resonance frequency.

For $u = 1$, two additional stationary solutions are present, an unstable node and an additional saddle point; see Fig. 3b. Qualitatively, the basin of attraction of both stable solutions remains the same. However, the domain for synchronous solution becomes larger. This also means that the interval for the initial phase difference gets bigger, where passing through resonance is possible during the run-up.

Increasing the value of u further, first, the lower stable node and the neighboring saddle point near each other and eventually vanish. From this point on, there is only one stable solution in the system, the overcritical synchronous solution, and the whole phase space is its basin of attraction. Second, the unstable node and the lower saddle point vanish, and the phase space gets the form in Fig. 3c, which is depicted for $u = 2$.

If the motor power is increased further, the last stationary solutions disappear, and the periodic solution wrapping around the phase cylinder is born through the infinite period bifurcation; see Fig. 3d. The stationary solutions of the averaged system correspond to motions where rotors have constant velocity, and the carrier oscillates with constant amplitude. The periodic solution corresponds to an oscillation of the carrier with slowly modulated amplitudes, where the rotor velocities oscillate between two nominal speeds $\lambda_1 = 1.5$ and $\lambda_2 = 1.7$. This behavior can be seen in Fig. 4, where the full and averaged system results are represented, which are in a reasonable agreement.

Exciting a system with almost identical rotors with sufficiently strong motors, it is possible to produce slowly modulated oscillations. This solution lies in the overcritical domain where neither capture into resonance nor self-synchronization is possible. The amplitude of the carrier is relatively small in this case, and therefore, its use as an exciter can be limited. An improved amplitude curve can be achieved if the periodic solution is around the resonance frequency. This case is discussed in the next section.



(a) $u = 0.5$ (b) $u = 1$

(c) $u = 2$ (d) $u = 5$

Fig. 3. Phase space for different parameter values u

4.2 System with Rotors of Different Unbalanced Masses

In this section, a system with rotors of different unbalanced masses is investigated. The considered parameters are:

$$\begin{aligned} &\varepsilon = 0.001, \quad \sigma = 0.2, \quad b_1 = 0.1, \quad s_1 = s_2 = 1, \quad \nu_1 = \nu_2 = 1, \\ &\mu_1 = 0.3, \quad \mu_2 = 0.1, \quad \lambda_1 = 1.5, \quad \lambda_2 = 1.7, \quad u_1 = u_2 = u \end{aligned} \tag{35}$$

Stationary solutions are derived analogously and depicted in Fig. 5. In this case, there is only one stable solution for any parameter value u, so the phase space will not be discussed further. Bifurcations also occur similarly. For weak

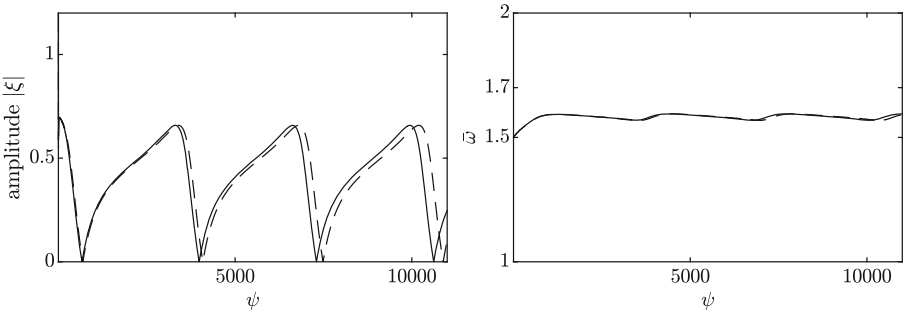


Fig. 4. Evaluation of the carrier amplitude (left) and mean rotor velocity (right) for motor characteristics slope $u = 5$, original (dashed) and averaged (full) system

motors ($u < u_1^* \approx 0.34$), capturing into resonance is the only stable solution. Self-synchronization is the only stable solution for sufficiently strong motors ($u > u_2^* \approx 0.66$). If the motors are even stronger, the periodic overcritical solution appears. All these three solutions are the same types investigated in the previous section. A new periodic solution exists in the range $u_1^* < u < u_2^*$, where rotor velocities oscillate around resonance frequency, and the carrier conducts slowly modulated vibrations. This behavior is shown in Fig. 6. It can only occur if the rotors are not identical.

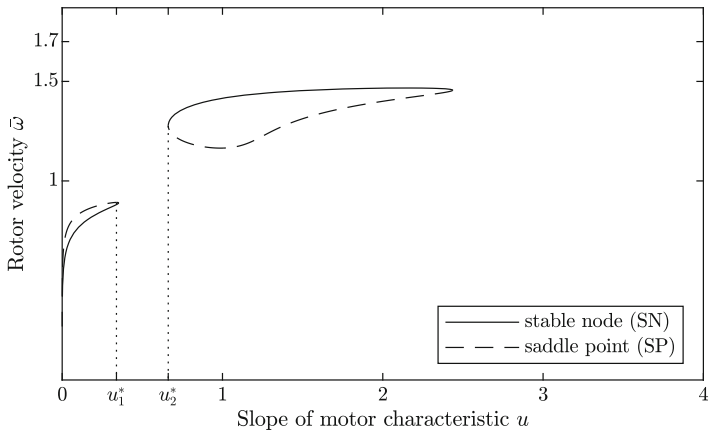


Fig. 5. Stationary rotor velocity $\bar{\omega}$ as a function of bifurcation parameter u for non-identical rotors

It is possible to investigate the equations for stationary solutions, $\bar{f}_1 + \bar{f}_2 = 0$ and $\bar{f}_2 - \bar{f}_1 = 0$, to derive existence conditions, but it is not possible to represent the effect of parameters on the existence of stationary solutions explicitly. These

modulated vibrations around resonance frequency can be described as nonstationary capture into resonance. It features relatively high amplitudes and is the ideal type of solution if a modulating excitation is desired.

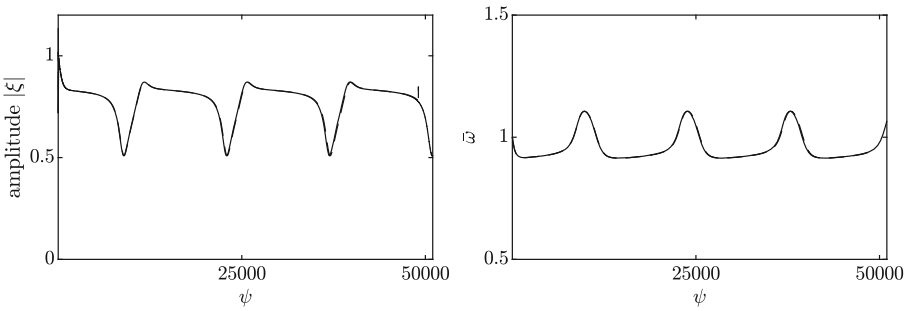


Fig. 6. Evaluation of the carrier amplitude (left) and mean rotor velocity (right) for motor characteristics slope $u = 0.5$, original (dashed) and averaged (full) system

5 Controlling the Modulated Vibrations

In this section, a parameter study is carried out in order to see the effect of certain parameters on the modulated vibrations. A system with parameters given in (35) is taken as a starting point, and individual parameters are varied.

First, the effect of motor power is investigated by changing the slope characteristic u. Results are shown in Fig. 7a. Increasing the parameter has multiple effects. Maximum vibration amplitudes increase. The modulation frequency increases as well. This means that changing this parameter can control the time intervals with small and large amplitudes in one modulation period. Next, the mass difference is investigated. For that purpose, the sum of mass ratios $\mu_1 + \mu_2 = 0.4$ is held constant, and the difference $\mu_1 - \mu_2$ is varied. Results are given in Fig. 7b. With a more significant mass ratio difference, the amplitudes decrease, and the modulation frequency increases. Damping seems to have a small impact on the qualitative behavior of this motion. It does not change the shape of the modulation. Figure 7c depicts the effect of inertial damping σ. Higher damping results in smaller amplitudes, as expected. Relative damping b_1 affects only modulation frequency; see Fig. 7d.

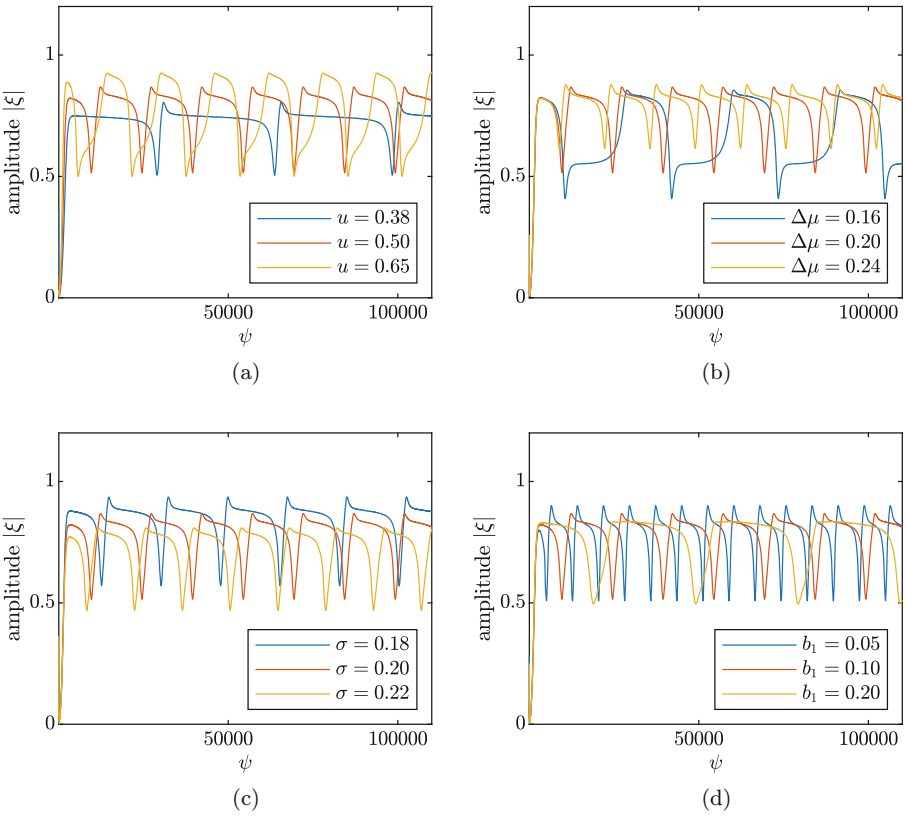


Fig. 7. Modulated vibrations for different values of (a) slope characteristic u, (b) mass ratio difference $\Delta\mu$, (c) inertial damping σ, and (d) relative damping b_1

6 Conclusions

A new vibration exciter is proposed. It consists of a pair of inertially coupled unbalanced rotors with an additional viscous coupling between them. Its slow dynamics are investigated using the averaging method for partially strongly damped systems. With this method, the original differential equation system of the sixth order is reduced to a system of the second order. The system is first investigated with almost identical rotors. Its stationary and periodic solutions are shown. Bifurcations and basins of attraction of coexisting solutions are presented by displaying the phase space. With a different set of parameters, it is shown that the system can possess different solution types: capture into resonance, self-synchronization, and slowly modulated vibrations in overcritical and resonant regimes. In the last case, only one stable solution exists for different motor powers, which might be beneficial for applications. A small parameter study is carried out to see different parameters' effects on controlling or modifying the slowly modulated vibrations. Motor power seems to be the most

influential factor on the shape of the modulation, which changes the relation of time intervals with small and large amplitudes over a modulation period. This is highly beneficial, for example, if such an exciter is used in a test rig to predict the fatigue life of mechanical parts.

Further investigation can focus on the realization of such a model and experimental verification of the obtained results or combining viscous coupling with other features like multiple rotors or multiple coaxial unbalances.

References

1. Balthazar, J.M., Brasil, R., Felix, J.L., Tusset, A.M., Piccirillo, V., Souza, L.: An overview of nonlinear dynamics of electro-mechanical engineering systems, excited by small motors, pp. 1–10 (2015)
2. Balthazar, J.M., Felix, J.L.P., Brasil, R.M.: Some comments on the numerical simulation of self-synchronization of four non-ideal exciters. Appl. Math. Comput. **164**(2), 615–625 (2005)
3. Blekhman, I.I.: Self-synchronization of certain vibratory devices. Eng. Trans. **16** (1953)
4. Blekhman, I.I.: Synchronization in Science and Technology. ASME press, New York (1988)
5. Blekhman, I.I., Fradkov, A.L., Tomchina, O.P., Bogdanov, D.: Self-synchronization and controlled synchronization: general definition and example design. Math. Comput. Simul. **58**(4–6), 367–384 (2002)
6. Drozdetskaya, O., Fidlin, A.: On the passing through resonance of a centrifugal exciter with two coaxial unbalances. Eur. J. Mech.-A/Solids **72**, 516–520 (2018)
7. Drozdetskaya, O., Fidlin, A.: Passing through resonance of the unbalanced rotor with self-balancing device. Nonlinear Dyn. **106**(3), 1647–1657 (2021). https://doi.org/10.1007/s11071-021-06973-4
8. Fidlin, A., Drozdetskaya, O.: On the averaging in strongly damped systems: the general approach and its application to asymptotic analysis of the sommerfeld effect. Procedia IUTAM **19**, 43–52 (2016)
9. Fidlin, A., Thomsen, J.J.: Non-trivial effects of high-frequency excitation for strongly damped mechanical systems. Int. J. Non-Linear Mech. **43**(7), 569–578 (2008)
10. Green, K., Champneys, A., Lieven, N.: Bifurcation analysis of an automatic dynamic balancing mechanism for eccentric rotors. J. Sound Vib. **291**(3–5), 861–881 (2006)
11. Horvath, R., Flowers, G.T., Fausz, J.: Influence of nonideaities on the performance of a self-balancing rotor system. In: ASME International Mechanical Engineering Congress and Exposition, vol. 42126, pp. 233–242 (2005)
12. Hou, Y., Fang, P., et al.: Investigation for synchronization of a rotor-pendulum system considering the multi-dof vibration. Shock Vibr. **2016** (2016)
13. Kong, X., Chen, C., Wen, B.: Composite synchronization of three eccentric rotors driven by induction motors in a vibrating system. Mech. Syst. Signal Process. **102**, 158–179 (2018)
14. Kong, X., Zhang, X., Chen, X., Wen, B., Wang, B.: Synchronization analysis and control of three eccentric rotors in a vibrating system using adaptive sliding mode control algorithm. Mech. Syst. Signal Process. **72**, 432–450 (2016)

15. Kuchle, A., Naunheimer, H., Bertsche, B., Ryborz, J., Novak, W., Fietkau, P.: Automotive Transmissions: Fundamentals Selection, Design and Application. Springer, Heidelberg (2010)
16. Neishtadt, A.I.: Passage through resonances in a two-frequency problem. Akademiia Nauk SSSR Doklady **221**, 301–304 (1975)
17. Neishtadt, A.I.: Capture into resonance and scattering on resonances in two-frequency systems. Trudy Matematicheskogo Instituta imeni VA Steklova **250**, 198–218 (2005)
18. Pechenev, A.V.: Averaging of systems with hierarchy of phase rotation speeds on substantially large time intervals. Dokl. Akad. Nauk SSSR **315**(1), 28–31 (1990)
19. Rodrigues, D., Champneys, A., Friswell, M., Wilson, R.: Automatic two-plane balancing for rigid rotors. Int. J. Non-Linear Mech. **43**(6), 527–541 (2008)
20. Ryzhik, B., Sperling, L., Duckstein, H.: Non-synchronous motions near critical speeds in a single-plane auto-balancing device. Technische Mechanik-Eur. J. Eng. Mech. **24**(1), 25–36 (2004)
21. Sommerfeld, A.: Beiträge zum dynamischen Ausbau der Festigkeitslehre. wydawca nieznany (1902)
22. Thearle, E.L.: A new type of dynamic-balancing machine. Trans. Am. Soc. Mech. Eng. **54**(2), 131–140 (1932)
23. Wen, B., Zhang, X., Zhao, C., Liu, S., Lin, X.: Vibratory synchronization modes and their applications based on a vibrating bedstand with coupled exciters. In: Proceedings of the 14th IFToMM World Congress, pp. 4–11 (2015)
24. Yüzbasioglu, T., Fidlin, A.: Passage through resonance of two coupled exciters. PAMM **21**(1), e202100121 (2021)
25. Zhang, X., Wen, B., Zhao, C.: Vibratory synchronization transmission of a cylindrical roller in a vibrating mechanical system excited by two exciters. Mech. Syst. Signal Process. **96**, 88–103 (2017)
26. Zimmermann, K., Zeidis, I.: Dynamical behavior of a mobile system with two degrees of freedom near the resonance. Acta. Mech. Sin. **27**, 7–17 (2011)

Phase-Amplitude Reduction of Limit-Cycling Networks for Optimal Synchronization

Petar Mircheski[1(✉)], Jinjie Zhu[2], and Hiroya Nakao[1]

[1] Department of Systems and Control Engineering Tokyo Institute of Technology, Tokyo 152-8552, Japan
mircheski.p@k.sc.e.titech.ac.jp

[2] State Key Laboratory of Mechanics and Control of Mechanical Structures College of Aerospace Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Abstract. We present a phase-amplitude reduction framework for networked dynamical systems collectively exhibiting stable limit-cycle oscillations. The framework takes into account both the dynamics along the limit cycle and deviations from the limit cycle, described by the phase variable and the amplitude variables, respectively. Phase-amplitude reduction allows us to analyze the dynamics of weakly interacting limit-cycling networks including synchronization. As an example, we optimize the phase locking of a network of FitzHugh-Nagumo elements by a periodic signal. By deriving the optimal injection signal and applying feedback control, we can stabilize the collective oscillations of the network without altering the phase dynamics and realize fast synchronization.

Keywords: networks · coupled dynamical systems · collective oscillations · phase-amplitude reduction · synchronization

1 Introduction

Collective oscillations of networked dynamical systems are commonly observed in the real world [1–6]. Phase reduction is a conventional dimensionality-reduction method for limit cycles that projects the dynamics of the system only on the phase direction along the autonomous limit cycle orbit and derives a simple phase equation [2,3,7–9].

Recently, a generalization of the phase reduction method has gained attention. The method, called phase-amplitude reduction, describes the network dynamics by using both the phase and amplitude variables of the system, respectively representing dynamics along the limit cycle and deviations from the limit cycle. In addition to the phase equation, amplitude equations characterized by Floquet exponents with negative real parts are used to describe deviations of the system state from the limit cycle [10–16]. Moreover, by keeping the phase and

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 245–254, 2025.
https://doi.org/10.1007/978-3-031-72794-8_18

the slowest decaying amplitude variables of the system, a pair of coupled one-dimensional phase-amplitude equations are obtained.Unlike the phase reduction method, which relies on the assumption that the input given to the system is weak enough and the deviation of the network state from the limit cycle can be neglected, the phase-amplitude reduction method can describe the underlying dynamics of the oscillator even under moderately strong inputs by considering the amplitudes.

In this article, we briefly review our recent formulation of the phase-amplitude reduction framework for limit-cycling networks exhibiting stable collective oscillations [18]. As a demonstration of the method, we consider a star network of coupled FitzHugh-Nagumo elements and study its synchronization properties. We use the phase-amplitude reduction framework to derive the optimal injection signal to achieve fast phase locking while suppressing the amplitude deviations.

2 Collectively Oscillating Network

2.1 Networked Dynamical Systems Exhibiting Collective Oscillations

We consider a network of N coupled dynamical elements described by:

$$\dot{\boldsymbol{x}}_i(t) = \boldsymbol{f}_i(\boldsymbol{x}) + \sum_{j=1}^{N} \boldsymbol{g}_{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j), \quad (i = 1, ..., N), \tag{1}$$

where $\boldsymbol{x}_i(t) \in \mathbb{R}^{n_i}$ represents a n_i-dimensional state of element i at time t, $\dot{\langle\rangle}$ represents the time derivative, $\boldsymbol{f}_i : \mathbb{R}^{n_i} \to \mathbb{R}^{n_i}$ represents individual dynamics of element i, and $\boldsymbol{g}_{ij} : \mathbb{R}^{n_i} \times \mathbb{R}^{n_j} \to \mathbb{R}^{n_i}$ describes the coupling, i.e., the effect that element j exerts on element i, respectively. Each element has no self coupling, i.e., $\boldsymbol{g}_{ii} = 0$ for all i, which can be assumed without loss of generality by appropriately defining the individual dynamics, and no element is isolated, i.e., the network is connected. The dimensionality n_i of each element and the coupling of the network can be arbitrary. The total dimension of the whole dynamical system is denoted by $M = \sum_{i=1}^{N} n_i$. We assume that this network has an exponentially stable limit-cycle solution representing stable collective oscillations.

We represent the network state collectively by a M-dimensional vector, $\boldsymbol{X}(t) = (\boldsymbol{x}_1(t), \boldsymbol{x}_2(t), \ldots \boldsymbol{x}_N(t)) \in \mathbb{R}^M$, and the exponentially stable limit-cycle solution by $\boldsymbol{X}^{(0)}(t) \in \mathbb{R}^M$ satisfying $\boldsymbol{X}^{(0)}(t) = \boldsymbol{X}^{(0)}(t + T)$, where T is the natural period of the limit cycle, i.e., the collective oscillations. The natural frequency of the limit cycle is given by $\omega = 2\pi/T$. The linear stability of the limit cycle solution $\boldsymbol{X}^{(0)}(t)$ is characterized by its Floquet exponents $\lambda_0, \lambda_1, \ldots, \lambda_{M-1}$, sorted in decreasing order of their real parts [17]. The first Floquet exponent is $\lambda_0 = 0$, which is associated with the tangent direction along the limit cycle. The other exponents $\lambda_1, \ldots \lambda_{M-1}$, which possess negative real parts and can in general be complex, are associated with the amplitude deviations from the limit cycle.

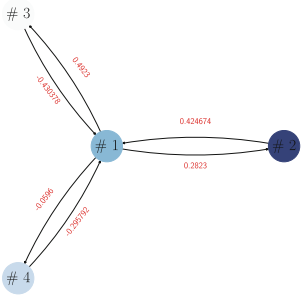


Fig. 1. Star network of FitzHugh-Nagumo elements ($N = 4$).

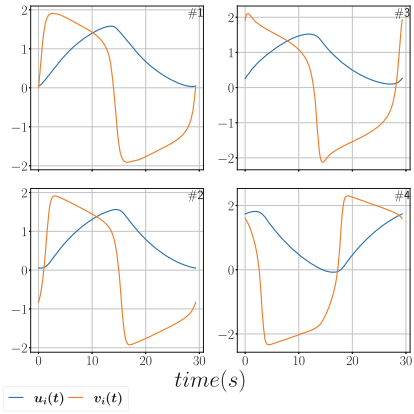


Fig. 2. One-period time evolution of $\boldsymbol{x}_i(t) = (u_i(t), v_i(t))^\top$ for the star network ($i = 1, 2, 3$, and 4).

2.2 Networks of FitzHugh-Nagumo Elements

In this article, we apply the phase-amplitude reduction framework to a star network of FitzHugh-Nagumo elements exhibiting collective oscillations (Fig. 1). The state of the ith element is represented by a column vector $\boldsymbol{x}_i(t) = (u_i(t), v_i(t))^\top$ for $i = 1, ..., N$, and the individual dynamics $\boldsymbol{f}_i(u_i, v_i)$ of the ith element is described by a two-dimensional vector field:

$$\begin{pmatrix} \dot{u}_i \\ \dot{v}_i \end{pmatrix} = \begin{pmatrix} \delta(v_i + a + bu_i) \\ v_i - \frac{v_i^3}{3} - u_i + I \end{pmatrix}, \tag{2}$$

where δ, a, b, and I are parameters. As the FitzHugh-Nagumo model represents a neuron model, interactions between elements are assumed to occur only through their membrane potential v_i. The coupling function between elements i and j ($i, j = 1, \ldots, N$) is defined as:

$$\boldsymbol{g}_{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j) = K_{ij} \begin{pmatrix} 0 \\ v_j - v_i \end{pmatrix}, \tag{3}$$

where K_{ij} represents the coupling strength.

The parameters $a = 0.7$, $b = 0.8$, $\delta = 0.08$, and $I = 0.8$ are identical for all elements, while the matrix K of the coupling strengths $\{K_{ij}\}$ is given by

$$K = \begin{pmatrix} 0.0 & 0.28230 & 0.49230 & -0.05960 \\ 0.424674 & 0.0 & 0.0 & 0.0 \\ -0.430378 & 0.0 & 0.0 & 0.0 \\ -0.295792 & 0.0 & 0.0 & 0.0 \end{pmatrix}. \tag{4}$$

As the elements are coupled in a star topology with random weights, they do not exhibit uniformly synchronized oscillations. However, all the elements exhibit synchronized collective oscillations in the sense that each element consistently displays periodic dynamics while maintaining a constant timing relationship with the other elements as shown in Fig. 2.

We have numerically evaluated the natural period of this network as $T \approx 29.31$ and accordingly the natural frequency as $\omega \approx 0.214$. The slowest-decaying Floquet exponent has been calculated as $\lambda_1 \approx -0.09$, and the second-slowest exponents as $\lambda_2 \approx -0.12$. The slowest exponent is real and the associated amplitude exhibits a simple exponential decay.

3 Phase-Amplitude Reduction Theory

In this section, we briefly review phase-amplitude reduction theory for networks developed in Ref. [18]. For an exponentially stable limit-cycle solution $\boldsymbol{X}^{(0)}(t) = (\boldsymbol{x}_1^{(0)}(t), \ldots, \boldsymbol{x}_N^{(0)}(t))$ of the network, we can introduce a phase function $\Theta : B \to [0, 2\pi)$ satisfying $\dot{\Theta}(\boldsymbol{X}) = \dot{\Theta}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N) = \sum_{i=1}^{N} \frac{\partial \Theta}{\partial \boldsymbol{x}_i} \cdot \dot{\boldsymbol{x}}_i = \omega$ and amplitude functions $R_m : B \to \mathbb{C}$ satisfying $\dot{R}_m(\boldsymbol{X}) = \dot{R}_m(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N) = \sum_{i=1}^{N} \frac{\partial R_m}{\partial \boldsymbol{x}_i} \cdot \dot{\boldsymbol{x}}_i = \lambda_m R_m$ $(m = 1, ..., M)$ on the limit cycle and in its basin of attraction $B \subset \mathbb{R}^M$, where $\partial/\partial \boldsymbol{x}_i$ represents the gradient with respect to $\boldsymbol{x}_i$ and $\langle \cdot \rangle$ represents a scalar product of two vectors. These phase and amplitude functions are closely related to the Koopman eigenfunctions of the system [11,12].

We introduce the phase and the amplitudes of the network state $\boldsymbol{X}$ by using the phase and amplitude functions defined above as $\theta = \Theta(\boldsymbol{X})$ and $r_m = R_m(\boldsymbol{X})$ $(m = 1, ..., M-1)$. By the definition of Θ and R_m, these phase and amplitude variables obey simple linear equations, $\dot{\theta} = \omega$ and $\dot{r}_m = \lambda_m r_m$, respectively, for arbitrary $\boldsymbol{X} \in B$ when the network is unperturbed. In what follows, we denote the limit cycle as $\tilde{\boldsymbol{X}}^{(0)}(\theta) = \boldsymbol{X}^{(0)}(\theta/\omega)$ as a function of the phase $\theta \in [0, 2\pi)$.

We now consider a network perturbed by a weak input $\boldsymbol{q}_i(t) \in \mathbb{R}^{n_i}$ given to each element i, described by

$$\dot{\boldsymbol{x}}_i(t) = \boldsymbol{f}_i(\boldsymbol{x}) + \sum_{j=1}^{N} \boldsymbol{g}_{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j) + \boldsymbol{q}_i(t), \quad (i = 1, ..., N). \tag{5}$$

Using the definitions of Θ and R_m, we can derive the following phase-amplitude equations for the phase θ and amplitudes $\{r_m\}$ of the network to the first-order approximation in $\{\boldsymbol{q}_i\}$ [18]:

$$\dot{\theta}(t) = \omega + \sum_{i=1}^{N} \boldsymbol{z}_i(\theta) \cdot \boldsymbol{q}_i(t), \tag{6}$$

$$\dot{r}_m(t) = \lambda_m r_m(t) + \sum_{i=1}^{N} \boldsymbol{I}_{m,i}(\theta) \cdot \boldsymbol{q}_i(t), \quad (m = 1, ..., M-1) \tag{7}$$

where we have neglected errors of $O(\|\boldsymbol{q}_i\|^2)$ ($\|\boldsymbol{q}_i\| = \sqrt{\boldsymbol{q}_i \cdot \boldsymbol{q}_i}$ is the norm of $\boldsymbol{q}_i$). Here, $\boldsymbol{z}_i(\theta) \in \mathbb{R}^{n_i}$ and $\boldsymbol{I}_{m,i}(\theta) \in \mathbb{R}^{n_i}$ are the phase sensitivity function (PSF) $\boldsymbol{z}_i(\theta) = \frac{\partial \Theta}{\partial \boldsymbol{x}_i}|^{(0)}_{\boldsymbol{X}=\tilde{\boldsymbol{X}}}(\theta)$ and mth amplitude sensitivity functions (ISFs) $\boldsymbol{I}_{m,i}(\theta) = \frac{\partial R_m}{\partial \boldsymbol{x}_i}|^{(0)}_{\boldsymbol{X}=\tilde{\boldsymbol{X}}}(\theta)$ of the ith element, evaluated at $\boldsymbol{X} = \tilde{\boldsymbol{X}}^{(0)}(\theta)$ on the limit cycle, respectively. These functions characterize linear response of the phase θ and amplitudes $\{r_m\}$ to small inputs $\{\boldsymbol{q}_i\}$ and can be numerically evaluated by solving the adjoint equations associated with Eq. (1) [18–20]. Furthermore, we can reduce the dimensionality of the system by keeping only the n ($< M-1$) slowest decaying modes of the system. In what follows, we set $n = 1$ and focus only on the slowest decaying mode $m = 1$; we omit the subscript m hereafter.

4 Optimal Synchronization of Oscillatory Networks

4.1 Phase Locking of the Network to Periodic Input Signals

One valuable application of phase-amplitude equations is the phase locking of a network through periodic input signals. We denote by $S \subseteq \{1, 2, ..., N\}$ the set of elements receiving the inputs and assume that each element $i \in S$ is subjected to a weak periodic input $\boldsymbol{p}_i(\Omega t) \in \mathbb{R}^{n_i}$ of $O(\epsilon)$. The period of the input signals is denoted as T_e and its frequency as $\Omega = 2\pi/T_e$. We assume that T_e is close to the natural frequency T of the network such that $\Omega - \omega = O(\varepsilon)$.

To analyze phase-locking in the standard way [2,7,9], we introduce a slow relative phase $\phi(t) = \theta(t) - \Omega t$, which satisfies

$$\dot{\phi}(t) = \Delta + \sum_{i \in S} \boldsymbol{z}_i(\phi + \Omega t) \cdot \boldsymbol{p}_i(\Omega t), \tag{8}$$

where $\Delta = \omega - \Omega$ is the frequency mismatch between the network and the periodic input. To eliminate the explicit time dependence, we approximately average the right-hand side over one period of oscillation to obtain:

$$\begin{aligned} \dot{\phi} &= \Delta + \Gamma(\phi), \\ \Gamma(\phi) &= \sum_{i \in S} \Gamma_i(\phi), \\ \Gamma_i(\phi) &= \frac{1}{T_e} \int_0^{T_e} \boldsymbol{z}_i(\phi + \Omega s) \cdot \boldsymbol{p}_i(\Omega s) ds, \end{aligned} \tag{9}$$

where $\Gamma_i(\phi)$ represents the individual phase coupling function of the element $i \in S$ to the periodic input signal [2] and $\Gamma(\phi)$ is their sum. If Eq. (9) possesses a stable fixed point, the collective oscillations of the network will be phase-locked to the periodic input signal, provided that the input is sufficiently weak and the approximation of phase reduction remains valid.

4.2 Optimization of the Periodic Input Signals

As discussed in Ref. [21–24] for a single oscillator, using the phase coupling function, we can formulate an optimization problem for the periodic input $\boldsymbol{p}_i(\Omega t)$ for $i \in S$. We seek the periodic input that (i) guarantees the existence of a phase-locking point ϕ^* to which the relative phase converges, given by $\Delta + \Gamma(\phi^*) = 0$, and (ii) maximizes the linear stability under a constraint on the average power of the input signal over one period,

$$\sum_{i \in S} \left[\|\boldsymbol{p}_i(\Omega t)\|^2 \right]_t = P, \tag{10}$$

where $P > 0$ is the power and $[f(t)]_t = (1/T_e) \int_0^{T_e} f(s)ds$ denotes an average of the function $f(t)$ over one period T_e. This optimization problem can be formulated as

$$\begin{aligned} \text{max:} \quad & -\Gamma'(\phi^*), \\ \text{s.t:} \quad & \sum_{i \in S} \left[\|\boldsymbol{p}_i(\Omega t)\|^2 \right]_t = P, \quad \Delta + \Gamma(\phi^*) = 0. \end{aligned} \tag{11}$$

When the periodic input signals are sufficiently weak, the optimal input signals obtained as described above lead to faster synchronization than simple sinusoidal inputs. However, when the input signals are not sufficiently weak, the network state can deviate from the unperturbed limit cycle, and the phase-only reduction method may fail to describe the collective oscillations of the network.

In Ref. [10], two methods, the amplitude penalty method and amplitude feedback method, were proposed for optimizing a single oscillator's behavior using the reduced phase-amplitude equations. In Ref. [18], these methods are extended to address the collective oscillations of the networked dynamical system and evaluate their efficiency. We briefly review the two methods and apply them to a network of FitzHugh-Nagumo elements with a star topology.

In the amplitude penalty method, the ISFs of the system are included in the cost function, penalizing deviations of the network state from the limit cycle. The optimization problem is formulated as:

$$\begin{aligned} \text{max:} \quad & -\Gamma'(\phi^*) - \sum_{i \in S} W_i \left[|\boldsymbol{I}_i(\phi^* + \Omega t) \cdot \boldsymbol{p}_i(\Omega t)|^2 \right]_t, \\ \text{s.t:} \quad & \sum_{i \in S} \left[\|\boldsymbol{p}_i(\Omega t)\|^2 \right]_t = P, \quad \Delta + \Gamma(\phi^*) = 0, \end{aligned} \tag{12}$$

where $W_i > 0$ represents the ith weight of the amplitude penalty. The amplitude penalty method seeks to maximize the linear stability $-\Gamma'(\phi^*)$ while controlling amplitude deviations from the limit cycle, balancing these aspects. The result of the optimization problem (12) is expected to give balanced inputs $\{\boldsymbol{p}_i(\Omega t)\}$ that maximize the linear stability of the phase-locking point as much as possible while avoiding large amplitude deviations of the network state from the unperturbed

limit cycle. We can solve this problem by using the method of Lagrange multipliers, where we can numerically determine the multipliers and obtain $\{\boldsymbol{p}_i(\Omega t)\}$ for $i \in S$ such that constraints given in Eq. (12) are satisfied [18].

In the amplitude feedback method, the optimization problem remains the same as in the phase-only case, Eq. (11). However, a feedback input is added to the optimized periodic input, taking the form:

$$\boldsymbol{p}_i^{fb}(t) = \boldsymbol{p}_i^{ls}(t) - \alpha \boldsymbol{y}_i(t), \quad (i \in S), \tag{13}$$

where $\alpha \geq 0$ represents the feedback gain, $\boldsymbol{p}_i^{ls}(t)$ is the solution of the phase-only optimization problem from Eq. (11), and $\boldsymbol{y}_i(t) = \boldsymbol{x}_i(t) - \boldsymbol{x}_i^{(0)}(\theta(t))$ with $\theta(t) = \Theta(\boldsymbol{x}_1(t), \ldots, \boldsymbol{x}_N(t))$ measures the deviation of the network state from the state on the unperturbed limit cycle with the same asymptotic phase $\theta(t)$. The amplitude feedback method allows for stronger periodic inputs while ensuring that the deviations of the network state from the limit cycle are suppressed.

It is worth noting that the amplitude penalty method is a feedforward control, whereas the amplitude feedback method requires continuous measurement and evaluation of the asymptotic phase of the network state. In the numerical simulations in Sect. 4.3, both the penalty weights $\{W_i\}$ and feedback gain α have been chosen empirically to achieve the desired synchronization dynamics.

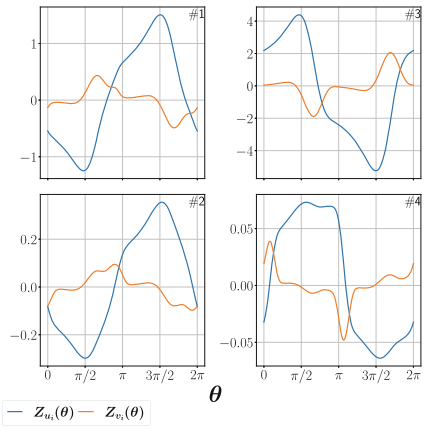


Fig. 3. Phase sensitivity functions (PSFs) $\boldsymbol{z}_i(\theta) = (z_i^u, z_i^v)$ ($i = 1, 2, 3$, and 4) of the star network.

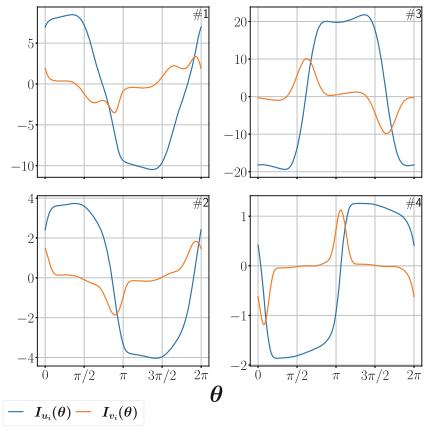


Fig. 4. Amplitude sensitivity functions (ISFs) $\boldsymbol{I}_i(\theta) = (\boldsymbol{I}_{i,u}, \boldsymbol{I}_{i,v})$ ($i = 1, 2, 3$, and 4) of the star network.

4.3 Phase Locking of a Star Network of FitzHugh-Nagumo Elements

We now illustrate the two methods of optimal phase locking with amplitude suppression or feedback by using a star network of FitzHugh-Nagumo elements

exhibiting collective limit-cycle oscillations (Figs. 1 and 2). We consider only a single amplitude, i.e., $n = 1$, in what follows. Numerically obtained PSFs $\{\boldsymbol{z}_i\}$ and ISFs $\{\boldsymbol{I}_i\}$ of the limit-cycle solution of the star network are shown in Figs. 3 and 4, respectively. We can observe that, reflecting the heterogeneity of the network topology, the PSFs and ISFs show distinct characteristics.

Figure 5 shows the results of optimal phase locking for the case where only the central element ($i = 1$) receives the optimal input signal. The average power of the input is set as $P = 0.001$ or $P = 0.01$. We assume there is no frequency mismatch, $\Delta = 0$, and set the target phase-locking point as $\phi^* = 0$. We set the weight of the amplitude penalty as $W_i = 10$ and use a feedback gain of $\alpha = 0.1$ for the weak input case and $\alpha = 10$ for the strong input case. The initial value of the relative phase is set as $\Delta\phi(0) = \pi/4$ to realize phase locking at $\phi^* = 0$.

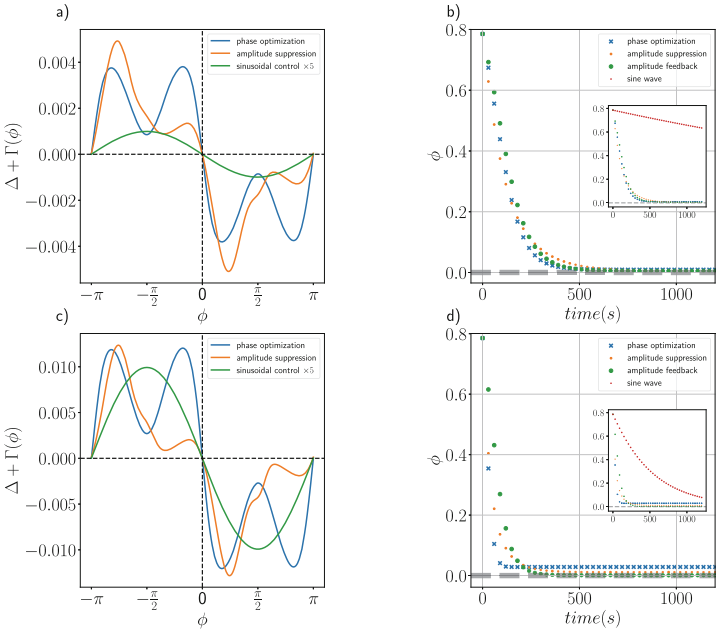


Fig. 5. Phase locking for two different input powers. (a, c) Phase coupling functions with phase-only optimization, amplitude suppression, and periodic sinusoidal forcing. (b, d) Evolution of phase differences (calculated at every period). The insets show the results with sinusoidal inputs with equal powers for comparison. (a, b) $P = 0.001$, (c, d) $P = 0.01$.

Figures 5(a, c) show the phase coupling functions with phase-only optimization, amplitude suppression, and periodic sinusoidal forcing. The phase coupling function for the sinusoidal forcing has been scaled five-fold to enhance clarity and proper visualization and to draw a clear comprehensible comparison with the other methods. As the average power level of the input is increased, the phase coupling function Γ for the case with the amplitude-penalty method shrinks in magnitude, in contrast to the case with the phase-only method. This reduction

in the magnitude occurs in compensation to the suppression of the amplitude and reduces the phase-locking range than the case with the phase-only method. We can observe that the linear stability of $\phi^* = 0$ is lower in the case with amplitude penalty than in the case without amplitude penalty. However, the suppression of amplitude is indeed important in realizing the target point for stronger input, as explained below.

Figures 5(b, d) show the results of direct numerical simulations. We observe that for the weaker input $P = 0.001$ the relative phase converges to the target phase-locking point $\phi^* = 0$ for all three methods correctly but slowly. In the insets of Figs. 5(b, d), the results for sinusoidal inputs of equivalent powers without amplitude suppression are shown as baselines in addition to the results with the optimized inputs. We also note that the convergence speed of the sinusoidal input is of magnitudes smaller compared to the convergence speed of all other three optimization methods. For the stronger input $P = 0.01$, the phase-only method yields deviation from the target point due to amplitude deviations, while the amplitude penalty and amplitude feedback methods successfully realize the target phase-locking point. The amplitude feedback method is slightly better at realizing the target phase-locking point.

5 Conclusion

We have briefly presented a comprehensive framework for phase-amplitude reduction of collective oscillations in networked dynamical systems developed in [18]. This framework has been applied to a network composed of interconnected FitzHugh-Nagumo elements in a star topology. Our findings demonstrate that the reduced phase-amplitude equations can serve as a valuable tool for deriving optimal input waveforms, enabling swift phase locking of collective oscillations. Our observations reveal that the application of the optimized inputs led to marked improvements in convergence, when compared with simple sinusoidal inputs. Furthermore, by implementing amplitude penalty or amplitude feedback methods, we could maintain the fast convergence speed while suppressing amplitude deviations even when using stronger inputs, facilitating efficient attainment of the target phase-locked state.

Our present framework can also be extended to address other optimization challenges, such as identifying optimal combinations of elements to achieve stable phase locking or determining element combinations that help avoid undesired phase-locked states.

References

1. Winfree, A.T.: The Geometry of Biological Time. Springer, New York (1980)
2. Kuramoto, Y.: Chemical Oscillations, Waves, and Turbulence. Springer (1984)
3. Ermentrout, G.B., Terman, D.H.: Mathematical Foundations of Neuroscience. Springer, New York (2010)
4. Pikovsky, A., Rosenblum, M., Kurths, J.: Synchronization: A Universal Concept in Nonlinear Science. Cambridge University Press (2003)

5. Strogatz, S.H.: Sync: How order emerges from chaos in the universe, nature, and daily life. Hachette, UK (2012)
6. Iatsenko, D., et al.: Evolution of cardiorespiratory interactions with age. Philos. Trans. Roy. Soc. Math. Phys. Eng. Sci. **371**(1997), 20110622 (2013)
7. Hoppensteadt, F.C., Izhikevich, E.M.: Weakly Connected Neural Networks, vol. 126. Springer, New York (1997)
8. Brown, E., Moehlis, J., Holmes, P.: On the phase reduction and response dynamics of neural oscillator populations. Neural Comput. **16**(4), 673–715 (2004)
9. Nakao, H.: Phase reduction approach to synchronisation of nonlinear oscillators. Contemp. Phys. **57**(2), 188–214 (2016)
10. Takata, S., Kato, Y., Nakao, H.: Fast optimal entrainment of limit-cycle oscillators by strong periodic inputs via phase-amplitude reduction and floquet theory. Chaos Interdisc. J. Nonlinear Sci. **31**(9), 093124 (2021)
11. Mauroy, A., Susuki, Y., Mezić, I.: Koopman Operator in Systems and Control. Springer (2020)
12. Shirasaka, S., Kurebayashi, W., Nakao, H.: Phase-amplitude reduction of limit cycling systems. Koopman Oper. Syst. Control: Concepts, Methodol. Appl. 383–417 (2020)
13. Wilson, D., Ermentrout, B.: Phase models beyond weak coupling. Phys. Rev. Lett. **123**(16), 164101 (2019)
14. Shirasaka, S., Kurebayashi, W., Nakao, H.: Phase-amplitude reduction of transient dynamics far from attractors for limit-cycling systems. Chaos: Interdisc. J. Nonlinear Sci., **27**(2) (2017)
15. Wilson, D., Moehlis, J.: Isostable reduction of periodic orbits. Phys. Rev. E **94**(5), 052213 (2016)
16. Mauroy, A., Mezić, I., Moehlis, J.: Isostables, isochrons, and Koopman spectrum for the action-angle representation of stable fixed point dynamics. Physica D: Nonlinear Phenomena **261**, 19–30 (2013)
17. Guckenheimer, J., Holmes, P.: Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields. Springer, New York (1983)
18. Mircheski, P., Zhu, J., Nakao, H.: Phase-amplitude reduction and optimal phase locking of collectively oscillating networks. Chaos Interdisc. J. Nonlinear Sci. **33**(10), 103111 (2023)
19. Nakao, H., Yamaguchi, K., Katayama, S., Yanagita, T.: Sparse optimization of mutual synchronization in collectively oscillating networks. Chaos Interdisc. J. Nonlinear Sci. **31**(6), 063113 (2021)
20. Nakao, H., Yasui, S., Ota, M., Arai, K., Kawamura, Y.: Phase reduction and synchronization of a network of coupled dynamical elements exhibiting collective oscillations. Chaos Interdisc. J. Nonlinear Sci. **28**(4), 045103 (2018)
21. Zlotnik, A., Chen, Y., Kiss, I., Tanaka, H.A., Li, J.-S.: Optimal waveform for fast entrainment of weakly forced nonlinear oscillators. Phys. Rev. Lett. **111**(2), 024102 (2013)
22. Tanaka, H.-A.: Optimal entrainment with smooth, pulse, and square signals inweakly forced nonlinear oscillators. Physica D: Nonlinear Phenomena **288**, 1–22 (2014)
23. Zlotnik, A., Nagao, R., Kiss, I.Z., Li, J.-S.: Phase-selective entrainment of nonlinear oscillator ensembles. Nat. Commun. **7**, 10788 (2016)
24. Kato, Y., Zlotnik, A., Li, J.-S., Nakao, H.: Optimization of periodic input waveforms for global entrainment of weakly forced limit-cycle oscillators. Nonlinear Dyn. **105**(3), 2247–2263 (2021)

Design of Artificial Limit-Cycle Oscillators with Global Synchronization Properties

Norihisa Namura(✉) and Hiroya Nakao

Department of Systems and Control Engineering, Tokyo Institute of Technology, Tokyo 152-8552, Japan
namura.n.aa@m.titech.ac.jp, nakao@sc.e.titech.ac.jp

Abstract. We design an artificial limit-cycle oscillator that can achieve global in-phase synchronization under linear mutual coupling. Using a method for designing the vector field of a dynamical system with a prescribed limit cycle that we proposed recently based on phase reduction theory, we design a limit-cycle oscillator with a complex trajectory but without high-harmonic components in the phase response properties, which yields a phase coupling function with a globally stable in-phase fixed point. We provide an example of the designed oscillator and demonstrate its global in-phase synchronization dynamics by numerical simulations.

Keywords: global synchronization · limit-cycle oscillators · phase reduction · phase response · optimization

1 Introduction

Synchronization of rhythmic systems is widely observed in the real world and often plays important functional roles [1]. Limit-cycle oscillators provide a fundamental mathematical model of stable periodic dynamics in rhythmic systems. Some examples of such rhythmic dynamics include circuit oscillations [2,3], animal gaits [4], and passive dynamic walking [5]. When a periodic input is given to a limit-cycle oscillator, entrainment or phase locking [6,7], where the oscillator synchronizes with the periodic input, can be observed. When two or more oscillators are coupled, they can exhibit mutual synchronization, in which their rhythms are aligned with each other.

The phase reduction is a useful method for analyzing synchronization phenomena of weakly coupled limit-cycle oscillators [6,8,9], which enables us to reduce the nonlinear multidimensional dynamics of a limit-cycle oscillator to a single-variable phase equation characterized by the natural frequency and phase response property of the oscillator. In engineering applications, synchronization with a specific phase relation is often required [10]. Therefore, it is important to ensure desired synchronization dynamics from arbitrary initial phase conditions.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 255–262, 2025.
https://doi.org/10.1007/978-3-031-72794-8_19

Recently, we proposed a method to design a limit-cycle oscillator with a desirable stable periodic trajectory and phase-response property based on phase reduction [11]. In this study, using our method, we design an artificial limit-cycle oscillator that does not possess high-harmonic components in the phase-response property and show that a pair of these designed oscillators can exhibit global in-phase synchronization when linearly coupled with each other.

2 Model of Mutual Synchronization

In this study, we focus on mutual synchronization of two oscillators. We consider a pair of identical limit-cycle oscillators with symmetric linear coupling, described by

$$\begin{aligned} \dot{\boldsymbol{X}}_1 &= \boldsymbol{F}(\boldsymbol{X}_1) + \varepsilon \boldsymbol{K}\boldsymbol{X}_2, \\ \dot{\boldsymbol{X}}_2 &= \boldsymbol{F}(\boldsymbol{X}_2) + \varepsilon \boldsymbol{K}\boldsymbol{X}_1, \end{aligned} \tag{1}$$

where $\boldsymbol{X}_{1,2}(t) \in \mathbb{R}^N$ is the state of each oscillator at time t, $\boldsymbol{F} : \mathbb{R}^N \to \mathbb{R}^N$ is a smooth vector field describing the dynamics of individual oscillators, $\dot{\boldsymbol{X}}$ is the time derivative of $\boldsymbol{X}$, and $\boldsymbol{K} \in \mathbb{R}^{N\times N}$ is a constant matrix that represents coupling coefficients between the vector components of the oscillator states. We assume that the individual oscillator described by $\dot{\boldsymbol{X}} = \boldsymbol{F}(\boldsymbol{X})$ has an exponentially stable limit-cycle solution $\tilde{\boldsymbol{X}}_0(t) = \tilde{\boldsymbol{X}}_0(t+T)$ with a period T and natural frequency $\omega = 2\pi/T$. The parameter ε is a sufficiently small positive value satisfying $0 < \varepsilon \ll 1$.

Here, we briefly explain the phase reduction theory [6,8,9]. We first introduce the asymptotic phase function in the basin $\mathcal{B}$ of the limit cycle where the basin is a region in the state space from which the oscillator state converges to the limit cycle. The asymptotic phase function $\Theta : \mathcal{B} \to [0, 2\pi)$ is defined to satisfy

$$\dot{\theta}(t) = \nabla\Theta(\boldsymbol{X}(t)) \cdot \boldsymbol{F}(\boldsymbol{X}(t)) = \omega, \tag{2}$$

where $\theta = \Theta(\boldsymbol{X})$ is the phase value of the state $\boldsymbol{X} \in \mathcal{B}$ of the oscillator. Here, the phase values 0 and 2π are considered identical. The state of the oscillator on the limit cycle can be expressed as $\boldsymbol{X}_0(\theta) = \tilde{\boldsymbol{X}}_0(t = \theta/\omega)$ as a function of the phase θ, where $\boldsymbol{X}_0$ is a function of period 2π (i.e., $\boldsymbol{X}_0(\theta) = \boldsymbol{X}_0(\theta + 2\pi)$).

If the perturbation applied to the oscillator is sufficiently weak and of $O(\varepsilon)$, the state of each oscillator can be approximated as

$$\boldsymbol{X}_{1,2}(t) = \boldsymbol{X}_0(\theta_{1,2}) + O(\varepsilon). \tag{3}$$

Substituting Eq. (3) into (1) and ignoring the errors of $O(\varepsilon^2)$, we can obtain the following phase equations:

$$\begin{aligned} \dot{\theta}_1 &= \omega + \varepsilon \boldsymbol{Z}(\theta_1) \cdot \boldsymbol{K}\boldsymbol{X}_0(\theta_2), \\ \dot{\theta}_2 &= \omega + \varepsilon \boldsymbol{Z}(\theta_2) \cdot \boldsymbol{K}\boldsymbol{X}_0(\theta_1), \end{aligned} \tag{4}$$

where $\boldsymbol{Z} : [0, 2\pi) \to \mathbb{R}^N$ is the phase sensitivity function (PSF) defined by the gradient of the phase function on the limit cycle as

$$\boldsymbol{Z}(\theta) = \nabla \Theta(\boldsymbol{X})|_{\boldsymbol{X}=\boldsymbol{X}_0(\theta)}. \tag{5}$$

The PSF represents linear response of the oscillator phase to a weak perturbation that is applied to the oscillator state near the limit cycle. It is represented as $\boldsymbol{Z}(\theta) = \tilde{\boldsymbol{Z}}(t = \theta/\omega)$, where $\tilde{\boldsymbol{Z}}$ can be obtained as a T-periodic solution to the following adjoint equation [12,13]:

$$\dot{\tilde{\boldsymbol{Z}}}(t) = -\boldsymbol{J}^\top(\tilde{\boldsymbol{X}}_0(t))\tilde{\boldsymbol{Z}}(t). \tag{6}$$

Here, $\boldsymbol{J}(\boldsymbol{X}) \in \mathbb{R}^{N\times N}$ is the Jacobian matrix of $\boldsymbol{F}$ at $\boldsymbol{X}$, i.e., $\boldsymbol{J}(\boldsymbol{X}) = \nabla \boldsymbol{F}(\boldsymbol{X})$. The PSF should satisfy $\tilde{\boldsymbol{Z}}(t) \cdot \boldsymbol{F}(\tilde{\boldsymbol{X}}_0(t)) = \omega$ as the normalization condition.

Mutual synchronization can be analyzed by using the averaging approximation [6,9,14]. Introducing the relative phase $\phi_{1,2}(t) = \theta_{1,2}(t) - \omega t$, we can rewrite the expression (4) as follows:

$$\begin{aligned} \dot{\phi}_1 &= \varepsilon \boldsymbol{Z}(\phi_1 + \omega t) \cdot \boldsymbol{K}\boldsymbol{X}_0(\phi_2 + \omega t), \\ \dot{\phi}_2 &= \varepsilon \boldsymbol{Z}(\phi_2 + \omega t) \cdot \boldsymbol{K}\boldsymbol{X}_0(\phi_1 + \omega t). \end{aligned} \tag{7}$$

Since the right-hand side of each equation is of $O(\varepsilon)$ and $\phi_{1,2}$ are slowly varying, we can average the right-hand side over one period of oscillation and obtain

$$\begin{aligned} \dot{\phi}_1 &= \varepsilon \Gamma(\phi_1 - \phi_2), \\ \dot{\phi}_2 &= \varepsilon \Gamma(\phi_2 - \phi_1), \end{aligned} \tag{8}$$

where Γ is the phase coupling function (PCF), which is a 2π-periodic function defined by

$$\Gamma(\varphi) = \langle \boldsymbol{Z}(\varphi + \psi) \cdot \boldsymbol{K}\boldsymbol{X}_0(\psi) \rangle_\psi. \tag{9}$$

Here, we denote the averaging of a function $f(\psi)$ over one period of oscillation as $\langle f(\psi) \rangle_\psi = \frac{1}{2\pi}\int_0^{2\pi} f(\psi) d\psi$. Introducing the phase difference $\varphi = \phi_1 - \phi_2$, we obtain the dynamics of the phase difference as

$$\dot{\varphi} = \varepsilon \Gamma_a(\varphi), \tag{10}$$

where $\Gamma_a(\varphi) = \Gamma(\varphi) - \Gamma(-\varphi)$ is the antisymmetric part of the PCF (APCF), which is 2π-periodic.

If the time derivative of the phase difference is zero (i.e., $\dot{\varphi} = 0$), the two oscillators are synchronized. The phase-locking point φ^* is stable if it satisfies the phase-locking condition (11) and stability condition (12):

$$\Gamma_a(\varphi^*) = 0, \tag{11}$$

$$\Gamma_a'(\varphi^*) < 0, \tag{12}$$

respectively, where $\Gamma_a'(\varphi^*)$ is the derivative of $\Gamma_a(\varphi)$ at φ^*. In the case of in-phase synchronization (i.e., $\varphi^* = 0$), the phase-locking condition is obviously satisfied from the antisymmetric property of the APCF.

3 Design of Stable Limit-Cycle Oscillators with Desired Periodic Trajectories and PSFs

Next, we briefly explain a method for designing stable two-dimensional limit-cycle oscillators with desired periodic trajectories and PSFs proposed in [11].

3.1 Conditions for the Periodic Trajectory and PSF

To design the oscillator dynamics with a prescribed non-intersecting periodic trajectory $\boldsymbol{p}(t) = [p_1(t)\ p_2(t)]^\top$, we first approximate the vector field of the oscillator by polynomials up to nth degree as

$$\boldsymbol{F}(\boldsymbol{X}) = \begin{bmatrix} F_1(\boldsymbol{X}) \\ F_2(\boldsymbol{X}) \end{bmatrix} \simeq \begin{bmatrix} \boldsymbol{U}^\top(\boldsymbol{X})\boldsymbol{\zeta}_1 \\ \boldsymbol{U}^\top(\boldsymbol{X})\boldsymbol{\zeta}_2 \end{bmatrix}, \tag{13}$$

where $\boldsymbol{X} = [x_1\ x_2]^\top$,

$$\boldsymbol{U}(\boldsymbol{X}) = \left[1\ \ \overline{x_1}\ \ \overline{x_2}\ \ \overline{x_1^2}\ \ \overline{x_1 x_2}\ \ \overline{x_2^2}\ \ \cdots\ \ \overline{x_2^n}\right]^\top \in \mathbb{R}^P \tag{14}$$

is a vector of standardized monomials, and $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2 \in \mathbb{R}^P$ are the coefficient vectors. Here, to adjust the scale of each monomial, we employ standardization denoted by the overline, $\overline{z} \coloneqq (z - \mu)/\sigma$, where μ and σ are the mean and the standard deviation of a variable z, respectively. These μ and σ are evaluated from each monomial composed of $\{\boldsymbol{p}(t_k)\}_{k=1}^L$, where $t_k = (k-1)\Delta t$ $(k = 1, \ldots, L)$ is the discretized time t with $\Delta t = T/L$.

We first impose a condition

$$\dot{\boldsymbol{p}}(t) = \boldsymbol{F}(\boldsymbol{p}(t)) \tag{15}$$

on the vector field. Introducing a coefficient vector $\boldsymbol{\xi} = \left[\boldsymbol{\zeta}_1^\top\ \boldsymbol{\zeta}_2^\top\right]^\top \in \mathbb{R}^{2P}$, we can express the difference between both sides of Eq. (15) at each time t_k under the polynomial approximation as

$$\boldsymbol{A}_{\boldsymbol{p},k}\boldsymbol{\xi} - \boldsymbol{b}_{\boldsymbol{p},k}, \tag{16}$$

where

$$\boldsymbol{A}_{\boldsymbol{p},k} = \begin{bmatrix} \boldsymbol{U}^\top(\boldsymbol{p}(t_k)) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}^\top(\boldsymbol{p}(t_k)) \end{bmatrix}, \quad \boldsymbol{b}_{\boldsymbol{p},k} = \begin{bmatrix} \dot{p}_1(t_k) \\ \dot{p}_2(t_k) \end{bmatrix}. \tag{17}$$

We next introduce the condition for the PSF into the vector field. We require the adjoint equation (6) to be satisfied so that the oscillator has a prescribed PSF $\tilde{\boldsymbol{Z}}(t) = [\tilde{Z}_1(t)\ \tilde{Z}_2(t)]^\top$. The adjoint equation at each time t_k can be expressed as

$$\begin{aligned} \tilde{Z}_1(t_k)\boldsymbol{U}_1^\top(\boldsymbol{p}(t_k))\boldsymbol{\zeta}_1 + \tilde{Z}_2(t_k)\boldsymbol{U}_1^\top(\boldsymbol{p}(t_k))\boldsymbol{\zeta}_2 &\simeq -\dot{\tilde{Z}}_1(t_k), \\ \tilde{Z}_1(t_k)\boldsymbol{U}_2^\top(\boldsymbol{p}(t_k))\boldsymbol{\zeta}_1 + \tilde{Z}_2(t_k)\boldsymbol{U}_2^\top(\boldsymbol{p}(t_k))\boldsymbol{\zeta}_2 &\simeq -\dot{\tilde{Z}}_2(t_k) \end{aligned} \tag{18}$$

under the polynomial approximation, where $\boldsymbol{U}_i(\boldsymbol{X}) = (\nabla \boldsymbol{U}(\boldsymbol{X}))_i$. As with the condition for the periodic trajectory, we can express the difference between both sides of Eq. (6) as

$$\boldsymbol{A}_{\boldsymbol{Z},k}\boldsymbol{\xi} - \boldsymbol{b}_{\boldsymbol{Z},k}, \tag{19}$$

where

$$\boldsymbol{A}_{\boldsymbol{Z},k} = \begin{bmatrix} \tilde{Z}_1(t_k)\boldsymbol{U}_1^\top(\boldsymbol{p}(t_k)) & \tilde{Z}_2(t_k)\boldsymbol{U}_1^\top(\boldsymbol{p}(t_k)) \\ \tilde{Z}_1(t_k)\boldsymbol{U}_2^\top(\boldsymbol{p}(t_k)) & \tilde{Z}_2(t_k)\boldsymbol{U}_2^\top(\boldsymbol{p}(t_k)) \end{bmatrix}, \quad \boldsymbol{b}_{\boldsymbol{Z},k} = \begin{bmatrix} -\dot{\tilde{Z}}_1(t_k) \\ -\dot{\tilde{Z}}_2(t_k) \end{bmatrix}. \tag{20}$$

In summary, we can express (the half of) the sum of the squared errors on both conditions as

$$E = \frac{1}{2}\|\boldsymbol{A}\boldsymbol{\xi} - \boldsymbol{b}\|^2, \tag{21}$$

where the matrices $\boldsymbol{A}$ and $\boldsymbol{b}$ are constructed by concatenating the matrices $\boldsymbol{A}_{\boldsymbol{p},k}$, $\boldsymbol{A}_{\boldsymbol{Z},k}$ and the vectors $\boldsymbol{b}_{\boldsymbol{p},k}$, $\boldsymbol{b}_{\boldsymbol{Z},k}$ for all k, respectively.

3.2 Condition for the Linear Stability

We also impose the condition for the stability of the periodic trajectory on the vector field. From the Floquet theory [15], the two Floquet exponents λ_1 and λ_2 characterizing the linear stability of the periodic trajectory $\tilde{\boldsymbol{X}}_0(t)$ satisfy

$$\lambda_1 + \lambda_2 = \frac{1}{T}\int_0^T \mathrm{tr}(\boldsymbol{J}(\tilde{\boldsymbol{X}}_0(t)))dt, \tag{22}$$

where tr denotes the trace of a matrix. Since the two Floquet exponents of the two-dimensional limit-cycle oscillator are real and the first Floquet exponent satisfies $\lambda_1 = 0$ from the condition (15) for the periodic trajectory, we only need to evaluate $\lambda = \lambda_2$, which should be negative for the stability.

To ensure that the periodic trajectory is sufficiently stable, we impose the condition on λ by

$$\lambda \leq \lambda_{\mathrm{tol}} \tag{23}$$

for some maximum tolerance value $\lambda_{\mathrm{tol}} < 0$. Discretizing the time t into $\{t_k\}$, the condition above can be expressed as

$$\boldsymbol{C}\boldsymbol{\xi} \leq \lambda_{\mathrm{tol}} \tag{24}$$

under the polynomial approximation, where

$$\boldsymbol{C} = \frac{1}{L}\sum_{k=1}^{L}\left[\boldsymbol{U}_1^\top(\boldsymbol{p}(t_k))\ \boldsymbol{U}_2^\top(\boldsymbol{p}(t_k))\right]. \tag{25}$$

3.3 Optimization Problem

Summarizing the conditions for the periodic trajectory, PSF, and linear stability, we can finally formulate the convex optimization problem for the coefficient vector $\boldsymbol{\xi}$ as

$$\boldsymbol{\xi}^* = \underset{\boldsymbol{\xi}}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{A}\boldsymbol{\xi} - \boldsymbol{b}\|^2 + \gamma\|\boldsymbol{\xi}\|^2 \quad \text{s.t.} \quad \boldsymbol{C}\boldsymbol{\xi} \leq \lambda_{\text{tol}}, \tag{26}$$

where $\gamma \geq 0$ is the weight of the regularization term $\|\boldsymbol{\xi}\|^2$ for preventing the vector field from becoming unnecessarily large.

In numerical implementation, we first determine the smallest maximum degree that can realize the desired oscillator dynamics by trying $n = 2, 3, \cdots$ in order without regularization. We then find a reasonable regularization parameter γ for the determined n by grid search.

4 Design of an Oscillator for Global Synchronization

As an example, we designed the vector field of a stable limit-cycle oscillator with the following periodic trajectory $\boldsymbol{X}_0(\theta)$ and PSF $\boldsymbol{Z}(\theta)$:

$$\boldsymbol{X}_0(\theta) = \begin{bmatrix} \frac{1}{2}\cos(\theta) - \frac{1}{6}\cos(3\theta) \\ \frac{3}{2}\sin(\theta) - \frac{1}{6}\sin(3\theta) \end{bmatrix}, \quad \boldsymbol{Z}(\theta) = \begin{bmatrix} -\sin(\theta) \\ \cos(\theta) \end{bmatrix}, \tag{27}$$

where $\boldsymbol{X}_0(\theta)$ and $\boldsymbol{Z}(\theta)$ satisfy the normalization condition explained previously. The system state is assumed to rotate with a constant frequency $\omega = 1$, namely, $T = 2\pi$ (in this case, we have $\theta = \omega t = t$). By using our method, we could successfully design such a limit-cycle oscillator when $L = 1000$, $n = 9$, $\gamma = 10^{-3}$, and $\lambda_{\text{tol}} = -2$, where the results are shown in Fig. 1.

The APCF between a pair of linearly coupled oscillators possessing the above properties given by Eq. (1) with $\boldsymbol{K} = \boldsymbol{I}$ (identity matrix) can be calculated as

$$\Gamma_a(\varphi) = -2\sin(\varphi), \tag{28}$$

which is shown in Fig. 2(a). Because this APCF has only a single stable fixed point at $\varphi = 0$, it is expected that the oscillators globally converge to the in-phase synchronized state. We note here that the trajectory of the above oscillator is not simple and possesses high-harmonic components, which typically yields PSFs with high-harmonic components in general cases, leading to APCFs with multiple fixed points that hamper global in-phase synchronization. The above result shows that we can artificially design an oscillator with a complex trajectory with high-harmonic components, yet with a simple PSF without high-harmonic components.

We performed numerical simulations of the coupled system of the above designed oscillators with the coupling strength $\varepsilon = 10^{-2}$. The dynamics of the phase difference are shown in Fig. 2(b). We can confirm that the designed oscillator can indeed achieve global in-phase synchronization from different initial conditions.

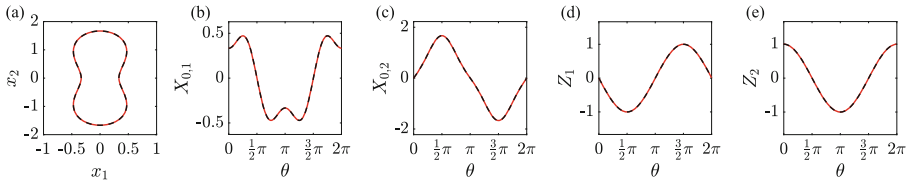


Fig. 1. (a) Limit-cycle trajectory. (b,c) Limit-cycle solution as a function of θ. (b) x_1 component, (c) x_2 component. (d,e) PSF without high-harmonic components. (d) x_1 component, (e) x_2 component. In each figure, the red line shows the designed functional form and the black line shows the original one, respectively.

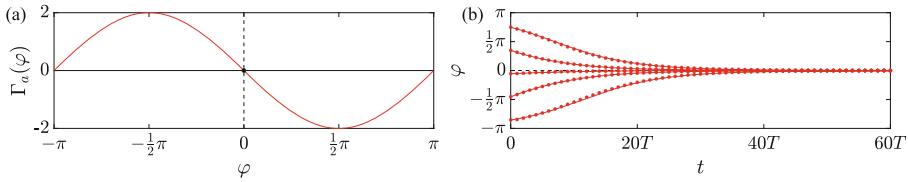


Fig. 2. (a) APCF with a globally stable in-phase fixed point. (b) Dynamics of the phase difference φ. The red solid lines represent the convergence of the phase difference φ described by Eq. (10) and the red dots represent the phase difference $\varphi = \theta_1 - \theta_2$ measured from directly-simulated $\boldsymbol{X}_{1,2}$ every cycle.

5 Concluding Remarks

We reviewed a method for designing limit-cycle oscillators with the desired periodic trajectory and phase response properties and designed a vector field possessing a stable limit cycle with a complex periodic trajectory but with a simple phase-response property. By numerical simulations, we demonstrated that global in-phase synchronization occurs when a pair of such designed oscillators are linearly and symmetrically coupled with each other. Although we considered a pair of identical oscillators in this article, global synchronization should also occur between nonidentical oscillators when the discrepancy between the designed oscillators is sufficiently small. Our results will be helpful in designing weakly coupled systems of limit-cycle oscillators that can achieve global synchronization.

References

1. Strogatz, S.H.: Nonlinear Dynamics and Chaos. CRC Press, Boca Raton (2015)
2. van der Pol, B., van der Mark, J.: Frequency demultiplication. Nature **120**(3019), 363–364 (1927)
3. van der Pol, B.: Forced oscillations in a circuit with non-linear resistance (Reception with reactive triode). London, Edinburgh, Dublin Philos. Mag. J. Sci. **3**(13), 65–80 (1927)

4. Collins, J.J., Stewart, I.N.: Coupled nonlinear oscillators and the symmetries of animal gaits. J. Nonlinear Sci. **3**(1), 349–392 (1993)
5. Garcia, M., Chatterjee, A., Ruina, A., Coleman, M.: The simplest walking model: stability, complexity, and scaling. J. Biomech. Eng. **120**(2), 281–288 (1998)
6. Kuramoto, Y.: Chemical Oscillations, Waves, and Turbulence. Springer, Berlin (1984)
7. Pikovsky, A., Rosenblum, M., Kurths, J.: Synchronization: A Universal Concept in Nonlinear Science. Cambridge University Press, Cambridge (2001)
8. Winfree, A.T.: The Geometry of Biological Time. Springer, New York (2001)
9. Nakao, H.: Phase reduction approach to synchronisation of nonlinear oscillators. Contemp. Phys. **57**(2), 188–214 (2016)
10. Dutta, S., et al.: Programmable coupled oscillators for synchronized locomotion. Nat. Commun. **10**(1), 3299 (2019)
11. Namura, N., Ishii, T., Nakao, H.: Designing 2-D limit-cycle oscillators with prescribed trajectories and phase-response characteristics. IEEE Trans. Autom. Control **69**(5), 3144–3157 (2024)
12. Brown, E., Moehlis, J., Holmes, P.: On the phase reduction and response dynamics of neural oscillator populations. Neural Comput. **16**(4), 673–715 (2004)
13. Ermentrout, G.B., Terman, D.H.: Mathematical Foundations of Neuroscience. Springer, New York (2010)
14. Hoppensteadt, F.C., Izhikevich, E.M.: Weakly Connected Neural Networks. Springer, New York (1997)
15. Guckenheimer, J., Holmes, P.: Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields. Springer, New York (1983)

Composite Materials and Soft Matter

Coarse-Graining for Bridging Spatio-Temporal Scales of Soft Matter Dynamics in Engineering Science

Itsuo Hanasaki(✉)

Tokyo University of Agriculture and Technology, Tokyo, Japan
hanasaki@cc.tuat.ac.jp

Abstract. Depending on the degree of nonlinearity in dynamical systems, the consequent dynamics can be dominated by the randomness due to chaos. This type of randomness is different from the external noise in measurements. The symbolic phenomenon is the well-known Brownian motion. The nonlinearity of physical characteristics can lead to the hierarchy of apparent physical behavior depending on the spatio-temporal scales of observation. The typical example is the suitable theoretical description on the basis of molecular discrete picture and continuum one. In this article, we would like to introduce examples of such situations by soft matters ranging from liquid crystals to nanopapers consisting of cellulose nanofibers (CNFs). The nanopaper is a transparent paper made of CNFs, and it is fabricated by drying aqueous dispersion of CNFs. The drying process undergoing gelation is experimentally analyzed by microscopy movie data analysis. The deformation characteristics of the whole specimen is designed by mechanical metamaterial approach. The former is based on the analysis of Brownian motion originating from the molecular nature of the system. The latter is based on the continuum mechanics, and the large deformation behavior makes use of bifurcation phenomenon. These distinctive approaches share the concept of connection between the parts and the whole, while the former is stochastic, and the latter is deterministic in the typical treatments. We would like to point out some important differences of existing studies of nonlinear dynamical systems and the applications of statistical mechanics for soft matter systems mainly originating from the stochasticity of the latter in the final section. Thereby, we propose and share the idea of the frontier to be explored rather than pedagogically explain the details of completed outcomes. This contribution is not a general review, but a perspective based on the authors' experience.

Keywords: Coarse Graining · Statistical Mechanics · Microscopy · Metamaterial

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 265–280, 2025.
https://doi.org/10.1007/978-3-031-72794-8_20

1 Coarse Graining in Simulation, Formulation, Data Analysis, and Structural Design

1.1 Atomistic and Molecular Picture with Simulation and Formulation

Atomistic or molecular motion can be described by the deterministic equations of motion for the whole system of interest. Therefore, molecular dynamics (MD) simulations have long been one of the valuable numerical analysis methods for atomistic and molecular systems [1]. The typical time resolution for the all-atom MD is in the order of a femtosecond i.e., 10^{-15} s [2,3]. The Newton's equation of motion for every atom in the system is smoothly solved to obtain the dynamics:

$$m_i \frac{\mathrm{d}^2 \boldsymbol{r}_i}{\mathrm{d}t^2} = \boldsymbol{f}_i \tag{1}$$

where t is the time and m_i, $\boldsymbol{r}_i$, and $\boldsymbol{f}_i$ are mass, position and force on i-th atom, respectively. The conservative force is derived from the potential energy determined by the atomic positions in the system of interest. When looking at fluid at a lower resolution, the Brownian motion of suspended tiny particles or fibers are described by the Langevin equation:

$$M_i \frac{\mathrm{d}^2 \boldsymbol{R}_i}{\mathrm{d}t^2} = \boldsymbol{f}_\mathrm{C} + \boldsymbol{f}_\mathrm{D} + \boldsymbol{f}_\mathrm{R}, \tag{2}$$

where M_i and $\boldsymbol{R}_i$ are mass and position of the particle of interest, and $\boldsymbol{f}_\mathrm{C}$, $\boldsymbol{f}_\mathrm{D}$, and $\boldsymbol{f}_\mathrm{R}$ are conservative force, dissipative force, and random force, respectively. The whole conservative force $\boldsymbol{f}_i$ consisting of the every atom in the system is split into the three components. The dissipative and random forces describes the interaction with the surrounding fluid, and related with each other by the fluctuation-dissipation theorem. Here, the Hamiltonian of the whole system is "coarse-grained" in the formulation [4]. Coarse graining is a technical term typically in statistical mechanics/physics and it means the reduction of description resolution from finer level to the coarser one while keeping essential properties. The decoupling of the particle/fiber of interest and the surrounding molecules is based on the different characteristic time scales [5]. In other words, departure from the linear Langevin equation is observed when these time scales are not so far from each other [6,7].

In particular, the difference of molecular and macroscopic drag emerges, which is directly related to the departure from Stokes-Einstein equation:

$$D = \frac{k_\mathrm{B} T}{3\pi\eta d_\mathrm{p}}, \tag{3}$$

where D and d_p are diffusion coefficient and the diameter of a particle of interest, and k_B is the Boltzmann's constant, T and η, are absolute temperature and viscosity of the surrounding fluid, respectively. The diffusion coefficient can be derived from the trajectory data of the particle of interest as follows:

$$D = \lim_{t\to\infty} \frac{1}{2n_\mathrm{d} t} \langle |\boldsymbol{R}(t) - \boldsymbol{R}(0)|^2 \rangle, \tag{4}$$

where $\langle|\boldsymbol{R}(t) - \boldsymbol{R}(0)|^2\rangle$ is the mean squared displacement (MSD) and n_d is the dimension of position vector $\boldsymbol{R}$ in terms of observation. In the regime of the linear Langevin equation, the diffusion coefficient is independent of the mass of the particle of interest, but it depends on the size of the particle besides the viscosity and temperature of the surrounding fluid. However, when the particle mass is rather close to each of the surrounding fluid molecules, the particle mass also slightly affects the diffusion coefficient [6]. This is related to the characteristics of friction coefficients in the Langevin equation. This friction coefficient corresponds to the drag in the macroscopic fluid dynamics. The macroscopic drag does not depend on the mass of the object but what matters is the size and shape. In contrast, the friction coefficient of a particle with a mass close to the surrounding fluid molecule can exhibit the difference of friction coefficients evaluated by the constant force condition and constant velocity condition [7].

The constant velocity condition is nontrivial to realize in experiment since it requires the firm fixture in a space or extremely fast feedback of the applied force. The constant force condition means the fluctuating particle velocity due to the thermal fluctuation. Based on the all-atom MD simulations, a Cl^- ion shows lower friction coefficient for constant force condition compared to the constant velocity condition [7]. The origin of this difference is confirmed by the force autocorrelation function in terms of time. Since the deviation is small while significant, the approximation by the Langevin equation is useful. However, if this small deviation is rather valuable, the approximation drops it. The mass dependence of friction coefficient or molecular drag is relevant to reduction of energy dissipation in the fluid dynamics. For example, there seems to be the room to exploit this molecular drag characteristics in the surface coating of the ships or planes.

MD provides huge amount of information as phase space trajectories. The evaluation of essential characteristics is a typical statistical mechanics. The drawback of fine resolution is the huge computational cost. Therefore, the quest of coarse graining is driven by the necessity of information amount reduction. It is still unrealistic to perform all-atom molecular dynamics (MD) simulation of macroscopic spatio-temporal scale today. The coarse-grained MD (CGMD) is therefore a straightforward option in theory [8,9]. The CGMD is a framework where the interaction potential is not of atomistic resolution but based on the interaction sites that represent multiple atoms. The rigorous governing equation corresponds to the Langevin equation, and the interaction potential is a kind of free energy, which is sometimes called the potential of mean force. Thereby, the resolution is lower and the computational cost is lower.

However, the reduction of the resolution is not an obvious procedure since it is usually impossible for the coarse-grained model to retain all the detailed characteristics of those with finer resolution. In particular, the smectic liquid crystals are sensitive to the details of force field. Therefore, one has to solve the multi-objective optimization problem with a suitable definition and procedure [10]. Once the sufficient quality of the force field is derived, the larger spatio-temporal scales beyond the reach of all-atom MD is realized. As far as the

molecular resolution is kept, the possible computational speed is limited. The dynamics of soft matter requires even longer time scale compared to the hard material besides the space scale. The obtained knowledge from CGMD can be employed for the basis of analytical scaling model, where sometimes renormalization group theory plays an important role [11].

There is a hierarchy of the description from equations of motion for every atom, the linear Langevin equation, and the continuum description of the whole system. The spatio-temporal scales and resolutions typically shows larger gaps between continuum picture and the Langevin equation or CGMD. Unless the thermal fluctuation is important, the system is described by the continuum picture. Otherwise, it is necessary to retain the information of thermal fluctuation to understand the dynamics of the system. Where there is a significant role of thermal fluctuation, the system is less dominated by mass or inertia, but the damping effect is dominant from macroscopic viewpoint. For example, the thermal fluctuation induces Brownian motion, and the Brownian motion leads to the diffusion of the system of interest. This makes contrast with the fluid dynamics with significant Reynolds numbers of macroscopic systems. In particlular, even if the Stokes number as well as the Reynolds number is small, small Péclet number causes the departure of suspended particle dynamics from the surrounding fluid at the inlet of constriction channel of which the width or height is comparable to the size of the particle of interest [12]. This phenomenon is experimentally observed in the situation of Lab on a Chip development [13]. Such phenomena cannot be explained or predicted if one assumes macroscopic equations without thermal fluctuation.

Besides the above-mentioned example, the stochastic nature of the dynamics sometimes shows counter-intuitive characteristics. For example, the escape time τ_{TST} of a tiny particle from trapping force field is usually estimated by the height of the potential barrier [14–16]

$$\tau_{\mathrm{TST}} = \tau_0 \exp\left(\frac{\Delta E}{k_{\mathrm{B}}T}\right), \tag{5}$$

where ΔE is the height of the potential barrier, and τ_0 is often derived from the diffusion coefficient D and the domain width d_{s} of the potential as

$$\tau_0 = \frac{d_{\mathrm{s}}^2}{8D}. \tag{6}$$

The trapping potential has not only height but also shape. The evaluation of escape time solely by the height of the potential appears as if it does not depend on the shape. However, it actually depends on the shape of the potential. The shape of the potential is defined by the generalized spring constant k_{s} as follows [17]:

$$k_{\mathrm{s}} = \frac{\alpha_{\mathrm{s}} k_{\mathrm{B}} T}{(d_{\mathrm{s}}/2)^{n_{\mathrm{s}}}}, \tag{7}$$

where α_{s} tunes the height of the potential while n_{s} can vary the steepness of the potential. Using this generalized spring constant, the potential barrier is

described as

$$\Delta E_n(r) = k_s r^n \tag{8}$$

for the domain with a width of d_s. Thereby, the shape is independent of the height of the potential. Then, the softer potential traps the Brownian particle for a longer time [17]. This is an important physical law when trying to trap a particle using the principles of optical tweezer [18], electrostatic [19], dielectrophoretic [20], and Paul trap [21–23], where there is always a technical limitation on the height of the trapping potential.

The trapping potential is experimentally evaluated by the position distribution of the trapped particles [24]. The mapping from position probability distribution into the potential of mean force is a link between the geometric data and the thermodynamics. Although the effect of mass is not typically dominant and the stochasticity hinders differentiation with respect to time, the link between the geometric information and the mechanics is provided by the fact that the thermal fluctuation is in the order of $k_B T$.

1.2 Big Data Analysis of Measurement Data from Microscopy

The above-mentioned scenario is mainly based on the theoretical work with computational and analytical modelling. MD is a direct simulation of statistical mechanics, and the interaction potential is the origin of physical characteristics, which is obtained by the evaluation of physical quantities from the sampled trajectories in phase space. As already mentioned, the advantage of high spatiotemporal resolution also means the drawback of high computational cost. An alternative approach is the direct sampling of trajectories, i.e., the microscopy movie data analysis, typically based on the single particle tracking (SPT). SPT is a versatile tool to understand the system of interest through microscopy observation. The standard algorithms are now established [25,26] while there is a room for sophistication. The tracking of agent of interest provides the trajectory information. In the case where special molecules in a biological cell or tissue is of interest, the fluorescent label molecules are attached to them. On the other hand, there are also cases where such labels are not available, and probe particles with observable sizes can be used for understanding the microscopic environment surrounding such probe particles.

Applications of SPT ranges from biology and medicine to physics and material science. Microscopy is originally a means to observe, but the objective understanding usually requires quantitative evaluation, i.e., measurement and analysis. The analysis of measurement data is more advantageous when the system of interest is hard to idealize or simplify for modelling. For example, a kind of transparent papers, "nanopaper" [27–30] is fabricated from cellulose nanofibers (CNFs) [31] by drying the aqueous dispersion of CNFs. CNFs are basically environmentally friendly materials, and nanopapers are promising for substrates of flexible devices [32–38]. The drying process of aqueous CNF dispersion affects the nanopaper properties from the same CNFs [39]. The rheology of the dispersion can be evaluated through the characteristics of Brownian motion of probe particles [40] even though CNFs themselves are too small to be directly observed by

optical microscope. The Brownian particles in the aqueous CNF dispersion show different characteristics in the random dynamics. The framework of the generalized diffusion [41] is useful in the microrheological evaluation of the media where probe particles explore:

$$\langle |\boldsymbol{R}(t) - \boldsymbol{R}(0)|^2 \rangle \approx 2n_{\mathrm{d}} D_\alpha t^\alpha, \tag{9}$$

where D_α is the generalized diffusion coefficient and α is the scaling exponent. When the material of interest has simply higher viscosity, the diffusion coefficient is lower. The quantitative correspondence is known as the Stokes-Einstein relation (cf. Eq. 3) as far as $\alpha = 1$ known as normal diffusion. When the material of interest is not a simple liquid but structural order exists in the liquid environment, the probe particles show subdiffusive behaviour corresponding to $\alpha < 1$. There are many other characteristics that helps understanding of the wet soft matter systems by SPT [40]. In the evaluation of experimental or numerical data, the evaluation of the generalized diffusion coefficient can be based on either frame-based (FB) average

$$D_{\mathrm{FB}\alpha} = \frac{1}{2n_{\mathrm{d}} \Delta t^\alpha \sum_{i=1}^{N_{\mathrm{I}}} (N_{\mathrm{F}i} - 1)} \sum_{i=1}^{N_{\mathrm{I}}} \sum_{j=1}^{N_{\mathrm{F}i}-1} |\boldsymbol{r}_i(t_{j+1}) - \boldsymbol{r}_i(t_j)|^2, \tag{10}$$

or inidividual-based (IB) average

$$D_{\mathrm{IB}\alpha} = \frac{1}{N_{\mathrm{I}}} \sum_{i=1}^{N_{\mathrm{I}}} \left[\frac{1}{2n_{\mathrm{d}} \Delta t^\alpha (N_{\mathrm{F}i} - 1)} \sum_{j=1}^{N_{\mathrm{F}i}-1} |\boldsymbol{r}_i(t_{j+1}) - \boldsymbol{r}_i(t_j)|^2 \right], \tag{11}$$

where $N_{\mathrm{F}i}$ is the number of consecutively tracked time steps for i-th particle, N_{I} is the number of individual particles (i.e., different trajectories), and $\boldsymbol{r}_i(t_j)$ indicates the j-th time step of the trajectory of i-th particle.

This microrheological approach is by far less invasive compared to the conventional shear-based rheometers. Therefore, the drying process of CNF dispersion can be analyzed in-process [42]. The drying process of aqueous CNF dispersion causes gelation from the sol state and then finally the dry film known as nanopaper is formed. This is a nonequilibrium phenomenon, and a higher initial CNF concentration does not merely mean the reduction of drying process [42]. The drying in a typical container at a typical thermodynamic condition proceeds at substantially slower time scale compared to that of the analysis of the generalized diffusion. Therefore, the generalized diffusion of probe particles is evaluated from a movie data of e.g., ten seconds consisting of a hundred frames, which is approximately regarded as in equilibrium.

Thus, the drying-induced gelation of the aqueous CNF dispersion is a phenomenon of multiple time scales of interest. This phenomenon is useful for engineering purpose besides the fundamental process of nanopaper fabrication. The inkjet printing requires sufficiently low viscosity of the ink that contains pigment particles so that the ink is successfully discharged from the nozzle of the printer.

On the other hand, the drying sessile colloidal droplet tends to exhibit so-called coffee-ring phenomenon when the droplet liquid has a low viscosity. The particles in a droplet are transported to the periphery of the droplet, leading to the ring structure that reminds us of coffee stains. One of the solutions is the addition of a small amount of CNFs in the ink [43,44]. When the ink is discharged, the viscosity is sufficiently low due to the low CNF concentration. The radial transport of particles to the droplet periphery is most activated at the late stage of drying process, which is sometimes called a rush hour [45]. The late stage of drying corresponds to the high concentration of CNFs where CNFs form network structure that hinders the flow transport of particles into the periphery. Thus, it is a 4D method designed with time as a parameter besides the 3D microscopic space of the droplet.

Since the small amount of CNFs can suppress the coffee-ring effect, it is rather counterintuitive when observing the intermittent and random flow velocity fields in the drying process of aqueous CNF dispersion in a container [46]. The intermittent dynamics originates from the intermittent displacements of air-water-sidewall interface in the drying process. In addition, random velocity field is observed in the gelation process. The intermediate state between the sol and gel is complex and the mechanism of gelation remains to be fully addressed. Furthermore, nanopaper shows gelation behaviour even when it is soaked in water, i.e., the nanopaper shows swelling-based gelation [47].

So far, the CNFs in water is introduced as an example of soft matter systems where multiple spatio-temporal scales are of great interest. The SPT is shown as a very useful means to analyze the wet system. In theory, a sufficiently long trajectory of random motion can be understood in terms of the large deviation principle [48–50], beyond the realm of the central limit theorem. When a finite time mean $\bar{u}(\tau)$ for a period τ of time-series data is of interest, the scaled cumulant-generating function $\phi(q)$ with a real-number parameter q [51] is defined by

$$\phi(q) = \lim_{\tau \to \infty} \frac{1}{\tau} \ln \langle \exp[q\tau \bar{u}(\tau)] \rangle. \tag{12}$$

$\phi(q)$ is related to the large deviation function $S(u)$ via

$$S(u(q)) = qu(q) - \phi(q) \tag{13}$$

for strictly convex and differentiable $\phi(q)$ and generally

$$S(u(q)) = \sup_q [qu(q) - \phi(q)], \tag{14}$$

where $S(u)$ is defined as

$$S(u) = -\lim_{\tau \to \infty} \frac{1}{\tau} \ln P(\bar{u} \in [u, u + du]), \tag{15}$$

and $P(\bar{u} \in [u, u + du])$ indicates the probability of $\bar{u}$ taking the value between u and $u + du$. The rate function $S(u)$ means the rate of convergence to the most

probable value with increasing samples. Characteristics of randomness in trajectories can include the information of the anisotropy of the particle of interest [48] and motility of bacteria [49]. The same diffusion coefficient D can consist of different nature of the elementary process [50].

On the other hand, typical microscopy movie data provide short trajectories while the number of samples can be increased. There are methodologies to evaluate the specific characteristics of the dynamics there. For example, multimodal diffusion can be detected by simple theoretical idea [52]. There are many situations where multiple diffusion modes coexist in a system. For example, biomolecules in the vicinity of glass substrate shows diffusion affected by the wall and there are some molecules trapped there. It is nontrivial to distinguish the influence by the trapping when the trapped molecules also show diffusion without completely fixed on the spots. The overlapping of the two distribution of the displacements hinders splitting. However, it is possible to detect the bimodal distribution and even the diffusion coefficients themselves by evaluation of the distribution of the logarithm of the squared displacements [53]. There are two peaks in the distribution for the logarithm whereas single peak is noticeable in the original displacement distribution. This simple idea of the logarithmic measure performs compatible with machine-learning based approach [52].

The potential applications of the logarithmic measure for the evaluation of multimodal diffusion are infinite, but we introduce some more examples here. When using micrometer-sized particles for probe of microrheological analysis of aqueous CNF dispersion, the combination of particle diameter and CNF concentration can lead to bimodal diffusion of the probe particles. This is because some particles are trapped by the network structure of the CNFs while some particles diffuse without getting trapped. The detection of bimodal diffusion can be used as the hallmark of partial structural order of the CNFs between the sol and gel states [54]. The logarithmic measure can also be used for collective order formation of the spindle [52] from the similar viewpoint. Furthermore, the collective dynamics of multiple particles trapped by the optical tweezer can be found by the logarithmic measure [55]. In this case, the circular alignment of the particles shows intermittent collective yet random motion besides the pure thermal fluctuation. This collective random motion is attributed to part of the optical force.

The rich information of random motion in a high-concentration systems can be resolved even without the tracking. The field of diffusion coefficient can be evaluated by particle image diffusometry (PID) [56,57]. Historically, the Brownian motion of tiny objects far below the wavelength of visible light has long been analyzed by the dynamic light scattering (DLS) [58]. Based on the Stokes-Einstein relation (cf. Eq. 3), diffusion coefficient is converted to the effective size of the object. However, the DLS cannot resolve the nonuniform field of diffusion coefficients but provides a representative value. On the other hand, SPT requires that the objects of interest are can be tracked. This is not the case when the objects are far below the wavelength of visible light and also it is not possible to label with fluorescent dye molecules. Furthermore, even if the fluorescent

labelling is possible, the objects cannot be tracked when they are located with too high concentrations.

In contrast to the history of the experimental evaluation of Brownian motion, there is a long history of particle image velocimetry (PIV), where the quantity of interest is not a diffusion or Brownian motion but the velocity of the flow field. PIV includes the methods based on the particle tracking just like SPT, and there is another well-known approach based on the correlation matching between the regions of interest (ROI). ROI is a subunit of an image captured in time series, and the matching between the time-sequential images provides the displacement of the ROI per frame interval. The PIV without particle tracking yields the flow velocity field, but they are restricted to the evaluation of velocity. In fact, the Brownian motion of probe particles has long been regarded as unnecessary "noise" in the context of PIV. However, it should be noted here that the Brownian motion originates from the system of interest itself, and hence it is not an external noise.

PID can evaluate the field of diffusion coefficients without tracking of particles even when the particles are highly concentrated [57]. The PID has successfully evaluated the intermediate state between the liquid and solid before crystal nucleation in solution [57]. From the viewpoint of specific applications, molecular crystallization from aqueous solution is of vital importance in the pharmaceutical science and industry since the complex molecular structure of biomolecules are determined by X-ray analysis of crystals and specific forms of crystals are necessary for a pharmaceutical molecules to function as expected. Crystallization is a symbolic example of high concentration, and there are so many situations in the soft matter physics and engineering where labelling is impossible, or concentrations of particles are too high to be tracked from microscopy movie data.

Understanding of physical law is more activated when the experimental data are sufficiently obtained more easily than before. The term "big data" is partly due to this basic trend. However, there is a stark difference when looking into the big data of mesoscale between atomistic and continuum. Not only the precision and resolution but also the amount of data is of essential importance. In other words, precision and resolution cannot replace the amount of data because of the stochasticity of the dynamics. The analysis requires some kinds of statistical approach. Microscopy movie data analysis is often a direct application of statistical mechanics just like the MD simulation is. The point of these microscopy movie data analysis for random motion is that they are in a sense the coarse graining from the detailed information of the big data.

1.3 Deformation of Continuum with Bifurcation and Beyond

The nanopaper itself is recognized as continuum, whereas the CNF dispersion is analysed from the basis of molecular picture of Brownian motion. The bridging of scales is realized by the microscopy movie data analysis. Furthermore, even within the picture of continuum mechanics, the design of unit structure can play an important role in the functionality of the whole system. The concept of mechanical metamaterial corresponds to this idea (e.g., [59–68]). The

stretchability of nanopaper is endowed by the periodic Kirigami structure [68], where buckling of the unit structure is employed for the conversion of bending of units into stretching of the whole structure. The buckling is a kind of bifurcation phenomena, and bifurcation is one of the typical situations where nonlinear dynamics is used for the engineering functionality of mechanical systems.

The bifurcation phenomena typically manifested as buckling is today an intensive target of functionality for mechanical metamaterials [59]. The Origami and Kirigami structures for sheet materials are the typical examples of such design strategies. The order of the effective number of variables in a system is affected by how firmly the constituent parts are related with each other. For example, solids and liquids with continuum description originally corresponds to infinite degrees of freedom, but the model description as the dynamical systems often reduces to drastically small number of degrees of freedom. The same solid is described with lots of degrees of freedom when the structure is designed as such [69].

The discussion of mechanical metamaterial is mainly based on the continuum description of materials where the top-down approach of processing/manufacturing is sufficiently feasible. The bottom-up approach is more focused on the characteristics of raw materials based on the self-assembly [24]. The above-mentioned approach for analysis of the nanopaper formation from aqueous CNF dispersion is suitable for the latter. The buckling itself is well-defined for continuum beam with simple characteristics of raw material. On the other hand, the buckling-like phenomena where undulation structures emerge are not always obvious when the molecular characteristics are directly related to the structural characteristics of membrane [9]. In such cases, not only microscopy movie data analysis but also the CGMD is a useful and reasonable means to understand and predict the phenomena and mechanism.

2 Perspective of Link Between the Behavior of Soft Matter and Nonlinear Dynamical Systems

The bridging across different spatio-temporal scales is one of the typical interests of statistical mechanics, where the coarse graining is a technical term of the corresponding conceptual approach. We have introduced some of the recent examples from our studies in the previous sections. One thing to note here is that there is a necessary decision of problem definition in the coarse graining. Namely, the focus to retain the precise characteristics in the modelling or measurement has to be defined to solve the problem. For example, one evaluates mean and variance from a set of sampled data in statistics, and the mean does not tell the variance in general. The examples that we have introduced in this article is a kind of soft mater systems, where statistical mechanics is already recognized as useful means to understand and design the system of interest. From fundamental theoretical point of view, statistical mechanics and dynamical systems has smooth connection. The well-established book of nonlinear dynamical

systems [70] had pointed out the frontier is in the increased number of variables to approach the continuum limit.

However, the link between the statistical mechanics of soft matter and dynamical systems theory is still relatively scarce. As the focus on unstable periodic orbit shed light in the studies of turbulence, we expect much more developments and link to the nonlinear dynamical systems theory for understanding and control of soft matter systems in view of engineering applications as concrete examples. The most fundamental difference of the above-mentioned systems from the typical existing case studies is the lack of smoothness in the trajectories of obtained data at the mesoscale between molecular and continuum picture, as is usual in the SPT data. The existing valuable theories and theorems of nonlinear dynamical systems to understand complex dynamics often assume the smoothness of the dynamics so that the target data are sufficiently differentiable with respect to time, which makes contrast with the stochasticity of mesoscale description.

Nevertheless, the starting point of full exploitation of microscopy movie data analysis in the above-mentioned example was the alternative means of MD in terms of sampling of phase space trajectories. Although the over-damped and random nature of the system leads to apparently the discussion on the trajectories of positions alone, the thermal fluctuations play vital role in the dynamics, and the analysis directly reflects this fact. The unit of normal diffusion coefficient is m^2/s whereas that of velocity is m/s. The generalized diffusion coefficient has a unit of m^2/s^α, where α is the scaling exponent for the time evolution of mean squared displacement. There seems to be some room to exploit theory of nonlinear dynamical systems here, while the measurement time resolution calls for the stochastic treatment. The link between geometric information and thermodynamics is mentioned above. Therefore, we consider that the application of the theoretical framework of stochastic thermodynamics [71] is promising while nontrivial.

Looking at the linear Langevin equation, the separation of time scales for inertial effect appears as if the thermal fluctuation is a noise for the slower dynamics. However, the advection-diffusion equation at the continuum picture is a coupling of different time scales at macroscopic space scale. The slow flow field in the drying process of a sol to form gel and dried film [46] might include intricate interplay between the stochastic nature with some slower dynamics. In either case, the system is fundamentally time dependent at least. One of the important differences of fully nonequilibrium dynamics with effectively thermal equilibrium system with stochastic dynamics is the path dependence of the consequent state [42]. The nonequilibrium nature such as path dependence can be the source of "memory" functionality in the materials [72]. Once the memory of the material is employed for functionality, the material behaviour will be smarter.

The above-mentioned examples of soft matter systems are under the physical and/or measurement conditions where the damping factor is dominant. On the other hand, the dynamics can be smooth even when the inertia does not play

significant role. At the continuum picture, soft robotics making use of elastomers and large deformations often realize smart response against input through the mechanics of materials by structural design and fabrication. The use of nonlinear dynamical systems theory in the soft matter dynamics seems promising for structural design of the mechanical metamaterials, where bifurcation phenomena are often employed for functionality. In the above-mentioned example of Kirigami-based mechanical metamaterial [68], the resilience against repeated deformation is accomplished by the simple design principle of stress defocusing, which has been well-known for avoiding fracture of rather brittle and hard materials. In fact, the existing reports of mechanical metamaterials made of planar parts has been focused rather on the geometric aspects of Origami as well as Kirigami. The full exploitation of material characteristics still remains a frontier.

The gap between molecular or Brownian regime and continuum one is sometimes employed as the room for distinctive designs for each regime to accomplish a desired functionality. The above-mentioned example to fabricate mechanical metamaterials made of CNFs consists of the designs of the accumulated structures of CNFs as the nanopaper and the Kirigami structures. The structural levels of CNFs and nanopaper is treated as distinctive problems of design while they are finally integrated. We would like to introduce another example where a static fluid structure is constructed in a fluid without resorting to viscosity or plasticity. While usual 3D printing technology assumes solid structure in space, design of ink consisting of colloidal particles in terms of Brownian dynamics and design of ink drawing motion in terms of macroscopic fluid dynamics enables the beverage 3D printing that sustains the structure for a party time [73]. The time scale as well as the space scale is essential when designing the dynamical system that ranges multiple scales of hierarchical descriptions.

References

1. Allen, M.P., Tildesley, D.J.: Computer Simulation of Liquids. Oxford Science Publications, Oxford (1986)
2. Hanasaki, I., Nakatani, A.: Hydrogen bond dynamics and microscopic structure of confined water inside carbon nanotubes. J. Chem. Phys. **124**, 174714 (2006)
3. Hanasaki, I., Nakatani, A.: Flow structure of water in carbon nanotubes: poiseuille type or plug-like? J. Chem. Phys. **124**, 144708 (2006)
4. Evans, D.J., Morriss, G.: Statistical Mechancis of Nonequilibrium Liquids. Cambridge University Press, Cambridge (2008)
5. Kubo, R., Toda, M., Hashitsume, N.: Statistical Physics II: Nonequilirbrium Stastistical Mechanics, 2nd edn. Springer, Heidelberg (1991)
6. Hanasaki, I., Nagura, R., Kawano, S.: Coarse-grained picture of Brownian motion in water: role of size and interaction distance range on the nature of randomness. J. Chem. Phys. **142**, 104301 (2015)
7. Hanasaki, I., Fujiwara, D., Kawano, S.: Departure of microscopic friction from macroscopic drag in molecular fluid dynamics. J. Chem. Phys. **144**, 094503 (2016)
8. Voth, G.A.: Coarse-Graining of Condensed Phase and Biomolecular Systems. CRC Press, Boca Raton (2009)

9. Hanasaki, I., Walther, J.H., Kawano, S., et al.: Coarse-grained molecular dynamics simulations of shear-induced instabilities of lipid bilayer membranes in water. Phys. Rev. E **82**, 051602 (2010)
10. Sakamoto, M., Hanasaki, I.: Derivation of coarse-grained force fields for bucklinginduced topological defects of liquid crystals. Phys. Rev. E **104**, 024704 (2021)
11. Sakamoto, M., Hanasaki, I.: Analytical model of critical buckling transition for smectic liquid crystal based on the viscoelastic scaling of coarse-grained molecular dynamics. Phys. Rev. E **107**, 014703 (2023)
12. Hanasaki, I., Walther, J.H.: Suspended particle transport through constriction channel with Brownian motion. Phys. Rev. E **96**, 023109 (2017)
13. Kazoe, Y., Mawatari, K., Kitamori, T.: Behavior of nanoparticles in extended nanospace measured by evanescent wave-based particle velocimetry. Anal. Chem. **87**, 4087 (2015)
14. Tanaka, Y., Yoshikawa, H., Masuhara, H.: Two-photon fluorescence spectroscopy of individually trapped pseudoisocyanine j-aggregates in aqueous solution. J. Phys. Chem. B **110**, 17906–17911 (2006)
15. Juan, M.L., Gordon, R., Pang, Y., et al.: Self-induced back-action optical trapping of dielectric nanoparticles. Nature Phys. **5**, 915–919 (2009)
16. Pelton, M., Liu, M., Kim, H.Y., et al.: Optical trapping and alignment of single gold nanorods by using plasmon resonances. Opt. Lett. **31**, 2075 (2006)
17. Hanasaki, I., Nemoto, T., Tanaka, Y.: Soft trapping lasts longer: dwell time of a Brownian particle varied by potential shape. Phys. Rev. E **99**, 022119 (2019)
18. Ashkin, A., Dziedzic, J.M., Bjorkholm, J.E., et al.: Observation of a single-beam gradient force optical trap for dielectric particles. Opt. Lett. **11**, 288–290 (1986)
19. Mojarad, N., Krishnan, M.: Measuring the size and charge of single nanoscale objects in solution using an electrostatic fluidic trap. Nat. Nanotechnol. **7**, 448–452 (2012)
20. Characterization of the stiffness of multiple particles trapped by dielectrophoretic tweezers in a microfluidic device. Langmuir **32**, 922–927 (2016)
21. Guan, W., Joseph, S., Park, J.H., et al.: Paul trapping of charged particles in aqueous solution. Proc Nat Acad Sci. **108**, 9326–9330 (2011)
22. Park, J.H., Krstic, P.S.: Thermal noise in aqueous quadrupole micro and nanotraps. Nanoscale Res. Lett. **7**, 156 (2012)
23. Park, J.H., Guan, W., Reed, M.A., et al.: Tunable aqueous virtual microscope. Small **8**, 907–912 (2012)
24. Hanasaki, I., Shoji, T., Tsuboi, Y.: Regular assembly of polymer nanoparticles by optical trapping enhanced with a random array of Si needles for reconfigurable photonic crysitals in liquid. ACS Appl. Nano Mater. **2**, 7637–7643 (2019)
25. Chenouard, N., Smal, I., de Chaumont, F., et al.: Objective comparison of particle tracking methods. Nat. Methods **11**(3), 281–289 (2014)
26. Sbalzarini, I.F., Koumoutsakos, P.: Feature point tracking and trajectory analysis for video imaging in cell biology. J. Struct. Biol. **151**(2), 182–195 (2005)
27. Wang, Q., Yao, Q., Liu, J., et al.: Processing nanocellulose to bulk materials: a review. Cellulose **26**, 7585–7617 (2019)
28. Gonzalez, I., Alcala, M., Chinga-Carrasco, G., et al.: From paper to nanopaper: evolution of mechanical and physical properties. Cellulose **212**, 2599–2609 (2014)
29. Barhoum, A., Samyn, P., Ohlund, T., et al.: Review of recent research on flexible multifunctional nanopapers. Nanoscale **9**, 15181–15205 (2017)
30. Benitez, A.J., Walther, A.: Cellulose nanofibril nanopapers and bioinspired nanocomposistes: a review to understand the mechanical perperty space. J. Mater. Chem. A **5**, 16003–16024 (2017)

31. Nechyporchuk, O., Belgacem, M.N., Bras, J.: Production of cellulose nanofibers: a review of recent advances. Ind. Crop. Prod. **93**, 2–25 (2016)
32. Russo, A., Ahn, B.Y., Adams, J.J., et al.: Pen-on-paper flexible electronics. Adv. Mater. **23**, 3426–3430 (2011)
33. Wang, L., Chen, D., Jiang, K., et al.: New insights and perspectives into biological materials for flexible electronics. Chem. Soc. Rev. **46**, 6764–6815 (2017)
34. Hsieh, M., Kim, C., Nogi, M., et al.: Electrically conductive lines on cellulose nanopaper for flexible electrical devices. Nanoscale **5**, 9289–9295 (2013)
35. Xu, X., Zhoua, J., Jiang, L., et al.: Highly transparent, low-haze, hybrid cellulose nanopaper as electrodes for flexible electronics. Nanoscale **9**, 12294–12306 (2016)
36. Jiang, Y., Liu, Z., Matsuhisa, N., et al.: Auxetic mechanical metamaterials to enhance sensitivity of stretchable strain sensors. Adv. Mater. **30**, 1706589 (2018)
37. Liao, C., Zhang, M., Yao, M.Y., et al.: Flexible organic electronics in biology: Materials and devices. Adv. Mater. **27**, 7493–7527 (2015)
38. Vosgueritchian, M., Lipomi, D.J., Bao, Z.: Highly conductive and transparent PEDOT:PSS films with a fluorosurfactant for stretchable and flexible transparent electrodes. Adv. Funct. Mater. **22**, 421–428 (2012)
39. Mizumura, D., Hanasaki, I., Ooi, Y., et al.: Improving the macroscopic uniformity of nanopaper by multi-step coating of cellulose nanofiber dispersion. Micro Nano Lett. **12**, 516–519 (2017)
40. Motohashi, R., Hanasaki, I.: Characterization of aqueous cellulose nanofiber dispersions from microscopy movie data of Brownian particles by trajectory analyss. Nanoscale Adv. **1**, 421 (2019)
41. Klages, R., Radons, G., Sokolov, I.M.: Anomalous transport: Foundations and applications. Wiley-VCH, Weinheim (2008)
42. Koitabashi, T., Hanasaki, I.: Drying path dependence in microrheological characteristics of cellulose nanofiber dispersion revealed by single particle tracking. J. Phys. D Appl. Phys. **54**, 295302 (2021)
43. Ooi, Y., Hanasaki, I., Mizumura, D., et al.: Suppresing the coffee-ring effect of colloidal droplets by dispersed cellulose nanofibers. Sci. Technol. Adv. Mater. **18**, 316–324 (2017)
44. Koyama, N., Hanasaki, I.: Spatio-temporally controlled suppression of the coffee-ring phenomenon by cellulose nanofibers. Soft Matter **17**, 4826–4833 (2021)
45. Marín, Á.G., Gelderblom, H., Lohse, D., et al.: Order-to-disorder transition in ringshaped colloidal stains. Phys. Rev. Lett. **107**, 085502 (2011)
46. Watanabe, H., Hanasaki, I.: Precursory flow in the formation of cellulose nanofiber films revealed by multiscale image analysis. Phys. Rev. Res. **5**, 023073 (2023)
47. Moriwaki, S., Hanasaki, I.: Swelling-based gelation of wet cellulose nanopaper evaluated by single particle tracking. Sci. Technol. Adv. Mater. **24**, 2153622 (2023)
48. Hanasaki, I., Isono, Y.: Detection of diffusion anisotropy due to particle asymmetry from single-particle tracking of brownian motion by the large-deviation principle. Phys. Rev. E **85**, 051134 (2012)
49. Hanasaki, I., Kawano, S.: Evaluation of bacterial motility from non-Gaussianity of finite-sample trajectories using the large deviation principle. J. Phys.: Condens. Matter **25**, 465103 (2013)
50. Hanasaki, I., Uehara, S., Kawano, S.: Role of time scales for the non-gaussianity of brownian motion combined with intermittent adsorption. J. Comp. Sci. **10**, 311–316 (2015)
51. Touchette, H.: The large deviation approach to statistical mechanics. Phys. Rep. **478**, 1–69 (2009)

52. Dalton, B.A., Sbalzarini, I.F., Hanasaki, I.: Fundamentals of the logarithmic measure for revealing multimodal diffusion. Biophys. J . **120**, 829–843 (2021)
53. Hanasaki, I., Uehara, S., Arai, Y., et al.: Threshold-free evaluation of near-surface diffusion and adsorption-dominated motion from single-molecule tracking data of single-stranded DNA through total internal reflection fluorescence microscopy. Jpn. J. Appl. Phys. **54**(12), 125601 (2015)
54. Shimizu, Y., Hanasaki, I.: Partial structural order of gel-forming material detected as multimodal subdiffusion by logarithmic measure. J. Phys.: Condens. Matter **33**, 455101 (2021)
55. Hanasaki, I., Hosokawa, C.: Non-uniform stochastic dynamics of nanoparticle clusters at a solid-liquid interface induced by laser trapping. Jpn. J. Appl. Phys. **58**, SDDK07 (2019)
56. Hanasaki, I., Ooi, Y.: Particle image diffusometry: resolving diffusion coefficient field from microscopy movie data without particle tracking. AIP Adv. **8**, 065014 (2018)
57. Hanasaki, I., Okano, K., Yoshikawa, H.Y., et al.: Spatiotemporal dynamics of laser-induced molecular crystal precursors visualized by particle image diffusometry. J. Phys. Chem. Lett. **10**, 7452–7457 (2019)
58. Berne, B.J., Pecora, R.: Dynamic Light Scattering. Dover Publications, Mineola (1976)
59. Rodriguez, S.E., Calius, E.P., Khatibi, A., et al.: Mechanical metamaterial systems as transformation mechanism. Extreme Mech. Lett. **61**, 101985 (2023)
60. Saxena, K.K., Das, R., Calius, E.P.: Three decades of auxetics research materials with negative Poisson's ratio: a review. Adv. Eng. Mater. **18**, 1847–1870 (2016)
61. Lakes, R.S.: Negative-Poisson's-ratio materials: auxetic solids. Annu. Rev. Mater. Res. **47**, 63–81 (2017)
62. Duncan, O., Shepherd, T., Moroney, C., et al.: Review of auxetic materials for sports applications: expanding options in comfort and protection. Appl. Sci. **8**, 941 (2018)
63. Evans, E., Alderson, A., thinking: Auxetic materials: functional materials and structures from lateral. Adv. Mater. **12**, 617–628 (2000)
64. Prawoto, Y.: Seeing auxetic mateirals from the mechanics point of view: a structural review on the negative Poisson's ratio. Comput. Mater. Sci. **58**, 140–153 (2012)
65. Kolken, H.M.A., Zadpoor, A.A.: Auxetic mechanical metamaterials. RSC Adv. **7**, 5111–5129 (2019)
66. Shyu, T.C., Damasceno, P.F., Dodd, P.M., et al.: A kirigami approach to engineering elasticity in nanocomposites through patterned defects. Nat. Mater. **14**, 785–789 (2015)
67. Sussman, D.M., Cho, Y., Castle, T., et al.: Algorithmic lattice kirigami: a route to pluripotent materials. PNAS **112**, 7449–7453 (2015)
68. Fujita, T., Nakagawa, D., Komiya, K., et al.: Resilient mechanical metamaterial based on cellulose nanopaper with Kirigami structure. Nanomaterials **12**, 2431 (2022)
69. Sylvestre, J., Morissette, J.F.: Neuromorphic metamaterial structures. Mater. Des. **210**, 11078 (2021)
70. Strogatz, S.H.: Nonlinear Dynamics and Chaos. Perseus Books Publishing, New York (1994)
71. Seifert, U.: Stochastic thermodynamics, fluctuation theorems and molecular machines. Rep. Prog. Phys. **75**, 126001 (2012)

72. Keim, N.C., Paulsen, J.D., Zeravcic, Z., et al.: Memory formation in matter. Rev. Mod. Phys. **91**, 035002 (2019)
73. Inoue, I., Hanasaki, I., Suetsugu, D., et al.: Drawing 3D fluid patterns in beverages using a robotic nozzle. J. Food Eng. **357**, 111555 (2023)

Two-Scale Tensile Analysis of Dovetail Joints for CFRP Fan Blades Considering Material Nonlinearity of Resins

Eiichiro Mori[1(✉)], Tetsuya Matsuda[1], Chiharu Tsujikawa[1], Naoki Morita[1], Gai Kubo[2], Masahiro Hojo[2], Toshihiko Hosaka[3], Shinya Fukushige[3], and Nobuhiro Yoshikawa[4]

[1] University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Japan
s2320894@u.tsukuba.ac.jp, matsuda@kz.tsukuba.ac.jp
[2] Japan Aerospace Exploration Agency (JAXA), 7-44-1 Jindaiji-Higashi-Machi, Chofu, Japan
[3] IHI Corporation, 3-1-1 Toyosu, Koto, Tokyo, Japan
[4] The University of Tokyo, 4-6-1 Komaba, Meguro, Tokyo, Japan

Abstract. In this study, two-scale tensile analysis of a dovetail joint for a fan blade made of a carbon fiber-reinforced plastic (CFRP) is conducted to investigate macroscopic and microscopic behaviors of the dovetail joint. To this end, the two-scale analysis method developed by the authors is implemented in the finite element analysis software Abaqus with its user subroutine function. The present method is then applied to the two-scale tensile analysis of a dovetail joint consisting of a dovetail of a multi-directional CFRP laminate and a dovetail slot of a metal. In this analysis, material nonlinearity (viscoplasticity) of the matrix resin is explicitly considered. It is shown from the analysis results that high mode I tensile stress and mode II shear stress respectively occur at the area just above the contact surface between the dovetail and slot and at the area around the upper end of the contact surface, both macroscopically and microscopically. It is also shown that, even though the macroscopic load-displacement relationship is almost linear, the epoxy resin clearly exhibits nonlinear viscoplasitc deformation microscopically, which can cause inelastic deformation-induced fatigue damages and delamination.

Keywords: CFRP · Composite · Fan blade · Dovetail joint · Nonlinear · Multiscale

1 Introduction

Carbon fiber-reinforced plastics (CFRP) are typical composite materials generally manufactured by stacking laminae (plies) which are submillimeter thick resin sheets unidirectionally reinforced by a number of carbon fibers with diameters of 5–10 μm. Since CFRP have some superior characteristics such as high specific strength and high specific stiffness, they are widely being used as structural materials for aircraft. Moreover, in these days, some jet engine manufactures have tried to use CFRP for fan blades of jet

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 281–292, 2025.
https://doi.org/10.1007/978-3-031-72794-8_21

engines (Fig. 1) to improve the engine efficiency [1]. In general, fan blades are subjected to cyclic and large loads due to centrifugal force, aerodynamic force, impact force by bird strikes [2], and so on. Thus, some experimental and numerical investigations for CFRP fan blades were already conducted targeting damages due to fatigue [3–5] and large dynamic loads [6, 7]. These investigations revealed that the damages can nonlinearly affect static and dynamic responses of CFRP fan blades. It is therefore important to understand the damage behavior of CFRP fan blades.

When investigating such behavior of CFRP fan blades, one has to pay special attention to "dovetail joints" (Fig. 1), because they are connecting sections between fan blades and rotating shafts, and are subjected to the above-mentioned loads most severely. As illustrated in Fig. 1, dovetail joints consist of tapered parts at the bottom end of fan blades called "dovetails" and grooves on rotating shafts called "dovetail slots". CFRP dovetails are made up of the root plies continuous from fan blades and the inserted plies to change the thickness of dovetails. Moreover, each ply consists of numerous carbon fibers and a resin. Thus, CFRP dovetails have significantly complex multiscale structures compared with conventional metal dovetails. Multiscale analysis is therefore promising to understand the mechanical behavior of CFRP dovetail joints. Especially, it is necessary to know how much (visco)plastic deformation of resins microscopically occurs in plies because it can lead to fatigue damages of resins, resulting in delamination between plies [8], which is known as a dominant damage mode of CFRP dovetails [9, 10]. Even though some experimental and numerical investigations were conducted to reveal the damage and fracture mechanisms of composite dovetails [9–12], there is no research focusing on multiscale inelastic analysis.

Our research group has developed an inelastic two-scale analysis method for CFRP [13]. This method is one of the multiscale analysis methods based on a homogenization theory [14, 15], which enables to simulate the macroscopic mechanical behavior of CFRP based on the microscopic mechanical properties of carbon fibers and resins including material nonlinearity such as viscoplasticity. Moreover, the inelastic two-scale analysis method has been implemented in the commercial finite element analysis software Abaqus to deal with more complex cases such as finite deformation and contact problems [16]. It is thus expected that this method is applicable to the multiscale analysis of CFRP dovetail joints.

In this study, inelastic two-scale tensile analysis of a dovetail joint consisting of a CFRP dovetail and a dovetail slot of steel is conducted using the inelastic two-scale analysis method implemented in Abaqus. In this analysis, material nonlinearity (viscoplasticity) of a matrix resin is explicitly considered. Then, macroscopic and microscopic behaviors of the CFRP dovetail such as stress and viscoplastic strain distributions are examined.

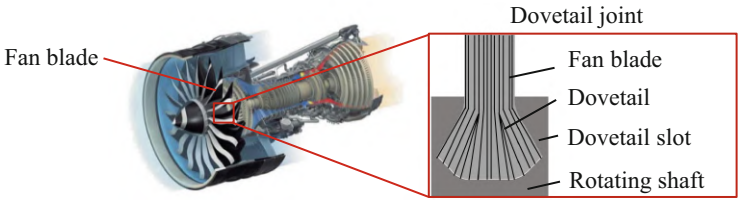


Fig. 1. CFRP fan blades and dovetail joints in jet engine [1]

2 Inelastic Two-Scale Analysis Method

Let us consider a macroscopic composite structure Ω with a boundary Γ_Ω on the macroscopic Cartesian coordinates x_i ($i = 1, 2, 3$) as shown in Fig. 2(a), and its microscopic unit cell Y considering a fiber and a matrix with a boundary Γ_Y on the microscopic Cartesian coordinates y_i ($i = 1, 2, 3$) as shown in Fig. 2(b). Ω is subjected to macroscopic constrains and loads.

2.1 Problem in Macro-Scale

For Ω, the equilibrium of macroscopic stress is described in a weak form as

$$\int_\Omega \dot{\Sigma}_{ij}\delta\dot{U}_{i,x_j}d\Omega - \int_{\Gamma_\Omega} \dot{\Sigma}_{ij}n_j\delta\dot{U}_i d\Gamma_\Omega = 0, \tag{1}$$

where Σ_{ij}, U_i and n_j indicate the macroscopic stress, macroscopic displacement, and unit vector outward normal to Γ_Ω, respectively. In addition, $(\ )_{,x_i}$ and $\dot{(\ )}$ represent the differentiation with respect to x_i and time t, respectively. Considering that Ω consists of elastic-viscoplastic materials, the macroscopic constitutive equation of Ω is expressed as

$$\dot{\Sigma}_{ij} = c^0_{ijkl}\left(\dot{E}_{kl} - \beta^0_{kl}\right), \tag{2}$$

where E_{ij}, c^0_{ijkl} and β^0_{kl} represent the macroscopic strain, macroscopic elastic stiffness and macroscopic viscoplastic function, respectively. On the other hand, macroscopic strain rate $\dot{E}_{ij}$ is expressed as

$$\dot{E}_{ij} = \frac{1}{2}\left(\dot{U}_{i,x_j} + \dot{U}_{j,x_i}\right). \tag{3}$$

By substituting Eqs. (2) and (3) into Eq. (1), a boundary value problem in the macroscale is obtained as follows:

$$\int_\Omega c^0_{ijpq}\dot{U}_{p,x_q}\delta\dot{U}_{i,x_j}d\Omega = \int_{\Gamma_\Omega} \dot{\Sigma}_{ij}n_j\delta\dot{U}_i d\Gamma_\Omega + \int_\Omega c^0_{ijkl}\beta^0_{kl}\delta\dot{U}_{i,x_j}d\Omega. \tag{4}$$

The above equation is generally solved by the finite element method, but c^0_{ijkl} and β^0_{kl} are unknown at the present stage. Thus, in the next section, they are determined by solving a problem in the micro-scale.

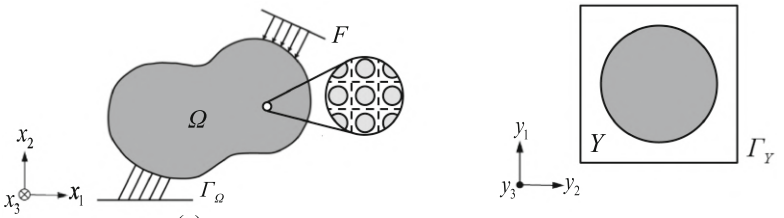


Fig. 2. (a) Macroscopic structure Ω with periodic microstructure, (b) unit cell Y.

2.2 Problem in Micro-Scale

For Y, the equilibrium of microscopic stress is expressed in a weak form by considering the periodicity of Y as [15]

$$\int_Y \dot{\sigma}_{ij} v_{i,y_j} dY = 0, \tag{5}$$

where σ_{ij} and v_i represent the microscopic stress and an arbitrary variation of perturbed velocity field in Y, respectively. Considering that the constituents of Y are elastic-viscoplastic materials, the microscopic constitutive equation is expressed as

$$\dot{\sigma}_{ij} = c_{ijkl}(\dot{\varepsilon}_{kl} - \beta_{kl}), \tag{6}$$

where ε_{kl}, c_{ijkl} and β_{kl} indicate the microscopic strain, microscopic elastic stiffness and microscopic viscoplastic function, respectively. Then, microscopic stress evolution in Y can be obtained as follows [15]:

$$\dot{\sigma}_{ij} = c_{ijpq}(\delta_{pk}\delta_{ql} + \chi^{kl}_{p,y_q})\dot{E}_{kl} - c_{ijkl}(\beta_{kl} - \varphi_{k,y_l}), \tag{7}$$

where χ^{kl}_i and φ_i are the periodic functions called "characteristic functions" determined by solving the following boundary value problems for Y using the finite element method:

$$\int_Y c_{ijpq}\chi^{kl}_{p,y_q} v_{i,y_j} dY = -\int_Y c_{ijkl} v_{i,y_j} dY, \tag{8}$$

$$\int_Y c_{ijpq}\varphi_{p,y_q} v_{i,y_j} dY = \int_Y c_{ijkl}\beta_{kl} v_{i,y_j} dY. \tag{9}$$

2.3 Coupling of Macro-Scale and Micro-Scale

By taking the volume average of Eq. (7) in Y, the following macroscopic constitutive equation of the composite can be obtained:

$$\dot{\Sigma}_{ij} = \left\langle c_{ijpq}\left(\delta_{pk}\delta_{ql} + \chi^{kl}_{p,y_q}\right)\right\rangle \dot{E}_{kl} - \left\langle c_{ijkl}\left(\beta_{kl} - \varphi_{k,y_l}\right)\right\rangle, \tag{10}$$

where $\langle\,\rangle$ denotes the volume average in Y defined as $\langle \# \rangle = |Y|^{-1} \int_Y \# dY$ in which $|Y|$ stands for the volume of Y. Comparison between Eqs. (2) and (10) gives the following relations:

$$c^0_{ijkl} = \left\langle c_{ijpq}\left(\delta_{pk}\delta_{ql} + \chi^{kl}_{p,y_q}\right)\right\rangle. \tag{12}$$

$$c^0_{ijkl}\beta^0_{ijkl} = \left\langle c_{ijkl}\left(\beta_{kl} - \varphi_{k,y_l}\right)\right\rangle. \tag{13}$$

By substituting Eqs. (12) and (13) into Eq. (4), the macroscopic boundary value problem can be rewritten as

$$\begin{aligned} &\int_\Omega \left\langle c_{ijpq}\left(\delta_{pk}\delta_{ql} + \chi^{kl}_{p,y_q}\right)\right\rangle \dot{U}_{p,x_q}\delta\dot{U}_{i,x_j}d\Omega \\ &= \int_{\Gamma_\Omega} \dot{\Sigma}_{ij}n_j\delta\dot{U}_i d\Gamma_\Omega + \int_\Omega \left\langle c_{ijkl}\left(\beta_{kl} - \varphi_{k,y_l}\right)\right\rangle \delta\dot{U}_{i,x_j}d\Omega\,. \end{aligned} \tag{14}$$

This boundary value problem can be solved by the finite element method, enabling to obtain the macroscopic behavior of the composite from the microscopic behavior of the unit cell including material nonlinearity of the constituents.

2.4 Implementation of Two-Scale Analysis Method into Abaqus

The inelastic two-scale analysis method described in the previous section is implemented in the commercial finite element analysis software Abaqus using the user subroutine "Umat". Umat enables to define mechanical behavior of materials by Fortran programing. In the analysis, a macroscopic finite element model for a composite structure and a microscopic finite element model for a unit cell are respectively defined. Then, using the macro-model, the analysis in the macro-scale is incrementally conducted based on Eq. (14). At every time step during the macroscopic analysis, Umat is called at every integration point in the macro-model, and the analysis in the micro-scale is conducted using the micro-model (unit cell model) to obtain macroscopic responses described in Eqs. (12) and (13) at the integration point. Using this computational framework, the inelastic two-scale analysis of composite structures can be performed using Abaqus.

3 Analysis Models and Conditions

As a macro-model, a dovetail joint shown in Fig. 3(a) was considered, where a CFRP dovetail and a dovetail slot made of steel contacted each other. A half region of the dovetail joint was modeled based on the symmetry, and the dovetail and slot were respectively discretized into eight-node isoparametric elements (dovetail: 12430 elements and 15654 nodes, dovetail slot: 1850 elements and 2460 nodes). On the other hand, a micro-model (unit cell model) consisting of a carbon fiber and epoxy resin was defined as shown in Fig. 3(b). This was a 2D model by assuming a hexagonal fiber array (60% fiber volume fraction) and the generalized plain strain condition, and a half region of the unit cell was employed as an analysis domain based on the point symmetry. The half unit cell was discretized into four-node isoparametric elements (99 elements and 116 nodes).

The CFRP dovetail consisted of unidirectional laminae (plies) stacked in the x_2-direction. The laminate configuration was set as shown in Fig. 4 to emulate actual CFRP dovetails, where the blue, green and red parts indicate 0°-, 45°- and -45°-plies, respectively. As mentioned in the Sect. 1, the lower part of the dovetail had the additional inserted plies to form the tapered shape. Unit cells assigned at the integration points in the 45°- and -45°-plies were respectively rotated by 45° and -45° with respect to the x_2 (y_2)-axis as illustrated in Fig. 4 to reproduce the fiber orientations in the plies. Moreover, at the tapered part, unit cells were rotated with respect to the x_1-axis in accordance with the taper angle.

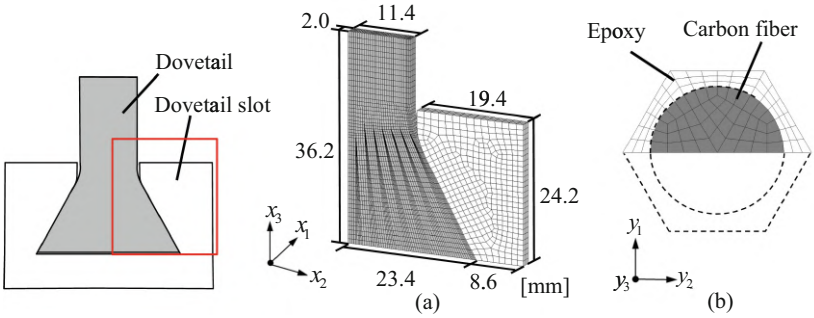


Fig. 3. Analysis models; (a) macro-model (dovetail joint), (b) micro-model (unit cell).

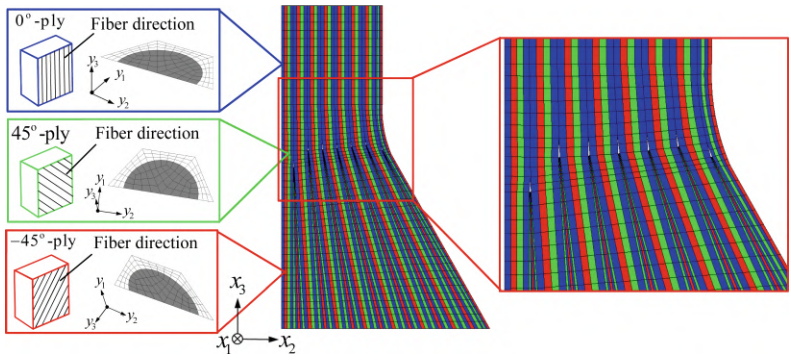


Fig. 4. Laminate configuration of CFRP dovetail.

Boundary conditions applied to the macro-model are shown in Fig. 5. The displacement in the x_2-direction on the left side surface of the dovetail was fixed. The displacement in the x_2-direction on the right side surface and the displacement in the x_3-direction of the bottom surface of the dovetail slot were also fixed. In addition, friction was defined on the contact surface between the dovetail and slot with the coefficients of static friction $\mu_s = 0.3$ and dynamic friction $\mu_d = 0.2$. Forced displacement in the x_3-direction at the displacement rate of 0.1 mm/min was applied to the top surface of the dovetail with target displacement of 0.15 mm.

Material constants of the carbon fibers, epoxy resin, and steel are shown in Table 1. They were regarded as transversely isotropic elastic materials, an isotropic elastic-viscoplastic material with a hardening function depending on the equivalent viscoplastic strain $\overline{\varepsilon}^{\mathrm{p}}$, and an isotropic elastic material, respectively. The material properties of the epoxy were determined from our experimental results of tensile tests using epoxy coupon specimens. Those of the carbon fibers, on the other hand, were determined from the data provided by the manufacturer, and were confirmed by our experimental results of tensile tests using CFRP coupon specimens.

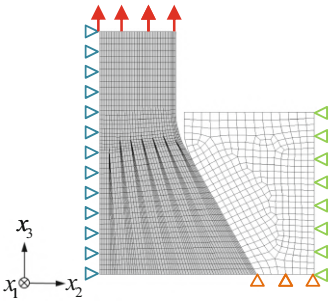


Fig. 5. Boundary conditions applied to macro-model.

Table 1. Material constants of carbon fibers, epoxy, and steel (S45C)

Carbon fibers	
Longitudinal Young's modulus $E_{\mathrm{L}}[GPa]$	235
Transverse Young's modulus $E_{\mathrm{T}}[GPa]$	23.5
Longitudinal Poisson's ratio ν_{LT}	0.26
Transverse Poisson's ratio ν_{TT}	0.49
Shearing stiffness $G_{\mathrm{LT}}[GPa]$	24.2
Epoxy	
Young's modulus $E_{\mathrm{m}}[GPa]$	2.67
Poisson's ratio ν_{m}	0.41
Hardening function $g(\overline{\varepsilon}^{\mathrm{p}})$	$85.0(\overline{\varepsilon}^{\mathrm{p}})^{0.15} + 10.0$
Steel (S45C)	
Young's modulus $E_{\mathrm{s}}[GPa]$	205
Poisson's ratio ν_{s}	0.3

4 Analysis Results

4.1 Macroscopic Results

First, the relationship between the macroscopic load and the displacement of the top surface of the dovetail is depicted in Fig. 6(a). It is found from the figure that the relationship is almost linear in this load/displacement range. Then, stress distributions

in the macro-model when the displacement of the top surface is 0.15 mm are examined. Because the macroscopic stress in the x_2-x_3 plane becomes dominant in this analysis, the distributions of macroscopic stress Σ_{33}, Σ_{22} and Σ_{23} are shown in Figs. 6(b)–(d), respectively. As seen from Fig. 6(b), high Σ_{33} occurs in the 0°-plies at the right side (outer side) of the dovetail. The stress level gradually decreases from the right side to the left side (inner side), and it becomes considerably low at the central region of the dovetail. This means that the applied load is mainly absorbed in the 0°-plies (carbon fibers) at the right side of the dovetail.

On the other hand, Σ_{22} and Σ_{23} shown in Figs. 6(c) and (d) are also important stresses because they can cause the mode I and II delaminations, respectively. As seen from Fig. 6(c), Σ_{22} at the lower part of the dovetail is negative (compressive) due to the constraint by the dovetail slot, whereas Σ_{22} just above the contact surface becomes positive (tensile). Then, it is seen from Fig. 6(d) that the absolute value of Σ_{23} is markedly high around the upper end of the contact surface. These results imply that the mode I delamination is more likely to occur in the area above the contact surface, and the mode II delamination is more likely to occur around the upper end of the contact surface, which is consistent with the areas where delaminations actually occured in the tensile tests of CFRP dovetail joints [9, 10].

4.2 Microscopic Results

As microscopic results, the microscopic stress σ_{22} and σ_{23} and microscopic equivalent viscoplastic strain $\overline{\varepsilon}^{\mathrm{p}}$ in unit cells are examined because they have a strong relation with delamination. Especially, these values in epoxy regions have to be paid attention to because epoxy resins are known to exhibit fatigue damages, leading to delamination as already stated in the Sect. 1, whereas carbon fibers generally exhibit no fatigue damage.

First, the distributions of σ_{22} and σ_{23} when the displacement of the top surface is 0.15 mm are shown in Figs. 7 and 8, respectively. The present method enables to examine the microscopic distributions of stress and strain in unit cells at all the integration points in the macro-model. In this case, however, typical four points are selected to check σ_{22} and σ_{23}. It can be seen from the left upper figure in Fig. 7 that high positive (tensile) σ_{22} concentrates in the epoxy. It is also seen from Fig. 8 that relatively high σ_{23} in the epoxy can be observed in all the unit cells. These stresses are lower than the strength of the epoxy, but they can cause fatigue damages of the epoxy and delamination during cyclic loading.

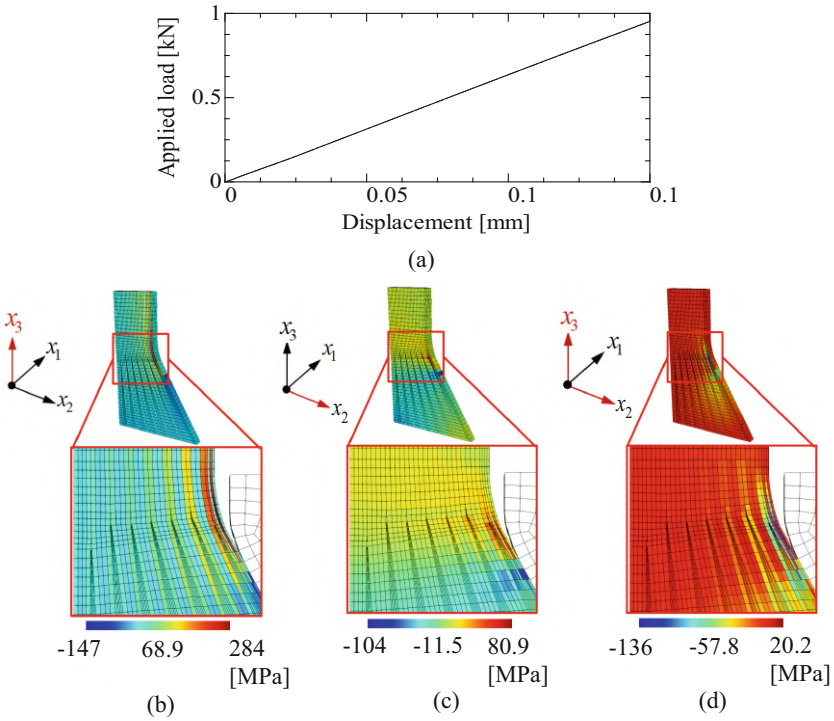


Fig. 6. (a) Load-displacement relationship on the top surface of dovetail, and distributions of macroscopic stress; (b) Σ_{33}, (c) Σ_{22}, (d) Σ_{23}.

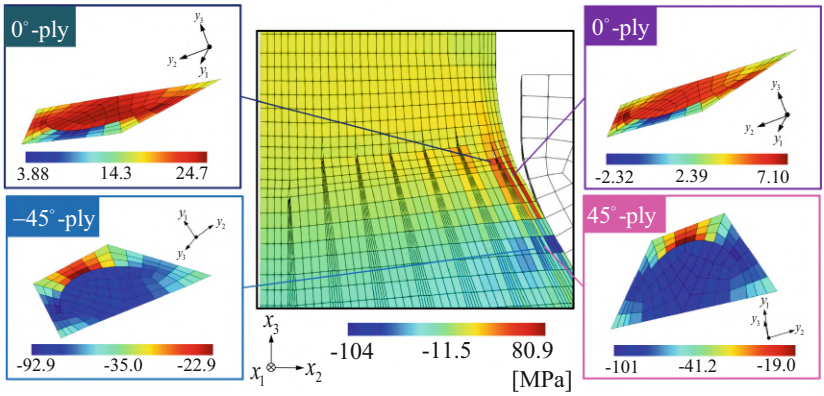


Fig. 7. Distributions of microscopic stress σ_{22} in unit cells (left and right four figures), and distribution of macroscopic stress Σ_{22} (center figure).

Then, the distributions of $\bar{\varepsilon}^{\mathrm{p}}$ in addition to the macroscopic equivalent stress $\overline{\Sigma}$ when the displacement of the top surface is 0.15 mm are shown in Fig. 9. This figure indicates that viscoplastic strain clearly occurs at the epoxy regions especially in the unit

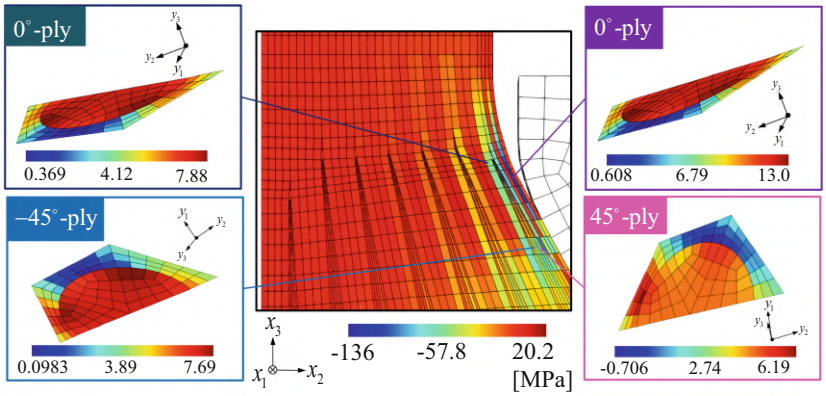


Fig. 8. Distributions of microscopic stress σ_{23} in unit cells (left and right four figures), and distribution of macroscopic stress Σ_{23} (center figure).

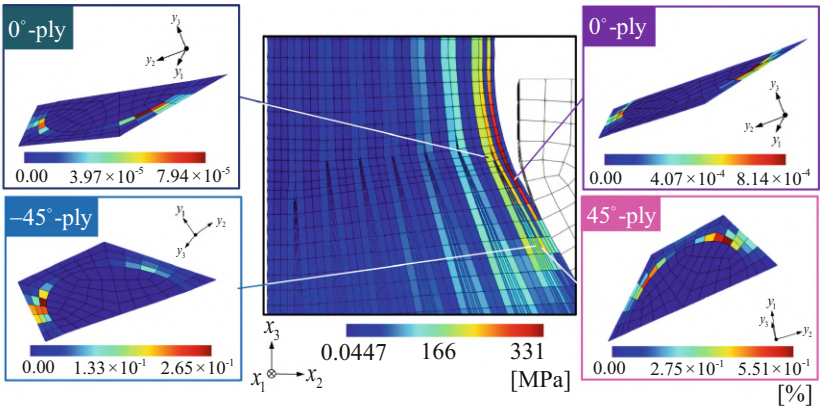


Fig. 9. Distributions of microscopic equivalent viscoplastic strain $\bar{\varepsilon}^{\mathrm{p}}$ (left and right four figures), and distribution of macroscopic equivalent stress $\overline{\Sigma}$ (center figure).

cells located near the contact surface. This means that, even though the macroscopic load-displacement relationship is almost linear as shown in Fig. 6(a), the epoxy exhibits material nonlinearity in the micro-scale, leading to inelastic deformation-induced fatigue damages of the epoxy, and resulting in delamination [8]. This causes the stiffness and strength degradation of CFRP dovetails, affecting the static and dynamic responses of whole CFRP fan blades. It should be noted that the high viscoplastic strain areas found in the present analysis are in good agreement with the areas where delaminations actually occured in the fatigue tests of CFRP dovetail joints conducted by the authors.

5 Conclusions

In this study, macroscopic and microscopic elastic-viscoplastic behaviors of a CFRP dovetail joint subjected to a tensile load were investigated using an inelastic two-scale analysis method implemented in Abaqus. It was shown that high mode I tensile stress and mode II shear stress respectively occurred at the area just above the contact surface between the dovetail and slot and at the area around the upper end of the contact surface, macroscopically and microscopically. It was also shown that microscopic viscoplastic strain occurred in resin sections near the upper end of the contact surface, even though the macroscopic response of the CFRP dovetail is almost linear. Moreover, the areas where these high stresses and viscoplastic strain were observed agreed well with those where delaminations occured in actual experiments of CFRP dovetail joints. The present method therefore has a possibility to predict damage occurrences of CFRP dovetails quantitatively from microscopic point of view, which can contribute to understand the static and dynamic responses of CFRP fan blades.

References

1. Fujimura, T., Nishikawa, S., Moriya, N., Imamura, M.: Development of GEnx turbofan engine. Tech. J. IHI **48**(3), 153–158 (2008). (in Japanese)
2. Amoo, L.M.: On the design and structural analysis of jet engine fan blade structures. Progr. Aerosp. Sci. **60**, 1–11 (2013)
3. Voudouris, G., Mario, D.D., Sever, I.A.: Experimental fatigue behavior of CFRP components under vibration and thermal loading. Int. J. Fatigue **140**, 105791 (2020)
4. Koch, I., Zscheyge, M., Tittmann, K., Gude, M.: Numerical fatigue analysis of CFRP components. Compos. Struct. **168**, 392–401 (2017)
5. Jin, L., Chen, Y., Tang, X., Zhang, J., Wang, Z.: A numerical study on damage characteristics in composite tapered laminates under cyclic loading with different stress ratios. Compos. Struct. **311**(1), 116777 (2023)
6. Takeda, T., et al.: Numerical modeling of adhesively-bonded CFRP/metal joints of composite fan blade. In: Proceedings of JCOSSAR 2019, pp. 201–204 (2019, in Japanese)
7. Yoshikawa, N., et al.: On dynamic delamination strength model of CFRP. In: Proceedings of JCOSSAR 2019, pp. 470–474 (2019, in Japanese)
8. Kozakai, R., et al.: Interlaminar fatigue life of CFRP: experiments using flatwise specimens and prediction based on inelastic two-scale analysis. In: Proceedings of Advanced Technology in Experimental Mechanics and International DIC Society Joint Conference 2023 (2023)
9. Zhang, B., Kawashita, L.F., Jones, M.I., Lander, J.K., Hallet, S.R.: An experimental and numerical investigation into damage mechanisms in tapered laminates under tensile loading. Compos. Part A **133**, 105862 (2020)
10. Noh, H., Go, M., Lim, J., Choi, Y., Kim, J.: Numerical modeling and experimental validation of lamina fracture and progressive delamination in composite dovetail specimens under tensile loading. Compos. Struct. **325**, 117578 (2023)
11. Kumar, R.S., Welsh, G.S.: Delamination failure in ceramic matrix composites: numerical predictions and experiments. Acta Materialia **60**(6–7), 2886–2900 (2012)
12. Zhang, S., Zhou, Y., Gao, X., Song, Y.: Influence of pore distribution on the failure behavior of CMC components. Ceram. Int. **49**(14), 22913–22923 (2023)
13. Matsuda, T., Toyomura, J., Ogaki, T., Arai, M.: Two-scale analysis of thermal behavior of CFRP laminates based on a thermoelastoviscoplastic homogenization theory. Key Eng. Mater. **725**, 433–438 (2017)

14. Suquet, P.M.: Introduction. Part IV elements of homogenization for inelastic solid mechanics. In: Sanchez–Palencia, E., Zaoui, A. (eds.) Homogenization Techniques for Composite Media. LNP, vol. 272, pp. 193–198. Springer, Heidelberg (1987). https://doi.org/10.1007/3-540-17616-0_15
15. Ohno, N., Matsuda, T., Wu, X.: A homogenization theory for elastic-viscoplasitc composites with point symmetry of internal distribution. Int. J. Solids Struct. **38**, 2867–2878 (2001)
16. Izumizaki, K., Miura, N., Matsuda, T., Kubo, G., Hojo, M.: Two-scale interlaminar tensile analysis of CFRP using FEA software. In: Proceedings of M&M 2021 Conference (2021, in Japanese)

Two-Scale Damage/Strength Analysis of CFRP Considering Randomness of Fiber Distribution

Yukinobu Shimura(✉) and Tetsuya Matsuda

University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Japan
s2320863@u.tsukuba.ac.jp, matsuda@kz.tsukuba.ac.jp

Abstract. In this study, we perform two-scale damage/strength analysis of carbon fiber-reinforced plastics (CFRP) that explicitly takes into account the randomness of fiber distribution. This analysis is conducted by combining a two-scale analysis method based on a homogenization theory with the Hoffman's damage determination criterion. To consider the randomness of fiber distribution, several unit cell models with random fiber distribution are prepared as microscopic models. Then, they are randomly assigned at integration points in a macroscopic model of CFRP to perform two-scale damage/strength analysis. The effects of random fiber distribution on the damage and strength properties of $[0]_4$ and $[90]_4$ unidirectional CFRP and a $[0/90]_s$ cross-ply CFRP are investigated by comparing the results with and without considering the randomness of fiber distribution, where the numbers in [] denote the fiber orientation angles, and the subscripts 4 and s respectively indicate the number of plies and the symmetric lamination. It is confirmed that the effects of random fiber distribution for $[0]_4$ are relatively small, whereas those for $[90]_4$ are remarkably large; the strength of $[90]_4$ with random fiber distribution becomes about 32% lower than that without randomness. In the case of $[0/90]_s$, initial damages are observed at quite lower stress levels when the randomness of fiber distribution is considered.

Keywords: CFRP · Fiber distribution · Randomness · Damage · Strength · Multiscale

1 Introduction

Carbon fiber-reinforced plastics (CFRP) are used as structural materials in aerospace and energy applications, because of their higher specific strength, specific stiffness, and corrosion resistance compared to common metal materials [1, 2]. For example, CFRP are used in the body and wings of Boeing 787, and CFRP and other composite materials account for approximately 50% of the weight of the aircraft [3]. Furthermore, the advanced use of CFRP in aircraft has been expanding in recent years, with a trend to incorporate CFRP into the structural components of jet engines.

When designing industrial products using CFRP such as aircraft components mentioned above, it is extremely important to accurately predict the nonlinear behavior of CFRP due to damages, because it can markedly affect static and dynamic responses of

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 293–304, 2025.
https://doi.org/10.1007/978-3-031-72794-8_22

CFRP products. For example, it is known that vibration characteristics such as resonance frequencies generally change when materials are damaged [4, 5], and the effects of damages on vibration characteristics of CFRP have been confirmed [6–8]. Also, a great deal of research has been carried out on the damage behavior of CFRP under dynamic loading [9–12].

In general, CFRP are manufactured by stacking laminae (plies) which are submillimeter thick resin sheets unidirectionally reinforced by a number of carbon fibers with diameters of 5–10 μm. Thus, when CFRP are subjected to large loads, micro-scale damages such as resin cracking, fiber/resin interface debonding and fiber fracture firstly occur in laminae, and then they progress and accumulate, leading to macro-scale stiffness degradation and failure [13]. Therefore, analyses considering microscopic damage behavior are required to predict damage/strength properties of CFRP [14, 15]. Hence, this research group has developed a two-scale damage analysis method [16, 17] by introducing damage criteria into a two-scale analysis method [18, 19] based on a homogenization theory [20, 21], which can analyze both microscopic damage propagation and macroscopic strength of CFRP. In these studies, however, the randomness of fiber distribution that occurs in actual CFRP was not considered, and thus they did not accurately capture actual damage behavior of CFRP.

In this study, two-scale damage/strength analysis of CFRP that takes into account the randomness of fiber distribution is performed. For this, several unit cell models with random fiber distribution are prepared as microscopic models. Then, they are randomly assigned at integration points in a macroscopic model of CFRP, and the two-scale tensile analysis of CFRP is performed. In the analysis, $[0]_4$ and $[90]_4$ unidirectional CFRP and a $[0/90]_s$ cross-ply CFRP are considered, and the effects of random fiber distribution on the damage and strength properties of the CFRP are investigated by comparing the results with and without considering the randomness of fiber distribution.

2 Two-Scale Damage Analysis Method Based on Homogenization Theory

2.1 Two-Scale Analysis Method Based on Homogenization Theory

Let us consider a CFRP structure Ω with a periodic microscopic structure consisting of elastic fibers and a resin, and its boundary Γ_Ω as shown in Fig. 1. Note that the microscopic periodic structure is assumed to be sufficiently small compared to Ω. Then the unit cell Y and Γ_Y are defined as the smallest unit of this microscopic periodic structure and its boundary as shown in Fig. 2, respectively. Now, the Cartesian coordinates x_i $(i = 1, 2, 3)$ and y_i $(i = 1, 2, 3)$ are defined for Ω (macro-scale) and Y (micro-scale), respectively.

First, let us consider the macroscopic problem for Ω. Using the summation convention, the equilibrium of macroscopic stress can be expressed in a weak form as

$$\int_{var\Omega} \dot{\Sigma}_{ij} \delta \dot{U}_{i,x_j} d\Omega - \int_{\Gamma_\Omega} \dot{\Sigma}_{ij} n_j \delta \dot{U}_i d\Gamma_\Omega = 0, \tag{1}$$

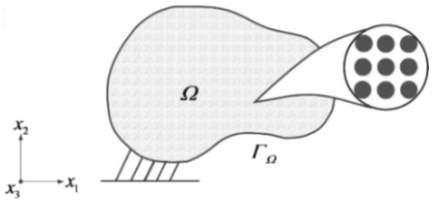


Fig. 1. Macroscopic CFRP structure with periodic microscopic structure.

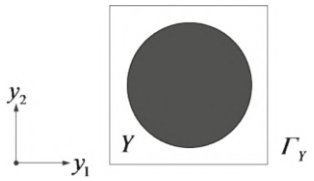


Fig. 2. Unit cell Y with carbon fiber and resin matrix.

where, Σ_{ij}, E_{ij} and U_i indicate the macroscopic stress, macroscopic strain and macroscopic displacement, respectively, and $\dot{(\,)}$ and $(\,)_{,x_i}$ denote the differentiation with respect to time t and x_i, respectively. Now, the macroscopic constitutive equation for Ω is expressed as follows:

$$\dot{\Sigma}_{ij} = c^0_{ijkl}\dot{E}_{kl}, \tag{2}$$

where, c^0_{ijkl} indicates the macroscopic elastic stiffness. In addition, the displacement-strain relationship is as follows:

$$\dot{E}_{ij} = \frac{1}{2}\left(\dot{U}_{i,x_j} + \dot{U}_{j,x_i}\right). \tag{3}$$

By substituting Eqs. (2) and (3) into Eq. (1), the macroscopic boundary value problem comes to the following equation

$$\int_\Omega c^0_{ijpq}\dot{U}_{p,x_q}\delta\dot{U}_{i,x_j}d\Omega - \int_{\Gamma_\Omega}\dot{\Sigma}_{ij}n_j\delta\dot{U}_i d\Gamma_\Omega = 0. \tag{4}$$

This problem can be generally solved using the finite element method (FEM); however it cannot be solved at present because c^0_{ijkl} is unknown. Therefore, let us obtain c^0_{ijkl} by considering the following microscopic problem for the unit cell Y.

According to the homogenization theory [20, 21], microscopic stress σ_{ij} can be expressed as follows:

$$\dot{\sigma}_{ij} = c_{ijpq}(\delta_{pk}\delta_{ql} + \chi^{kl}_{p,y_q})\dot{E}_{kl}, \tag{5}$$

where $(\,)_{,y_j}$ and δ_{ij} indicate the differentiation with respect to y_i and Kronecker's delta, and χ^{kl}_i is the characteristic function determined by solving the following microscopic boundary value problem with the periodic boundary condition [20, 21]

$$\int_Y c_{ijpq}\chi^{kl}_{p,y_q}v_{i,y_j}dY = -\int_Y c_{ijkl}v_{i,y_j}dY, \tag{6}$$

where v_i is an arbitrary variation of perturbed velocity. By taking the volume average of Eq. (5), the homogenized constitutive equation is derived as follows:

$$\dot{\Sigma}_{ij} = \left\langle c_{ijpq}\left(\delta_{pk}\delta_{ql} + \chi^{kl}_{p,y_q}\right)\right\rangle \dot{E}_{kl}, \tag{7}$$

where $\langle\ \rangle$ signifies the volume average in Y. Comparison between Eq. (2) and Eq. (7) gives the following equation

$$c^0_{ijkl} = \left\langle c_{ijpq}\left(\delta_{pk}\delta_{ql} + \chi^{kl}_{p,y_q}\right)\right\rangle. \tag{8}$$

Considering Eq. (8), the macroscopic problem, Eq. (4), eventually results in the following equation

$$\int_{\Omega}\left\langle c_{ijpq}\left(\delta_{pk}\delta_{ql} + \chi^{kl}_{p,y_q}\right)\right\rangle \dot{U}_{p,x_q}\delta\dot{U}_{i,x_j}d\Omega = \int_{\Gamma_\Omega}\dot{\Sigma}_{ij}n_j\delta\dot{U}_i d\Gamma_\Omega. \tag{9}$$

By solving Eq. (9) using FEM, deformation behavior of the macrostructure with the microstructure can be calculated.

2.2 Failure Criterion

In this study, the Hoffman's criterion [22] is used as a failure criterion for the carbon fibers and resin matrix in the microscopic problem mentioned above. In this criterion, the damage parameter F is calculated for each FE element using the following equation

$$\begin{aligned} F = {} & C_1(\sigma_{\mathrm{T}} - \sigma_{\mathrm{Z}})^2 + C_2(\sigma_{\mathrm{Z}} - \sigma_{\mathrm{L}})^2 + C_3(\sigma_{\mathrm{L}} - \sigma_{\mathrm{T}})^2 \\ & + C_4\sigma_{\mathrm{L}} + C_5\sigma_{\mathrm{T}} + C_6\sigma_{\mathrm{Z}} + C_7\tau^2_{\mathrm{TZ}} + C_8\tau^2_{\mathrm{ZL}} + C_9\tau^2_{\mathrm{LT}}, \end{aligned} \tag{10}$$

where σ_{L}, σ_{T}, σ_{Z}, τ_{LT}, τ_{TZ} and τ_{ZL} represent the microscopic stresses which can be determined from σ_{ij} described in the previous section, and subscripts L, T and Z denote the fiber direction, transverse direction and perpendicular direction to the L- and T-directions, respectively. In addition, $C_1,\ C_2,\ \ldots, C_9$ is the material parameters determined by the following equations

$$\begin{gathered} C_1 = \frac{1}{2}\left(\frac{1}{F^{\mathrm{t}}_{\mathrm{T}}F^{\mathrm{c}}_{\mathrm{T}}} + \frac{1}{F^{\mathrm{t}}_{\mathrm{Z}}F^{\mathrm{c}}_{\mathrm{Z}}} - \frac{1}{F^{\mathrm{t}}_{\mathrm{L}}F^{\mathrm{c}}_{\mathrm{L}}}\right), \\ C_2 = \frac{1}{2}\left(\frac{1}{F^{\mathrm{t}}_{\mathrm{Z}}F^{\mathrm{c}}_{\mathrm{Z}}} + \frac{1}{F^{\mathrm{t}}_{\mathrm{L}}F^{\mathrm{c}}_{\mathrm{L}}} - \frac{1}{F^{\mathrm{t}}_{\mathrm{T}}F^{\mathrm{c}}_{\mathrm{T}}}\right), \\ C_3 = \frac{1}{2}\left(\frac{1}{F^{\mathrm{t}}_{\mathrm{L}}F^{\mathrm{c}}_{\mathrm{L}}} + \frac{1}{F^{\mathrm{t}}_{\mathrm{T}}F^{\mathrm{c}}_{\mathrm{T}}} - \frac{1}{F^{\mathrm{t}}_{\mathrm{Z}}F^{\mathrm{c}}_{\mathrm{Z}}}\right), \\ C_4 = \frac{1}{F^{\mathrm{t}}_{\mathrm{L}}} - \frac{1}{F^{\mathrm{c}}_{\mathrm{L}}},\ C_5 = \frac{1}{F^{\mathrm{t}}_{\mathrm{T}}} - \frac{1}{F^{\mathrm{c}}_{\mathrm{T}}},\ C_6 = \frac{1}{F^{\mathrm{t}}_{\mathrm{Z}}} - \frac{1}{F^{\mathrm{c}}_{\mathrm{Z}}}, \\ C_7 = \left(\frac{1}{F^{\mathrm{s}}_{\mathrm{TZ}}}\right)^2,\ C_8 = \left(\frac{1}{F^{\mathrm{s}}_{\mathrm{ZL}}}\right)^2,\ C_9 = \left(\frac{1}{F^{\mathrm{s}}_{\mathrm{LT}}}\right)^2, \end{gathered} \tag{11}$$

where F^{t}, F^{c} and F^{s} indicate the tensile, compressive and shear strengths, respectively. If F in Eq. (10) becomes equal to or greater than 1, the element is to be damaged.

3 Two-Scale Damage/Strength Analysis of CFRP

3.1 Analysis Models

In the present analysis, two-scale tensile analysis of coupon-shaped CFRP was conducted. For the macroscopic model, a CFRP with four plies shown in Fig. 3 was considered, and was divided into eight-node linear isoparametric hexahedral elements (240 elements and 385 nodes). For the microscopic models, on the other hand, three random array unit cell models with four carbon fibers and an epoxy resin were prepared as shown in Figs. 4(a)–(c). They were two-dimensional by assuming the generalized plain strain condition, and were divided into four-node linear isoparametric rectangular elements, where the numbers of elements and nodes were 734 and 794, 765 and 834, and 801 and 868, respectively. These random array models were generated using the method proposed by Wongsto et al. [23]. The minimum distance between fibers was 0.03 times the fiber radius (3.5 μm). For comparison, a hexagonal array semi-unit cell model was also prepared as shown in Fig. 4(d), and was divided into four-node linear isoparametric rectangular elements (117 elements and 135 nodes). All the unit cell models shown in Figs. 4(a)–(d) had 56% fiber volume fraction. It is noted that, in this study, the sizes of the unit cells were relatively small, each of which had only four fibers, due to the limitation of computational resources. Although these kinds of unit cells have difficulty in reproducing large-size fiber-rich or matrix-rich regions, they are to some extent reasonable in this case because the fibers spontaneously distribute randomly but evenly in plies due to their relatively high fiber volume fraction (56%). In addition, this study is targeted at recent high quality CFRP with well-distributed fibers. In future work, however, it is more desirable to consider the appropriate unit cell size which depends on the fiber volume fraction, the ratio between Young's moduli of fibers and a matrix, etc. [24].

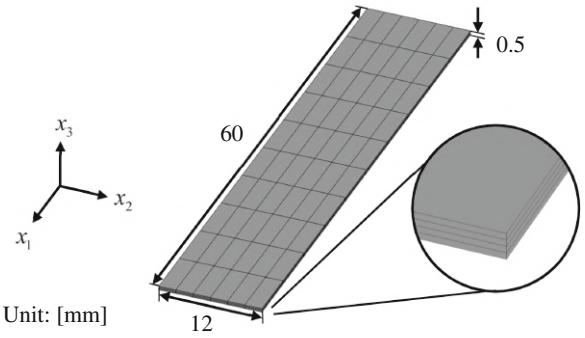


Fig. 3. Macroscopic model of CFRP.

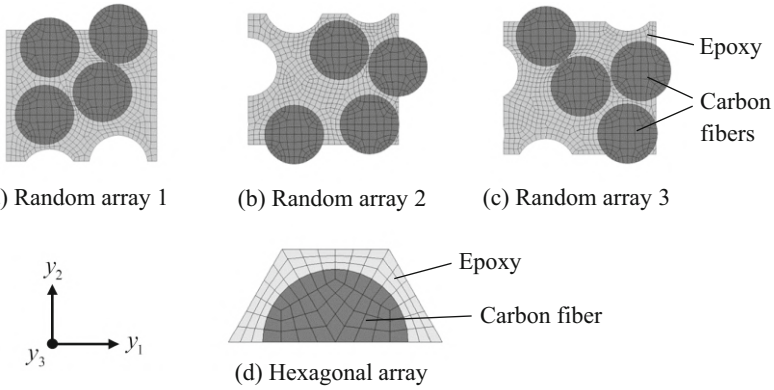


Fig. 4. Microscopic models of CFRP; (a)–(c) random arrays, (d) hexagonal array.

3.2 Material Constants

The carbon fibers were regarded as a transversely isotropic elastic material, while the epoxy resin was an isotropic elastic material. The material parameters for the fibers and epoxy are given in Tables 1 and 2 [25], respectively, assuming IM7/8552 carbon fiber/epoxy CFRP.

Table 1. Material properties of carbon fibers.

Longitudinal Young's modulus	$E_{\mathrm{L}}[GPa]$	276
Transverse Young's modulus	$E_{\mathrm{T}}[GPa]$	19.0
Longitudinal Poisson's ratio	ν_{LT}	0.20
Transverse Poisson's ratio	ν_{TZ}	0.20
Shear modulus	$G_{\mathrm{LT}}[GPa]$	27.0
Tensile Strength	$F_{\mathrm{L}}^{\mathrm{t}}[GPa]$	5.18
Compressive strength	$F_{\mathrm{L}}^{\mathrm{c}}[GPa]$	3.20

Table 2. Material properties of epoxy.

Young's modulus	$E_{\mathrm{m}}[GPa]$	4.08
Poisson's ratio	ν_{m}	0.38
Tensile Strength	$F^{\mathrm{t}}[MPa]$	99.0
Compressive strength	$F^{\mathrm{c}}[MPa]$	130.0
Shear strength	$F^{\mathrm{s}}[MPa]$	57.0

3.3 Analysis Conditions

In the two-scale analysis with the random fiber arrays shown in Fig. 4(a)–(c), one of the three unit cell models was randomly selected and assigned to each integration point in the macroscopic model shown in Fig. 3. In total, 1920 unit cells were assigned to whole the macroscopic model because each hexahedral element in the macroscopic model had eight integration points. Moreover, when assigning, a random rotation angle with respect to the y_3-axis was given to each unit cell. On the other hand, in the two-scale analysis with the hexagonal fiber array shown in Fig. 4(d), the unit cell was assigned to all integration points in the macroscopic model without rotation.

Three kinds of CFRP with different laminate configurations were considered, i.e., $[0]_4$ and $[90]_4$ unidirectional CFRP, and a $[0/90]_s$ cross-ply CFRP. It is noted that the reference axis of the fiber orientation angle was the x_1-axis, meaning that the fiber direction in $[0]_4$ was coincident with the x_1-direction, while that in $[90]_4$ was coincident with the x_2-direction. In the case of $[0/90]_s$, on the other hand, the fiber directions in the outer and inner two plies were coincident with the x_1- and x_2-directions, respectively. These laminates were subjected to the forced displacement in the x_1-direction at the side of $x_1 = 60$ mm as shown in Fig. 5. The displacement rate was $\dot{U}_1 = 6.0 \times 10^{-2}$mm/s. The displacement in the x_2 and x_3-directions at the side of $x_1 = 60$ mm and the displacement in all the directions at the side of $x_1 = 0$ mm were fixed.

Note that the analysis with the random arrays is carried out three times, in each of which a different set of randomness was employed with regard to the assignment of the unit cell models to the integration points in the macroscopic model.

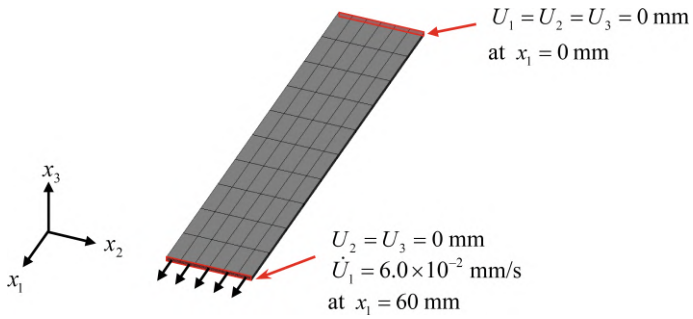


Fig. 5. Boundary conditions for macroscopic model.

3.4 Analysis of $[0]_4$ Unidirectional CFRP

The macroscopic stress-strain relationships in the x_1-direction for the $[0]_4$ unidirectional CFRP are shown in Fig. 6. It can be seen from the figure that, in the case of the hexagonal array, the stress-strain relationship is linear, and reaches the final failure (sudden stress drop) at 1.74% macroscopic strain, resulting in the strength (maximum stress) of 2717 MPa. The stress-strain relationships for the three random cases are almost similar to that for the hexagonal case, though the strength is about 5% higher in the random cases

2 and 3. It can be therefore concluded that the randomness of fiber distribution, including the random selection and rotation of unit cells, little affects the damage/strength properties of $[0]_4$ unidirectional CFRP, which is reasonable from the perspective of the rule of mixture.

3.5 Analysis of $[90]_4$ Unidirectional CFRP

The macroscopic stress-strain relationships for the $[90]_4$ unidirectional CFRP are shown in Fig. 7. This figure shows that the similar shape to $[0]_4$ can be found in the case of the hexagonal array, i.e., the linear relationship and the sudden stress drop at 0.97% macroscopic strain, although the strength (93 MPa) is much smaller than $[0]_4$. In contrast, in the random cases, the first stress drops occur around 0.6% macrostrain, followed by the second ones around 0.7% macrostrain and the third ones around 0.8% macrostrain. This tendency is almost identical for the three random cases. As a result, the strength in the random cases becomes around 63 MPa, which is about 32% lower than in the hexagonal case, indicating that the strength of $[90]_4$ significantly decreases when the randomness of fiber distribution is considered. This means that the hexagonal array can overestimate the strength of $[90]_4$, and it is important to consider the random fiber distribution for safer design of CFRP structures.

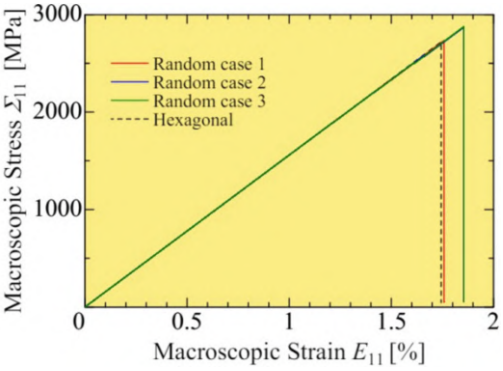


Fig. 6. Macroscopic stress-strain relationships of [0]4 unidirectional CFRP.

To investigate the reason for the strength difference between the random and hexagonal arrays mentioned above, damage distributions in the macro- and micro-scales are examined. First, the stiffness reduction ratio in the x_1-direction in the macroscopic models is shown in Fig. 8. As shown in Fig. 8(a), in the hexagonal case, the stiffness reduction ratio suddenly and uniformly increases from 0% to almost 100% at 0.97% macroscopic strain (final failure). In contrast, the stiffness reduction is considerably different from element to element in the random case as shown in Fig. 8(b), and the stiffness has decreased at much earlier stage (0.69% macroscopic strain) compared with the hexagonal case.

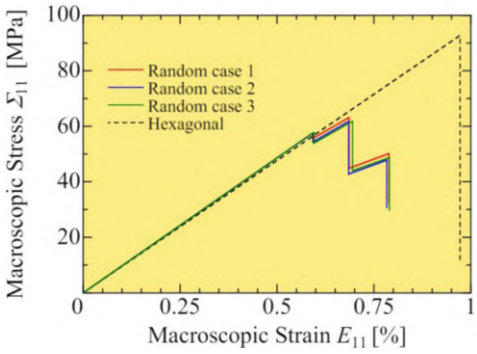


Fig. 7. Macroscopic stress-strain relationships of [90]4 unidirectional CFRP.

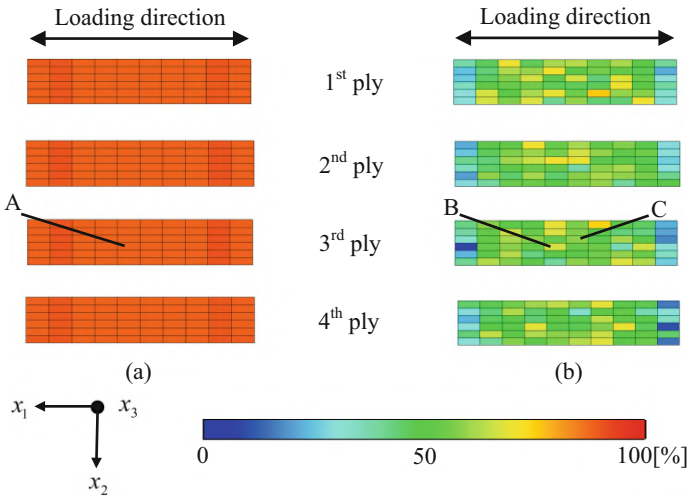


Fig. 8. Stiffness reduction ratio in x_1-direction in macroscopic model ([90]4); (a) hexagonal case (0.97% macroscopic strain), (b) random case 1 (0.69% macroscopic strain).

Furthermore, the microscopic damage distributions in the unit cells assigned at the point A, B and C in the macroscopic models, which are indicated in Fig. 8, are depicted in Fig. 9. In the hexagonal case, the microscopic damages shown in Fig. 9(a) occurred in the epoxy, and the same damage distribution was observed at not only A but also everywhere in the CFRP evenly, resulting in the high strength and the sudden and uniform failure as shown in Figs. 7 and 8(a). In contrast, it can be found from Figs. 9(b) and 9(c) that, in the random cases, the damage distributions of the epoxy become completely different depending on the position (i.e., assigned unit cell and its rotation angle) in the macroscopic model, causing the variability of stiffness reduction shown in Fig. 8(b). It is also found from the figures that the damages concentrate on the epoxy areas where the inter-fiber distance in the loading direction is small. Such behaviors of unit cells at all integration points were averaged over the macroscopic model, resulting in the

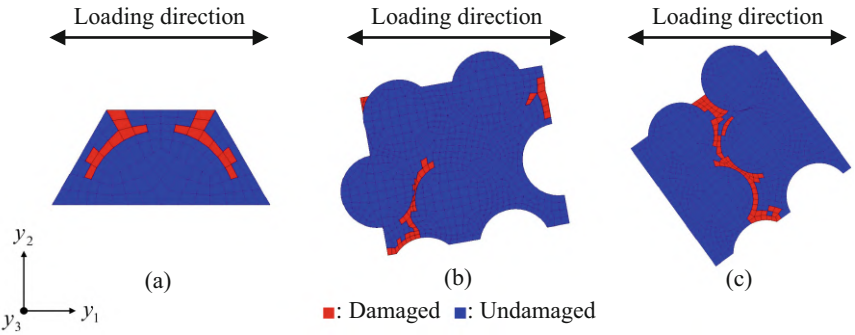


Fig. 9. Damage distributions in unit cells assigned at (a) point A, (b) point B, (c) point C ([90]4).

earlier stiffness degradation and the lower strength in the random cases compared to the hexagonal case as shown in Fig. 7.

3.6 Analysis of $[0/90]_s$ Cross-ply CFRP

The macroscopic stress-strain relationships for the $[0/90]_s$ cross-ply CFRP laminate are shown in Fig. 10. In this case, nonlinearity of the macroscopic stress-strain relationships becomes stronger; several small stress drops can be observed prior to the final failure around 1.8% macroscopic strain, resulting from the damages in the 90°-plies. This behavior agrees well with that commonly observed in actual tensile tests of cross-ply CFRP. A close look at the figure reveals that the initial damage for the hexagonal array occurs at 1.0% macroscopic strain, whereas that for the random cases occurs at much earlier stage around 0.6% macrostrain, which corresponds well to the difference in strength of $[90]_4$ depending on the fiber distribution described in the previous section. This affects the final failure of $[0/90]_s$, and the strength of the random cases becomes about 5% lower than that of the hexagonal case, which is reversed results compared to $[0]_4$. These results suggest that the effect of randomness of fiber distribution on the initial damage of $[0/90]_s$ is significant, although that on the stress level of final failure

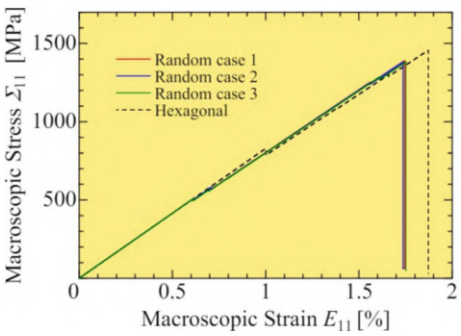


Fig. 10. Macroscopic stress-strain relationships of [0/90]s cross-ply CFRP.

is not so large. This can become important when careful attention to the initiation of damages in CFRP products rather than their final failure is required.

4 Conclusions

In this study, two-scale damage/strength analysis of CFRP was performed in consideration of the randomness of fiber distribution. To this end, three kinds of unit cell models with randomly arranged carbon fibers and an epoxy resin were prepared. For comparison, a hexagonal array unit cell was also considered. The results show that the randomness of fiber distribution has significant influence on the damage and strength properties of CFRP. It is therefore important to consider the randomness of fiber distribution for more precise and safer prediction of nonlinear behavior of CFRP due to damages, leading to deeper understanding of static and dynamic responses of CFRP products.

Acknowledgments. This work was partly supported by JSPS Grant-in-Aid for Scientific Research (B) Grant Number 20H02034.

References

1. Msaoubi, R., et al.: High performance cutting of advanced aerospace alloys and composite materials. CIRP Ann. **64**(2), 557–580 (2015)
2. Rajak, D.K., Pagar, D.D., Menezes, P.L., Linul, E.: Fiber-reinforced polymer composites: manufacturing, properties, and applications. Polymers **11**(10), 1667 (2019)
3. AERO - Boeing 787 from the Ground Up. https://www.boeing.com/commercial/aeromagazine/articles/qtr_4_06/article_04_1.html. Accessed 17 Oct 2023
4. Doebling, S.W., Farrar, C.R., Prime, M.B., Shevitz, D.W.: Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review. Los Alamos National Laboratory Report LA-13070-MS. Los Alamos National Laboratory, United States (1996)
5. Salawu, O.S.: Detection of structural damage through changes in frequency: a review. Eng. Struct. **19**(9), 718–723 (1997)
6. Capozucca, R.: Vibration of CFRP cantilever beam with damage. Compos. Struct. **116**, 211–222 (2014)
7. Gomes, G.F., Mendéz, Y.A.D., da Cunha Jr, S.S., Ancelotti, A.C., Jr.: A numerical-experimental study for structural damage detection in CFRP plates using remote vibration measurements. J. Civ. Struct. Health Monit. **8**, 33–47 (2018)
8. Capozucca, R., Bonci, B.: Notched CFRP laminates under vibration. Compos. Struct. **122**, 367–375 (2015)
9. Waimer, M., Kohlgrüber, D., Hachenberg, D., Voggenreiter, H.: Experimental study of CFRP components subjected to dynamic crash loads. Compos. Struct. **105**, 288–299 (2013)
10. Waimer, M., Siemann, M.H., Feser, T.: Simulation of CFRP components subjected to dynamic crash loads. Int. J. Impact Eng **101**, 115–131 (2017)
11. Yoshikawa, N., et al.: On dynamic delamination strength model of CFRP. In: Proceedings of The Ninth Japan Conference on Structural Safety and Reliability, pp. 470–474 (2019). (in Japanese)

12. Takeda, T., et al.: Numerical modeling of adhesively-bonded CFRP/metal joints of composite fan blade. In: Proceedings of The Ninth Japan Conference on Structural Safety and Reliability, pp. 201–204 (2019). (in Japanese)
13. Hobbiebrunken, T., Hojo, M., Adachi, T., Jong, C.D., Fiedler, B.: Evaluation of interfacial strength in CF/epoxies using FEM and in-situ experiments. Compos. Part A Appl. Sci. Manuf. **37**(12), 2248–2256 (2006)
14. Okabe, T., Nishikawa, M., Toyoshima, H.: A periodic unit-cell simulation of fiber arrangement dependence on the transverse tensile failure in unidirectional carbon fiber reinforced composites. Int. J. Solids Struct. **48**(20), 2948–2959 (2011)
15. Kumagai, Y., Onodera, S., Salviato, M., Okabe, T.: Multiscale analysis and experimental validation of crack initiation in quasi-isotropic laminates. Int. J. Solids Struct. **193–194**, 172–191 (2020)
16. Kohashi, T., Matsuda, T.: Two-scale progressive damage analysis of CFRP laminates. In: Proceedings of 14th World Congress in Computational Mechanics and ECCOMAS Congress (2021)
17. Akashi, K., Matsuda, T.: Two-scale damage propagation/strength analysis of CFRP laminates considering variability of fiber strength. In: Proceedings of 15th World Congress in Computational Mechanics and APCOM Congress (2022)
18. Matsuda, T., Toyomura, J., Ogaki, T., Arai, M.: Two-scale analysis of thermal behavior of CFRP laminates based on a thermoelastoviscoplastic homogenization theory. Key Eng. Mater. **725**, 433–438 (2016)
19. Matsui, K., Terada, K., Yuge, K.: Two-scale finite element analysis of heterogeneous solids with periodic microstructures. Comput. Struct. **82**(7–8), 593–606 (2004)
20. Suquet, P.M.: Introduction. Part IV elements of homogenization for inelastic solid mechanics. In: Sanchez–Palencia, E., Zaoui, A. (eds.) Homogenization Techniques for Composite Media. LNP, vol. 272, pp. 193–198. Springer, Heidelberg (1987). https://doi.org/10.1007/3-540-17616-0_15
21. Ohno, N., Matsuda, T., Wu, X.: A homogenization theory for elastic-viscoplasitc composites with point symmetry of internal distributions. Int. J. Solids Struct. **38**(16), 2867–2878 (2001)
22. Hoffman, O.: The brittle strength of orthotropic materials. J. Compos. Mater. **1**(2), 200–206 (1967)
23. Wongsto, A., Li, S.: Micromechanical FE analysis of UD fibre-reinforced composites with fibres distributed at random over the transverse cross-section. Compos. Part A Appl. Sci. Manuf. **36**(9), 1246–1266 (2005)
24. Pingaro, M., Reccia, E., Trovalusci, P., Masiani, R.: Fast statistical homogenization procedure (FSHP) for particle random composites using virtual element method. Comput. Mech. **64**, 197–210 (2019)
25. Kaddour, A.S., Hinton, M.J., Smith, P.A., Li, S.: Mechanical properties and details of composite laminates for the test cases used in the third world-wide failure exercise. J. Compos. Mater. **47**(20–21), 2427–2442 (2013)



Pattern Formation in Tape-Peeling Traces of Deformed Adhesives

Keisuke Taga[1(✉)], Hiroya Nakao[2], and Yoshihiro Yamazaki[1]

[1] Department of Physics, School of Advanced Science and Engineering, Waseda University, Tokyo 169-8555, Japan
tagaksk@akane.waseda.jp

[2] Department of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology, Tokyo 152-8552, Japan

Abstract. We can find fractal patterns in the peel traces of adhesive tapes. These tape-peeling patterns resemble the patterns obtained from a certain class of reaction-diffusion equations, and it is expected that the tape-peeling dynamics can also be modeled by a reaction-diffusion equation. Several models have been proposed previously for tape-peeling patterns, but these models contradict Newton's law of action and reaction. To address this problem, we have recently proposed a new reaction-diffusion model for the tape-peeling pattern formation based on Newton's equation of motion for the peel front. Although the physical mechanism described by our proposed model differs from that of the previous models, both can reproduce the experimental results. In this paper, we compare our new model with one of the previous models and discuss their relationship and differences.

Keywords: soft matter · pattern formation · reaction-diffusion system

1 Introduction

Many kinds of spatiotemporal patterns appear around us, and some of them have been modeled mathematically with reaction-diffusion equations. The reaction-diffusion equation is a nonlinear partial differential equation that consists of a nonlinear local dynamics term and a linear spatial diffusion term, given as follows:

$$\frac{\partial u}{\partial t} = F(u) + \Delta_x u.$$

Here, u is a field variable that depends on time t and position x, and Δ_x is a diffusion operator.

Reaction-diffusion equations can exhibit various kinds of patterns and have been used as models for explaining the patterns of fish skin, chemical waves in the Belouzov-Zhabotinsky reaction, etc. Some of the reaction-diffusion models show fractal patterns that resemble the Sierpinski gaskets, as reported in the

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 305–311, 2025.
https://doi.org/10.1007/978-3-031-72794-8_23

Bonhoffer-van der Pol type equation [1], Gray-Scott equation [2], and complex Ginzburg-Landau equation [3]. In this paper, we discuss a new reaction-diffusion model [4] for tape-peeling traces that exhibits fractal random patterns similar to the Sierpinski gaskets [5–7].

Looking closely at the experimental peel front, there occur two types of states, with a tunnel structure (state A) and without a tunnel structure (state B). Only state A is stable for slow peeling, and only state B is stable for fast peeling. At intermediate speeds, the peel front switches between these two states chaotically. The peel traces corresponding to states A and B are different from each other; thus, such chaotic switching yields intermittent spatiotemporal patterns, as shown in Fig. 1(c).

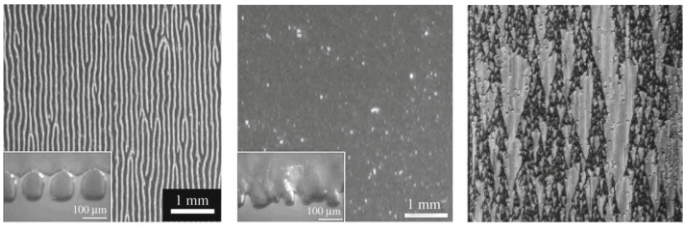


Fig. 1. Peel traces of (a) state A and (b) state B. Peeling proceeds from top to bottom. The small image at the bottom left of each figure shows an enlarged image of the peel front. (c) Fractal coexisting pattern of peel states A and B (the actual size is 25 mm × 25 mm). White and black regions represent the peel states A and B, respectively. (Reproduced from [6] with permission. © 2006 Elsevier)

Several mathematical models have been proposed previously to explain the pattern formation of peel traces [6–11]. These models consider state transitions of the peel front. One characteristic property shared by these models is an asymmetric interaction that takes into account the asymmetry in the experimental observation, such that state A, which is adjacent to state B, can easily make a transition to state B, but not vice versa.

However, considering the relationship between the state transitions and the dynamics of the peel-front structures, the assumed asymmetric interaction contradicts Newton's law of action and reaction. Recently, we proposed a new model based on Newton's equation of motion for the peel front [4]. In this paper, we briefly review the newly proposed model, compare it with one of the previous models [8], and discuss their similarity and differences.

2 Models of Tape-Peeling Traces

2.1 Previous Model

First, we review one of the previous tape-peeling models introduced in Refs. [8, 10]. It is given by

$$\begin{cases} \dfrac{d\phi_i}{dt} = -\alpha\phi_i(\phi_i - 0.5)(\phi_i - 1) \\ \qquad\qquad + D\{\theta(\phi_{i+1} - \phi_i) + \theta(\phi_{i-1} - \phi_i)\} - v + \xi_i, \quad i = 1, \dots N, & \text{(1a)} \\ \tau\dfrac{dv}{dt} = (\bar{\phi} - V) - v, \quad \tau = \dfrac{N}{k}, \quad \bar{\phi} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\phi_i. & \text{(1b)} \end{cases}$$

This model is a phase-field-type model that describes the dynamics of the state at the peel front. The peel front is discretized into N units whose size corresponds to the characteristic size of the tunnel structure shown in Fig. 1. Thus, the variable ϕ_i represents the state of the ith unit at the peel front. It is assumed that $\phi_i = 0$ corresponds to state A and $\phi_i = 1$ corresponds to state B, respectively.

In the experiment, one end of the tape is pulled with a spring, which is used to investigate the effect of the elasticity of the tape [7]. Thus, the variable v is introduced to represent the velocity of the peel front induced by the spring, and the parameter V is the pulling speed of the spring. Spatial inhomogeneity of the adhesives is represented by uniformly distributed random variables $\xi_i \in [-0.1, 0.1]$ independently given to the individual units. The function θ in Eq. (1a) represents the asymmetric spatial interaction,

$$\theta(x) = \begin{cases} 0 & x < 0, \\ x & x \geq 0, \end{cases}$$

which is a characteristic feature common to the previous models. This asymmetric interaction is essentially important in reproducing the experimental results by this model. However, this asymmetric term contradicts the law of action and reaction, because it represents a nonreciprocal action, i.e., $\theta(x - y) \neq \theta(y - x)$. Thus, we introduced a new model based on Newton's equation of motion for the peel front as explained in the next subsection.

2.2 Proposed Model

In Ref. [4], we recently introduced a new model of tape peeling. Our model in a general form is given by the following Newton's equation of motion:

$$\begin{aligned} \frac{d^2u_i}{dt^2} = &-\left.\frac{dU}{du}\right|_{u=u_i} - b\left(V + \frac{du_i}{dt}\right) + \frac{a}{1 + c(u_i - d)^2}\frac{du_i}{dt} \\ &+ D_1(u_{i+1} + u_{i-1} - 2u_i) + D_2\left(\frac{du_{i+1}}{dt} + \frac{du_{i-1}}{dt} - 2\frac{du_i}{dt}\right). \end{aligned} \tag{2}$$

This model describes the dynamics of the peel front. The peel front is discretized as in Eq. (1), where u_i represents the deformation of the ith unit at the peel front. From the experimental observation, the deformation of the peel front in state A is larger than that in state B [7]. We thus assume that the unit is in state A when u is relatively large and in state B when u is relatively small.

The meaning of each term on the right-hand side is as follows. The first term $-\frac{dU}{du}\big|_{u=u_i}$ represents a double-well potential, where the two bottoms of the well correspond to states A and B, respectively. The parameter V is the peel speed, $-b\frac{du_i}{dt}$ is a linear viscous term, and $\frac{a}{1+c(u_i-d)^2}\frac{du_i}{dt}$ is a negative viscous term which has the effect of destabilizing state B. In addition, the interaction terms between the units are introduced to incorporate the viscoelasticity of the adhesives; the term $D_1(u_{i+1}+u_{i-1}-2u_i)$ is the elastic interaction and $D_2\left(\frac{du_{i+1}}{dt}+\frac{du_{i-1}}{dt}-2\frac{du_i}{dt}\right)$ is the viscous interaction. See Ref. [4] for a more detailed discussion of these terms.

For the numerical simulations, we use the following explicit form of Eq. (2),

$$\begin{aligned}\frac{d^2u_i}{dt^2} =& -3(u_i-1)^2(u_i-2)-\left(V+\frac{du_i}{dt}\right)+\frac{2}{1+20(u_i-1)^2}\frac{du_i}{dt}\\ &+(u_{i+1}+u_{i-1}-2u_i)+0.1\left(\frac{du_{i+1}}{dt}+\frac{du_{i-1}}{dt}-2\frac{du_i}{dt}\right),\quad i=1,\ldots,N.\end{aligned} \tag{3}$$

In this case, we consider the potential bottom at $u\approx 2$ as state A and that at $u\approx 1$ as state B, respectively. We consider a system of size $N=1000$ and assume periodic boundary conditions. We impose $u_i=0$ and $\frac{du_i}{dt}=0$ as an initial condition.

We numerically integrated Eq. (3) by using the 4th Runge-Kutta method with a time step $dt=0.01$. To incorporate the spatial inhomogeneity of the adhesives, we reset the deformation u_i of each unit to 0 randomly with a probability $p=0.01dt$ during the numerical time evolution.

Figure 2 shows the patterns obtained from Eq. (3). As shown in Figs. 2(a) and (e), state A is stable for small peeling speed V and state B is stable for large V, resulting in uniform patterns. At intermediate values of V, states A and B switch with each other chaotically, and we obtain fractal random patterns similar to the Sierpinski gaskets as shown in Figs. 2(b), (c), and (d). In Ref. [4], we also investigated the scaling properties (such as fractal dimensions) of the random patterns obtained from the model and compared them with experimental results.

We can transform Eq. (2) into the following Liénard system by introducing a new parameter $w_i\equiv\frac{du_i}{dt}+bu_i-a'\arctan(c'(u_i-d))-D_2(u_{i+1}+u_{i-1}-2u_i)$:

$$\begin{cases}\dfrac{du_i}{dt} = -w_i-bu_i+a'\arctan(c'(u_i-d))+D_2(u_{i+1}+u_{i-1}-2u_i), & \text{(4a)}\\ \dfrac{dw_i}{dt} = \left.\dfrac{dU}{du}\right|_{u=u_i}+bV-D_1(u_{i+1}+u_{i-1}-2u_i), & \text{(4b)}\end{cases}$$

where a' and c' are related to the parameters a and c in Eq. (2) as $a' = a/\sqrt{c}$ and $c' = \sqrt{c}$. If we further replace the double-well potential with a single-well potential, the reaction term of Eq. (4) is almost the same as that of the Bonhoffer-van der Pol type reaction-diffusion equation for chemical waves analyzed in Ref. [1], which has been reported to show Sierpinski-gasket-like patterns.

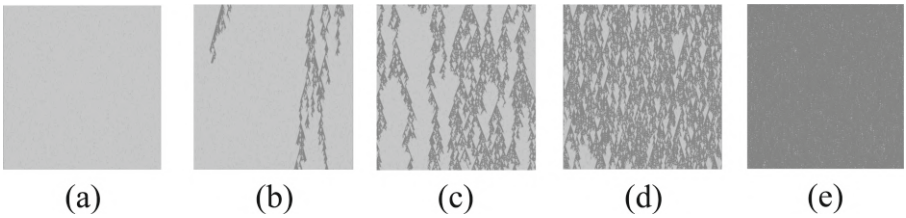


Fig. 2. Typical spatiotemporal patterns obtained from Eq. (3). Time proceeds from top to bottom ($500 \leq t \leq 1200$). (a) $V = 0.29$, (b) $V = 0.30$, (c) $V = 0.309$, (d) $V = 0.32$, (e) $V = 0.50$. (Reproduced from [1] with permission. © 2023 The Physical Society of Japan)

3 Comparison of the Models

Although the previous models and Eq. (2) describe different physical mechanisms, they can both reproduce the experimental results. Thus, we are interested in the mathematical relations between those models.

Here, we compare our proposed model, Eq. (2), with one of the previous models, Eq. (1). For comparison, we transform the variables of Eq. (1) as

$$\phi_i \to -\phi_i, \quad v \to -v.$$

Then, we obtain

$$\begin{cases} \dfrac{d\phi_i}{dt} = -a\phi_i(\phi_i + 0.5)(\phi_i + 1) \\ \qquad\qquad + D\{\hat{\theta}(\phi_{i+1} - \phi_i) + \hat{\theta}(\phi_{i-1} - \phi_i)\} - v - \xi_i, \ i = 1, \ldots N, & \text{(5a)} \\ \tau \dfrac{dv}{dt} = (\bar{\phi} + V) - v. \ \tau = \dfrac{N}{k}, \ \bar{\phi} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \phi_i, & \text{(5b)} \end{cases}$$

where $\hat{\theta}(x) = -\theta(-x)$. By this transformation, $\phi_i = 0$ corresponds to state A and $\phi_i = -1$ corresponds to state B, respectively. Namely, we can regard large ϕ_i as state A and small ϕ_i as state B.

Though our new model based on Newton's equation, described by Eq. (2) or Eq. (4), seem very different from the previous phase-field-type model, Eq. (1) or Eq. (5), we can find that many terms actually represent similar effects. Here, we qualitatively compare the terms in Eq. (4) and Eq. (5).

The correspondence of the variables between the two models is ϕ_i vs. u_i and v vs. w_i. We note that the variable v in Eq. (5) is a global variable while w_i is a local variable in Eq. (4), but we ignore this difference in the following discussion; this corresponds to the case where many units take similar values of w_i.

First, we compare the right-hand side between Eqs. (4a) and (5a). Both $-bu_i + a' \arctan(c'(u_i - d))$ and $-\alpha\phi_i(\phi_i + 0.5)(\phi_i + 1)$ represent inverted-N-shaped functions, and $-w_i$ and $-v$ have the same form. Both equations have interaction terms, $D\{\hat{\theta}(\phi_{i+1} - \phi_i) + \hat{\theta}(\phi_{i-1} - \phi_i)\}$ and $D_2(u_{i+1} + u_{i-1} - 2u_i)$, respectively, but the former is asymmetric while the latter is symmetric.

Regarding the right-hand sides of Eqs. (4b) and (5b), the term $\frac{dU}{du}\big|_{u=u_i}$ comes from the double-well potential, which gives a N-shaped function, but ϕ is a linearly increasing function. The terms bV and V share the same form. The term that corresponds to $-D_1(u_{i+1} + u_{i-1} - 2u_i)$ in Eq. (4b) does not occur in Eq. (5b). Also, the term that corresponds to $-v$ in Eq. (5b) does not occur in Eq. (4b).

As compared above, Eq. (4) and Eq. (5) share some similarities, but there also exist some differences. In the following, we discuss the relationship between these models focusing on their asymmetric properties.

Firstly, if we change the interaction term in Eq. (5) to a symmetric interaction like $D_2(u_{i+1} + u_{i-1} - 2u_i)$ in Eq. (4), the black region does not propagate into the white region nor diffuse [8]. This suggests that the asymmetric interaction accelerates the propagation of state B (black region). In Eq. (4), the propagation of state B is realized by the term $D_1(u_{i+1} + u_{i-1} - 2u_i)$. Combining the left-hand side of Eq. (2) and $D_1(u_{i+1} + u_{i-1} - 2u_i)$ in the right-hand side, we obtain a wave equation. Thus, this term facilitates the propagation of state B.

Secondly, the asymmetric interaction also describes the state transition from A to B, namely, it reproduces the experimental fact that state B propagates into state A but state A does not propagate into state B. The existence of the double-well potential in Eq. (4) realizes this asymmetric phenomenon.

From the above discussion, we find that the physical details of both models are different from each other, in particular, Eq. (2) does not describe the dynamics of a spring, but the two models are qualitatively similar to each other. By comparing their differences, we find that the terms needed to realize the propagation of state B into state A and the asymmetric transition between states A and B in one model are substituted by other terms in the other model.

4 Summary

In Ref. [4], we proposed a new model for pattern formation in the peeling trace of the adhesive tapes. The patterns obtained from our model were similar to those observed in the previous experiments, and the scaling properties of the patterns were consistent with those observed experimentally [4].

In this paper, we compared the proposed model with one of the previous models. By transforming the proposed model into a Liénard system, we showed that there are some similarities between the two models, but several points

are clearly different, for example, the interaction terms. We discussed how the new terms in our present model describe the asymmetric properties that were described by the asymmetric interaction assumed in the previous model.

This paper presented only a rough discussion of the proposed model, and we leave a more precise analysis of the proposed model for future work. Also, we recently found a relationship between the spatiotemporal patterns exhibited by our model and the directed percolation universality class in nonequilibrium phase transition [12]. A more detailed discussion of this relationship will also be discussed in future work.

Acknowledgments. We acknowledge JST CREST JP-MJCR1913 for financial support.

References

1. Hayase, Y., Ohta, T.: Sierpinski gasket in a reaction-diffusion system. Phys. Rev. Lett. **81** 1726 (1998). https://doi.org/10.1103/PhysRevLett.81.1726
2. Hayase, Y., Ohta, T.: Self-replicating pulses and Sierpinski gaskets in excitable media. Phys. Rev. E **62** 5998 (2000). https://doi.org/10.1103/PhysRevE.62.5998
3. Chaté, H., Pikovsky, A., Rudzick, O.: Forcing oscillatory media: phase kinks vs. synchronization. Physica D **131** 17 (1999). https://doi.org/10.1016/S0167-2789(98)00215-2
4. Taga, K., Yamazaki, Y.: A tape-peeling model for spatiotemporal pattern formation by deformed adhesives. J. Phys. Soc. Jpn. **92** 043001 (2023). https://doi.org/10.7566/JPSJ.92.043001
5. Yamazaki, Y., Toda, A.: Dynamical-morphological property of adhesive tape in peeling. J. Phys. Soc. Jpn. **71** 1618 (2002). https://doi.org/10.1143/jpsj.71.1618
6. Yamazaki, Y., Toda, A.: Stability of tunnel structure and relationship between peel load and spatiotemporal pattern by deformed adhesive during peeling. J. Phys. Soc. Jpn. **73** 2342 (2004). https://doi.org/10.1143/jpsj.73.2342
7. Yamazaki, Y., Toda, A.: Pattern formation and spatiotemporal behavior of adhesive in peeling. Physica D **214** 120 (2006). https://doi.org/10.1016/j.physd.2005.12.009
8. Yamazaki, Y.: Collective behavior of bistable units with global and asymmetric local interactions. Prog. Theor. Phys. **125** 641 (2011). https://doi.org/10.1143/PTP.125.641
9. Yamazaki, Y., Yamamoto, K., Kadono, D., Toda, A.: Spatiotemporal patterns formed by the dynamics of bistable units with global and asymmetric local interactions. J. Phys. Soc. Jpn. **81** 043002 (2012). https://doi.org/10.1143/JPSJ.81.043002
10. Yamazaki, Y.: Dynamical transitions between stick-slip and steady motions of bistable units with global and asymmetric local interactions. J. Phys. Soc. Jpn. **86** 043001 (2017). https://doi.org/10.7566/JPSJ.86.043001
11. Ohmori, S., Yamazaki, Y.: Comments on statistical properties for cellular automaton models with probabilistic global and asymmetric local rules. J. Phys. Soc. Jpn. **88** 105001 (2019). https://doi.org/10.7566/JPSJ.88.105001
12. Hinrichsen, H.: Non-equilibrium critical phenomena and phase transitions into absorbing states. Adv. Phys. **49** 815 (2000). https://doi.org/10.1080/00018730050198152

Modeling of Porous Hyperelastic Materials Containing Compressible Voids

Xiao Li[1] and Akihiro Matsuda[2(✉)]

[1] Graduate School of Science and Technology, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8577, Ibaraki, Japan
[2] Institute of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8577, Ibaraki, Japan
a_matsuda@kz.tsukuba.ac.jp

Abstract. In this study, a general continuum level constitutive model of porous hyperelastic materials containing compressible voids was developed. A specific expression for the strain energy density function was derived using a representative volume element in the form of a hollow sphere, by assuming a specific deformation state. The strain energy density function of porous hyperelastic materials was found to be related to the initial porosity, the properties of the incompressible hyperelastic matrix material, and the final macroscopic deformation state. Based on the strain energy density function, the relationship between stress and strain was also investigated. A representative example of porous hyperelastic materials was calculated using a Neo-Hookean matrix material with compressible voids to predict the stress-strain behavior. The stress-strain behavior for porous hyperelastic materials is proven to be related to the initial porosity and the final macroscopic deformation state. The effect of compressible voids on the stress-strain behavior of the porous hyperelastic matrix materials under the specific stretching was clarified.

Keywords: Porous hyperelastic materials · Constitutive model · Stress-strain behavior

1 Introduction

Due to the excellent structural and functional properties such as good large deformation capability and low density and energy absorption, porous materials have a wide range of applications in aerospace, electronics, biomedicine and other fields [1, 2]. The impact of voids on the mechanical deformation behavior of porous materials presents a significant engineering concern. Two primary research approaches address this issue. The first involves the generation of modulus-porosity curves through experimental methods. This is coupled with theoretical analyses that employ micromechanics models to predict the mechanical response, ultimately elucidating the influence of porosity on the mechanical behavior of porous materials [3]. The second approach centers on the development of a constitutive model for porous hyperelastic materials [4]. Through this model, a theoretical solution for understanding the effect of porosity on mechanical behavior is derived, grounded in the principles of continuum mechanics. This obviates the need for repetitive experiments on various target materials.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 312–320, 2025.
https://doi.org/10.1007/978-3-031-72794-8_24

Notably, previous studies have not comprehensively explored the influence of the air within voids on the mechanical properties of porous materials. Consequently, there remains an insufficient quantitative evaluation of the impact of compressible voids on the mechanical behavior of porous hyperelastic materials.

In this work, a constitutive model applicable to porous materials was developed. This model accounts for the influence of compressible voids and was designed to be compatible with a wide range of matrix hyperelastic materials. A hollow sphere containing compressible air served as the representation volume element (RVE) for the porous hyperelastic material. Specific assumptions about the final deformation were made, facilitating the derivation of the deformation gradient tensor, the first and second stretch invariants, and the strain energy density function. For illustrative purposes, a Neo-Hookean solid was selected as a computational example to demonstrate how the compressible air inside affects the stress-strain behavior of porous hyperelastic material, both qualitatively and quantitatively.

2 Representative Volume Element

In order to investigate the deformation and mechanical properties of porous hyperelastic materials, a spherical shell was established as the representative volume element, illustrated in Figs. 1 and 2. This type of spherical shell volume element has been commonly utilized in previous studies to model porous materials under external pressure [4, 5]. The influence of the voids diminishes as material points move outward from the center of the spherical shell. When situated at a sufficient distance from the center of the spherical shell, the matrix material exhibits behavior akin to that of an isotropic solid. In this study, the voids exist independently, and it is not assumed that the voids come close and connect to each other. Thus, by arranging and combining such representative volume elements along with isotropic solids, it becomes feasible to simulate porous hyperelastic materials for practical application scenarios.

The initial configuration of the shell is characterized by a void volume fraction f_0, with inner diameter and outer diameter denoted as A and B, respectively, given by $f_0 = A^3/B^3$. The hyperelastic matrix material was considered pointwise incompressible, while the air within the void is presumed to be compressible.

The macroscopic deformation of this sphere was determined by three pairs of external symmetric stretches in the radius direction perpendicular to each other, which are imposed simultaneously. The macroscopic deformation is characterized by these three pairs of stretches collectively, denoted as $\left\{\overline{\lambda}_1, \overline{\lambda}_2, \overline{\lambda}_3\right\}$. In response to these stretches, the spherical shell transforms into an ellipsoid, as depicted in Fig. 2.

In a purely hydrostatic expansion, every material point in the spherical shell undergoes radial movement. However, in general, when the spherical shell transforms into an ellipsoid shell, material point displacements will not be exclusively radial. Nevertheless, obtaining an exact solution for the displacement of each point in the deformed sphere shell is highly challenging. In a study on cavitation in hyperelastic solids, Hang-Sheng et al. [6] proposed a law describing the motion of points within a solid that incorporates internal voids.

$$x_i = \psi_i X_i (nosumoni) \tag{1}$$

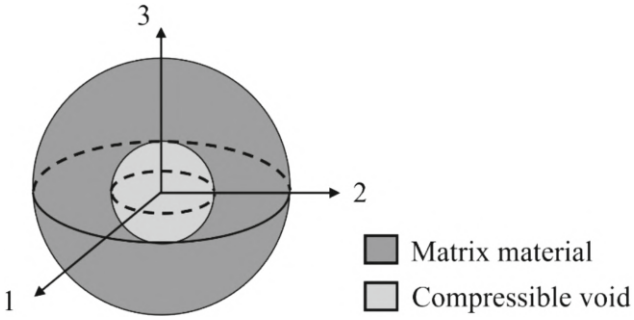


Fig. 1. The undeformed configuration of the RVE

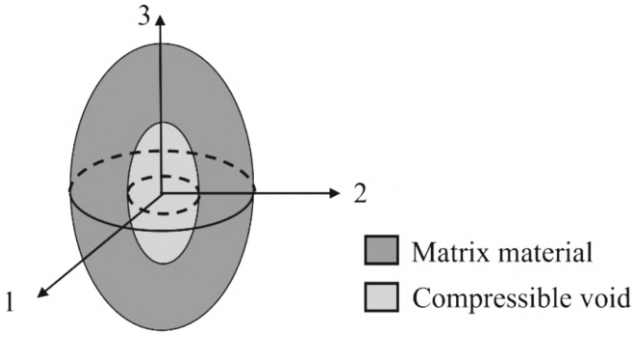


Fig. 2. The deformed configuration of the RVE

Here, x_i represents the coordinate after deformation in a specific direction within a spherical coordinate system, with the sphere's center serving as the origin, while X_i denotes the initial coordinate. ψ_i is a function with respect to R, where R is the distance from the origin. Moreover, in a spherical coordinate system, the coordinates of any point are defined as

$$X_1 = R\,sin\theta\,sin\phi \tag{2}$$

$$X_2 = R\,sin\theta\,cos\phi \tag{3}$$

$$X_3 = R\,cos\theta \tag{4}$$

Using the Eqs. 2, 3 and 4, the deformation gradient tensor $\boldsymbol{F}$ can be derived as

$$[\boldsymbol{F}]_{ij} = \frac{\partial x_i}{\partial X_j} = \frac{X_i X_j}{R}\frac{d\psi_i}{dR} + \psi_i \delta_{ij} \tag{5}$$

where δ_{ij} is the Kronecker delta. It is further assumed that the volume change of air $J_{air} = 1 + P/k$, is associated with the inner air pressure P and the bulk modulus of air k. For illustrative purposes, an arbitrary point situated within the matrix material of

the representative volume element is selected, located at a distance R from the sphere's center. Thus, the local volume change can be given by

$$J \equiv \det \boldsymbol{F} = 1 + \frac{PA^3}{kR^3} \tag{6}$$

Considering Eqs. 2, 3 and 4, Eq. 5 is inserted into Eq. 6 to obtain a differential equation defining the function ψ_i.

$$\psi_1\psi_2\psi_3 + R\psi_1\psi_2\psi_3^{'} + \frac{X_1^2}{R}\left(\psi_1^{'}\psi_3 - \psi_3^{'}\psi_1\right) + \frac{X_2^2}{R}\left(\psi_2^{'}\psi_3 - \psi_3^{'}\psi_2\right) = 1 + \frac{PA^3}{kR^3} \tag{7}$$

Furthermore, Eq. 7 remains valid regardless of the point's coordinates, suggesting that the third and fourth terms of Eq. 7 are always equal to zero, and enabling the development of a system of differential equations concerning ψ_i.

$$\psi_1\psi_2\psi_3 + R\psi_1\psi_2\psi_3^{'} = 1 + \frac{PA^3}{kR^3} \tag{8}$$

$$\psi_1^{'}\psi_3 - \psi_3^{'}\psi_1 = 0 \tag{9}$$

$$\psi_2^{'}\psi_3 - \psi_3^{'}\psi_2 = 0 \tag{10}$$

To address the solution of this system of equations, the boundary condition, which pertains to the microscopic deformation $\overline{J}$ is introduced.

$$\overline{J} = 1 + \frac{PA^3}{kB^3} = \overline{\lambda}_1\overline{\lambda}_2\overline{\lambda}_3 \tag{11}$$

Therefore, it would be possible to express the parameters P and k, in relation to the air, in terms of microscopic stretches. In this manner, utilizing macroscopic stretches and macroscopic volume change enables the determination of ψ_i at different locations on the matrix material. This establishes a theoretical connection between macroscopic and microscopic quantities.

$$\psi_i = \frac{\overline{\lambda}_i}{\overline{J}^{\frac{1}{3}}}\left(1 + \frac{(\overline{J}-1)B^3}{R^3}(1 + ln\frac{R^3}{B^3})\right)^{\frac{1}{3}} \tag{12}$$

Consequently, the microscopic deformation gradient tensor F_{ij} and the microscopic left Cauchy-Green strain tensor B_{ij} for the representative volume element are expressed as follows:

$$F_{ij} = \frac{\overline{\lambda}_i}{\overline{J}^{\frac{1}{3}}} \cdot \left(\frac{\eta - \psi^3}{\psi^2}\frac{X_iX_j}{R^2} + \delta_{ij}\psi\right) \tag{13}$$

$$B_{ij} \equiv F\boldsymbol{F}^T = \frac{\overline{\lambda}_i\overline{\lambda}_j}{\overline{J}^{\frac{2}{3}}}\left\{\left(\frac{\eta - \psi^3}{\psi^2}\right)^2\frac{X_iX_jX_k^2}{R^4} + 2\frac{\eta - \psi^3}{\psi^2}\frac{X_iX_j}{R^2} + \delta_{ij}\psi^2\right\} \tag{14}$$

where $\psi = \overline{\lambda}_i\left(1 + \left(\overline{J} - 1\right)B^3/R^3(1 + lnR^3/B^3)\right)^{\frac{1}{3}}/\overline{J}^{\frac{1}{3}}$, $\eta = \left(\overline{J} - 1\right)B^3/R^3 + 1$. The first and second stretch invariants are defined as

$$I_1 \equiv tr(\boldsymbol{B}) = \frac{1}{\overline{J}^{\frac{2}{3}}}\left\{\left(\frac{\eta^2}{\psi^4} - \psi^2\right)\frac{\overline{\lambda}_1^2X_1^2 + \overline{\lambda}_2^2X_2^2 + \overline{\lambda}_3^2X_3^2}{R^2} + \overline{I}_1\psi^2\right\} \tag{15}$$

$$I_2 \equiv \frac{I_1^2 - tr(\boldsymbol{B}^2)}{2} = \overline{J}^{\frac{2}{3}} \left\{ \frac{\bar{I}_2 \eta^2}{\overline{J}^2 \psi^2} + \frac{1}{R^2} \left(\frac{X_3^2}{\bar{\lambda}_3^2} + \frac{X_2^2}{\bar{\lambda}_2^2} + \frac{X_1^2}{\bar{\lambda}_1^2} \right) \left(\psi^4 - \frac{\eta^2}{\psi^2} \right) \right\} \tag{16}$$

Here, $\bar{I}_1 = \bar{\lambda}_1^2 + \bar{\lambda}_2^2 + \bar{\lambda}_3^2$ is the first invariant of macroscopic stretch, and $\bar{I}_2 = \bar{\lambda}_1^2 \bar{\lambda}_2^2 + \bar{\lambda}_2^2 \bar{\lambda}_3^2 + \bar{\lambda}_3^2 \bar{\lambda}_1^2$ is the second invariant of macroscopic stretch.

Utilizing the stretch invariants, the homogenized strain energy density function $\overline{W}$ of the representative volume element can be deduced through Eqs. 14 and 15.

$$\overline{W} = \frac{1}{V} \int_A^B \int_0^{2\pi} \int_0^{\pi} W(I_1, I_2) R^2 \sin\theta \, d\theta d\phi \, dR \tag{17}$$

Here, $W(I_1, I_2)$ represents the point-by-point strain energy density function of matrix materials. $W(I_1, I_2)$ is characterized by the specific material model, such as the Neo-Hookean solid and Mooney-Rivlin solid. The homogenized strain energy density function $\overline{W}$ for the representative volume element is derived by integrating over the matrix material pointwise and subsequently normalizing by the total volume $V = 4\pi B^3/3$.

Consequently, the expression of the macroscopic homogenized Cauchy stress tensor is established by relying on the stretch invariants and the macroscopic volume change $\overline{J}$.

$$\overline{\boldsymbol{T}} = \frac{2}{\overline{J}} \frac{\partial \overline{W}}{\partial \overline{\boldsymbol{B}}} \overline{\boldsymbol{B}} = \frac{2}{\overline{J}} \frac{\partial \overline{W}}{\partial \bar{I}_1} \overline{\boldsymbol{B}} + \frac{2}{\overline{J}} \frac{\partial \overline{W}}{\partial \bar{I}_2} \left(\bar{I}_1 \overline{\boldsymbol{B}} - \overline{\boldsymbol{B}}^2 \right) + \frac{\partial \overline{W}}{\partial \overline{J}} \boldsymbol{I} \tag{18}$$

where $\boldsymbol{I}$ is the second order identity tensor. Up to this juncture, through the utilization of a hollow sphere characterized by void volume fraction f_0 that serves as a representative volume element, the constitutive model of porous hyperelastic materials is delineated. Particularly, it becomes evident that the mechanical properties are intrinsically linked to the ultimate deformation and void volume fraction f_0, with no dependence on the properties of the confined air.

3 Results Using a Neo-Hookean Solid

Based on the theoretical relationship established in the previous section between stress and deformation, it is imperative to provide a more specific description of the mechanical properties of the porous hyperelastic materials. As an illustrative case for subsequent calculations, the Neo-Hookean solid is chosen to determine the explicit formulation of the Cauchy stress and the von-Mises stress.

Additionally, the strain energy density function corresponding to the Neo-Hookean solid is represented by

$$W_{NH} = C_1(I_1 - 3) = \frac{\mu}{2}(I_1 - 3) \tag{19}$$

where μ denotes the shear modulus. Through the amalgamation of Eqs. 17 and 18, the homogenized strain energy density function $\overline{W}_{NH}$ for the representative volume element is derived as follows:

$$\overline{W}_{NH} = \frac{\mu \bar{I}_1}{6\overline{J}^{\frac{2}{3}}} \left(\int_{f_0}^{1} \frac{\left(\frac{\overline{J}-1}{u} + 1 \right)^2 + 2\left(1 + \frac{\overline{J}-1}{u}(1 + lnu) \right)^2}{\left(1 + \frac{\overline{J}-1}{u}(1 + lnu) \right)^{\frac{4}{3}}} dx \right) - \frac{3\mu}{2}(1 - f_0) \tag{20}$$

where $u = R^3/B^3$. While computing the triple integral, the integrand proves elusive. To address this challenge, an approximation function is proposed for computation as an alternative to utilizing Eq. 18.

$$\overline{W}_{NH} = \frac{\mu \bar{I}_1}{6\bar{J}^{\frac{2}{3}}}\left(\int_{f_0}^{1} A(\bar{J})\mathrm{e}^{B(\bar{J})u} + C(\bar{J})\mathrm{e}^{D(\bar{J})u}du\right) - \frac{3\mu}{2}(1-f_0) \tag{21}$$

where A, B, C and D are thoughtfully selected as elementary functions, each dependent on the macroscopic volume change $\bar{J}$. Furthermore, due to the discontinuity inherent in Eq. 18, there is a limitation on the range of the void volume fraction f_0 for the representative volume element. Therefore, the theoretical solution $\overline{W}_{NH}$ remains valid exclusively when the void volume fraction f_0 exceeds e^{-1}, approximately 0.3679, where e is the Napier's constant.

Ultimately, with the approximation function mentioned in Eq. 19, the Cauchy stress $\overline{\boldsymbol{T}}$ can be expressed as a function of the initial void volume fraction f_0 and microscopic deformation state $\bar{J}$.

$$\overline{\boldsymbol{T}} = \frac{2}{\bar{J}}\frac{\partial\left\{\frac{\mu}{2}\frac{\bar{I}_1}{3\bar{J}^{\frac{2}{3}}}\left(1-e^{f_0}\right)\left(\frac{Ae^B}{B}+\frac{Ce^D}{D}\right)\right\}}{\partial\overline{\boldsymbol{B}}}\overline{\boldsymbol{B}} = \frac{2}{\bar{J}}\frac{\partial\overline{W}}{\partial\bar{I}_1}\overline{\boldsymbol{B}} + \frac{\partial\overline{W}}{\partial\bar{J}}\boldsymbol{I} \tag{22}$$

$$\frac{\partial\overline{W}}{\partial\bar{I}_1} = \frac{\mu}{6\bar{J}^{\frac{2}{3}}}\left(1-e^{f_0}\right)\left(\frac{Ae^B}{B}+\frac{Ce^D}{D}\right) \tag{23}$$

$$\frac{\partial\overline{W}}{\partial\bar{J}} = \frac{\mu}{6\bar{J}^{\frac{2}{3}}}\left(1-e^{f_0}\right)\left(\frac{A'B-AB'+AB'B}{B^2}\mathrm{e}^B + \frac{C'D-CD'+CD'D}{D^2}\mathrm{e}^D\right)$$

$$-\frac{2\mu}{9\bar{J}^{\frac{5}{3}}}\left(\frac{Ae^B}{B}+\frac{Ce^D}{D}\right) \tag{24}$$

where $A\prime$, $B\prime$, $C\prime$ and $D\prime$ represent differentiation with respect to $\bar{J}$. For investigating the stress-strain behavior under varying deformation conditions and its correlation with the initial void volume fraction f_0, a special microscopic deformation $\left\{\bar{\lambda}_1 = \lambda, \bar{\lambda}_2 = 1, \bar{\lambda}_3 = 1\right\}$ is imposed on the representative volume element.

In Fig. 3, under the aforementioned microscopic deformation, in conjunction with Eqs. 22,23, and 24, the relationship between the macroscopic Cauchy stress and the true strain of the theoretical constitutive model is delineated. Additionally, three different levels of initial void volume fraction, 0.5, 0.6, and 0.8, are investigated to demonstrate how the void volume fraction affects the mechanical properties. The solid line depicts the macroscopic Cauchy stress-strain correlation in the axial direction, where the principal macroscopic stretch is denoted as $\bar{\lambda}_1 = \lambda$, while the dashed line illustrates the relationship in the other two constrained directions, with the principal macroscopic stretch identified as $\bar{\lambda}_2 = \bar{\lambda}_3 = 1$.

Firstly, in a general context, the stresses induced in both these deformation directions are influenced by the interplay of air pressure and matrix material stiffness. At the onset of initial deformation, significant stresses rapidly emerge in the axial and constrained directions, primarily due to the influence of air pressure. Subsequently, with the progression of deformation, the air pressure gradually diminishes, leading to a reduction in its impact on the stresses in both directions. Meanwhile, the contribution of the matrix material stiffness becomes increasingly prominent. In this context, the stress generated in the tensile direction experiences a progressive increase, while the stress in the constraint direction gradually recovers to the stress due to Poisson's effect only, and thus becomes smaller and smaller and stabilizes.

Secondly, increased initial void volume fraction evidently results in decreased porous hyperelastic material stiffness. Consequently, lower stress levels can be achieved under equivalent deformation conditions for a larger initial void volume fraction f_0. This effect is prominently noticeable in both the axial and constraint directions.

Drawing upon the findings concerning the Cauchy stress, the data pertaining to the von-Mises stress against true strain is presented in Fig. 4. Analogous to the insights provided by Fig. 3, as void volume fraction f_0 rises, the elastic compliance of the representative volume element diminishes, resulting in a more linear alteration of the von-Mises stress.

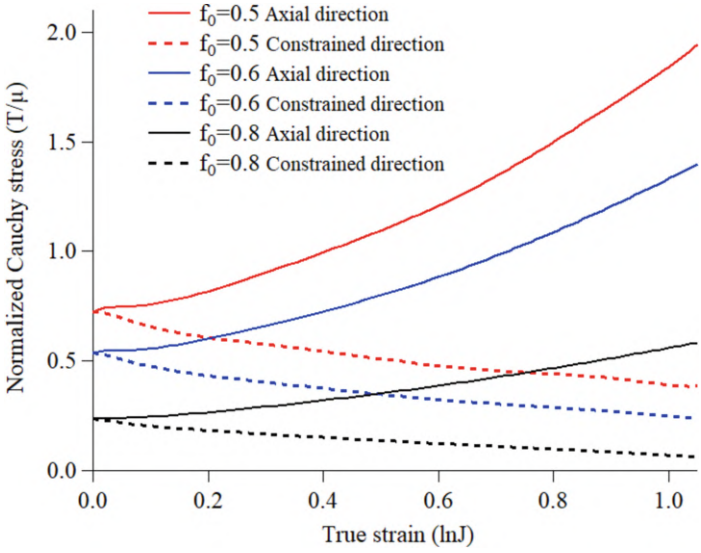


Fig. 3. The Cauchy stress-strain relationship under different void volume fraction f_0

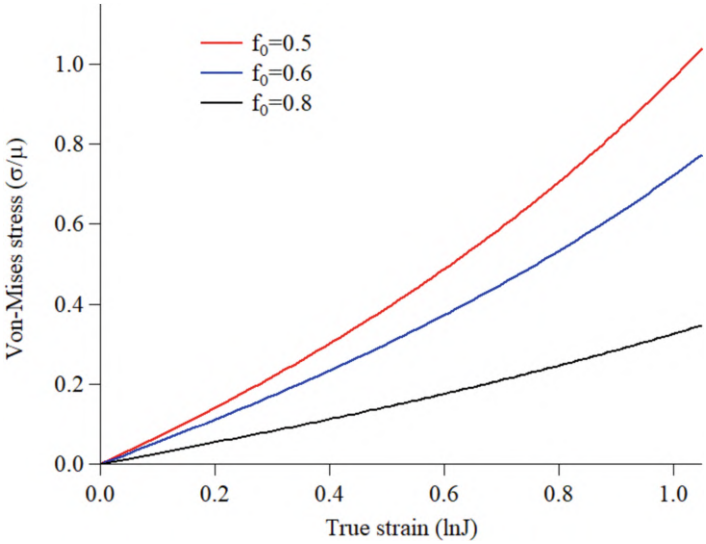


Fig. 4. The von-Mises stress-strain relationship under different void volume fraction f_0

4 Conclusion

In this study, a constitutive model for porous hyperelastic materials has been proposed. This model exhibits versatility, accommodating a wide spectrum of matrix materials, including Neo-Hookean and Mooney-Rivlin solids, among others. A hollow sphere containing compressible air was adopted as the representative volume element. Through an assumption of the final microscopic deformation state and utilizing the approaches outlined in [5] for characterizing every point during the deformation, we successfully derived the deformation gradient tensor, first and second stretch invariants, and formulated the strain energy density function. Our findings describe a detailed correlation between the strain energy density function, the final deformation, and the void volume fraction.

For a detailed demonstration, the Neo-Hookean solid was chosen as an illustrative computational example. The strain energy density function was subsequently employed to establish the relationship between the Cauchy stress and true strain. While this study does not encompass the entire void porosity range from 0 to 1, it nonetheless enables qualitative and quantitative insights into the impact of internal compressible air on the stress-strain behavior of porous hyperelastic materials, particularly at higher void porosity levels.

References

1. Liu, P., Chen, G.F.: Porous materials: processing and applications. Elsevier (2014)
2. Ba, A., Kovalenko, A., Aristégui, C., Mondain-Monval, O., Brunet, T.: Soft porous silicone rubbers with ultra-low sound speeds in acoustic metamaterials. Sci. Rep. **7**(1), 40106 (2017)

3. Lu, G., Xiao, Z.M.: Mechanical properties of porous materials. J. Porous Mater. **6**(4), 359–368 (1999)
4. Danielsson, M., Parks, D.M., Boyce, M.C.: Constitutive modeling of porous hyperelastic materials. Mech. Mater. **36**(4), 347–358 (2004)
5. Kakavas, P.A.: Influence of the cavitation on the stress–strain fields of compressible Blatz–Ko materials at finite deformation. Int. J. Solids Struct. **39**(3), 783–795 (2002)
6. Hang-Sheng, H., Abeyaratne, R.: Cavitation in elastic and elastic-plastic solids. J. Mech. Phys. Solids **40**(3), 571–592 (1992)

Sports Dynamics and Biomechanics

Head Speed Generating Mechanisms During Turning Phases of Athletic Hammer Throwing

Sekiya Koike[1](✉), Tempei Tominaga[1], Keigo B. Ohyama[1], and Alexander P. Willmott[2]

[1] University of Tsukuba, 1-1-1 Ten-nodai, Tsukuba, Ibaraki, Japan
koike.sekiya.fp@u.tsukuba.ac.jp
[2] University of Lincoln, Lincoln LN6 7TS, Lincolnshire, UK

Abstract. The purpose of this study was to quantify the hammer head speed generating mechanisms of male hammer throwers, using a simplified model for the thrower and hammer. Fourteen male hammer throwers of differing skill level performed maximum effort throws in an indoor experimental facility. The dynamic contributions to the generation of hammer head centre of gravity speed from (1) joint moments, (2) gravity, (3) a generalised initial velocity term, and (4) an acceleration constraint term relating to the proximal end of the simplified system were derived from the equations of motion for the individual segments and constraint conditions arising from the connection of adjacent segments. The motion-dependent term (MDT), a nonlinear term consisting of centrifugal force, Coriolis force and gyroscopic effect moment components, was the greatest contributor to hammer head speed, with the centrifugal force component especially important. The components of the MDT would be key factors in differentiating between throwers of different skill level.

Keywords: Multi-body Dynamics · Forward Dynamics Contribution Analysis · Equations of Motion · Simplified Modelling · Multi-revolutionary Motion

1 Introduction

The hammer, used in athletics, consists of three parts: a grip handle, a flexible wire, and a metal ball - usually called the hammer head - whose mass is 7.26 kg for men and 4.0 kg for women. The individual parts are connected to their adjacent parts via passive joints which do not exert torque. The hammer thrower grasps the handle with both hands and swings the hammer around the body several times before moving into three or four turning motions of the whole body [1–6]. The dynamics of high-speed sporting motions can be very complex [7, 8], and the head speed generating mechanisms in hammer throwing are still poorly understood from the viewpoint of the dynamics of a multi-body system that includes the human body, the passive joints of the hammer and a flexible wire.

Sporting motions are characterized by high speed and/or acceleration primarily driven by joint torques [9, 10]. The key determinants of human movement are joint

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 323–334, 2025.
https://doi.org/10.1007/978-3-031-72794-8_25

torques, sometimes quantified as joint moments. Kinematic data are commonly analysed through an inverse dynamics process, to quantify the individual joint torques. In this calculation method, the joint forces and joint torques required for the motion are calculated sequentially from the terminal segment. While kinetic analysis does provide information on the joint torques during human movements, it does not directly reveal how torque affects variables such as segment velocities, acceleration of the whole-body centre of gravity, and joint forces during human motion.

The equations of motion for articulated systems, such as the human body, enable the analysis of their dynamics. Causal relationships between joint torques and the resultant motion can be established, ensuring more objective analysis. However joint torques are not the only factor that cause movement. Since the dynamics of multi-segment systems, such as the combined hammer-thrower system, can be determined in detailed by the equations of motion for the system, speed generating mechanisms can be investigated directly. The purpose of this study was, therefore, to use a simplified modelling of the thrower and the hammer to quantify the hammer head-speed generating mechanisms of male hammer throwers, including nationally high-ranked athletes, and to explore how these mechanisms vary between throwers of different skill level.

2 Methods

2.1 Acquisition and Modelling of Motion Data for Hammer Throwing

Fourteen male hammer throwers participated in this study, performing maximum effort hammer throws in an indoor experimental facility (Fig. 1). Three-dimensional coordinate data for 59 reflective markers (47 attached to the individual participant's body and 12 to the hammer) and the ground reaction forces exerted by the legs were measured with a

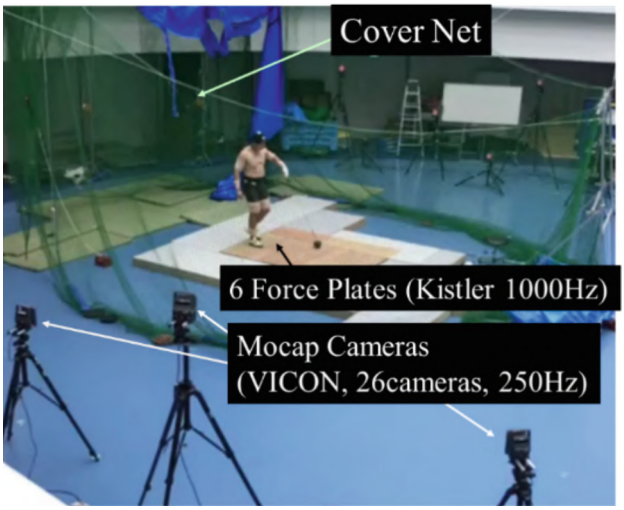


Fig. 1. A photo of the indoor facility used for motion data collection.

motion capture system with 26 infrared cameras (VICON-MX, 250 Hz, Vicon Motion Systems Ltd) and six force platforms (Kistler Inc., 1000 Hz, Winterthur, Switzerland).

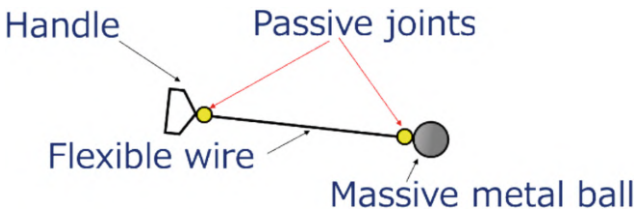


Fig. 2. The structure of an athletics hammer.

As mentioned above, the three segments of the hammer connect with their adjustment segments via passive joints that do not exert torque (Fig. 2). The hammer can be modeled as three linked segments without actuators, in what is sometimes called an under-actuated system.

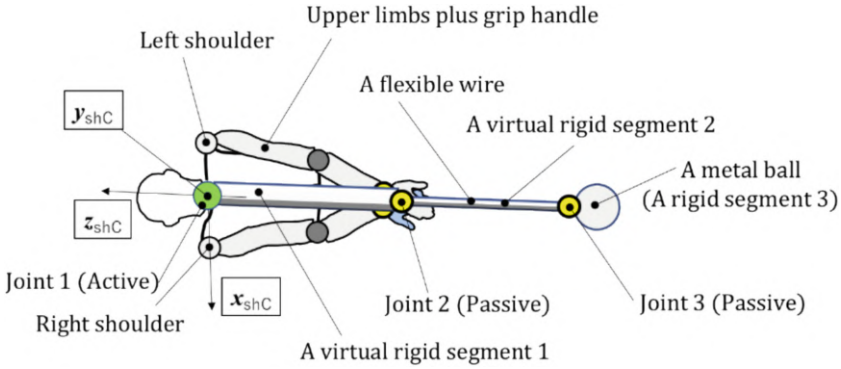


Fig. 3. A detailed model for the combined hammer-thrower system.

During each individual turn, the thrower constructs a fixed triangular structure, formed from the upper limbs and the handle, that maintains its shape against the large tensile force applied along the wire during the turning motions. A simplified model of the thrower's upper extremities plus hammer was created as follows:

1) The upper arms and handle form a rigid segment (Fig. 3) whose inertial parameters were determined from body segment parameters for the thrower's upper limbs. The proximal end of virtual rigid segment 1 (in Fig. 4) is constrained by the acceleration of the centre point of the shoulder joints.
2) The flexible wire (segment 2 in Fig. 4) is assumed to be a lightweight rigid segment because the wire is highly tensive during the turning motions, due to the large centrifugal force exerted along the wire.
3) Joints 2 and 3 are assumed to be passive joints that do not exert joint torques, due to the structure of the hammer (as shown in Fig. 2).

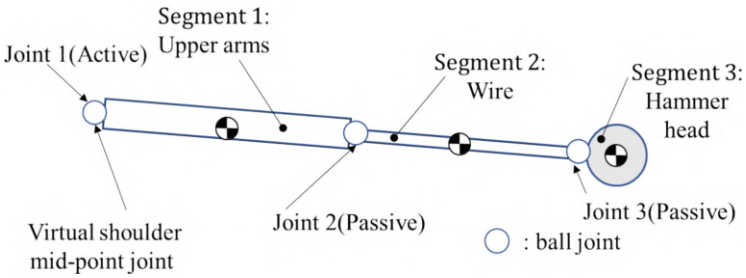


Fig. 4. A simplified model for the combined hammer-thrower system.

2.2 Derivation of Equations of Motion for the Simplified System

The dynamical equations for individual segments can be expressed in a matrix form with respect to all segments as follows [9],

$$\boldsymbol{M}\dot{\boldsymbol{V}} = \boldsymbol{P}\boldsymbol{F} + \boldsymbol{Q}\boldsymbol{N} + \boldsymbol{H} + \boldsymbol{G} \tag{1}$$

where $\boldsymbol{M}$ is the inertia matrix, and $\boldsymbol{V}$ is the generalised velocity vector consisting of linear velocity vectors and angular velocity vectors for all the segments. $\boldsymbol{P}$ is the coefficient matrix for vector $\boldsymbol{F}$, which contains all joint force vectors. $\boldsymbol{Q}$ is the coefficient matrix for vector $\boldsymbol{N}$, which contains all joint moment vectors. $\boldsymbol{H}$ is the gyroscopic effect moment vector, and $\boldsymbol{G}$ is the vector representing the gravitational force. The individual vectors $\boldsymbol{V}, \boldsymbol{F}, \boldsymbol{N}, \boldsymbol{H}, \boldsymbol{G}$, and matrices $\boldsymbol{M}, \boldsymbol{P}, \boldsymbol{Q}$, are written as follows:

$$\boldsymbol{V} = \begin{bmatrix} \dot{\boldsymbol{x}}_1^{\mathrm{T}} & \boldsymbol{\omega}_1^{\mathrm{T}} & \dot{\boldsymbol{x}}_2^{\mathrm{T}} & \boldsymbol{\omega}_2^{\mathrm{T}} & \dot{\boldsymbol{x}}_3^{\mathrm{T}} & \boldsymbol{\omega}_3^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \boldsymbol{F} = \begin{bmatrix} \boldsymbol{f}_1^{\mathrm{T}} & \boldsymbol{f}_2^{\mathrm{T}} & \boldsymbol{f}_3^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \boldsymbol{N} = \begin{bmatrix} \boldsymbol{n}_1^{\mathrm{T}} & \boldsymbol{n}_2^{\mathrm{T}} & \boldsymbol{n}_3^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \tag{2}$$

$$\boldsymbol{M} = \text{block diag}\{m_1\boldsymbol{E}, \hat{\boldsymbol{I}}_1, m_2\boldsymbol{E}, \hat{\boldsymbol{I}}_2, m_3\boldsymbol{E}, \hat{\boldsymbol{I}}_3\} \tag{3}$$

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{E} & -\boldsymbol{E} & \boldsymbol{O} \\ \left[\boldsymbol{r}_{1,\overline{\text{cg}-\text{P}}}\times\right] & -\left[\boldsymbol{r}_{1,\overline{\text{cg}-\text{D}}}\times\right] & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{E} & -\boldsymbol{E} \\ \boldsymbol{O} & \left[\boldsymbol{r}_{2,\overline{\text{cg}-\text{P}}}\times\right] & -\left[\boldsymbol{r}_{2,\overline{\text{cg}-\text{D}}}\times\right] \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{E} \\ \boldsymbol{O} & \boldsymbol{O} & \left[\boldsymbol{r}_{3,\overline{\text{cg}-\text{P}}}\times\right] \end{bmatrix}, \boldsymbol{Q} = \begin{bmatrix} \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{E} & -\boldsymbol{E} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{E} & -\boldsymbol{E} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{E} \end{bmatrix},$$

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{O}_{3\times1} \\ -\boldsymbol{\omega}_1 \times (\hat{\boldsymbol{I}}_1\boldsymbol{\omega}_1) \\ \boldsymbol{O}_{3\times1} \\ -\boldsymbol{\omega}_2 \times (\hat{\boldsymbol{I}}_2\boldsymbol{\omega}_2) \\ \boldsymbol{O}_{3\times1} \\ -\boldsymbol{\omega}_3 \times (\hat{\boldsymbol{I}}_3\boldsymbol{\omega}_3) \end{bmatrix}, \boldsymbol{G} = \begin{bmatrix} m_1\boldsymbol{g} \\ \boldsymbol{O}_{3\times1} \\ m_2\boldsymbol{g} \\ \boldsymbol{O}_{3\times1} \\ m_3\boldsymbol{g} \\ \boldsymbol{O}_{3\times1} \end{bmatrix}, [\boldsymbol{r}\times] = \begin{bmatrix} 0 & -r_z & r_y \\ r_z & 0 & -r_x \\ -r_y & r_x & 0 \end{bmatrix}, \tag{4}$$

where m_i is the mass of the i-th segment, $\boldsymbol{x}_i$ is the position vector of the segment's centre of gravity (CG), $\boldsymbol{\omega}_i$ is the angular velocity vector of the i-th segment, and the

vectors $\boldsymbol{r}_i$ with barred subscripts cg-P and cg-D denote position vectors running from the CG of the segment to its proximal and distal endpoints, respectively. Matrices $\boldsymbol{O}$ and $\boldsymbol{E}$ are a 3 × 3 sized zero matrix and unit matrix, respectively. $\hat{\boldsymbol{I}}_i$ is the inertia matrix of the i-th segment expressed in the global reference coordinate system, and $\boldsymbol{g}$ is the gravitational acceleration vector. Matrix $[\boldsymbol{r}\times]$ denotes the outer product of a vector $\boldsymbol{r}$ whose components are r_x r_y and r_z, as shown in the last part of Eq. (4).

The equations for the constraint conditions arising from the connection of adjacent segments at joints can be shown as follows [10]:

$$\boldsymbol{C}\dot{\boldsymbol{V}} + \dot{\boldsymbol{C}}\boldsymbol{V} = \ddot{\boldsymbol{\eta}} + \boldsymbol{B}\ddot{\boldsymbol{x}}_{\mathrm{ShC}}, \tag{5}$$

where $\boldsymbol{C}$ is the coefficient matrix for vector $\boldsymbol{V}$, $\dot{\boldsymbol{\eta}}$ is the vector consisting of the differences between the distal and proximal point velocity vectors at individual joints [9]. Matrix $\boldsymbol{B}$ is the coefficient matrix of the virtual shoulder mid-point joint constraint acceleration $\ddot{\boldsymbol{x}}_{\mathrm{ShC}}$.

Vectors $\boldsymbol{r^*}_i$ with subscripts cg-D and cg-P represent the velocity vectors, expressed in the i-th segment coordinate system, from a segment's CG to its distal and proximal ends with respect to extension and contraction of the position vector of $\boldsymbol{r}_i$.

$$\boldsymbol{C} = \begin{bmatrix} -\boldsymbol{E} & \left[\boldsymbol{r}_{1,\overline{\mathrm{cg-P}}}\times\right] & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{E} & -\left[\boldsymbol{r}_{1,\overline{\mathrm{cg-D}}}\times\right] & -\boldsymbol{E} & \left[\boldsymbol{r}_{2,\overline{\mathrm{cg-P}}}\times\right] & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{E} & -\left[\boldsymbol{r}_{2,\overline{\mathrm{cg-D}}}\times\right] & -\boldsymbol{E} & \left[\boldsymbol{r}_{3,\overline{\mathrm{cg-P}}}\times\right] \end{bmatrix},$$

$$\boldsymbol{\eta} = \begin{bmatrix} \ddot{\boldsymbol{r}}^*_{1,\overline{\mathrm{cg-P}}} \\ \ddot{\boldsymbol{r}}^*_{2,\overline{\mathrm{cg-P}}} - \ddot{\boldsymbol{r}}^*_{1,\overline{\mathrm{cg-D}}} \\ \ddot{\boldsymbol{r}}^*_{3,\overline{\mathrm{cg-P}}} - \ddot{\boldsymbol{r}}^*_{2,\overline{\mathrm{cg-D}}} \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} \boldsymbol{E} \\ \boldsymbol{O} \\ \boldsymbol{O} \end{bmatrix} \tag{6}$$

Combining Eqs. (1) and (5), the equation of motion for the simplified system can be derived as follows.

$$\dot{\boldsymbol{V}} = \boldsymbol{A}_{\boldsymbol{N}}\boldsymbol{N} + \overline{\boldsymbol{A}}_{\boldsymbol{V}}\boldsymbol{V} + \boldsymbol{A}_{\boldsymbol{G}}\boldsymbol{G} + \boldsymbol{A}_{\boldsymbol{nr}}\boldsymbol{n}_{\boldsymbol{r}} + \boldsymbol{A}_{\boldsymbol{\eta}}\ddot{\boldsymbol{\eta}} + \boldsymbol{A}_{\mathrm{Acc,ShC}}\ddot{\boldsymbol{x}}_{\mathrm{ShC}} \tag{7}$$

where the coefficient matrices $\boldsymbol{A}_{\boldsymbol{N}}, \overline{\boldsymbol{A}}_{\boldsymbol{V}}, \boldsymbol{A}_{\boldsymbol{G}}, \boldsymbol{A}_{\boldsymbol{nr}}, \boldsymbol{A}_{\boldsymbol{\eta}}$, and $\boldsymbol{A}_{\mathrm{Acc,ShC}}$ are shown as follows.

$$\left.\begin{aligned} \boldsymbol{A}_{\boldsymbol{N}} &= \boldsymbol{M}^{-1}\boldsymbol{Q} - \boldsymbol{M}^{-1}\boldsymbol{P}\left(\boldsymbol{C}\boldsymbol{M}^{-1}\boldsymbol{P}\right)^{-1}\boldsymbol{C}\boldsymbol{M}^{-1}\boldsymbol{Q}, \\ \overline{\boldsymbol{A}}_{\boldsymbol{V}} &= \boldsymbol{M}^{-1}\left[\overline{\boldsymbol{H}} - \boldsymbol{P}\left(\boldsymbol{C}\boldsymbol{M}^{-1}\boldsymbol{P}\right)^{-1}\left(\boldsymbol{C}\boldsymbol{M}^{-1}\overline{\boldsymbol{H}} + \dot{\boldsymbol{C}}\right)\right], \\ \boldsymbol{A}_{\boldsymbol{G}} &= \boldsymbol{M}^{-1} - \boldsymbol{M}^{-1}\boldsymbol{P}\left(\boldsymbol{C}\boldsymbol{M}^{-1}\boldsymbol{P}\right)^{-1}\boldsymbol{C}\boldsymbol{M}^{-1}, \\ \boldsymbol{A}_{\boldsymbol{nr}} &= \boldsymbol{M}^{-1}\boldsymbol{Q}_{\boldsymbol{r}} - \boldsymbol{M}^{-1}\boldsymbol{P}\left(\boldsymbol{C}\boldsymbol{M}^{-1}\boldsymbol{P}\right)^{-1}\boldsymbol{C}\boldsymbol{M}^{-1}\boldsymbol{Q}_{\boldsymbol{r}}, \\ \boldsymbol{A}_{\boldsymbol{\eta}} &= \boldsymbol{M}^{-1}\boldsymbol{P}\left(\boldsymbol{C}\boldsymbol{M}^{-1}\boldsymbol{P}\right)^{-1}, \\ \boldsymbol{A}_{\mathrm{Acc,ShC}} &= \boldsymbol{M}^{-1}\boldsymbol{P}\left(\boldsymbol{C}\boldsymbol{M}^{-1}\boldsymbol{P}\right)^{-1}\boldsymbol{B} \end{aligned}\right\} \tag{8}$$

where $\boldsymbol{A}_{\mathbf{N}}$ is the coefficient matrix for active moment vector $\boldsymbol{N}$, and matrix $\overline{\boldsymbol{A}}_{\boldsymbol{V}}$ is the coefficient matrix for vector $\boldsymbol{V}$, which is the generalised velocity vector consisting of the linear and angular velocity vectors. $\overline{\boldsymbol{A}}_{\boldsymbol{V}}$ is a function of the angular velocity vectors

$\boldsymbol{\omega}_i (i = 1, 2, 3)$ that are components of vector $\boldsymbol{V}$, because it contains the time derivative of $\boldsymbol{C}$, as shown in Eq. (8). Matrix $\boldsymbol{A}_{\mathbf{G}}$ is the coefficient matrix for the gravitational force vector $\boldsymbol{G}$, matrix $\boldsymbol{A}_{\boldsymbol{nr}}$ is the coefficient matrix for the residual error joint moment vector $\boldsymbol{n}_{\boldsymbol{r}}$, matrix $\boldsymbol{A}_{\boldsymbol{\eta}}$ is the coefficient matrix for the vector $\ddot{\boldsymbol{\eta}}$, $\boldsymbol{A}_{\mathrm{Acc,ShC}}$ is the coefficient matrix for the virtual shoulder mid-point joint constraint acceleration vector $\ddot{\boldsymbol{x}}_{\mathrm{ShC}}$, and matrix $\boldsymbol{Q}_{\boldsymbol{r}}$ is the coefficient matrix for vector $\boldsymbol{n}_{\boldsymbol{r}}$.

2.3 Head-Speed Generating Mechanisms in Hammer Throwing

Forward dynamic contributions to the generation of Δs_{HH} (the increment in hammer head CG speed s_{HH} with respect to the initial hammer head CG speed) from (1) the joint moment term, (2) the motion-dependent term, (3) the gravitational force term, (4) the residual error joint moment term, (5) a segment length fluctuation term, and (6) an acceleration constraint term relating to the proximal end of the simplified system at the virtual mid-shoulder were derived from the combination of Eq. (1), the equations of motion for the individual segments, and Eq. (5), the equations for the constraint conditions for joint accelerations arising from the connection of adjacent segments [10] as follows:

$$\Delta s_{\mathrm{HH}} = C^{\#}{}_{\boldsymbol{N}} + C^{\#}{}_{\mathrm{MDT}} + C^{\#}{}_{\boldsymbol{G}} + C^{\#}{}_{\boldsymbol{nr}} + C^{\#}{}_{\boldsymbol{\eta}} + C^{\#}{}_{\mathrm{Acc,ShC}} \tag{9}$$

where the individual contributions $C^{\#}{}_{\boldsymbol{N}}$, $C^{\#}{}_{\mathrm{MDT}}$, $C^{\#}{}_{\boldsymbol{G}}$, $C^{\#}{}_{\boldsymbol{nr}}$, $C^{\#}{}_{\boldsymbol{\eta}}$, and $C^{\#}{}_{\mathrm{Acc,ShC}}$ are forward dynamics contributions induced by the direct effects of the shoulder joint active moment vector, the motion-dependent term, the gravitational force, the residual error of the joint moments, segment length fluctuation, and the virtual shoulder mid-point acceleration constraint terms, respectively. The superscript # on the values $C^{\#}$ representing contributions in Eq. (9) is provided to distinguish them from matrix $\boldsymbol{C}$ in Eq. (5).

$$\left.\begin{aligned} C^{\#}{}_{\boldsymbol{N}} &= \int \boldsymbol{e}_{\mathrm{HH}}^{\mathrm{T}} \boldsymbol{S}_{\mathrm{HH}} \boldsymbol{A}_{\underline{\boldsymbol{N}}} \boldsymbol{N} \mathrm{d}t, \\ C^{\#}{}_{\mathrm{MDT}} &= \int \boldsymbol{e}_{\mathrm{HH}}^{\mathrm{T}} \boldsymbol{S}_{\mathrm{HH}} \boldsymbol{A}_{\boldsymbol{V}} \boldsymbol{V} \mathrm{d}t, \\ C^{\#}{}_{\boldsymbol{G}} &= \int \boldsymbol{e}_{\mathrm{HH}}^{\mathrm{T}} \boldsymbol{S}_{\mathrm{HH}} \boldsymbol{A}_{\boldsymbol{G}} \boldsymbol{G} \mathrm{d}t, \\ C^{\#}{}_{\boldsymbol{nr}} &= \int \boldsymbol{e}_{\mathrm{HH}}^{\mathrm{T}} \boldsymbol{S}_{\mathrm{HH}} \boldsymbol{A}_{\boldsymbol{nr}} \boldsymbol{n}_{\boldsymbol{r}} \mathrm{d}t, \\ C^{\#}{}_{\boldsymbol{\eta}} &= \int \boldsymbol{e}_{\mathrm{HH}}^{\mathrm{T}} \boldsymbol{S}_{\mathrm{HH}} \boldsymbol{A}_{\boldsymbol{\eta}} \ddot{\boldsymbol{\eta}} \mathrm{d}t, \\ C^{\#}{}_{\mathrm{Acc,ShC}} &= \int \boldsymbol{e}_{\mathrm{HH}}^{\mathrm{T}} \boldsymbol{S}_{\mathrm{HH}} \boldsymbol{B} \ddot{\boldsymbol{x}}_{\mathrm{ShC}} \mathrm{d}t \end{aligned}\right\} \tag{10}$$

where $\boldsymbol{e}_{\mathrm{HH}}$ denotes the unit vector for the linear velocity vector of the hammer-head segment's CG, and matrix $\boldsymbol{S}_{\mathrm{HH}}$ is the selective matrix extracting the linear velocity vector of the hammer-head segment's CG from the generalised velocity vector $\boldsymbol{V}$

$$\boldsymbol{S}_{\mathrm{HH}} = \text{block diag}\{\boldsymbol{O}, \boldsymbol{O}, \boldsymbol{O}, \boldsymbol{O}, \boldsymbol{E}, \boldsymbol{O}\} \tag{11}$$

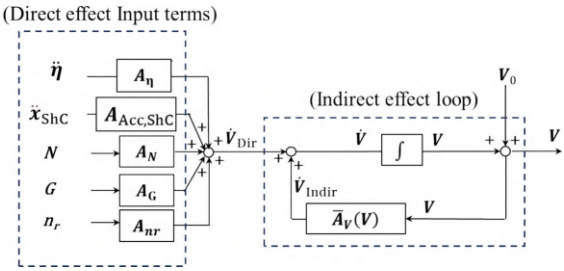


Fig. 5. A block diagram representing the relationships between accelerations arising from the direct and indirect effects of inputs (e.g., joint moments, gravity and other terms) and the generalised velocity vectors. Plain arrows indicate multiplication of the input vector with the matrix inside the box to which the arrow is pointing.

2.4 Nonlinear Phenomena in Head-Speed Generating Mechanisms

When the motion-dependent term (MDT) showed significant contributions to the generation of the hammer head CG speed, a conversion algorithm [10] was applied to divide the MDT into its components. The MDT, including the effects of the centrifugal force and Coriolis force, is a highly nonlinear term and nonlinearity is crucial in the head-speed generating mechanisms in hammer throwing. In the conversion algorithm [10], a forward-difference approximation formula was applied to quantify the indirect effects of input terms to the generation of the generalised velocity vector for the initial generalised velocity term in the head speed generating mechanism, as described below (see Appendix 2 of [10] for details).

In order to quantify the indirect effect of the individual input terms, the generalised velocity vector $\boldsymbol{V}(k)$ in the matrix $\bar{\boldsymbol{A}}_{\boldsymbol{V}}(\boldsymbol{V}(k))$ in Eq. (2) must be replaced with the generalised velocity vector $\tilde{\boldsymbol{V}}\ (k)$ measured at the k-th time instant:

$$\boldsymbol{V}(k+1) = \Delta t \dot{\boldsymbol{V}}_{\mathrm{Dir}}(k) + \tilde{\boldsymbol{\Psi}}_{\boldsymbol{V}}(k)\boldsymbol{V}(k), \tag{12}$$

where matrix $\tilde{\boldsymbol{\Psi}}_{\boldsymbol{V}}$ is the coefficient matrix for the generalised velocity vector $\boldsymbol{V}(k)$, written as follows:

$$\tilde{\boldsymbol{\Psi}}_{\boldsymbol{V}}(k) = \boldsymbol{E}_{18} + \Delta t \bar{\boldsymbol{A}}_{\boldsymbol{V}}\left(\tilde{\boldsymbol{V}}\ (k)\right) \tag{13}$$

where $\boldsymbol{E}_{18}$ denotes a unit matrix with eighteen rows and columns.

With use of the recurrence formula Eq. (12), the generalised velocity vector at time instant k can be expressed by the following equation:

$$\boldsymbol{V}(k) = \Delta t \dot{\boldsymbol{V}}_{\mathrm{Dir}}(k-1) + \Delta t \sum_{h=1}^{k-1} \left[\left\{ \prod_{j=h}^{k-1} \tilde{\boldsymbol{\Psi}}_{\boldsymbol{V}}(j) \right\} \dot{\boldsymbol{V}}_{\mathrm{Dir}}(h-1) \right] + \left\{ \prod_{j=0}^{k-1} \tilde{\boldsymbol{\Psi}}_{\boldsymbol{V}}(j) \right\} \boldsymbol{V}(0), \tag{14}$$

where the function Π denotes the factorial function shown as

$$\prod_{j=h}^{k} \tilde{\boldsymbol{\Psi}}_{\boldsymbol{V}}(j) = \tilde{\boldsymbol{\Psi}}_{\boldsymbol{V}}(k)\tilde{\boldsymbol{\Psi}}_{\boldsymbol{V}}(k-1)\tilde{\boldsymbol{\Psi}}_{\boldsymbol{V}}(k-2)\cdots\tilde{\boldsymbol{\Psi}}_{\boldsymbol{V}}(h), \tag{15}$$

and Δt is the discrete time interval. Vector $\dot{\boldsymbol{V}}_{\mathrm{Dir}}(k)$ is the sum of input terms at time instant k, expressed as follows (as shown in Fig. 5):

$$\begin{aligned}\dot{\boldsymbol{V}}_{\mathrm{Dir}}(k) = {} & \boldsymbol{A}_{\boldsymbol{N}}(k)\boldsymbol{N}(k) + \boldsymbol{A}_{\boldsymbol{G}}(k)\boldsymbol{G}(k) + \boldsymbol{A}_{\boldsymbol{nr}}(k)\boldsymbol{n}_{\boldsymbol{r}}(k) + \boldsymbol{A}_{\boldsymbol{\eta}}(k)\ddot{\boldsymbol{\eta}}(k) \\ & + \boldsymbol{A}_{\mathrm{Acc,ShC}}(k)\ddot{\boldsymbol{x}}_{\mathrm{ShC}}(k)\end{aligned} \tag{16}$$

From Eqs. (12), (13) and (14), the contribution from the initial generalised velocity vector at the start of the analysis to the generation of head-speed due to the indirect effect of the initial generalised velocity vector $\boldsymbol{V}(0)$ can be obtained using the following equation:

$$s_{\mathrm{HH}}(k) = \boldsymbol{e}_{\mathrm{HH}}^{\mathrm{T}}(k)\boldsymbol{S}_{\mathrm{HH}}\left\{\prod_{j=0}^{k-1}\tilde{\boldsymbol{\Psi}}_{\boldsymbol{V}}(j)\right\}\boldsymbol{V}(0), \tag{17}$$

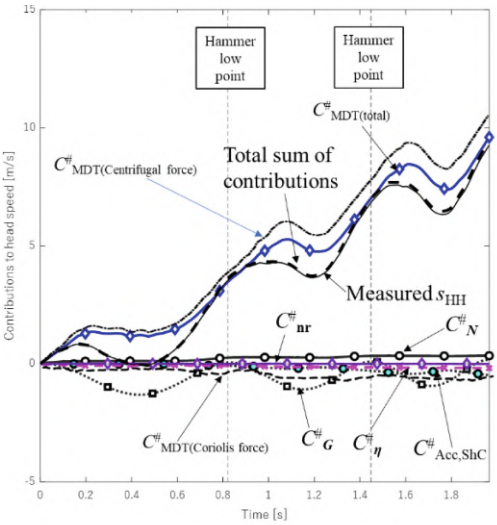
With regard to Eq. (13), since the matrix $\overline{\boldsymbol{A}}_{\boldsymbol{V}}\left(\tilde{\boldsymbol{V}}\ (k)\right)$ is not a zero matrix, and the components of the matrix are determined by the factors related to the configuration of the simplified model consisting of the three links, the factorial function expressed by Eq. (17) shows strong nonlinearity over time for the initial generalised velocity vector.

3 Results

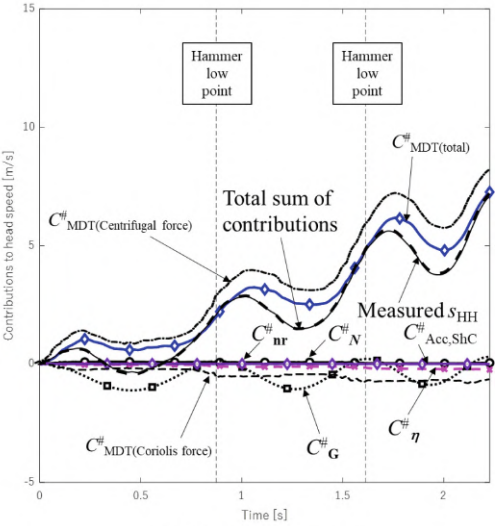
The MDT was the dominant contributor to the generation of the hammer head speed for all throwers (lines labelled with $C^{\#}{}_{\mathrm{MDT(total)}}$ in Figs. 6a and b). As the sum of the individual contributions matched the measured hammer-head speed (see the lines labelled with 'Measured S_{HH}' and 'Total sum of contributions'), it can be concluded that both the forward dynamic contribution analysis and the conversion of MDT into its decomposition were successful (Figs. 6a,b and 7a,b).

The gravitational term made small negative contributions to the hammer head speed regardless of thrower skill level (see the lines labelled with $C^{\#}{}_{\mathbf{G}}$ in Figs. 6a,b). When decomposing the MDT into its centrifugal force and Coriolis force components, the main contributor was the centrifugal force (see lines labelled with $C^{\#}{}_{\mathrm{MDT(centrifugal\ force)}}$ in Figs. 6a,b).

Considering the indirect effects of the same factors, the individual contributions $C^{*}{}_{\mathbf{N}}$, $C^{*}{}_{\mathbf{G}}$, $C^{*}{}_{\mathbf{nr}}$, $C^{*}{}_{\boldsymbol{\eta}}$, $C^{*}{}_{\mathrm{Acc,ShC}}$ and $C^{*}{}_{\mathbf{V}_0}$ in Figs. 7a,b are forward dynamics contributions to the generation of s_{HH} induced by the indirect effects of the virtual shoulder mid-point joint active moment vector, the gravitational force, the residual error of the joint moments, segment length fluctuation, the virtual shoulder mid-point acceleration constraint, and the initial generalised velocity vector, respectively. The superscript * on values C^{*} indicates



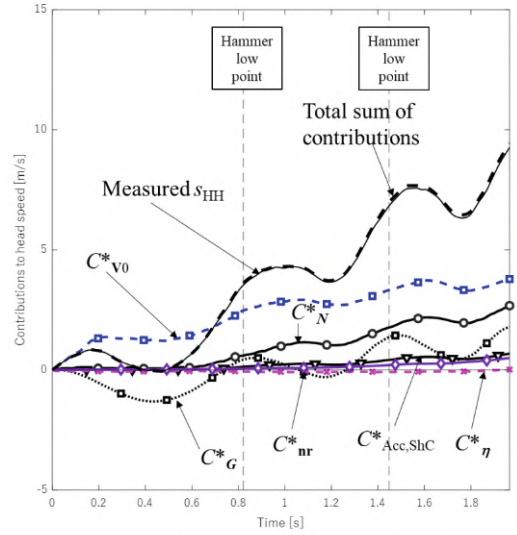
(a) Nationally ranked elite thrower



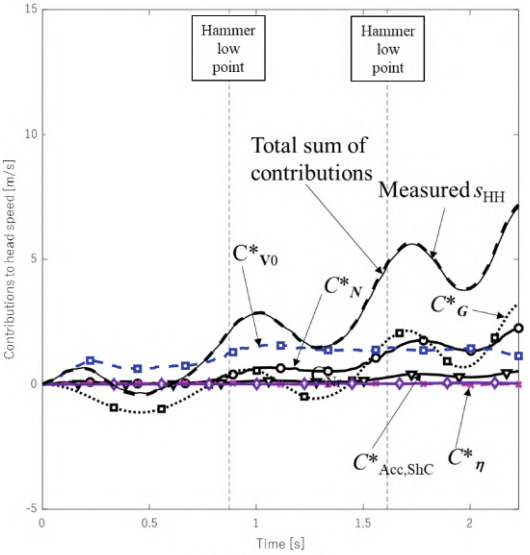
(b) Collegiate thrower

Fig. 6. Forward dynamic contributions to the generation of hammer-head speed during the turning phases arising from the direct effects of the factors indicated. Note that terms that did not show a significant contribution have been omitted, for clarity.

that these are contributions arising from the indirect effects of the factors, to distinguish them from those contributions arising from the direct effects of the factors (which have superscript #).



(a) Nationally ranked elite thrower



(b) Collegiate thrower

Fig. 7. Forward dynamic contributions to the generation of hammer-head speed during the turning phases with the contributions from MDT further broken down into components based on their origin (with indirect effects from the shoulder joint active moment vector, the gravitational force, the residual error of the joint moments, segment length fluctuation, the virtual shoulder mid-point acceleration constraint, and the initial generalised velocity). Note that terms that did not show a significant contribution have been omitted, for clarity.

Whereas the MDT made crucial contributions, both the direct and indirect effects of the active moment vector at the virtual shoulder mid-point joint made smaller contributions to the generation of hammer head speed regardless of the thrower's skill level (see the lines labelled with $C^{\#}{}_{\mathrm{MDT(total)}}$ and $C^{\#}{}_{\boldsymbol{N}}$ in Figs. 6a,b and those labelled with $C^{*}{}_{\boldsymbol{N}}$ in Figs. 7a,b).

In addition, following decomposition of the MDT into its generating factors, large differences in the indirect contributions from the initial generalised velocity vector were observed between the throwers of different skill level (see lines labelled with $C^{*}{}_{\mathbf{V}0}$ in Figs. 7a and b).

4 Discussion and Conclusion

The purpose of this study was to identify which terms contribute most to the generation of hammer-head CG speed during the turning phases. The MDT, which is a nonlinear term consisting of centrifugal force, Coriolis force and gyroscopic effect moment components, was the largest contributor to the generation of hammer head speed during the turning phases regardless of the thrower skill level (as shown by the lines labelled with $C^{\#}$ with subscripts "MDT (total)", "MDT (centrifugal force)", and "MDT (Coriolis force)" in Figs. 6a,b), and the individual components of the MDT are key factors in distinguishing between throwers of different skill level (as seen, for example, in the lines labelled with $C^{*}{}_{\mathbf{V}0}$ in Figs. 7a,b)). Comparing Figs. 6 and 7, the direct effects of the virtual mid-shoulder joint moment made a smaller contribution to the increase in head speed (as shown by the lines labelled with $C^{\#}{}_{\boldsymbol{N}}$ in Figs. 6a,b), than the indirect effects from the same joint moment (as shown by the lines labelled with $C^{*}{}_{\boldsymbol{N}}$ in Figs. 7a,b). These results indicate that a non-intuitive mechanism due to nonlinear effects plays an important role in hammer throwing. Modelling the body-hammer system as a multi-body system has also revealed that a motion-dependent term associated with the centrifugal force contributes significantly to the mechanism of increasing hammer head speed.

Since the contribution from the constraint acceleration of the mid-shoulder joint is small, parameter excitation does not play a significant role in speed generation in hammer throwing (as shown by the lines labelled with $C^{*}{}_{\mathrm{Acc,shC}}$ in Figs. 7a,b) Similarly, with respect to the nationally ranked elite thrower, the contribution to head speed generation from the indirect effect of the joint moment inputs at the virtual shoulder joint midpoint is approximately half the size of the contribution to head speed generation from the indirect effect of the initial state velocity vector, indicating that voluntary inputs do not substantially increase hammer head speed.

The generalised velocity vector of the initial state is determined at the first instant of the start of the first turning phase, and its indirect effect increases gradually with time. The coupled system operates as a self-excited system excited by the initial velocity vector at the first instant, because there are no other external inputs to the system.

Through the derivation of equations of motion for a simplified model and the identification of causal relationships between inputs and outputs, mechanisms could be extracted even from limited data. This approach also enables the observation of individual differences.

It is evident that the adjustment of joint torque and acceleration of the constraint point at the mid-shoulder joint, which can be modified at the thrower's discretion, do

not significantly increase the speed of the hammer head. However, the initial state of the generalised velocity vector at the beginning of the turn has a more notable impact on the speed of the hammer head.

This study has revealed that the adjustments a thrower could make voluntarily during the turning phases, including joint moment and acceleration input at the midpoint of the shoulder joint constraint, did not directly result in a significant increase in hammerhead speed. Conversely, the initial state of the generalised velocity vector at the start of the turn was found to be more influential. The findings suggest that modifications to the active input during the turning phases do not enhance hammer head speed. Rather, the speed is determined by the initial state at the start of the turn. It would be beneficial to mimic this speed-enhancing mechanism by limiting the simulation to a simplified 3-link model without a lower body (i.e., with the trunks and legs omitted) on a plane to avoid collapse of the lower body during the simulation.

Fundings. This work was supported by a KAKENHI grant from the Japan Society for the Promotion of Science [grant number 20H04068].

References

1. Brice, S.M., Ness, K.F., Rosemond, D., Lyons, K., Davis, M.: Development and validation of a method to directly measure the cable force during the hammer throw. Sports Biomech. **7**, 274–287 (2008)
2. Brice, S.M., Ness, K.F., Rosemond, D.: An analysis of the relationship between the linear hammer speed and the thrower applied forces during the hammer throw for male and female throwers. Sports Biomech. **10**, 174–184 (2011)
3. Dapena, J.: The pattern of hammer speed during a hammer throw and influence of gravity on its fluctuations. J. Biomech. **17**, 553–559 (1984)
4. Dapena, J.: A kinematic study of center of mass motions in the hammer throw. J. Biomech. **19**, 147–158 (1986)
5. Dapena, J., Feltner, M.E.: Influence of the direction of the cable force and of the radius of the hammer path on speed fluctuations during hammer throwing. J. Biomech. **22**, 565–575 (1989)
6. Dapena, J., Gutiérrez-Dávila, M., Soto, V.M., Rojas, F.J.: Prediction of distance in hammer throwing. J. Sports Sci. **21**, 21–28 (2003)
7. Maronski, R.: Optimal distance from the implement to the axis of rotation in hammer and discus throws. J. Biomech. **24**, 999–1005 (1991)
8. Murofushi, K., Sakurai, S., Umegaki, K., Takamatsu, J: Hammer acceleration due to thrower and hammer movement patterns. Sports Biomech. **6**, 301–314 (2007)
9. Koike, S., et al.: Modelling error distribution in the ground reaction force during an induced-acceleration analysis of running in rear-foot strikers. J. Sports Sci. **37**(9), 968–979 (2019)
10. Koike, S., et al.: Direct and indirect effects of joint torque inputs during an induced speed analysis of a swinging motion. J. Biomech. **86**, 8–16 (2019)

Nonlinear Material Modeling of Human Skeletal Muscle with Muscle Contraction

Tomotaka Hamajima[1] and Akihiro Matsuda[2(✉)]

[1] Graduate School of Science and Technology, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8577, Ibaraki, Japan
[2] Institute of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8577, Ibaraki, Japan
a_matsuda@kz.tsukuba.ac.jp

Abstract. In this study, a material model of human skeletal muscle which represents force exertion and anisotropy due to muscle contraction was developed. The mechanical properties of the skeletal muscle were divided into passive and active stresses. The passive stress was caused by external deformation, while the active stress was generated by the muscle contraction. The strain energy function proposed by Arya and Singh was used to model skeletal muscle [1]. In this model, the total deformation gradient tensor was divided into elastic and active component to represent the passive and active stress. The active deformation was determined by the degree of muscle activity and the orientation angle of the muscle fibers, while the elastic deformation is determined indirectly from the total deformation gradient tensor and the active deformation gradient tensor. The applicability of the proposed material model was investigated by varying the muscle activity and the fiber orientation.

Keywords: Skeletal muscle · Muscle contraction · Anisotropy

1 Introduction

Understanding the mechanical properties of biological soft tissues is crucial in biomechanics, medical engineering, and rehabilitation. These findings can also significantly contribute to the design of artificial muscles and joints [2, 3]. Soft tissues in the body are composed of closely packed tissue fibers with varying orientations, and muscle contractions enable the body to facilitate and restrict movements. To understand the mechanical properties of skeletal muscles during movement, it is essential to develop a material model that considers both the fiber orientation and muscle activity. Moreover, applying material modeling to skeletal muscles with nonlinear mechanical properties and finite deformation problems is a highly important research topic. Several material models have been proposed to accurately describe the passive and active behavior of skeletal muscle. These models include Hill's model, which treats elastic and contractile elements in series [4], as well as techniques involving the addition of strain energy functions [5, 6], and the separation of active and passive deformations [7]. In this study,

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 335–342, 2025.
https://doi.org/10.1007/978-3-031-72794-8_26

the strain energy function proposed by Arya and Singh was modified based on the right Cauchy-Green deformation tensor to calculate symmetric second Piola-Kirchhoff stress tensor. The partial derivatives of isotropic and anisotropic invariants with respect to the right Cauchy-Green deformation tensor in the modified strain energy function were shown. Three-dimensional finite element analysis was performed to show applicability and robustness of our proposed formulations. The work of Arya and Singh focused exclusively on theoretical solutions for uniaxial deformations and calculated stresses by deriving the first Piola-Kirchhoff stress tensor [1]. The first Piola-Kirchhoff stress tensor is an asymmetric matrix and is not suitable for finite element analysis. The second Piola-Kirchhoff stress tensor, which is a symmetric matrix, is more suitable for finite element analysis. In this study, the second Piola-Kirchhoff stress tensor and the tangent gradient tensor were calculated using the modified strain energy function. The modified second Piola-Kirchhoff stress tensor was applied to a three-dimensional finite element analysis program. The finite element analysis was performed for different fiber orientation angles relative to the tensile direction. Muscle contraction was also considered in the analysis to investigate the effect of muscle contraction on deformation and stress.

2 Material Model

The deformation gradient tensor $\boldsymbol{F}$, which represents the deformation of the material was multiplicatively decomposed into the elastic deformation gradient tensor $\boldsymbol{F}^e$ and the active deformation gradient tensor $\boldsymbol{F}^a$ [8].

$$\boldsymbol{F} = \boldsymbol{F}^e \boldsymbol{F}^a \tag{1}$$

In this research, active deformation gradient tensor proposed by Riccobelli and Ambrosi [7] was introduced to our finite element analysis.

$$\boldsymbol{F}^a = (1-\gamma)\boldsymbol{M} + \frac{1}{\sqrt{(1-\gamma)}}(\boldsymbol{I} - \boldsymbol{M}) \tag{2}$$

Here, γ is the activation parameter and $\boldsymbol{M}$ is the structure tensor, determined by the muscle fiber orientation vector $\boldsymbol{m}$ and written as $\boldsymbol{M} = \boldsymbol{m} \otimes \boldsymbol{m}$. The modified right Cauchy-Green deformation tensor $\overline{\boldsymbol{C}}$ and the modified deformation gradient tensor $\overline{\boldsymbol{F}}$ are introduced [9], utilizing the volume ratio J, to account for incompressibility.

$$\overline{\boldsymbol{F}} = J^{-1/3}\boldsymbol{F} \tag{3}$$

$$\overline{\boldsymbol{C}} = \overline{\boldsymbol{F}}^T \overline{\boldsymbol{F}} = J^{-2/3}\boldsymbol{F}^T\boldsymbol{F} = J^{-2/3}\boldsymbol{C} \tag{4}$$

The strain energy function was adopted from the research by Ehret et al. [6]. To consider the two separate deformations of the tissue that do not contribute to muscle activity and the contribution of muscle fibers, the strain energy function consisting of isotropic and anisotropic parts was calculated. This strain energy function is described as a function of the symmetric tensor, the modified right Cauchy-Green deformation tensor $\overline{\boldsymbol{C}}$, as follows:

$$W = \overline{W}_{iso}\left(\overline{\boldsymbol{C}}\right) + \overline{W}_{ani}\left(\overline{\boldsymbol{C}}^e\right) + p(J-1) \tag{5}$$

Here, $\overline{W}_{iso}$ and $\overline{W}_{ani}$ are expressed as follows, respectively[1]. The third term is a component that considers the constant volume condition using Lagrange multiplier p, which corresponds to hydrostatic pressure.

$$\overline{W}_{iso}\left(\overline{\boldsymbol{C}}\right) = \frac{\mu}{4}\left[\frac{1}{\alpha}\left\{e^{\alpha\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}\right)-1\right)}-1\right\}+\frac{1}{\beta}\left\{e^{\beta\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{-1}\right)-1\right)}-1\right\}\right] \tag{6}$$

$$\begin{aligned}\overline{W}_{ani}\left(\overline{\boldsymbol{C}}^{e}\right) = \frac{\mu}{4}\Bigg[&\frac{1}{\alpha}\left\{e^{\alpha\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e}\right)+w_p\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e}\boldsymbol{M}\right)-1\right)}-e^{\alpha\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e}\right)-1\right)}\right\}\\&+\frac{1}{\beta}\left\{e^{\beta\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e-1}\right)+w_p\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e-1}\boldsymbol{M}\right)-1\right)}-e^{\beta\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e-1}\right)-1\right)}\right\}\Bigg]\end{aligned} \tag{7}$$

Here, $\mu > 0$, $\alpha > 0$, and $\beta > 0$ are material parameters, and w_0 and $w_p = 1 - w_0$ are weight factors for isotropic tissue and muscle fibers. The strain energy function of the skeletal muscle described using the modified right Cauchy-Green deformation tensor $\overline{\boldsymbol{C}}$ was proposed.

The second Piola-Kirchhoff stress tensor $\boldsymbol{S}$ is obtained by partial differentiation of this strain energy function W by the right Cauchy-Green deformation tensor $\boldsymbol{C}$.

$$\boldsymbol{S} = 2\frac{\partial W}{\partial \boldsymbol{C}} = 2\left[\frac{\partial \overline{W}_{iso}\left(\overline{\boldsymbol{C}}\right)}{\partial \boldsymbol{C}}+\frac{\partial \overline{W}_{ani}\left(\overline{\boldsymbol{C}}^{e}\right)}{\partial \boldsymbol{C}}+\frac{\partial\{p(J-1)\}}{\partial \boldsymbol{C}}\right] \tag{8}$$

The second Piola-Kirchhoff stress tensor $\boldsymbol{S}$ is described with the right Cauchy-Green deformation tensor $\boldsymbol{C}$, leading to symmetric partial derivatives. The derivation of symmetric partial derivatives is an engineering advantage of our proposed model. Similarly, the fourth-order tangent gradient tensor became also symmetric. Here, the partial derivatives of each term in Eq. (8) are expressed using the chain rule of differentiation as follows:

$$\begin{aligned}\frac{\partial \overline{W}_{iso}\left(\overline{\boldsymbol{C}}\right)}{\partial \boldsymbol{C}} = &\frac{\mu}{4}e^{\alpha\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}\right)-1\right)}\left[J^{-2/3}\frac{w_0}{3}\boldsymbol{I}-\frac{1}{3}\left\{\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}\right)\right\}\boldsymbol{C}^{-1}\right]\\&+\frac{\mu}{4}e^{\beta\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{-1}\right)-1\right)}\left[\frac{1}{3}\left\{\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{-1}\right)\right\}\boldsymbol{C}^{-1}-J^{2/3}\frac{w_0}{3}\boldsymbol{C}^{-1}\boldsymbol{C}^{-1}\right]\end{aligned} \tag{9}$$

$$\begin{aligned}\frac{\partial \overline{W}_{ani}\left(\overline{\boldsymbol{C}}^{e}\right)}{\partial \boldsymbol{C}} = &\frac{\mu}{4}e^{\alpha\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e}\right)+w_p\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e}\boldsymbol{M}\right)-1\right)}\left[-\frac{1}{3}\left\{\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e}\right)+w_p\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e}\boldsymbol{M}\right)\right\}\boldsymbol{C}^{-1}\right.\\&\left.+J^{e-2/3}\left(\frac{w_0}{3}\boldsymbol{C}^{a-1}+w_p\boldsymbol{F}^{a-1}\boldsymbol{M}\boldsymbol{F}^{a-T}\right)\right]\\&-\frac{\mu}{4}e^{\alpha\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e}\right)-1\right)}\left[-\frac{1}{3}\left\{\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e}\right)\right\}\boldsymbol{C}^{-1}+J^{e-2/3}\frac{w_0}{3}\boldsymbol{C}^{a-1}\right]\\&+\frac{\mu}{4}e^{\beta\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e-1}\right)+w_p\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e-1}\boldsymbol{M}\right)-1\right)}\left[\frac{1}{3}\left\{\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e-1}\right)+w_p\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e-1}\boldsymbol{M}\right)\right\}\boldsymbol{C}^{-1}\right.\end{aligned}$$

$$-J^{e\,2/3}\left(\frac{w_0}{3}\boldsymbol{C}^{-1}\boldsymbol{C}^{a}\boldsymbol{C}^{-1}+w_p\boldsymbol{C}^{-1}\boldsymbol{F}^{a\,T}\boldsymbol{M}\boldsymbol{F}^{a}\boldsymbol{C}^{-1}\right)\Big]$$

$$-\frac{\mu}{4}e^{\beta\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e\,-1}\right)-1\right)}\left[\frac{1}{3}\left\{\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e\,-1}\right)\right\}\boldsymbol{C}^{-1}-J^{e\,2/3}\frac{w_0}{3}\boldsymbol{C}^{-1}\boldsymbol{C}^{a}\boldsymbol{C}^{-1}\right] \tag{10}$$

The second Piola-Kirchhoff stress tensor $\boldsymbol{S}$ was obtained by the above calculations. For finite element analysis, the second Piola-Kirchhoff stress tensor $\boldsymbol{S}$ is more suitable than the first Piola-Kirchhoff stress tensor $\boldsymbol{P}$ due to it being an asymmetric matrix. These calculations allow the model in Eqs. (5), (6), and (7) to be implemented in a three-dimensional finite element analysis. The Cauchy stress tensor $\boldsymbol{T}$ was calculated by using the second Piola-Kirchhoff stress tensor $\boldsymbol{S}$. The derived Cauchy stress tensor $\boldsymbol{T}$ was shown as follows:

$$\begin{aligned}\boldsymbol{T}=\frac{1}{J}\boldsymbol{F}\boldsymbol{S}\boldsymbol{F}^{T}=\frac{2}{J}\Bigg[&\frac{\mu}{4}e^{\alpha\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}\right)-1\right)}\left\{-\frac{1}{3}\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}\right)\right)\boldsymbol{I}+J^{-2/3}\frac{w_0}{3}\boldsymbol{B}\right\}\\&+\frac{\mu}{4}e^{\beta\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{-1}\right)-1\right)}\left\{\frac{1}{3}\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{-1}\right)\right)\boldsymbol{I}-J^{2/3}\frac{w_0}{3}\boldsymbol{B}^{-1}\right\}\\&+\frac{\mu}{4}e^{\alpha\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e}\right)+w_p\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e}\boldsymbol{M}\right)-1\right)}\left\{-\frac{1}{3}\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e}\right)+w_p\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e}\boldsymbol{M}\right)\right)\boldsymbol{I}\right.\\&\left.+J^{e\,-2/3}\left(\frac{w_0}{3}\boldsymbol{F}\boldsymbol{C}^{a\,-1}\boldsymbol{F}^{T}+w_p\boldsymbol{F}\boldsymbol{F}^{a\,-1}\boldsymbol{M}\boldsymbol{F}^{a\,-T}\boldsymbol{F}^{T}\right)\right\}\\&-\frac{\mu}{4}e^{\alpha\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e}\right)-1\right)}\left\{-\frac{1}{3}\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e}\right)\right)\boldsymbol{I}+J^{e\,-2/3}\frac{w_0}{3}\boldsymbol{F}\boldsymbol{C}^{a\,-1}\boldsymbol{F}^{T}\right\}\\&+\frac{\mu}{4}e^{\beta\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e\,-1}\right)+w_p\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e\,-1}\boldsymbol{M}\right)-1\right)}\left\{\frac{1}{3}\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e\,-1}\right)+w_p\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e\,-1}\boldsymbol{M}\right)\right)\boldsymbol{I}\right.\\&\left.-J^{e\,2/3}\left(\frac{w_0}{3}\boldsymbol{F}^{-T}\boldsymbol{C}^{a}\boldsymbol{F}^{-1}+w_p\boldsymbol{F}^{-T}\boldsymbol{F}^{a\,T}\boldsymbol{M}\boldsymbol{F}^{a}\boldsymbol{F}^{-1}\right)\right\}\\&-\frac{\mu}{4}e^{\beta\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e\,-1}\right)-1\right)}\left\{\frac{1}{3}\left(\frac{w_0}{3}\mathrm{tr}\left(\overline{\boldsymbol{C}}^{e\,-1}\right)\right)\boldsymbol{I}-J^{e\,2/3}\frac{w_0}{3}\boldsymbol{F}^{-T}\boldsymbol{C}^{a}\boldsymbol{F}^{-1}\right\}+p\boldsymbol{I}\Bigg]\end{aligned} \tag{11}$$

Here, $\boldsymbol{B}$ is the left Cauchy-Green deformation tensor and calculated as $\boldsymbol{B}=\boldsymbol{F}\boldsymbol{F}^{T}$.

Using this model, finite element analysis is conducted that considers muscle contraction and directions other than the fiber direction.

3 Calculation Results

An original finite element analysis program was developed using Fortran90 based on the proposed equations. Numerical simulations were conducted varying the fiber orientation and activation degree of the muscle. The material parameters μ, α and β were determined by experiments conducted by Arya et al. [1] and Lim et al. [10]. The analytical model had cubic dimensions with sides of 10 mm each. Each side was divided by 10 elements. The

total number of finite elements was 1,000. For the tensile analysis, boundary conditions were applied to constrain the top and bottom surfaces of the analytical model in the tensile direction. The angle between the tensile direction and the fiber orientation vector $\boldsymbol{m}$ was defined as θ. The analytical results for different fiber orientation angles θ are shown in Fig. 1 and Fig. 2. In this analysis, displacement correspond to the nominal strain of 0.7 was applied to all nodes of the top surface of the finite element model. Figure 1 shows the stress-strain relationship for various θ values, while Fig. 2 displays the equivalent stress distribution and deformation resulting from finite element analysis. When the fiber orientation angle θ was varied, the stress distribution and the behavior of the stress-strain curve changed. The equivalent stress distribution and deformation show stress concentration because of the anisotropy of the muscle fibers. The analytical results for varying the activation parameter γ are shown in Fig. 3, 4. Figure 3 shows the experimental results of Lim et al. and the stress-strain curves of the finite element analysis that reproduces the experiments. In this analysis, displacement was also applied to the top surface of the model. The activation parameter γ was adjusted to fit the experimental results. The analysis results showed good agreement with the experimental results. Figure 4 shows the deformation and equivalent stress distribution. In this analysis, the vertical displacement was constrained, and the activation parameter γ was varied to generate muscle contraction. The activation parameter γ is defined as 0 at the edges and a maximum value is applied in the center. The activation parameter γ between the center and the edges are linearly increased. In this analysis, the axial length was changed to 30 mm to observe the axial behavior in detail. The magnitude of stress and deformation varied with the value of the activation parameter γ. These results show that this model is possible to take account for muscle anisotropy and activation.

Figure 5 shows the deformation and equivalent stresses when a 90° torsion is applied to the top surface to confirm the robustness and stability of our proposed model and programs. The top and bottom surfaces of the analytical model were fully constrained and

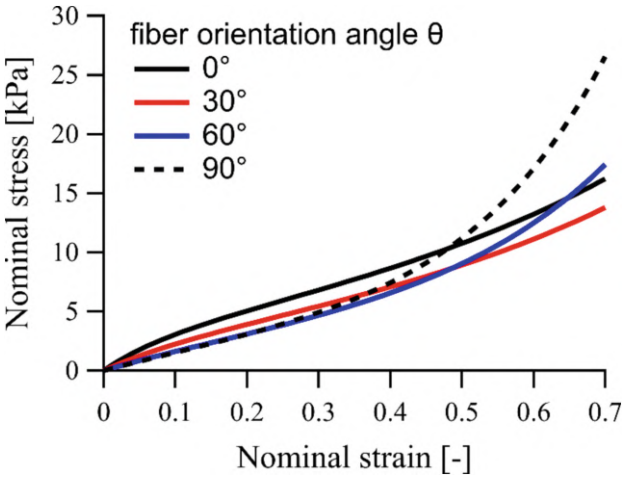


Fig. 1. Stress-strain relationship with different fiber orientation angles θ

displacements corresponding to 90° torsional deformation were applied. The activation parameter γ was set to 0.3 for all elements. The finite element analysis performed in this study allows us to consider arbitrary boundary conditions and force exertion characteristics of the muscles for the three-dimensional analytical model.

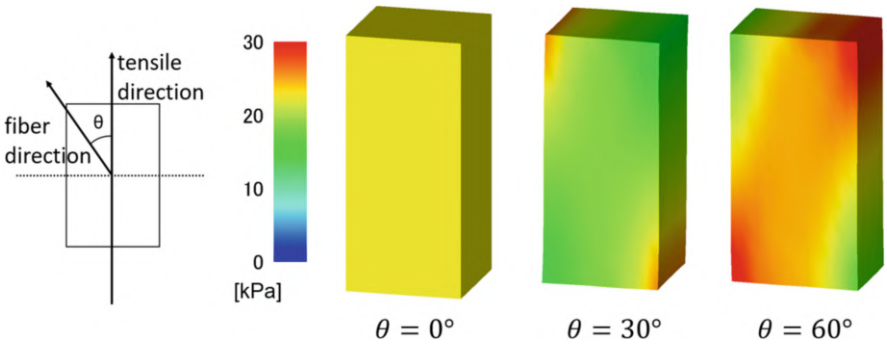


Fig. 2. Deformation and equivalent stress with different fiber orientation angles θ

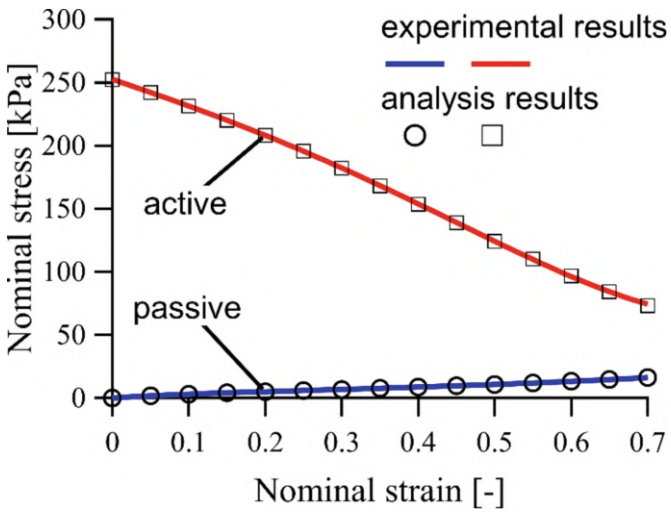


Fig. 3. Stress-strain relationship in passive and active conditions

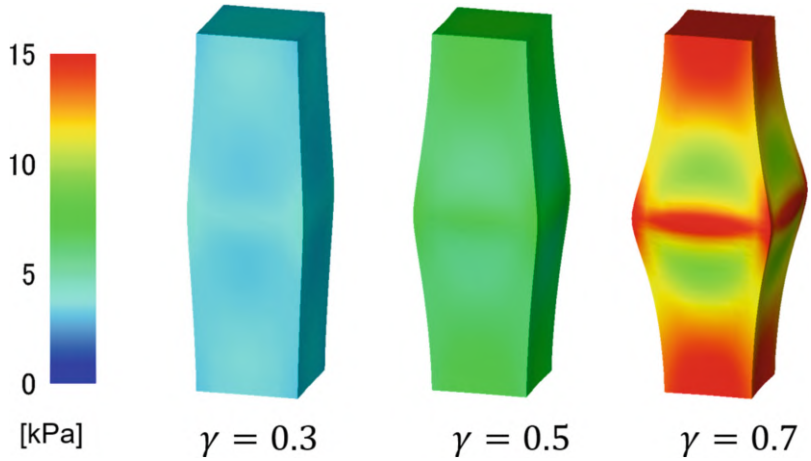


Fig. 4. Deformation and equivalent stress with muscle contraction ($\theta = 0°$)

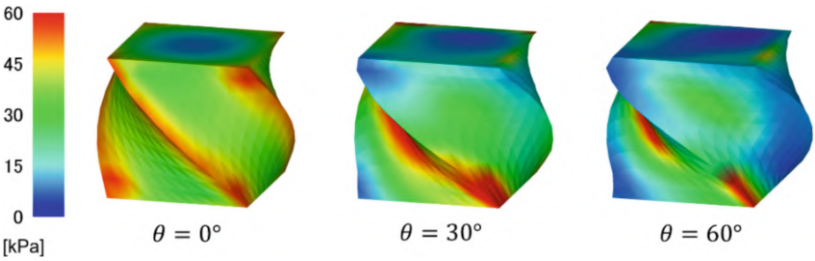


Fig. 5. Deformation and equivalent stress with muscle contraction (torsional deformation)

4 Conclusion

In this study, a three-dimensional finite element analysis was conducted, considering the anisotropy of skeletal muscles and muscle contraction. The formulation of the finite element analysis was based on the model proposed by Arya and Singh. This model was made feasible for three-dimensional finite element analysis by describing it using the symmetric tensor, the second Piola-Kirchhoff stress tensor. This allowed the analysis to not only address the stress-strain relationship in a single direction but also to model variations in fiber orientations and material deformation due to muscle activity. Different mechanical properties were observed by altering the fiber orientation angle and activation parameter, and the experimental results were accurately replicated. These analytical findings demonstrate the applicability of this finite element model for evaluating the mechanical properties of skeletal muscle.

References

1. Arya, K., Singh, A.: Active behavior of skeletal muscle and modeling sarcopenia with a mixture active strain approach based constitutive model. Int. J. Solids Struct. **248**, 111669 (2022)
2. Tondu, B., Ippolito, S., Guiochet, J., Daidié, A.: A seven-degrees-of-freedom robot-arm driven by pneumatic artificial muscles for humanoid robots. Int. J. Robot. Res. **24**(4), 257–274 (2005)
3. Waycaster, G., Wu, S.K., Shen, X.: Design and control of a pneumatic artificial muscle actuated above-knee prosthesis. J. Med. Devices **5**(3), 9 (2011)
4. Hill, A.: The heat of shortening and the dynamic constants of muscle. Proc. Royal Soc. London Seri. B-Biol. Sci. **126**, 843 (1938)
5. Ehret, A.E., Itskov, M.: A polyconvex hyperelastic model for fiber-reinforced materials in application to soft tissues. J. Mater. Sci. **42**, 8853–8863 (2007)
6. Ehret, A.E., Böl, M., Itskov, M.: A continuum constitutive model for the active behaviour of skeletal muscle. J. Mech. Phys. Solids **59**, 625–636 (2011)
7. Riccobelli, D., Ambrosi, D.: Activation of a muscle as a mapping of stress-strain curves. Extreme Mech. Lett. **28**, 37–42 (2019)
8. Lee, E.H.: Elastic-plastic deformation at finite strains. J. Appl. Mech. **36**(1), 1–6 (1969)
9. Flory, P.J.: Thermodynamic relations for high elastic materials. Trans. Faraday Soc. **57**, 829–838 (1961)
10. Lim, J.Y., Choi, S.J., Widrick, J.J., Phillips, E.M., Frontera, W.R.: Passive force and viscoelastic properties of single fibers in human aging muscles. Eur. J. Appl. Physiol. **119**, 2339–2348 (2019)

An Instrumented Pole for Measuring the Individual Hand Forces in Pole Vaulting

Rinri Uematsu[1(✉)], Kiyoshi Hirose[2], Osamu Takeda[1], Alexander P. Willmott[3], and Sekiya Koike[1]

[1] University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki Pref., Japan
polevault510@gmail.com, s1930487@u.tsukuba.ac.jp
[2] Kurume Institute of Technology, Kamitsumachi, Kurume 2228-66, Fukuoka Pref., Japan
[3] University of Lincoln, Lincoln, Lincolnshire LN6 7TS, UK

Abstract. The purpose of this study was to develop a method for measuring the forces exerted by the individual hands during a pole-vaulting motion via the use of an instrumented pole. Thirty-two strain gauges, specialised for Fiber Reinforced Plastic (FRP), were attached to the pole. A quasi-static calibration of the strain gauges was conducted under a large pole-deformation condition. Two miniature wireless strain amplifier systems were used to record output signals from the strain gauges. The estimated hand forces and moments passed through an extended Kalman filter matched well with those calculated from the pole tip (plug) forces obtained with a force plate situated below the pole vault box. From these results, use of this instrumented pole would make it possible to estimate the forces exerted by the individual hands during the pole-vaulting motion, even without a force plate situated below the box.

Keywords: Strain Gauge · Large Deformation · Fiber Reinforced Plastic · Extended Kalman Filter · Closed-loop problem

1 Introduction

Pole vaulters transform the kinetic energy gained during the run-up into elastic energy by bending long and flexible poles as much as possible [1, 2]. Subsequently, the resulting significant elastic forces lift their bodies, enabling them to clear the bar. Since the material of vaulting poles was changed from wood/metal to Fiber Reinforced Plastic (FRP), large deformations of the pole have become possible. An effective technique for controlling the magnitude and timing of the bending of the pole is, therefore, essential for achieving the greatest heights. In the pole-bending phase, the distribution of forces between the left and right hand arising from the body movements influences the pole-bending technique. In the subsequent pole-straightening phase, use of the pole reaction forces is important for generating actions such as inversion and clearance. The combinations of forces from the two hands that can deform the pole are infinite and have kinetic redundancy. The lower tip of the pole makes point contact with the pole vault box, which is a trapezoidal indentation into which vaulters plant their pole, and it cannot exert any moment. Pole vaulters must

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 343–352, 2025.
https://doi.org/10.1007/978-3-031-72794-8_27

manipulate a substantially curved pole to sustain equilibrium and ultimately clear the bar. Previous studies of the dynamics of pole vault motion have used a model in which the vaulter applied a force and a moment that were the resultants from the two upper limbs combined [3, 4]; this approach does not permit analysis of the magnitude and timing of application of the individual hand forces. Knowledge of the individual hand forces exerted on the pole would provide us with new technical insights into pole vaulting and help to elucidate strong nonlinear dynamics in pole vaulting. However, the forces exerted by the individual hands on the pole during the vault cannot be determined by inverse dynamics analysis because the upper limbs and the pole form a mechanical closed-loop system.

As mentioned above, due to the kinetic redundancy, the force exerted by each hand on sports equipment cannot be uniquely determined from only its movement, an issue known as a 'closed-loop problem' [5]. Strain gauges, very lightweight foil sensors capable of detecting bending moments and compressive or tensile forces, have been used as a means of measuring the forces applied to manipulate sports equipment such as skies, snowboards, bicycle pedals, baseball bats, and golf clubs [6–10]. The use of strain gauges to directly measure the forces applied to the equipment has proven particularly effective in addressing the closed-loop problems in double-handed hitting sport motions such as baseball batting and golf swing motions [9, 10].

With respect to instrumented hitting tools such as baseball bats and golf clubs, the strain gauges and amplifier can be connected via electrical wires to form electric circuits. However, directly applying this method to pole vaulting is challenging, as a very long electrical cable would be required to maintain the circuit. Therefore, in applying the force measurement methods of double-handed hitting sport motions previously used with wired connections, a wireless amplifier was attempted to be used to create an instrumented pole.

The purpose of this study was to propose and build a wireless instrumented pole that can measure the forces and moments acting on the individual hand grip regions, and to verify its accuracy.

2 Device and Methods

2.1 Structure of the Instrumented Pole

A total of five Wheatstone bridge circuits, with 16 strain gauges, were attached just below each grip (left- and right- hand grips) region of the vaulting pole (UCS Spirit, US, 14 ft, 145 lb). The positions of the strain gauges were aligned relative to the bending direction of the pole; vaulting poles have a pre-bend, which determines the direction of the main deformation in pole vaulting. The main bending axis was the $\boldsymbol{x}_{\mathrm{p}}$ axis, the direction of the long axis of the pole segment was the $\boldsymbol{z}_{\mathrm{p}}$ axis, and the $\boldsymbol{y}_{\mathrm{p}}$ axis was orthogonal to the $\boldsymbol{x}_{\mathrm{p}}$ axis and the $\boldsymbol{z}_{\mathrm{p}}$ axis. The specific procedure for setting up the axes using reflective markers and motion capture is described in a later section.

Two sets of strain gauges designed for composite materials (BFLAB-5–8,Tokyo Measuring Instruments Lab, Japan) were used to construct two-sided, two-active strain gauge connections in order to detect the bending moment exerted about the $\boldsymbol{x}_{\mathrm{p}}$ axis at locations 1, 2, 5, and 6 (Figs. 1 and 2). Similarly, two further double sets of strain

gauges were added to detect the bending moment exerted about the $\boldsymbol{y}_p$ axis at locations 3, 4, 7 and 8 (Figs. 1 and 2). Finally, a set of strain gauges was used for detection of the longitudinal force, adopting a four-gauge method consisting of eight strain gauges with two gauges connected in parallel in order to reduce the crosstalk due to bending at gauge locations from 9 to 16. Two miniature wireless data logger type multi-channel dynamic strain amplifiers (6-CH, Tech-Gihan, Japan), fixed near the lower grip, were synchronized electrically, and were used to record the strain gauge signals. The force measured by the bridge circuits near the left hand grip was taken as representative of the resultant force of both hands, while the force measured by the bridge circuits near the right hand grip represented the force from the right hand; since the strain gauges are attached on the surface of the pole, placing them at the precise grip locations would be difficult for the athlete. A coordinate system for the pole handle segment ($\sum_p \boldsymbol{x}_p \boldsymbol{y}_p \boldsymbol{z}_p$) will be described in detail in a subsequent section.

2.2 A Quasi-Static Sensor Calibration

A quasi-static calibration of the strain gauges under the condition of large deformation of the pole was conducted using the calibration setup shown in Fig. 3. The setup consisted of a hand winch, a force plate (Tech-Gihan, Kyoto, Japan, 100 Hz), an aluminum frame acting as a guide rail, and a slider designed to run along the rail that was connected to the winch. The force plate and the lower end of the pole (plug) made contact on the jig with a semi-circular recess, and no moments were exerted on the plug.

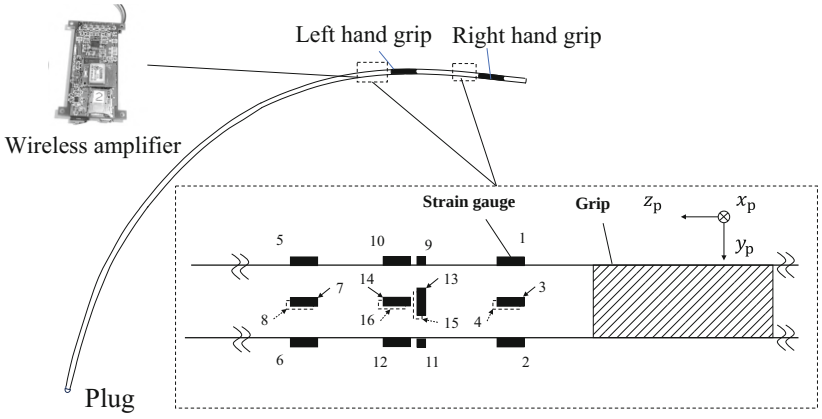


Fig. 1. A schematic representation of the locations of strain gauges on the instrumented pole.

Twenty-four reflective markers attached at 39 cm intervals along the pole allowed its position to be tracked in the global coordinate system. In order to define the coordinate system of the individual pole segment and to determine the position vectors from the plug to the strain gauges, the marker positions were measured using an optoelectronic marker-based motion capture system (26-cameras; 100 Hz; Vicon Motion Systems Ltd., Oxford, UK). A pole handle segment coordinate system ($\sum_p \boldsymbol{x}_p \boldsymbol{y}_p \boldsymbol{z}_p$; as shown in Fig. 4)

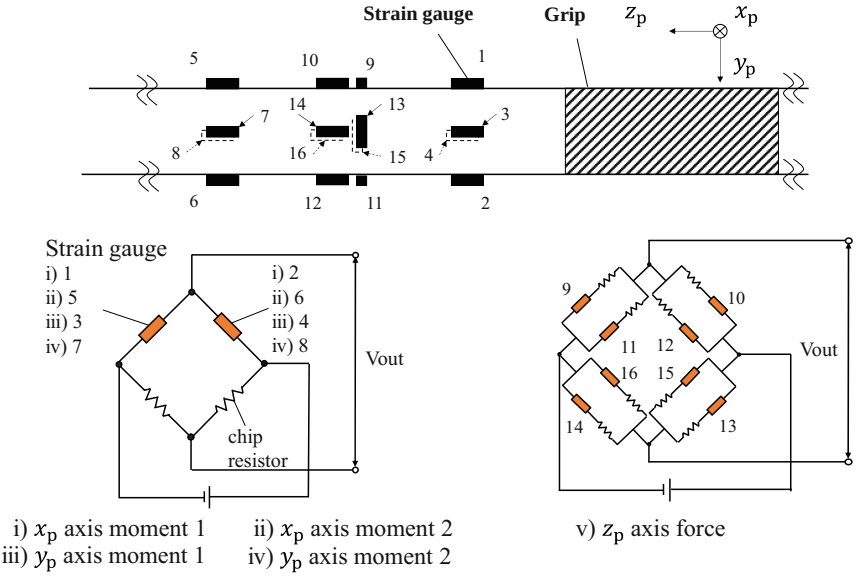


Fig. 2. A schematic representation of the strain gauge bridge circuits for detection of bending moments (left) and pole long axis forces (right).

was defined as a right-handed orthogonal reference frame in which the vector z_p was the unit vector running from the mid-point of the top markers (M1, M2) to the mid-point of the two markers (M3, M4), a temporary support vector s_p was defined as the unit vector pointing from the head-side left marker (M2) to the head-side right marker (M1), the unit vector y_p was obtained from the cross-product of z_p and s_p, and the unit vector x_p was the cross-product of the unit vectors y_p and z_p..The instrumented pole was used as a sensor providing kinetic values, such as bending moments and long axial force, and its output signals were measured with a wireless miniature dynamic strain amplifier (Tech-Gihan, Japan, 100Hz). The moments and forces arising from the force at the plug, expressed in the pole segment coordinate system, were calculated as follows:

$$^{\Sigma p}\boldsymbol{M}_{\text{sensor}} = {}^{\Sigma O}\boldsymbol{R}_{\text{p}}^{\text{T}}({}^{\Sigma O}\boldsymbol{r}_{\text{LS}} \times {}^{\Sigma O}\boldsymbol{F}_{\text{plug}}) \tag{1}$$

$$^{\Sigma p}\boldsymbol{F}_{sensor} = {}^{\Sigma O}\boldsymbol{R}_{\text{p}}^{\text{T}}\,{}^{\Sigma O}\boldsymbol{F}_{\text{plug}} \tag{2}$$

where the vector $^{\Sigma p}\boldsymbol{M}_{\text{sensor}}$ is the moment vector in the pole coordinate system, $^{\Sigma O}\boldsymbol{R}_{\text{p}}$ is the orientation matrix of the pole segment with the attached strain gauges, $^{\Sigma O}\boldsymbol{r}_{\text{LS}}$ is the position vector from the plug to the lower hand's sensor unit, and $^{\Sigma O}\boldsymbol{F}_{\text{plug}}$ and $^{\Sigma p}\boldsymbol{F}_{\text{sensor}}$ are the force vectors at the plug and sensor unit, respectively.

2.3 Motion Data Acquisition of Pole Vaulting

A pole vaulter performed three vaults with sub-maximal approach runs in the indoor experimental setup shown in Fig. 5. Kinematic data (47 markers on the body; 24 markers

on the poles) were captured with a motion capture system (Vicon Motion Systems Ltd., UK, 250 Hz). The ground reaction force at the pole end was recorded by a force plate (Kistler, Switzerland, 1000 Hz) situated below the vaulting box. The vaulter used the

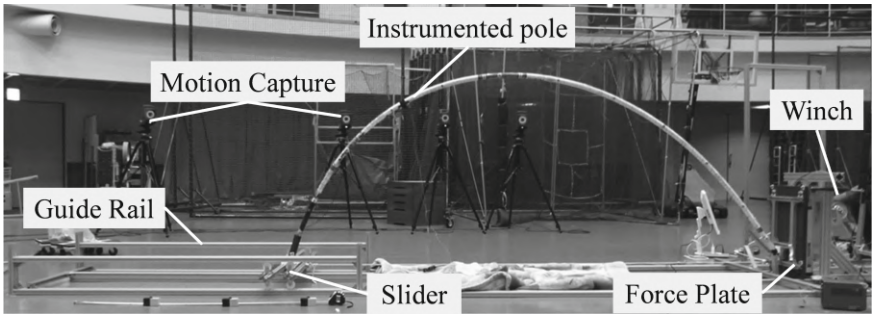


Fig. 3. The experimental set up for the quasi-static pole calibration.

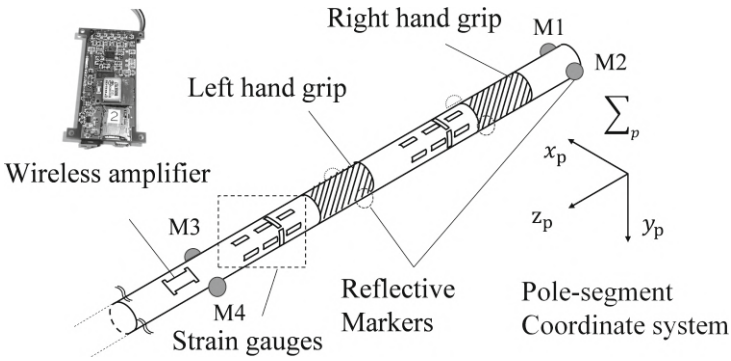


Fig. 4. Definition of the pole segment coordinate system.

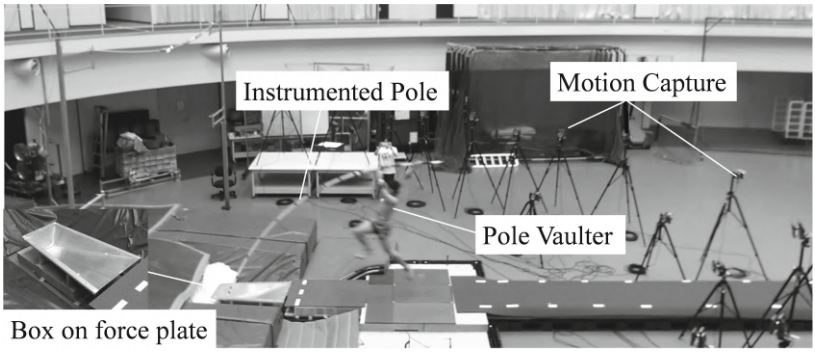


Fig. 5. The experimental setup for pole vaulting.

instrumented pole to allow measurement of the individual forces. Informed consent was obtained before conducting the experiments.

2.4 Calculation of Calibration Parameters for the Strain Gauge Sensors on the Instrumented Pole

The calibration of the instrumented pole was carried out in two steps. First, the calibration parameters were estimated based on the relationship between sensor output and the force exerted on the pole in the quasi-static calibration. Next, the correction matrix for the calibration parameters was estimated using an extended Kalman filter with the sensor outputs and the exerted force from one pole vault. This approach was necessary due to the limitation of applying only small loads around the $\boldsymbol{y}_{\mathrm{p}}$ axis in the quasi-static calibration. The pole coordinate system changed in relation to the sensor position due to a slight deviation in the position of the marker attached to the pole between the quasi-static calibration and pole vaulting trial.

The quasi-static calibration parameters were determined from the force and moments calculated using the vaulting box reaction force and the motion capture data in a quasi-static calibration procedure, as follows:

$$\boldsymbol{F}_{\mathrm{p}} = \boldsymbol{C}\boldsymbol{u} \tag{3}$$

where the vector $\boldsymbol{F}_{\mathrm{p}}$ is the generalized force vector consisting of the force components calculated from the force plate and motion capture outputs information as $\boldsymbol{F}_{\mathrm{p}} = \left[F_x F_y F_z\right]^{\mathrm{T}}$, and the vector $\boldsymbol{u}$ consists of the strain gauge signal outputs as $\boldsymbol{u} = \left[u_{mx1} u_{mx2} u_{my1} u_{my2} u_{fz}\right]^{\mathrm{T}}$**.** The matrix $\boldsymbol{C}$, whose 3×5 components represent the quasi-static calibration parameters used to calculate forces and moments from the strain gauge output signals, was estimated using a pseudo-inverse matrix as in the following equation:

$$\boldsymbol{C} = \boldsymbol{F}_{\mathrm{p,cal}} \left(\boldsymbol{u}_{\mathrm{cal}}^{\mathrm{T}} \boldsymbol{u}_{\mathrm{cal}}\right)^{-1} \boldsymbol{u}_{\mathrm{cal}}^{\mathrm{T}} \tag{4}$$

where the vector $\boldsymbol{F}_{\mathrm{p,cal}}$ and $\boldsymbol{u}_{\mathrm{cal}}$ are $\boldsymbol{F}_{\mathrm{p}}$ and $\boldsymbol{u}$ which were measured by a quasi-static calibration.

2.5 Modification of Calibration Parameters Based on Extended Kalman Filter

In order to improve the accuracy of force vector $\boldsymbol{F}_{\mathrm{p}}$ calculated from Eqs. (3) and (4), an extended Kalman filter, a non-linear Kalman filter that performs optimal estimation by approximately linearizing a non-linear function derived by partial differentiation [11], was used to estimate the correction coefficients matrix. The voltages from the strain gauges, which include noise and crosstalk due to bending of other axes, required the application of a robust estimation method. An extended Kalman filter was used as a Bayesian estimator of the parameters, which is effective for parameter estimation with observations containing vibration caused by box-pole contacts and pole sensor noise. The calibration parameters were sequentially updated using frame-by frame observed values for the sensor outputs and the exerted force from one pole vault trial, with the

quasi-static calibration parameters taken as the initial value. The updated calibration parameters were referred to as the modified calibration parameters.

$$\boldsymbol{x} = [x_{1,1}\, x_{1,2}\, x_{1,3}\, x_{2,1}\, x_{2,2}\, x_{2,3}\, x_{3,1}\, x_{3,2}\, x_{3,3}]^{\mathrm{T}} \tag{5}$$

$$\boldsymbol{y}_t = [F_{\mathrm{mea},x}\, F_{\mathrm{mea},y}\, F_{\mathrm{mea,z}}\, 1]^{\mathrm{T}} \tag{6}$$

$$\boldsymbol{H}(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{X}\boldsymbol{F}_{\mathrm{p}} \\ det\boldsymbol{X} \end{bmatrix} \tag{7}$$

$$\boldsymbol{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,1} & x_{2,2} & x_{2,3} \\ x_{3,1} & x_{3,2} & x_{3,3} \end{bmatrix} \tag{8}$$

$$\boldsymbol{F}_{\mathrm{p}} = \boldsymbol{C}\boldsymbol{u}_{\mathrm{pv}} \tag{9}$$

where $\boldsymbol{x}$ is a vector with the values of modified calibration parameters from Eq. (5). $\boldsymbol{y}_t$ is a vector with the values of the forces from Eq. (6). Matrix $\boldsymbol{C}$ contains the quasi-static calibration parameters, derived from Eq. (4), in Eq. (9). Vector $\boldsymbol{u}_{\mathrm{pv}}$ consists of strain gauge signal outputs as $\boldsymbol{u}_{\mathrm{pv}} = \left[u_{\mathrm{pv},mx1}\, u_{\mathrm{pv},mx2}\, u_{\mathrm{pv},my1}\, u_{\mathrm{pv},my2}\, u_{\mathrm{pv},fz}\right]^{\mathrm{T}}$.

The algorithm for the extended Kalman filter has the following two steps.

Step 1) Prediction of coefficient and covariance matrix:
The prediction step as time invariant is as shown in Eq. (10)

$$\boldsymbol{P}_t^- = \boldsymbol{P}_{t-1} + \boldsymbol{Q} \tag{10}$$

where $\boldsymbol{P}_t^-$ is an error covariance matrix before the update and $\boldsymbol{Q}$ is the covariance matrix for process noise.

Step 2) Update:
The measurement update step is as shown in Eqs. (11)–(13) as follows:

$$\boldsymbol{K}_t = \boldsymbol{P}_t^- \boldsymbol{h}(\boldsymbol{x})^{\mathrm{T}} (\boldsymbol{h}(\boldsymbol{x}) \boldsymbol{P}_t^- \boldsymbol{h}(\boldsymbol{x})^{\mathrm{T}} + \boldsymbol{R})^{-1} \tag{11}$$

$$\boldsymbol{x} = \boldsymbol{x} + \boldsymbol{K}_t (\boldsymbol{y}_t - \boldsymbol{H}(\boldsymbol{x})) \tag{12}$$

$$\boldsymbol{P}_t = (\boldsymbol{I} - \boldsymbol{K}_t \boldsymbol{h}(\boldsymbol{x})) \boldsymbol{P}_t^- \tag{13}$$

where $\boldsymbol{K}_t$ is the Kalman gain, and $\boldsymbol{I}$ is the unit matrix. Kalman gain defines the updating weight between the new measurements and the prediction value. $\boldsymbol{h}(\boldsymbol{x})$ is the partial differential of system matrix $\boldsymbol{H}(\boldsymbol{x})$ within Eq. (7), and $\boldsymbol{R}$ is the covariance matrix for observation noise. The elements of the coefficient matrix were estimated through iterative computation, with the values of $\boldsymbol{x}$ and $\boldsymbol{P}_t$ updated using the results obtained from Eqs. (12) and (13). The modified calibration parameters were iteratively calculated until

convergence was achieved. The modified force vectors $\widehat{\boldsymbol{F}}_{\mathrm{p}}$ were calculated by using $\boldsymbol{X}_{,}$, which is the correction matrix after iteration calculation in Eq. (14).

$$\widehat{\boldsymbol{F}}_{\mathrm{p}} = \boldsymbol{X}\boldsymbol{F}_{\mathrm{p}} \tag{14}$$

The forces calculated from Eqs. (9) and (14) are compared with the forces, expressed in the instrumented pole-segment coordinate system, as the true values. The force in the pole-segment coordinate system (${}^{\Sigma p}\boldsymbol{F}_{\mathrm{sensor}}$**:** the local coordinate system) calculated from the exerted ground reaction force (${}^{\Sigma O}\boldsymbol{F}_{\mathrm{plug}}$:: in the world coordinate system) measured with the box-force-plate from Eq. (2).

3 Results

The measurement accuracy of the instrumented pole, expressed as root mean square error (RMSE) between the resultant force estimated from the sensor outputs and the values calculated from the ground reaction force, is shown in Table 1. The RMSE for each axis of the force in two trials for accuracy verification decreased when using calibration parameters corrected with the extended Kalman filter. The time curves for the individual components of the resultant force vector exerted below the left-hand grip position are shown in Fig. 6 for two pole vaulting trials. Temporal data were normalised between 0.02 s after the box contact (0%) and pole release (100%) to avoid the impact of the box contact. Among these components of the force vector, ${}^{\Sigma p}F_y$ and ${}^{\Sigma p}F_z$ predominated. ${}^{\Sigma p}F_y$ increased gradually after take off, and then reached its peak value at around 40% of the normalised time, before subsequently decreasing as the pole extended (Fig. 6 c,d). Conversely, ${}^{\Sigma p}F_z$ showed an immediate increase to the instant of 10% normalised time and then maintenance of this level until approximately 70% of the normalised time (Fig. 6 e,f).

Table 1. The root mean square error of the resultant force is presented between estimated from sensor output and calculated from ground reaction force.

		RMSE [N]			%RMSE [%]		
		${}^{\Sigma p}F_x$	${}^{\Sigma p}F_y$	${}^{\Sigma p}F_z$	${}^{\Sigma p}F_x$	${}^{\Sigma p}F_y$	${}^{\Sigma p}F_z$
Calculated value using Eq. (9)	Trial01	63.5	1177.5	260.1	53.1	151.5	45.7
	Trial02	65.9	1181.5	286.1	72.8	148.6	51.0
Calculated value using Eq. (14)	Trial01	31.6	31.5	56.8	26.4	4.1	9.9
	Trial02	25.7	46.9	62.8	28.1	5.9	11.2

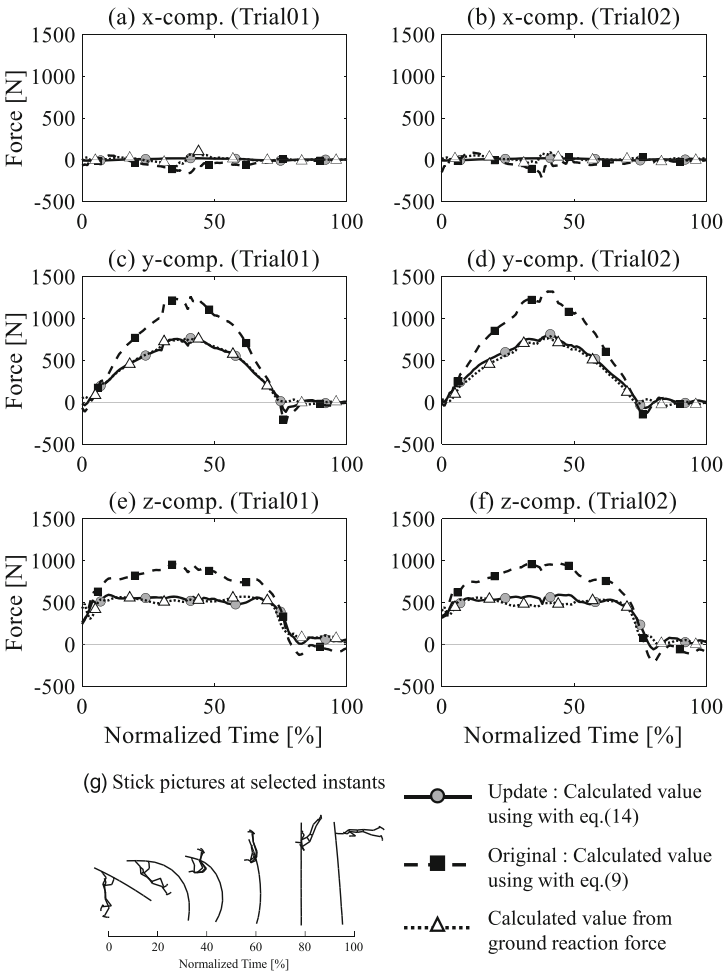


Fig. 6. The three components of the resultant force vector exerted below the left-hand grip position obtained through the individual equations for two pole vaulting trials.

4 Discussion and Conclusions

The purposes of this study were to propose and build a wireless instrumented pole that can measure the forces and moments acting close to the individual hand grip regions and to verify its accuracy. A total of ten sets of Wheatstone bridge circuits consisting of strain gauges have enabled detection of the moments and forces acting on the lower side of the grip regions. Since the output signals were amplified with a lightweight miniature data-logger type dynamic strain amplifier being attached on the pole, vaulters used the instrumented pole in vaulting trials without being constrained by long electrical wires.

The quasi-static calibration parameters obtained from the quasi-static calibration showed a good match when applied to the strain gauge signal outputs during the calibration trial, but the quasi-static parameters did not lead to good agreement between the

calculated and measured forces and moments during the pole vaulting trials. By contrast, the modified calibration parameters estimated using the extended Kalman filter reduced error in the calculated and measured forces and moments during the pole vaulting trials. This will allow more robust estimation of the conversion coefficients even when there is vibration, such as when the pole plug contacts the box, or other noise in the vaulting trials. Pole vaulting is a highly non-linear discipline, and jump height is not solely determined by pole parameters such as length and stiffness, or initial conditions like takeoff speed and angle; the force exerted by the pole vaulter on the pole after takeoff is also a crucial factor. However, it is difficult to estimate each hand force, especially the distribution ratio of their long-axis components, due to the large deformation of the pole. Therefore, measuring the forces exerted by the pole vaulter with high accuracy is necessary to contribute to better understanding of the nonlinear dynamics in pole vaulting.

The future work is to measure the forces and moments accurately by the instrumented pole without collection using the pole tip force measured with a force plate situated below the vaulting box.

Funding. This work was supported by a KAKENHI grant from the Japan Society for the Promotion of Science [grant number 20H04068].

References

1. Arampatizis, A., Schade, F., Brüggemann, G.-P.: Effect of the pole–human body interaction on pole vaulting performance. J. Biomech. **37**, 1353–1360 (2004)
2. Schade, F., Arampatizis, A., Brüggemann, G.-P.: Reproducibility of energy parameters in the pole vault. J. Biomech. **39**, 1464–1471 (2006)
3. McGinnis, P.M., Bergman, L.A.: An inverse dynamic analysis of the pole vault. Int. J. Sport Biomech. **2**, 186–201 (1986)
4. Morlier, J., Mesnard, M.: Influence of the moment exerted by the athlete on the pole in pole-vaulting performance. J. Biomech. **40**, 2261–2267 (2007)
5. Vaughan, C.L., Hay, J.G., Andrews, J.G. : Closed loop problems in biomechanics. Part II-An optimization approach. J. Biomech. **15**(3), 201–210 (1982)
6. Saito, A., Doki, H., Kondo, A., Hirose, K.: An attempt for developing the measurement system of reaction force from snow surface for private ski boots by compact force sensors. Procedia Eng. **112**, 326–331 (2015)
7. Kondo, A., Doki, H., Hirose, K.: An attempt for measuring snowboarding performance using compact force sensors. Procedia Eng. **72**, 86–91 (2014)
8. Reiser, R.F., Peterson, M.L., Broker, J.P.: Instrumented bicycle pedals for dynamic measurement of propulsive cycling loads. Sports Eng. **6**, 41–48 (2003)
9. Koike, S., Iida, H., Kawamura, T., Fujii, N., Ae, M.: An instrumented bat for simultaneous measurement of forces and moments exerted by the hands during batting. Eng. Sport **5**, 194–199 (2004)
10. Koike, S., Iida, H., Shiraki, H., Ae, M.: An instrumented grip handle for golf clubs to measure forces and moments exerted by each hand during swing motion. Eng. Sport **6**, 137–142 (2006)
11. Lennart, L.: Asymptotic Behavior of the extended Kalman filter as a parameter estimator for linear systems. IEEE Trans. Autom. Control **24**(1), 36–50 (1979)

Fluid Mechanics and Fluid-Structure Interaction

Coupled Bluff Body Energy Harvesters: Vortex Influenced Dynamics

Khawar Zamman Wani[1], Manoj Pandey[1(✉)], and Balakumar Balachandran[2]

[1] Department of Mechanical Engineering, Indian Institute of Technology Madras, Chennai 600036, Tamil Nadu, India
{me21d406,mpandey}@smail.iitm.ac.in

[2] Department of Mechanical Engineering, University of Maryland, College Park, MD 20742, USA
balab@umd.edu

Abstract. When placed in a fluid flowing with a certain critical speed, bluff bodies can undergo vortex-induced vibrations. These vortex-induced vibrations can be harnessed for conversion of fluid energy into electrical energy. Here, a cantilever beam with a cylindrical tip and a piezoelectric patch attached to the base of the cantilever is considered. The structure is placed in fluid flow. Experiments with two such bluff body oscillators placed in close proximity show an amplified response that can be attributed to the interaction between the wakes of the cylinders. This results in an enhanced power output. In this work, the authors model and simulate single and two oscillator systems by using finite element models for fluid-structure interactions. For the single oscillator, results from a wake oscillator based reduced order model are used to compare and verify the finite element simulation results. For the coupled oscillator case, the effect of varying the gap to diameter (G/D) ratio on the response is analysed. System dynamics and the potential for energy harvesting are discussed.

Keywords: Bluff body · Vortex-induced vibrations · Energy harvester

1 Introduction

In recent years, microelectromechanical sensors have seen a notable decrease in size and power consumption. Having an appropriate battery powered system can help a sensor stay off-grid and be used as an autonomous monitoring system.

The power available in a fluid flow is proportional to the density of the fluid and the cube of the fluid speed according to the relation proposed by Betz [1]:

$$P_{max} = c_B \frac{A_f}{2} \rho U_\infty^3 \tag{1}$$

This power can be harnessed by employing various methods. Conventional means include devices such as wind turbines, which use an electric motor consisting of a rotor-stator arrangement to convert the mechanical energy into electrical

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 355–369, 2025.
https://doi.org/10.1007/978-3-031-72794-8_28

energy. However, at lower length scales, the viscous effects dominate resulting in tremendous losses. Therefore, alternate device designs that can be used to harvest energy even on small scales are needed. These energy sources could trickle charge the batteries and enable continuous operation even in remote and inaccessible places. Energy harvesting technology has been explored as an enabling technology for powering unmanned autonomous systems, such as wireless sensor networks [2–5].

Vortex-induced vibrations are a common phenomenon observed in nature. Bluff bodies, when placed in a fluid with steady flow and a speed greater than a critical speed, shed vortices in the wake of the body. These vortices are shed alternately from each side of the structure, forming a von Kármán vortex street. The shedding frequency is related to a dimensionless number called the Strouhal number, given by:

$$St = \frac{f_s L}{U} \tag{2}$$

where f_s is the frequency of vortex shedding, L is a characteristic length, and U is the flow speed. These vortices create fluctuations in the pressure distribution along the surface of the structure, leading to the development of lift forces, resulting in motions transverse to the flow. Such fluctuating pressures can also be observed on thin flexible structures, such as a beam placed in the wake of the bluff body, resulting in vortex-induced vibrations. Structures subjected to vortex-induced vibrations can experience various types of responses, including limit cycle oscillations, chaos, and qualitative changes due to bifurcations [6–10]. These vibrations can be used for energy harvesting purposes by converting the associated mechanical energy into storable electrical form by using suitable energy transduction mechanisms based on active materials such as piezoelectric materials. For a piezo instrumented cantilever beam, the power harvested is given by the following relation[11].

$$P \propto y_t f^3 \tag{3}$$

where P is the harvested power, y_t is the tip displacement of the cantilevered beam, and f is the frequency of oscillation. Since the generated power is proportional to the tip displacement, the magnitude of the displacement can be used as a measure of the energy harvested.

Akaydin *et al.* [12] presented and experimentally studied a design with a short piezoelectric cantilever beam, vibrating in the wake of a stationary cylindrical bluff body, under steady inlet flow. However, because of the low aeroelastic coupling coefficient, only low amplitude linear beam vibrations are observed, except in a narrow band of flow speeds, where the beam vibration frequency matches with the vortex shedding frequency of the bluff body, resulting in a lock-in condition. A substantial increase in aeroelastic coupling was achieved by directly attaching a large aspect ratio cylinder to the free end of the cantilever beam [13]. This resulted in large amplitude oscillations and a high level of energy harvested. The flow speed required for maximum power generation was also reduced. Dai *et al.* [14] developed a reduced order model for the improved energy harvester

design proposed by Akaydin *et al.* The lift on the cylinder was modelled as a van der Pol oscillator by using the wake oscillator model proposed by Facchinetti *et al.* [15]. The cantilever beam and cylinder arrangement was modelled as an Euler-Bernoulli beam with a tip mass.

In the experimental study conducted by Azadeh-Ranjbar *et al.* [16], these authors compared the performance of the linear and nonlinear fluid energy harvesters proposed by Akaydin *et al.*. as well as an array of two nonlinear fluid energy harvesters stacked on top of each other normal to the flow direction. They concluded that the nonlinear array had a significantly larger power output and a wider operating speed range compared to the other cases. This improvement is attributable to the flow interactions and the resulting coupling between the two oscillators. In the same study, Azadeh-Ranjbar *et al.* presented an experimental comparison of the energy harvested in the two oscillator case for varying gap (G) over diameter (D) ratios. For large (G/D) ratios, the oscillators behaved like independent harvesters. However, for a small G/D ratio, the authors noticed an increase in the operational speed range and a large maximum power output (approximately ten times for $G/D \approx 0$) compared to that of an isolated harvester.

Various configurations of the oscillators would result in different flow interactions and varying coupling between them, possibly resulting in higher energy harvesting for optimal configurations. A robust computational model can help to carry out a parameter optimisation analysis. However, no such existing model for coupled bluff body oscillator system has been found in the literature. Even for the single oscillator case, a model that does not require parameters fitted from the experiments is not available. Here, the authors address this research gap by presenting a finite element (FE) model for fluid-structure interactions in this work. The computational model can be used to predict the behaviour of single and coupled nonlinear oscillator cases. The results are benchmarked against the experimental results presented by Azadeh-Ranjbar *et al.* [16].

2 Modelling

Due to strong coupling between the fluid and the structural motions, especially, in the case of multiple oscillators, a fully coupled fluid structure interaction based modelling scheme is employed. The fluid-structure interaction (FSI) problem considered here, which has been implemented through the finite element package Abaqus [17], consists of an incompressible Newtonian fluid flow interacting with an elastic solid undergoing large deformations. In Abaqus, one uses a sequentially coupled scheme in which non-overlapping computational domains are used to model the fluid (Ω_f) and structural (Ω_s) regions, with a conforming interface (Γ_{I}) as shown in Fig. 1. The structural solver is used to determine the motion of the structure, the movement of the fluid boundary, as well as to determine the velocity BC on the same. The fluid is used to solve for velocity and pressure fields in the fluid domain, and the resulting force on the boundary is applied as an external force to the structural domain in the next time step. A body-fixed

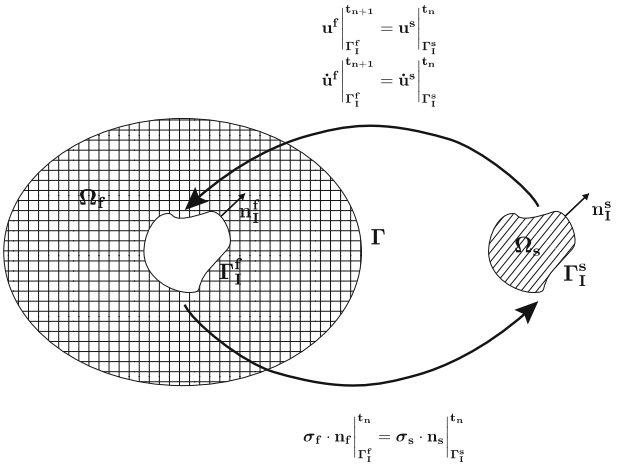


Fig. 1. Sequential coupled FSI approach.

Lagrangian mesh is used for structure, while a space-fixed Eulerian mesh is more suitable for fluid modelling, due to large deformations involved in the same. The motion of the structure results in the motion of the interface Γ_{I}, which causes the boundary mesh in the fluid to distort. In Abaqus, the arbitrary Lagrangian Eulerian (ALE) approach is employed to remesh the fluid domain and solve the governing equations on the moving mesh.

The material point displacement $\mathbf{u}^s$ in the structure is obtained from the momentum equation, written in the Lagrangian description as:

$$\rho^s \ddot{\mathbf{u}}^s - \boldsymbol{\nabla} \cdot (J\boldsymbol{\sigma}^s \boldsymbol{F}^{-\top}) + \mathbf{b}^s = \mathbf{0} \quad \text{in } \Omega_s \tag{4}$$

Here, $\boldsymbol{\sigma}^s$ is the Cauchy stress, $\boldsymbol{F} = \boldsymbol{I} + \boldsymbol{\nabla}\mathbf{u}^s$ is the deformation gradient tensor, and J is the third invariant of the deformation gradient tensor. Linear elastic material behaviour and geometrically nonlinear deformation are captured by using the St. Venant-Kirchhoff constitutive law, given by

$$\boldsymbol{\sigma}^s = \frac{1}{J}\boldsymbol{F}(\lambda^s \operatorname{tr}(\boldsymbol{E})\boldsymbol{I} + 2\mu^s \boldsymbol{E})\boldsymbol{F}^{\top} \tag{5}$$

where $\boldsymbol{E} = \dfrac{1}{2}(\boldsymbol{F}^{\top}\boldsymbol{F} - \boldsymbol{I})$. λ^s, and μ^s, are the Lamé constants.

The velocity field $\mathbf{v}^f$ in the fluid domain is obtained by using the Navier-Stokes equation, written in the Eulerian description as

$$\rho^f \frac{\partial \mathbf{v}^f}{\partial t} + \mathbf{v}^f \cdot \boldsymbol{\nabla}\mathbf{v}^f - \boldsymbol{\nabla} \cdot \boldsymbol{\sigma}^f + \mathbf{b}^f = \mathbf{0} \quad \text{in } \Omega_f \tag{6}$$

Here, ρ^f is the fluid density, $\mathbf{b}^f$ is the body force, and $\boldsymbol{\sigma}^f$ is the fluid stress. The fluid is assumed to be Newtonian, giving

$$\boldsymbol{\sigma}^f = -p\boldsymbol{I} + \eta(\boldsymbol{\nabla}\mathbf{v}^f + \boldsymbol{\nabla}\mathbf{v}^{f\top}) \tag{7}$$

where p is the thermodynamic pressure and η is the dynamic viscosity.

The fluid is considered to be incompressible, and hence, the Pressure-Poisson equation is used to determine the pressure from the velocity field.

$$\boldsymbol{\nabla}^2 p = -\boldsymbol{\nabla}\mathbf{v}^f : \boldsymbol{\nabla}\mathbf{v}^f \tag{8}$$

At the same time, the displacement and traction continuity at the interface is maintained by using the following conditions.

$$\begin{aligned} \mathbf{v}^f\Big|_{\Gamma_f} &= \frac{d\mathbf{u}^s}{dt}\Big|_{\Gamma_s} \\ \boldsymbol{\sigma}^f : \mathbf{n}\Big|_{\Gamma_f} &= \boldsymbol{\sigma}^s : \mathbf{n}\Big|_{\Gamma_s} \end{aligned} \tag{9}$$

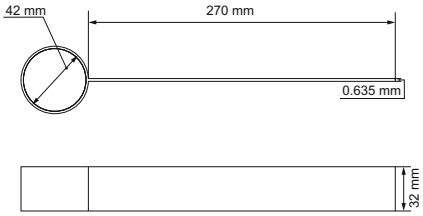
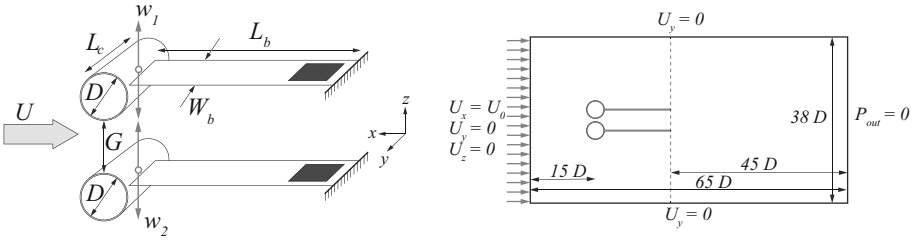


Fig. 2. Dimensions of the harvester.



(a) (b)

Fig. 3. (a) Schematic of the two oscillator piezoelectric fluid energy harvester. (b) Schematic of the fluid domain.

2.1 Finite Element Model

A two-dimensional (2D) approximation for the flow based energy harvester is employed in Abaqus to keep the computation tractable. The vibrating structure consisted of the cantilever beam and the attached cylinder. The dimensions of the harvester are shown in Fig. 2. The schematic of the two oscillator harvester is shown in Fig. 3a. The cylinder is considered to be rigid and the beam material is modelled as Saint Venant-Kirchhoff material. The beam is made of aluminium. The beam is constrained to move in the $x - y$ plane. The harvester model parameters are given in Table 1. C3D8R elements in Abaqus, which are linear hexahedral elements with reduced integration (one integration point), are used to model the structure. A rectangular fluid domain with boundaries at least $15D$ away from the FSI interface is used, as shown in Fig. 3b, in order to avoid boundary effects. The fluid enters the domain normal to the boundary from the inlet on the left and is discharged at zero pressure through the outlet on the right. Symmetric boundary conditions are applied to the upper and lower walls of the domain. The fluid is air with a density of 1.168 kg m^{-3} and viscosity of 1.8×10^{-5} $m^2 s^{-1}$. F3D8R elements in Abaqus, which are linear, reduced integration hexahedral elements, are used to model the fluid domain. Implicit dynamic simulations have been performed for a flow speed of $U_0 = 0.1$ ms^{-1} corresponding to $Re \approx 270$, until a steady state is achieved. A low Re number is chosen, as the 2D assumption is not valid at high Re numbers [18].

For enforcing the 2D approximation, the out-of-plane displacement and velocity are set to zero. A very fine mesh is used in the proximity of the FSI interface, and this mesh gradually became coarser towards the domain boundaries as shown in Fig. 4a. To determine the optimal mesh, a mesh convergence study was conducted. The convergence study involved simulating the system for a total of 15 s, corresponding to ≈ 3 vortex shedding cycles and assuming the finest mesh solution to be the correct one. The mean squared error in velocity is plotted against the number of elements, and the coarsest mesh in the tapered region of the plot is chosen as the optimal one as shown in Fig. 5. The beam has four elements through the thickness of the cantilever to accurately capture the bending stress variation, and three elements across the width as shown in Fig. 4b. To successfully capture the fluid-structure interaction between the structure and the fluid, the FSI boundary in the fluid domain should match exactly with the FSI boundary in the structural domain. Since a two-dimensional domain is considered, the model corresponds to a cylinder and a cantilever of equal width. Due to this, the mass of the cylinder needs to be artificially corrected by increasing the density of the cylinder.

The 2D approximation in the finite element model also results in a flow field that is representative of a flow field over an infinite cylinder with a splitter plate attached throughout its entire length [19]. This has an effect on the generated lift and is discussed in the following section.

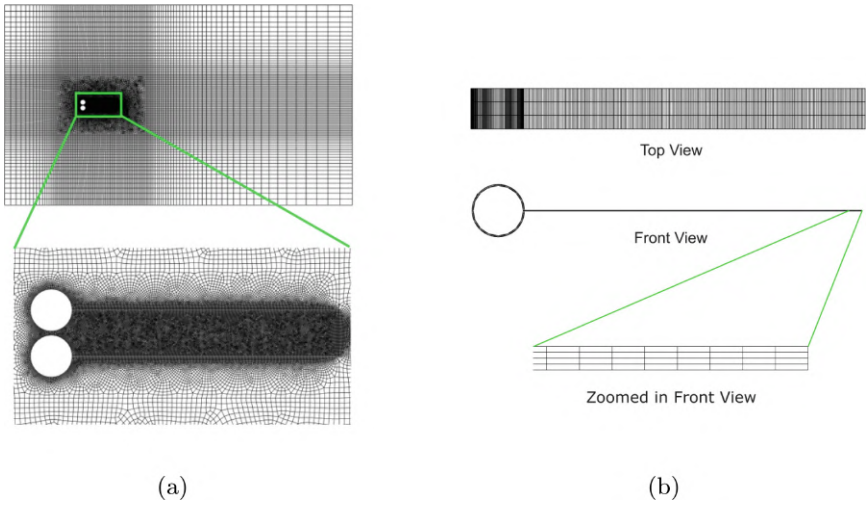


Fig. 4. (a) Fluid domain mesh. (b) Harvester mesh.

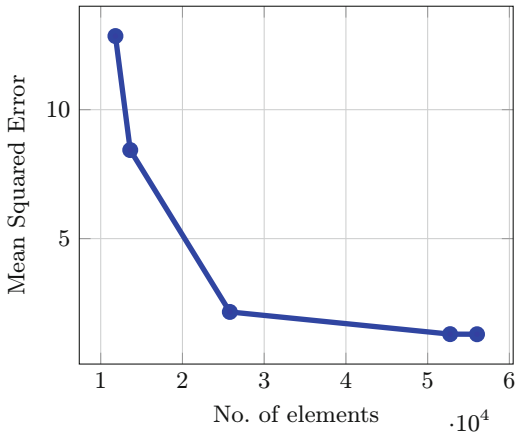


Fig. 5. Variation of mean squared error with number of elements.

2.2 Reduced Order Model

Results obtained from reduced order model (ROM) developed by Dai *et al.* [14] are used to verify the Finite Element Analysis results. The device is modelled as an Euler-Bernoulli cantilever beam in the ROM. The displacement $w(x,t)$ of the harvester is expressed in terms of time-varying coefficients $r_i(t)$ and linear undamped eigenfunctions $\varphi_i(x)$:

Table 1. Model Parameters [16]

	Beam	Cylinder
Length L_b, L_c	270 mm	32 mm
Width W_b	32 mm	–
Thickness t_b	0.635 mm	–
Diameter D	–	42 mm (outer), 40 mm (inner)
Mass Density ρ_b, ρ_c	2730 kg m^{-3}	3640 kg m^{-3}
Mass m_b, m_c	15 g	15 g
Young's Modulus E_b, E_c	73 GPa	∞
Polar Moment of Inertia J_c	–	2.205×10^{-6} kg m^2

$$w(x,t) = \sum_{i=1}^{n} r_i(t)\varphi_i(x) \tag{10}$$

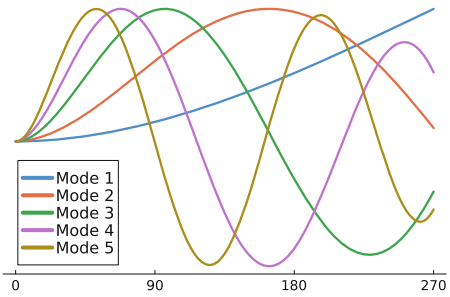


Fig. 6. Analytically obtained mode shapes of the harvester.

The mode shapes are modified to take into account the addition of the cylinder at the free end, which is considered to be a lumped mass at a distance of $D/2$ from the free end [20]. The mode shapes, which are shown in Fig. 6, are given by:

$$\varphi_i(x) = A_i \left[\cos\left(\frac{\beta_i}{L_b}x\right) - \cosh\left(\frac{\beta_i}{L_b}x\right) + \xi_i \left(\sin\left(\frac{\beta_i}{L_b}x\right) - \sinh\left(\frac{\beta_i}{L_b}x\right) \right) \right] \tag{11}$$

where

$$\xi_i = \frac{\sin(\beta_i) - \sinh(\beta_i) + \kappa_1\beta_i\left[\cos(\beta_i) - \cosh(\beta_i)\right] - \kappa_1\kappa_2\beta_i^2\left[\sin(\beta_i) + \sinh(\beta_i)\right]}{\cos(\beta_i) + \cosh(\beta_i) - \kappa_1\beta_i\left[\sin(\beta_i) - \sinh(\beta_i)\right] - \kappa_1\kappa_2\beta_i^2\left[\cos(\beta_i) - \cosh(\beta_i)\right]} \tag{12}$$

β_i are the roots of the equation:

$$1 + \cos(\beta)\cosh(\beta) + \kappa_1\beta[\cos(\beta)\sinh(\beta) - \cosh(\beta)\sin(\beta)] - 2\kappa_1\kappa_2\beta^2\sin(\beta)\sinh(\beta) - \frac{4}{3}\kappa_1\kappa_2^2\beta^3[\cos(\beta)\sinh(\beta) + \cosh(\beta)\sin(\beta)] + \frac{1}{3}\kappa_1^2\kappa_2^2\beta^4[1 - \cos(\beta)\cosh(\beta)] = 0 \tag{13}$$

$$\kappa_1 = \frac{m_c}{m_b} \text{ and } \kappa_2 = \frac{R_c}{L_b} = \frac{D}{2L_b}$$

The wake oscillator model proposed by Facchinetti *et al.* [15] is used. In this model, the fluctuating lift in the wake of the harvester is simulated by using a van der Pol oscillator. In Eq. (15), q is a variable associated with the fluctuating lift on the cylinder. The effect of structural oscillations on the fluid is modelled by an inertial forcing term. After carrying out a Galerkin projection, the following reduced order model is obtained:

$$\ddot{r}_i + 2\zeta_i\omega_i\dot{r}_i + \frac{C_D\rho_0 D U_0 L_c}{2}\left(\varphi_i(L_b) + \frac{D}{2}\varphi_i'(L_b)\right) \times \left[\sum_{j=1}^{n}\left(\varphi_j(L_b) + \frac{D}{2}\varphi_j'(L_b)\right)\dot{r}_j\right] + \omega_i^2 r_i = \alpha_i q \tag{14}$$

$$\ddot{q} + \lambda\omega_s(q^2 - 1)\dot{q} + \omega_s^2 q = \frac{A}{D}\sum_{i=1}^{n}\left(\varphi_i(L_b) + \frac{D}{2}\varphi_i'(L_b)\right)\ddot{r}_i \tag{15}$$

Here ζ_i and ω_i are respectively the damping ratio and natural frequency of the i^{th} structural mode, C_D is the mean drag coefficient, ω_s the vortex shedding frequency, and λ and A are empirical coefficients, respectively. The prime ($\prime$) notation denotes differentiation with respect to x.

The coupling coefficient, α_i, is given by

$$\alpha_i = C_{L_0}\rho_0 D U_0^2 L_c\left[\varphi_i(L_b) + \frac{D}{2}\varphi_i'(L_b)\right] \Big/ 4 \tag{16}$$

where C_{L_0} is the lift coefficient. C_D, λ, and A are assigned the same values as proposed by Dai *et al.* [14] with $C_D = 1.2$, $\lambda = 0.24$, and $A = 12$. The authors determined the Strouhal number by measuring the vortex-shedding frequency in the near wake region. A value of $St = 0.0785$ is obtained compared to 0.135, as measured experimentally by Azadeh-Ranjbar *et al.* [16], giving $\omega_s = 1.13$ rad s^{-1}. This discrepancy can be attributed to the 2D model considered here as a representation of a cylinder with a splitter plate. Similarly, the total lift in the 2D approximation used here would be higher compared to the actual system, and therefore $C_{L_0} = 0.5$ is obtained by fitting the ROM to the FE model, which is higher than $C_{L_0} = 0.3$ as considered in [14].

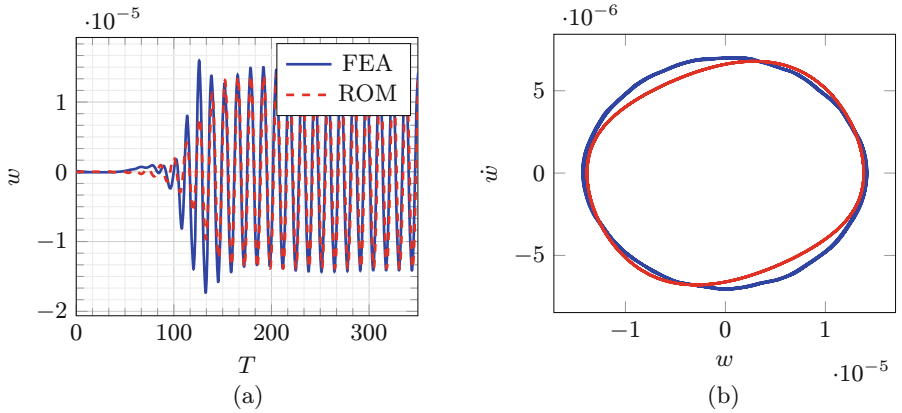


Fig. 7. Response of the single oscillator: (a) Displacement and (b) Phase portrait.

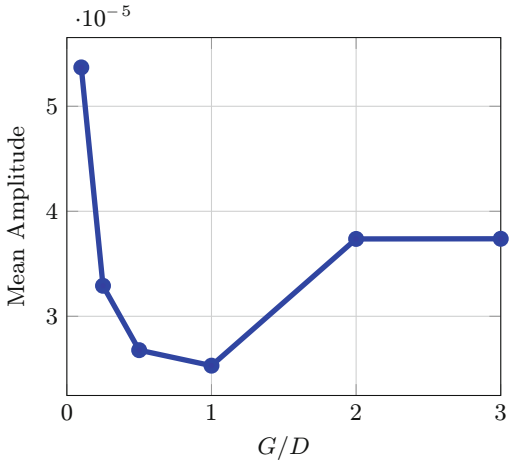


Fig. 8. Variation of maximum displacement with G/D ratio.

3 Results and Discussion

In Fig. 7, the displacement response of the tip of the harvester (point of attachment of the beam to the cylinder), obtained from finite element simulations is plotted along with the response obtained by simulating the reduced order model for the system parameters. The displacement response, simulation time, and velocity response are normalized by dividing with D, D/U_0, and U_0, respectively, where D is the diameter of the cylinder, and U_0 is the flow speed. A good match is observed between the time histories and phase portrait obtained from the two models. Only one mode response was considered in this study, as the Reynolds number is fairly low and the response is mostly due to the first mode vibration of the beam. For the model considered, the vortex shedding

frequency $\omega_s = 1.13$ rad s^{-1} and the fundamental frequency of the harvester $\omega_1 = 18.41$ rad s^{-1}. Hence, the system is operating away from the lock-in region. The fundamental frequency has been obtained from modal analysis of the finite element model and is found to match the frequency determined experimentally in earlier work [16]. For the reduced order model, the assumed initial condition is $[r\ \dot{r}\ q\ \dot{q}]^\top = [0\ 0\ 0.001\ 0]^\top$.

Finite element simulations of the two oscillator system were performed for the same flow and boundary conditions as for the single oscillator case. The effect of varying the G/D ratio was studied by plotting the mean peak-to-peak amplitude of the displacement response of the tips of the two oscillators against the G/D ratio. It is evident from Fig. 8 that for very small G/D ratios; that is, $G/D < 0.25$, an amplified response is observed. For $G/D > 2$, the response approaches the response of an isolated harvester. However, in the range $0.25 < G/D < 2$, the coupled oscillator harvester is found to perform worse than an isolated single harvester. The minimum vibration amplitude is observed for $G/D = 1$. These results are consistent with the experimental findings given in reference [16].

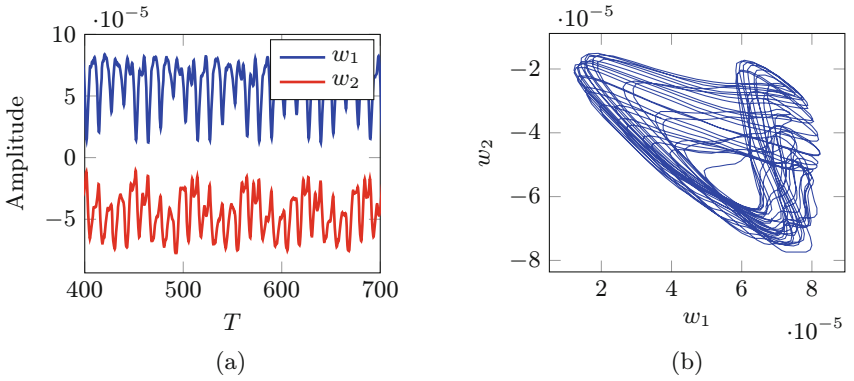


Fig. 9. Oscillator dynamics for $G/D = 0.1$: (a) Time histories. (b) Phase plane plots.

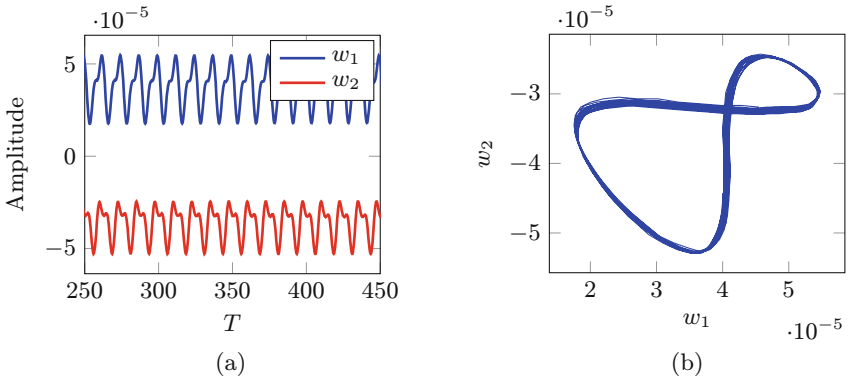


Fig. 10. Oscillator dynamics for $G/D = 0.25$: (a) Time histories. (b) Phase plane plots.

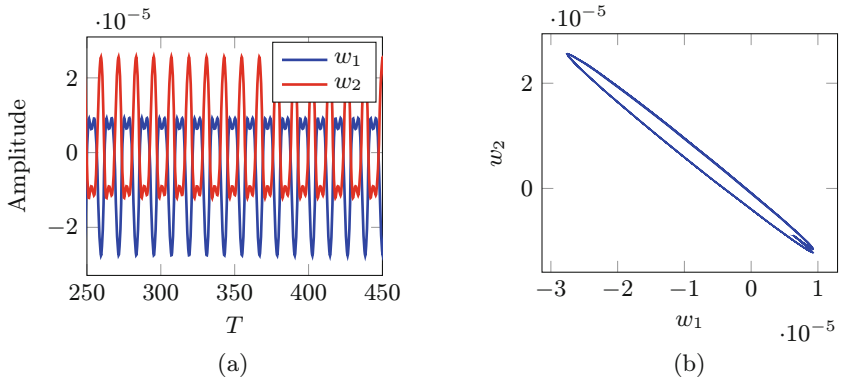


Fig. 11. Oscillator dynamics for $G/D = 2$: (a) Time histories. (b) Phase plane plots.

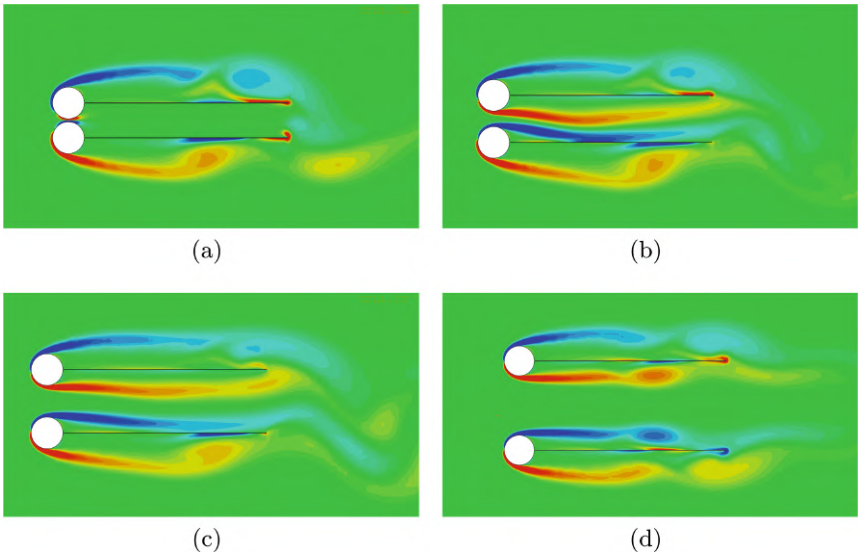


Fig. 12. Vorticity contours: (a) $G/D = 0.1$, (b) $G/D = 0.5$, (c) $G/D = 1$, and (c) $G/D = 2$.

The time histories and phase plane plots for the responses obtained for various G/D ratios are plotted in Figs. 9, 10, and 11. For $G/D >= 2$, the responses of the two harvesters are out of phase, as evident in Fig. 11. As G/D is reduced, the response becomes asymmetric, and for $G/D = 0.1$, the response appears to be chaotic, as seen in Fig. 9. In all cases, the two harvesters oscillate over a non-zero mean value.

Vorticity contours suggest that there is a bias in the flow direction in the two cylinder case for an intermediate spacing as shown in Figs. 12(b) and (c). This is due to the fact that the harvesters are in the biased flow regime as described

in earlier work [21–31]. The asymmetrical flow pattern is characterised by a gap flow biased towards one of the two cylinders. The cylinder toward which the flow is biased has a narrow near wake, higher frequency of vortex shedding, and a higher drag coefficient in contrast to the other cylinder which has a wider near wake, lower frequency vortex shedding, and lower drag coefficient.

4 Summary

The authors have developed a finite element computational model to study fluid-structure interactions associated with vortex-induced vibrations of a cylindrical bluff body based energy harvester. Apart from a finite element model, a reduced order model is also developed to explore the nonlinear dynamics of the considered bluff body oscillators for a single oscillator case. The model results are compared with experimental results from the literature. The computational model is used to simulate a stack consisting of two oscillators arranged perpendicular to the flow direction. The trends obtained from the simulations for varying the gap matches the experimental observations. The response is found to transition from an out-of-phase periodic motion for the case of a large gap to an aperiodic (chaotic) motion of large amplitude for a small gap, going through a minima for intermediate gaps. The vorticity contours are found to show a biased flow towards either of the oscillators, with a decreasing gap, resulting in differential lift on the two oscillators. In the future, the authors intend to derive a reduced order model for the two oscillator system and model the interaction effects with appropriate coupling terms in the model.

Acknowledgments. Partial support received for this work by Professor B. Balachandran of the University of Maryland, College Park, USA through the U.S. National Science Foundation through Grant No. CMMI2131594 is gratefully acknowledged.

References

1. Betz, A.: Das maximum der theoretisch möglichen ausnützung des windes durch windmotoren. Zeitschrift für das gesamte Turbinenwesen (1920)
2. Paradiso, J.A., Starner, T.: Energy scavenging for mobile and wireless electronics. IEEE Pervasive Comput. **4**(1), 18–27 (2005)
3. Bogue, R.: Energy harvesting and wireless sensors: a review of recent developments. Sens. Rev. **29**(3), 194–199 (2009)
4. Khaligh, A., Zeng, P., Zheng, C.: Kinetic energy harvesting using piezoelectric and electromagnetic technologies-state of the art. IEEE Trans. Industr. Electron. **57**(3), 850–860 (2009)
5. Rancourt, D., Tabesh, A., Fréchette, L.G.: Evaluation of centimeter-scale micro windmills: aerodynamics and electromagnetic power generation. Proc. PowerMEMS **20079**, 93–96 (2007)
6. Nayfeh, A.H., Mook, D.T.: Nonlinear oscillations. Wiley (2008)
7. Paidoussis, M.P.: Fluid-structure interactions: slender structures and axial flow, vol. 1. Academic Press (1998)

8. Dowell, E.H., Tang, D.: Nonlinear aeroelasticity and unsteady aerodynamics. AIAA J. **40**(9), 1697–1707 (2002)
9. Liu, L., Wong, Y.S., Lee, B.H.K.: Application of the centre manifold theory in non-linear aeroelasticity. J. Sound Vib. **234**(4), 641–659 (2000)
10. Vasconcellos, R., Abdelkefi, A.: Nonlinear dynamical analysis of an aeroelastic system with multi-segmented moment in the pitch degree-of-freedom. Commun. Nonlinear Sci. Numer. Simul. **20**(1), 324–334 (2015)
11. Mitcheson, P.D., Yeatman, E.M., Rao, G.K., Holmes, A.S., Green, T.C.: Energy harvesting from human and machine motion for wireless electronic devices. Proc. IEEE **96**(9), 1457–1486 (2008)
12. Akaydin, H., Elvin, N., Andreopoulos, Y.: Wake of a cylinder: a paradigm for energy harvesting with piezoelectric materials. Exp. Fluids **49**, 291–304 (2010)
13. Akaydin, H.D., Elvin, N., Andreopoulos, Y.: The performance of a self-excited fluidic energy harvester. Smart Mater. Struct. **21**(2), 025007 (2012)
14. Dai, H.L., Abdelkefi, A., Wang, L.: Theoretical modeling and nonlinear analysis of piezoelectric energy harvesting from vortex-induced vibrations. J. Intell. Mater. Syst. Struct. **25**(14), 1861–1874 (2014)
15. Facchinetti, M.L., De Langre, E., Biolley, F.: Coupling of structure and wake oscillators in vortex-induced vibrations. J. Fluids Struct. **19**(2), 123–140 (2004)
16. Azadeh-Ranjbar, V., Han, Y., Elvin, N., Andreopoulos, Y.: Non-linear piezoelectric fluidic energy harvesters: the mutual interaction of two oscillating cylinders. J. Intell. Mater. Syst. Struct. **31**(20), 2378–2389 (2020)
17. Dassault Systémes. Abaqus 6.14 documentation–theory guide (2015)
18. Williamson, C.H.K.: Three-dimensional wake transition. J. Fluid Mech. **328**, 345–407 (1996)
19. Song, J., Hu, G., Tse, K.T., Li, S.W., Kwok, K.C.S.: Performance of a circular cylinder piezoelectric wind energy harvester fitted with a splitter plate. Appl. Phys. Lett. **111**(22) (2017)
20. Krylov, S., Maimon, R.: Pull-in dynamics of an elastic beam actuated by continuously distributed electrostatic force. J. Vib. Acoust. **126**(3), 332–342 (2004)
21. Spivack, H.M.: Vortex frequency and flow pattern in the wake of two parallel cylinders at varied spacing normal to an air stream. J. Aeronaut. Sci. **13**(6), 289–301 (1946)
22. Hori, E.: Experiments on flow around a pair of parallel circular cylinders. In: Proceedings of 9th Japan National Congress for Applied Mechanics, pp. 231–234 (1959)
23. Ishigai, S., Nishikawa, E., Nishimura, K., Cho, K.: Experimental study on structure of gas flow in tube banks with tube axes normal to flow - 1. karman vortex flow from two tubes at various spacings. Bull JSME **15**(86), 949–956 (1972)
24. Bearman, P.W., Wadcock, A.J.: The interaction between a pair of circular cylinders normal to a stream. J. Fluid Mech. **61**(3), 499–511 (1973)
25. Williamson, C.H.K.: Evolution of a single wake behind a pair of bluff bodies. J. Fluid Mech. **159**, 1–18 (1985)
26. Kim, H.J., Durbin, P.A.: Investigation of the flow between a pair of circular cylinders in the flopping regime. J. Fluid Mech. **196**, 431–448 (1988)
27. Sumner, D., Wong, S.S.T., Price, S.J., Paidoussis, M.P.: Fluid behaviour of side-by-side circular cylinders in steady cross-flow. J. Fluids Struct. **13**(3), 309–338 (1999)
28. Zhiwen, W., Zhou, Y., Li, H.: Flow-visualization of a two side-by-side cylinder wake. J. Flow Vis. Image Process. **9**(2&3), 123–138 (2002)
29. Zhou, Y., Zhang, H.J., Yiu, M.W.: The turbulent wake of two side-by-side circular cylinders. J. Fluid Mech. **458**, 303–332 (2002)

30. Alam, M.M., Moriya, M., Sakamoto, H.: Aerodynamic characteristics of two side-by-side circular cylinders and application of wavelet analysis on the switching phenomenon. J. Fluids Struct. **18**(3–4), 325–346 (2003)
31. Xu, S.J., Zhou, Y., So, R.M.C.: Reynolds number effects on the flow structure behind two side-by-side cylinders. Phys. Fluids **15**(5), 1214–1219 (2003)

Dynamical Modeling and Analysis of Fluid-Structure Interaction Phenomena and Bifurcation Occurring in the Inducer of a Turbopump for Rocket Engines

Tsuyoshi Inoue[1(✉)], Hayate Okawa[1,2], and Hironori Horiguchi[2]

[1] Nagoya University, Nagoya, Japan
inoue.tsuyoshi@nagoya-u.jp
[2] Osaka University, Osaka, Japan

Abstract. In rocket engine turbopump inducers, the occurrence of unstable phenomena such as cavitation surge, rotating cavitation (RC), and attached asymmetric cavitation (AAC) are known. Particularly, in a three-blade inducer, the occurrence of RC and the transition from RC to AAC have been observed in experiment. However, no analytical explanation of the transition from RC to AAC has yet been provided. In this paper, transition phenomenon from RC to AAC is investigated and explained by considering the interaction between the fluid and the rotor systems. A coupled model of the one-dimensional flow path model of the inducer, a rotor model, and a two-dimensional flow model in a cascade was constructed. The transition from RC to AAC was observed and explained when the amplitude of the inducer's synchronous whirling motion increases.

Keywords: Attached asymmetric cavitation · Fluid-Structure Interaction Phenomena · Coupled Analysis · Rocket engine turbopump inducer

1 Introduction

Inducers are often used in rocket engine turbopumps to improve suction performance. It is known that fluid flow instabilities such as cavitation surge, rotating cavitation (RC), and attached asymmetric cavitation (AAC) may occur in inducers [1]. Many studies using CFD, experimental, and analytical methods on RC have been reported. However, there are few studies on the transition phenomena from RC to AAC. Yoshida et al. [2] reported a detailed investigation of the transition from RC to AAC when the cavitation number, namely inlet pressure, was decreased in an experiment on a three-blade inducer. Yoshida et al. [3] considered that this transition from RC to AAC is one of the self-excited oscillations caused by the coupling of the fluid and rotor systems. Kobayashi et al. [4] conducted experiments on a three-blade inducer with changing the magnitude of rotor imbalance and clarified the effect of the amplitude of synchronous whirling motion of the inducer on the occurrence of AAC.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 370–382, 2025.
https://doi.org/10.1007/978-3-031-72794-8_29

The authors have attempted to explain this RC to AAC transition phenomenon theoretically by considering the interaction between the fluid and rotor systems [5]. A coupled 1D flow path model of an inducer and a rotor model was used, and the transition was investigated. They assumed the sign change of the pump dynamic coefficient, which consists of cavitation compliance and mass flow gain factor, in one of the three flow paths between the blades based on the Yoshida's result [2], and showed that the transition from RC to AAC can be reproduced. However, a theoretical explanation of this assumed sign change has not yet been given.

2 Theoretical Model and Governing equation

2.1 Coupled Model

Figure 1 shows a schematic diagram of the coupled rigid-rotor system, one-dimensional flow path system [5], and two-dimensional cascade system [6]. The coupled system of rigid-rotor model and one-dimensional flow path model uses pump dynamic characteristic coefficients (cavitation compliance and mass flow gain factor) as inputs, and outputs the volume flow rate at inlet of the three flow paths. The two-dimensional flow model in a cascade system takes the inlet flow rate as input and outputs the pump dynamic characteristic of each flow path by analyzing the cavity length and thickness of the three paths. The pump dynamic characteristic coefficients of the three paths are again used as inputs to the couples system of rigid-rotor model and the one-dimensional flow path model. The coupled analysis of these three models attempts to verify the transition from rotating cavitation to attached asymmetric cavitation without assuming the characteristics in cavity geometry variation [5].

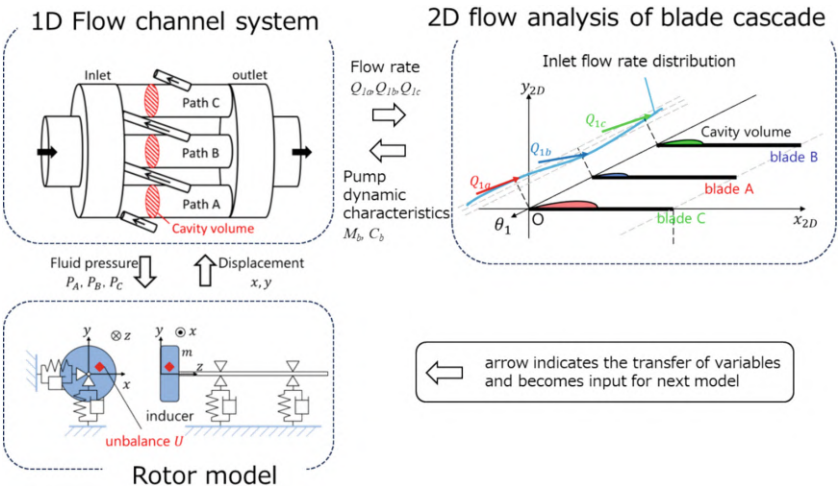


Fig. 1. Theoretical model of coupled system of 1D flow path model, rigid-rotor model, and 2D cascade model

2.2 1D Flow Path Model

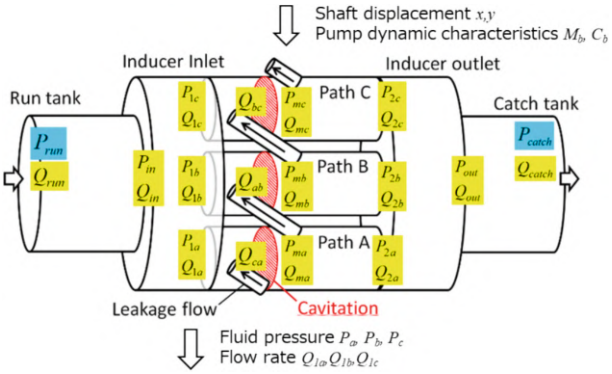


Fig. 2. One dimensional flow path model of three bladed inducer [5]

Figure 2 shows a one-dimensional flow path model of the inducer flow. The detail of governing equation of this model is shown in previous paper [5], and only the brief explanation is given here.

The independent variables are pressure P_{in} and flow rate Q_{in} at the inducer inlet; pressure P_{out} and flow rate Q_{out} at the inducer outlet; inlet pressure $P_{1i}(i = a, b, c)$ and flow rate $Q_{1i}(i = a, b, c)$; middle part pressure $P_{mi}(i = a, b, c)$ and flow rate $Q_{mi}(i = a, b, c)$; outlet pressure $P_{2i}(i = a, b, c)$ and flow rate $Q_{2i}(i = a, b, c)$; and flow rate Q_{ab}, Q_{bc}, Q_{ca} in the leakage flow paths. The supply pressure P_{run} and discharge pressure P_{catch} of the turbo pump are given as boundary conditions. The working fluid is considered incompressible. The inducer model flow shown in Figure 2 has been formulated using a set of nine fundamental equations [5]. As an example, the equation of motion in the first half of the flow path is shown as follows:

$$
\begin{aligned}
&P_{1*} - P_{in} + \rho L_{1*} \frac{dQ_{1*}}{dt} + R_{1*} Q_{1*}^2 = 0 \\
&L_{1*} = \frac{l_{1*}}{A_{1*}}, \quad R_{1*} = \frac{\rho l_{1*}}{2A_{1*}^2 D_{1*}} \lambda_{1*} \quad (* = a, b, c)
\end{aligned}
\tag{1}
$$

The $L_{1*}(* = a, b, c)$ indicates the inertance of each flow path, and the $R_{1*}(* = a, b, c)$ indicates the resistance of each flow path. In the similar manner, the equation of motion in the latter half of the flow path, equation for the dynamics of the fluid at the upstream pipe, equation for the dynamics of the fluid at the downstream pipe, equation for the flow rate dynamics due to cavity volume fluctuation, and so on are derived in detail in the previous paper [5].

In this modeling, the pump dynamic characteristic coefficients [7, 8] (cavitation compliance *Cb* and mass flow gain factor *Mb*) have been used to represent the flow rate fluctuation due to the cavity volume change, and they are considered important quantities for the transition phenomenon from RC to AAC. Cavitation compliance *Cb* represents the rate of change in cavity volume with respect to changes in inlet pressure in the flow path, and mass flow gain factor *Mb* represents the rate of change in cavity volume with respect to changes in inlet flow rate in the flow path. In this paper, these values are analyzed by introducing and analyzing 2D cascade model [6].

2.3 Rigid-Rotor Model

The equation of motion of the rotor system in the dimensional form is as follows [5]:

$$\begin{aligned} m_1\ddot{x} &= -c\dot{x} - kx + \left\{F_{ind_x}(x,y) + F_{cnt_x}\right\} + U\omega^2\cos\omega t \\ m_1\ddot{y} &= -c\dot{y} - ky + \left\{F_{ind_y}(x,y) + F_{cnt_y}\right\} + U\omega^2\sin\omega t \end{aligned} \tag{2}$$

Here, U is unbalance, F_{ind_x} and F_{ind_y} are the radial fluid forces acting on the inducer calculated by integrating the pressure $P_{1i}(i = a, b, c)$, $P_{mi}(i = a, b, c)$ and $P_{2i}(i = a, b, c)$ obtained in 1D flow path model for all flow paths at each time step. Force F_{cnt_x} and F_{cnt_y} are the contact forces when the rotor contacts the casing, which generally does not occur. The used parameter values of this model are shown in Table 1.

Table 1. Parameter values of rotor system

Variables	Value	Unit
m_1	6.48	kg
k	8.023×10^7	N/m
c	456.02	N · s/m
Rotational speed ω	42500	rpm
Unbalance U	0.5×10^{-4}	kg · m

2.4 2D Cascade Model

In the conventional studies, the pump dynamic characteristics were are given constant values which varied only for inlet pressure, and the analysis of rotating cavitation have been demonstrated. The values used were same for all flow rate and essentially determined based on the experimental result. However, authors clarified in the previous paper [5] that the variation of the pump dynamic characteristics for each flow path is important.

In this paper, the pump dynamic characteristics of each flow path are investigated theoretically to explain the validity of the assumptions made for the pump dynamic characteristics for each flow path in the previous paper [5]. For this purpose, the two-dimensional flow model in a cascade constructed by Horiguchi et al [6] is introduced, and an extended coupled system is constructed as shown in Fig. 1. This extended coupled system consists of the one-dimensional flow path model of the inducer, a rotor model, and a two-dimensional flow model in a cascade. When the inducer whirls synchronously, there are variations in the flow rate of flow paths obtained from the one-dimensional flow path model. These variations in flow rate were considered as inlet flow distortion of the two-dimensional flow model in a cascade as shown in Fig. 1, and the two-dimensional cavitating flow in a cascade was analysed.

Horiguchi et al. have analyzed cavitation occurring on the blade surface in a two-dimensional flow model in a cascade as shown in Fig. 3 [6]. As shown in Fig. 3, the case where the flow with mean flow velocity U_{2D} and attack angle α_{2D} enter the flat plate cascade with chord length C, pitch h_{2D}, and stagger angle β_{2D}, resulting in a cavity of length l and thickness η on the blade surface is considered. In this case, the flow field is represented by a superposition of (1) a uniform flow, (2) a source q_n representing the cavity, and (3) vortices γ_{1n} and γ_{2n} representing the blade.

In this paper, as in the previous paper [6], the source q_n and vortices γ_{1n} and γ_{2n} on the blade surface are obtained by assuming the cavity length l and thickness η and by giving the cavitation number σ at the inlet, the inlet flow velocity (u, v) at the attack angle α_{2D}. In this study, the inlet flow velocity is not a uniform flow as in the previous paper [6], but a velocity that simulates inlet distortion, in which the velocity distribution varies sinusoidally in the direction of blade row movement. The calculations are then repeated until the obtained source q_n and vortices γ_{1n} and γ_{2n} satisfy the boundary conditions (flow velocity condition on the cavity surface, flow velocity condition on the blade surface, Bernoulli's condition between the infinite upstream and the cavity surface, cavity closure condition, and Kutta's condition) to obtain the cavity length l and thickness η.

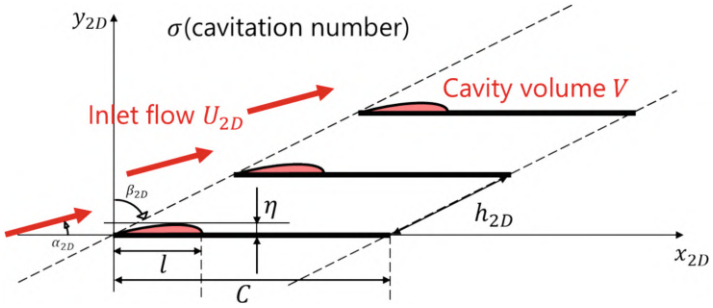


Fig. 3. Theoretical model of 2D cascade model

2.5 Influence of Synchronous Motion on the Inlet Flow Rate

The transition phenomenon to attached asymmetric cavitation occurs when the inducer with some amount of unbalance is in a synchronous whirling motion [4]. In this case, the gap between the inducer and casing varies with respect to θ_1, as shown in Fig. 4(a). Here, it should be noted that this gap distribution does not change with time when the shaft whirls in synchronous whirling motion.

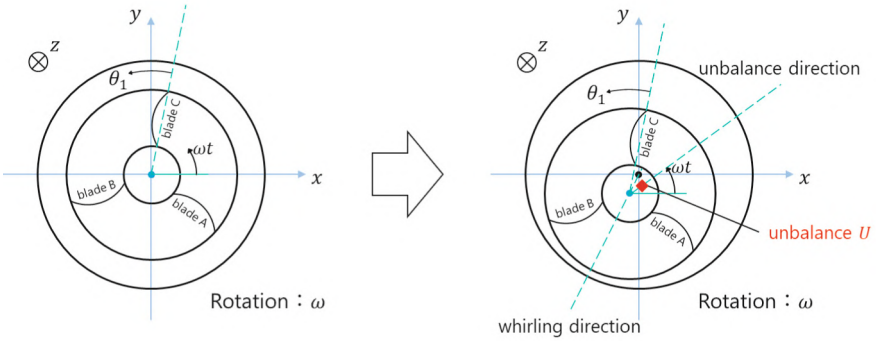
When the gap varies with respect to θ_1, the flow rate leaking through the gap between paths also varies with respect to θ_1. As a result, the inlet flow rates Q_{1a}, Q_{1b}, and Q_{1c} of each path take different values as shown in Fig. 4(b), influenced by the non-uniformity of the inducer leakage flow rate. Then, this inlet flow distribution affected by the shaft's synchronous whirling motion is given to the two-dimensional flow model in a cascade as shown in Fig. 4(c). This inlet flow distribution is represented as follows:

$$u_a = U_{2D}\cos(\alpha_{2D} + \beta_{2D}) + N_d \sin\left(2\pi \frac{r_2}{3h_{2D}} - \theta_0\right) \tag{3}$$

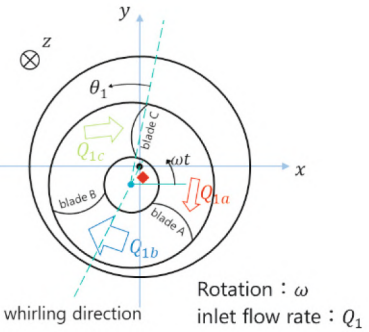
$$u_\theta = U_T$$

The circumferential flow velocity is constant at u_θ, indicating that only the axial flow velocity needs to be considered when considering inlet flow rate distribution. Considering these points, these values of N_d and θ_0 are determined by the inlet flow rates Q_{1a}, Q_{1b}, and Q_{1c}, and used as input for the analysis of two-dimensional flow model in a cascade explained in Sect. 2.4 to obtain cavity length l and thickness η. The pump dynamic characteristic coefficients (cavitation compliance Cb and mass flow gain factor Mb) for all flow paths are also obtained, and then they are used as input for 1D flow analysis explained in Sect. 2.2. Solving the two-dimensional flow model in a cascade again yields the inlet deflected flow.

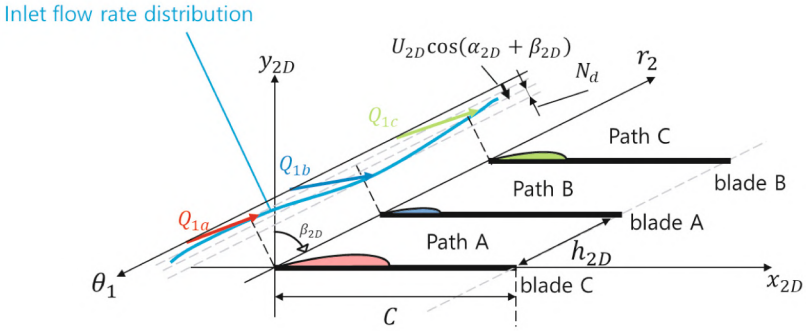
Finally, summary of this section is shown again. Figure 1 shows a schematic diagram of the coupled rigid-rotor system, one-dimensional flow path system [5], and two-dimensional flow model in a cascade [6]. The coupled system of rigid-rotor model and one-dimensional flow path model uses pump dynamic characteristic coefficients (cavitation compliance and mass flow gain factor) as inputs, and outputs the volume flow rate at inlet of the three flow paths. When there is synchronous whirling motion, the inlet flow rate distribution is generated. The two-dimensional flow model in a cascade takes the inlet flow rate with inlet flow rate distribution as input and outputs the pump dynamic characteristic of each flow path by analyzing the cavity length and thickness of the three paths individually. The pump dynamic characteristic coefficients of the three paths are again used as inputs to the coupled system of rigid-rotor model and the one-dimensional flow path model. The coupled analysis of these three models attempts to verify the transition from rotating cavitation to attached asymmetric cavitation without assuming the characteristics in cavity geometry variation [5].



(a)



(b)



(c)

(a) Synchronous whirling motion caused by shaft unbalance
(b) Different inlet flow rates caused by synchronous whirling motion
(c) Inlet flow rate distribution caused by synchronous whirling motion

Fig. 4. Theoretical model of inlet flow rate distribution for 2D cascade model (a) Synchronous whirling motion caused by shaft unbalance (b) Different inlet flow rates caused by synchronous whirling motion (c) Inlet flow rate distribution caused by synchronous whirling motion

3 Results and discussion

3.1 2D flow analysis of blade cascade model with inlet flow rate distribution

First, 2D flow analysis of blade cascade model with inlet flow rate distribution is demonstrated to confirm the validity of assumption made in the previous paper [5].

Figure 5 shows the variation of the cavity length l with respect to the cavitation number (representing the magnitude of inducer inlet pressure) when the inlet flow rate distribution is given for the case when the amplitude of synchronous whirling motion of the inducer is about 50% of the tip clearance. It indicates that the cavity length in one flow path (path C) changes negatively (shown with symbol ●) as the cavitation number decreases. In the previous paper [5], authors assumed this change of the cavity length l in negative direction in one flow path, and explained the transition from RC to AAC. So, Fig. 5 indicates the validity of the assumption made in the previous paper [5].

Figure 6 shows the variation of the pump dynamic coefficients (cavitation compliance *Cb* and mass flow gain factor *Mb*) with respect to the cavitation number when the inlet flow rate distribution is given for the case when the amplitude of synchronous whirling motion of the inducer is about 50% of the tip clearance. The cavitation compliance and mass flow gain factor become negative in one flow path when the cavity length in one flow path (path C) changes negatively in Fig. 5, which corresponds the summary given in the previous paper [5] that the transition from RC to AAC occurs only when both cavitation compliance *Cb* and mass flow gain factor *Mb* becomes negative in one flow path.

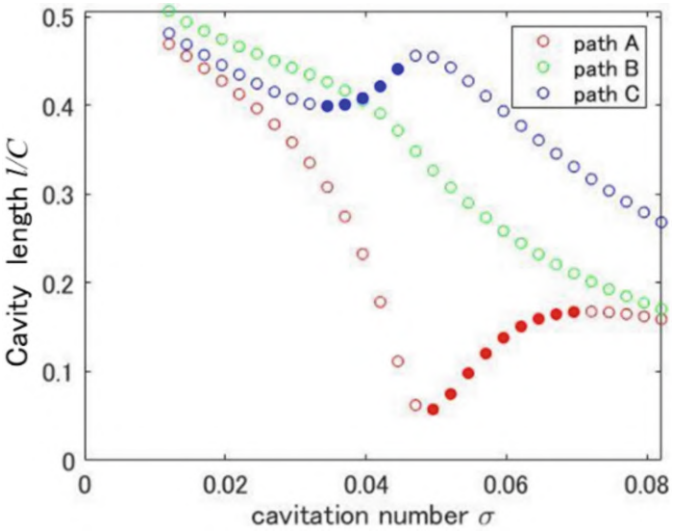


Fig. 5. Theoretical results of the variation of cavity length caused by the whirling motion of the rotor synchronized with the rotor rotation:

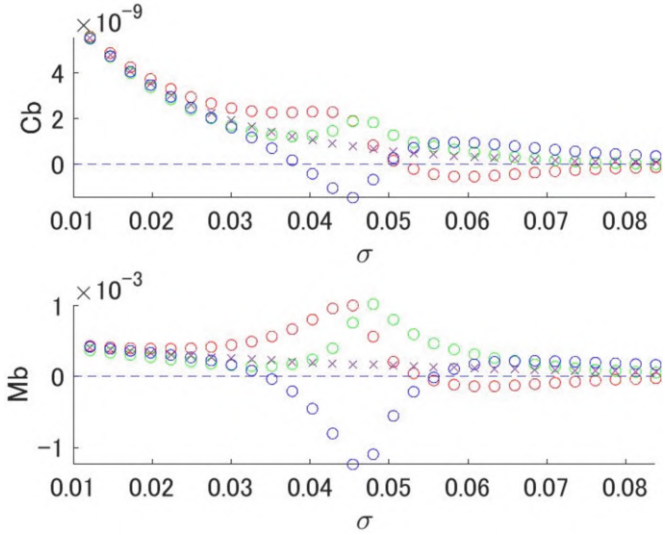


Fig. 6. Theoretical results of the variation of pump dynamic coefficients (cavitation compliance and mass flow gain factor) caused by the whirling motion of the rotor synchronized with the rotor rotation:

3.2 Coupled analysis of One-Dimensional Flow Path Model of the Inducer, a Rigid-Rotor Model, and a Two-Dimensional Flow Model in a Cascade

Figure 7 shows the experimental results [5] of the transition from RC to AAC when the cavitation number decreases. Figure 7(a) shows the waterfall diagram. Rotational speed component 100 Hz (1X component), and -1X, +2X and −2X components were observed. When the cavitation number σ decreased to 0.08, RC component with a frequency of about 120 Hz (about 1.2X component) were generated. Then, the transition from RC to AAC were observed when cavitation number σ decreased at around 0.045. At this transition, RC vibration component was converged to the synchronous (1X) vibration component, and the amplitude of synchronous (1X) vibration component increased at the transition to AAC. In the AAC condition, the high speed camera movie showed that the cavity lengths of 3 flow paths were long-long-short pattern as shown in Fig. 7(b), which is typical characteristics of AAC [4].

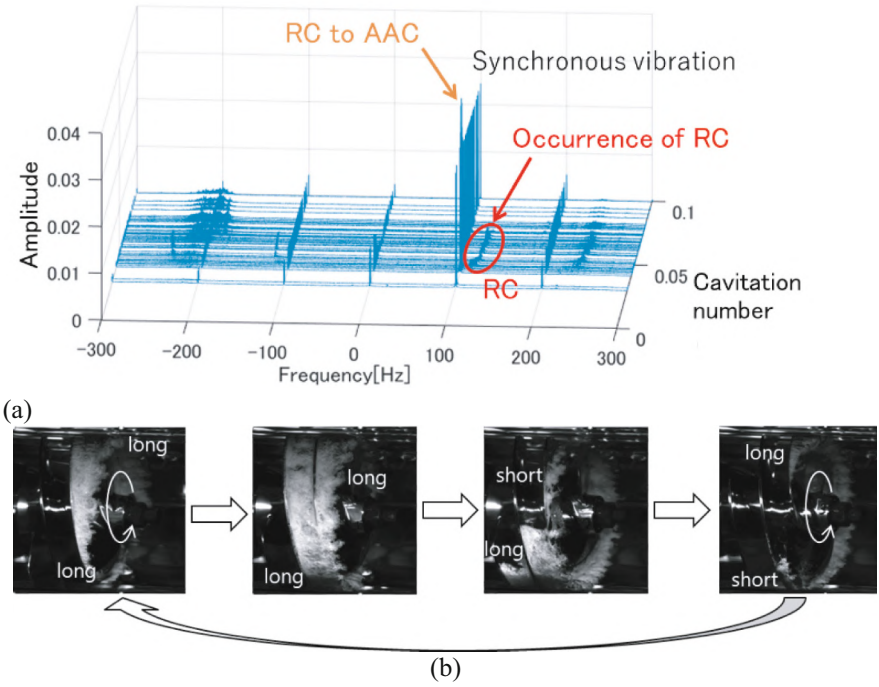
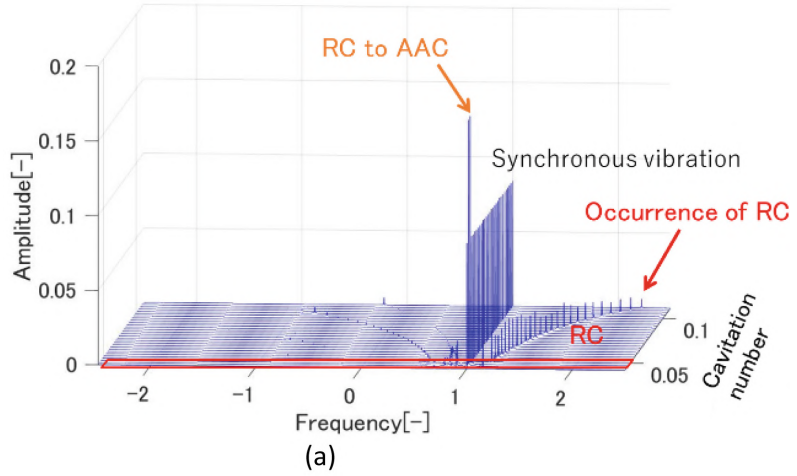


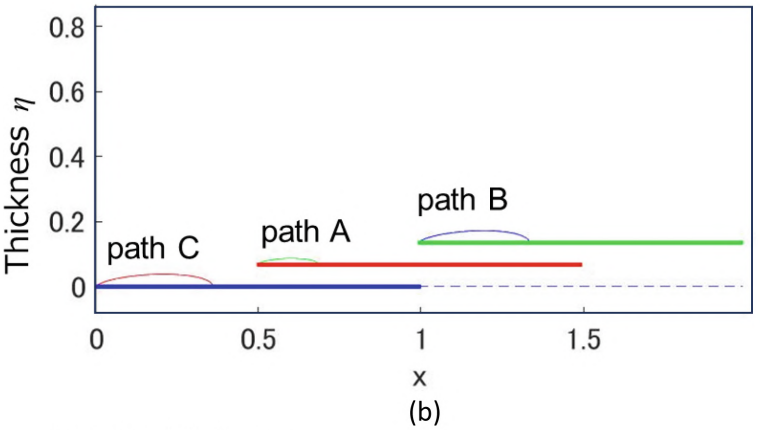
Fig. 7. Experimental and theoretical results of occurrence of RC and transition from RC to AAC [5]

Coupled analysis of one-dimensional flow path model of the inducer, a rigid-rotor model, and a two-dimensional flow model in a cascade are demonstrated to simulate this transition phenomena from RC to AAC. The coupled 1D flow path model and rigid-rotor model are analysed taking into account the variation of the pump dynamic characteristics, and they are calculated two-dimensional flow model in a cascade taking into account the inlet flow rate distribution. These coupled calculation of these two analyses are demonstrated with decreasing the cavitation number.

Figure 8 shows the theoretical result of the coupled analysis. Figure 8(a) shows the waterfall diagram corresponding to Fig. 7. It reproduces the occurrence of RC, and then the transition from RC to AAC almost exactly as observed in the experimental results shown in Fig. 7. Fig. 8(b) shows the cavity shape including the information of cavity length l and thickness η. Cavities on three blades shows the cavity lengths of long-long-short pattern which is corresponding to the experimental observation of Fig. 7 and reference [4].

Figure 9 shows the theoretical results of transition of flow rates in all flow paths at the transition from RC to AAC. It shows the periodic variation of flow rates in three flow paths before the transision, which corresponding to RC. At t= 26sec, cavitation number is decreased and the transition from RC to AAC occurs. At that time, the flow rates in three flow paths tend to be constant and the value of $Q_{1i}(i = a, b, c)$ converses to the





(a) waterfall diagram
(b) Cavity shapes of three blades at the transition to AAC

Fig. 8. Theoretical results of occurrence of RC and transition from RC to AAC: (a) waterfall diagram (b) Cavity shapes of three blades at the transition to AAC

pattern of small - large – small, which corresponds to the prediction from unbalance distribution given to the rigid-rotor.

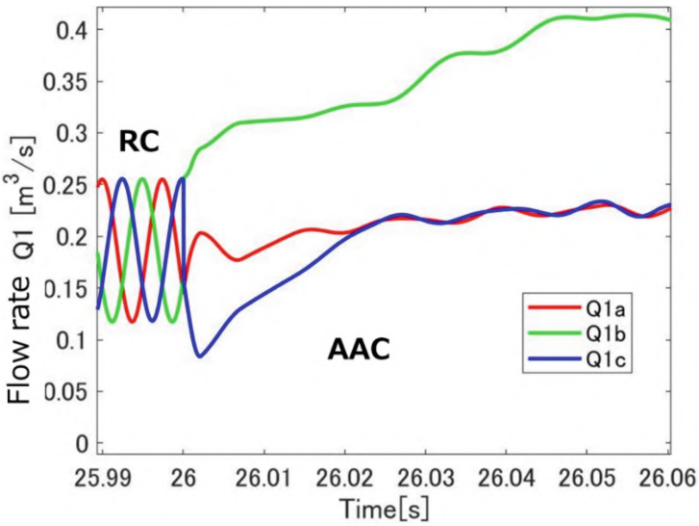


Fig. 9. Theoretical results of transition of flow rates in all flow paths at the transition from RC to AAC

4 Conclusions

In this study, the transition phenomena from rotating cavitation (RC) to attached asymmetric cavitation (AAC) were investigated using a coupled model of a one-dimensional flow paths system, a rigid-rotor system, and a two-dimensional flow model in a cascade, and the following conclusions were obtained.

The effect of synchronous whirling motion due to the shaft unbalance was focused. The analytical result of two-dimensional flow model in a cascade showed that the pump dynamic characteristic coefficients (cavitation compliance *Cb* and mass flow gain factor *Mb*) take different values for each flow path when the inlet flow rate distortion due to the synchronous whirling motion is taken into account. Then, it was found that cavitation compliance *Cb* and mass flow gain factor *Mb* of one of the three flow paths turned negative when the influence of synchronous whirling motion increased due to an increase in the shaft unbalance.

A coupled model of a one-dimensional channel system, a rigid rotor system, and a two-dimensional flow model in a cascade is developed, and a cavitation analysis method that includes the analysis of the pump dynamic characteristic coefficients (cavitation compliance *Cb* and mass flow gain factor *Mb*) for each flow path is proposed. By using this coupled analytical model, the transition from rotating cavitation to attached asymmetric cavitation is numerically demonstrated. The validity of the model and analytical result is explained through comparison with experimental results.

References

1. Tsujimoto, Y., Kamijo, K., Brennen, C.: Unified treatment of flow instabilities of turbomachines. J. Propulsion Power **17**(3), 636–643 (2001)

2. Yoshida, Y.: Rotordynamic force acting on three-bladed inducer under supersynchronous/synchronous rotating cavitation. Trans. ASME, J. Fluids Eng. **132**(6), 061105-1–061105-9 (2010)
3. Yoshida, Y., Nagaura, K., Kazami, Y.: Interaction between uneven cavity length and shaft vibration at the inception of synchronous rotating cavitation (in Japanese). Turbomachinery **36**(8), 465–472 (2008)
4. Kobayashi, S.: Effects of shaft vibration on occurrence of asymmetric cavitation in inducer. JSME Int. J. Ser. B Fluids Thermal Eng. **49**(4), 1220–1225 (2006)
5. Inoue, T., et al.: Investigation of transition phenomenon from rotating cavitation to attached asymmetric cavitation by coupled analysis of one-dimensional flow path model of 3-blade inducer and shaft vibration system. J. Vibr. Eng. Technol. Accept. (2023)
6. Horiguchi, H., Watanabe, S., Tsujimoto, Y., Aoki, M.: A theoretical analysis of alternate blade cavitation in inducer. Trans. ASME, J. Fluids Eng. **122**(1), 156–163 (2000)
7. Shimura, T.: Geometry effects in the dynamic response of cavitating LE-7 liquid Oxygen pump. J. Propulsion Power **11**, 330–336 (1995)
8. Tsujimoto, Y., Kamijo, K., Yoshida, Y.: A theoretical analysis of RC in inducers. Trans. ASME, J. Fluids Eng. **115**, 135–141 (1993).https://doi.org/10.1115/1.2910095

Experimental Study on Non-planar Motions of a Standing Pipe Conveying Fluid

Buana Mahkota Raja(✉) and Kiyotaka Yamashita

Fukui University of Technology, Fukuishi, Fukui 915-8505, Japan
mm23005rm@edu.fukui-ut.ac.jp

Abstract. Lateral vibration of pipes conveying fluid reveals various dynamical features, and has been the subject of research as an essential model of unstable phenomena in a continuous system for many years. Many studies have focused on the dynamics of pipes with intermediate spring supports. One of the reasons why this system fascinates many researchers is that the presence of double degeneracy (the pitchfork and Hopf bifurcation points) can produce a sequence of bifurcating motions. Standing pipe conveying fluid is also a system that can give rise to such double degeneracy. In this study, we consider a standing pipe conveying fluid. An additional mass is attached at the upper free end to easily produce the double degeneracy. We experimentally investigated the effects of the additional mass on the non-planar pipe motions and observed some non-planar motions.

Keywords: non-planar motion · standing pipe · experiment

1 Introduction

Long term efforts have been poured to the study on dynamics of pipes conveying fluid by many researchers [1]. Intense interest has been shown towards non-planar motions recently [2–5]. The effects of spring supports, damping and additional masses on the stability of elastic pipes conveying fluid have been investigated as they are representative of elastic stability problems in non-conservative continuous systems [6,7]. It has been pointed out that these conditions have a great influence not only on the linear stabilities of a pipe, but also on the post-critical nonlinear motions.

Many studies have focused on the dynamics of pipes with additional spring supports [8]. In this system, the presence of double degeneracy (the pitchfork and Hopf bifurcation points) can produce a sequence of bifurcating motions. Païdoussis and Semler [9] theoretically and numerically investigated the nonlinear planar dynamics of a cantilevered pipe with an intermediate spring support. They pointed out that the system reveals a very rich bifurcation behavior in the vicinity of double degeneracy points. Païdoussis and his research group [10–12]

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 383–393, 2025.
https://doi.org/10.1007/978-3-031-72794-8_30

extensively conducted a series of studies considering non-planar pipe motions. They numerically and experimentally revealed bifurcating complex non-planar motions.

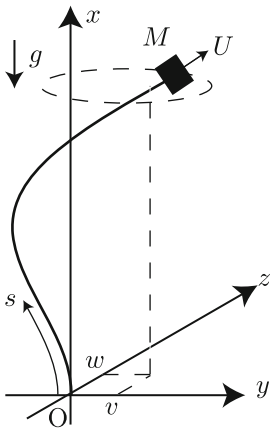


Fig. 1. Non-planar motion of a standing pipe.

Standing pipe conveying fluid is also a system that can give rise to such double degeneracy. Li and Païdoussis [13] theoretically investigated nonlinear planar dynamics of standing pipe conveying fluid and obtained reduced equations governing essential dynamical behavior of the system. There are a few studies that have quantitatively measured the non-planar motions of pipe conveying fluid in experiments. As shown in Fig. 1, we consider non-planar motions of a standing pipe conveying fluid. We focus on the nonlinear interactions between buckling and flutter instabilities. In particular, as a first step to clarify such nonlinear interactions, we conduct experiment involving a silicone rubber pipe conveying water to qualitatively confirm the non-planar motions caused by the interactions between buckling and flutter. An additional mass M is attached at the upper free end to easily produce the double degeneracy. For suspended pipes under gravity, the additional mass M is known to have a significant influence on the generation of complex non-planar motions [14].

2 Basic Equation

As shown in Fig. 1, we consider non-planar motions of a standing cantilevered pipe conveying fluid. When the pipe is short, the stable static equilibrium state of the pipe without flow coincides with the x-axis whose direction is opposite to gravity. The y- and z-axes lie on the horizontal plane. s is the curvilinear coordinate along the centerline of the pipe. Pipe is considered as a beam of overall length ℓ, mass per unit length m and bending rigidity EI. Let S, ρ and g

be the bore area of the pipe, fluid density and acceleration of gravity. The fluid is discharged into the atmosphere at the upper free end of the pipe. The flow velocity relative to the pipe motion is U. Let $v(s,t)$ and $w(s,t)$ be the lateral pipe deflections in the y- and z-directions, respectively, where t is time.

Usual assumptions are used to derive the governing equations of non-planar pipe motions. Non-dimensional variables $t^* = \sqrt{EI/(m+\rho S)\ell^4}t$, $s^* = s/\ell$, $v^* = v/\ell$, and $w^* = w/\ell$ are introduced. Governing equation and its boundary condition in $x-y$ plane are expressed in the following dimensionless form:

$$
\begin{aligned}
&\ddot{v} + 2\sqrt{\beta}V\dot{v}' + \gamma\left[(\alpha+1-s)\,v'\right]' + V^2 v'' + v'''' \\
&= -v''\int_s^1 \left[-\int_0^s (\ddot{v}'v' + \dot{v}'^2 + \ddot{w}'w' + \dot{w}'^2)ds + v'\ddot{v} + w'\ddot{w}\right]ds + \frac{1}{2}v'^2\ddot{v} \\
&-\frac{1}{2}\left[2v'\int_0^s (\ddot{v}'v' + \dot{v}'^2 + \ddot{w}'w' + \dot{w}'^2)ds - \ddot{w}\left\{v'w' - \int_0^s (v'w'' - v''w')ds\right\}\right] \\
&-\frac{1}{2}\left[v'\left(v'v'' + w'w''\right) + w''\int_0^s \left(v'w'' - v''w'\right)ds\right]'' + v''\int_s^1 (v''v''' \\
&+w''w''')ds - \alpha v''\left[v_1'\ddot{v}_1 + w_1'\ddot{w}_1 - \int_0^1 (\ddot{v}'v' + \dot{v}'^2 + \ddot{w}'w' + \dot{w}'^2)ds\right] \\
&-\frac{1}{2}V^2\left[v'\left(v'v'' + w'w''\right) + w''\int_0^s \left(v'w'' - v''w'\right)ds\right] \\
&-\sqrt{\beta}V\left[v'\left(v'\dot{v}' + w'\dot{w}'\right) + \dot{w}'\int_0^s \left(v'w'' - v''w'\right)ds\right] \\
&-\frac{1}{2}\alpha\gamma\left[v'\left(v'v'' + w'w''\right) + w''\int_0^s \left(v'w'' - v''w'\right)ds - v''\left(v_1'^2 + w_1'^2\right)\right] \\
&-\frac{1}{2}(1-s)\gamma\left[v'\left(v'v'' + w'w''\right) + w''\int_0^s \left(v'w'' - v''w'\right)ds\right] \\
&+\frac{1}{2}\gamma v''\int_s^1 \left(v'^2 + w'^2\right)ds + \frac{1}{2}\gamma w'\int_0^s (v'w'' - v''w')ds,
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
&s=0: \quad v=0, \quad v'=0, \\
&s=1: \quad v''=0, \\
&v''' - \alpha\ddot{v} + \alpha\gamma v' = \frac{1}{2}\alpha\left[2v'\int_0^s \left(\ddot{v}'v' + \dot{v}'^2 + \ddot{w}'w' + \dot{w}'^2\right)ds\right. \\
&\left.-\ddot{v}v'^2 - \ddot{w}\left\{v'w' - \int_0^1 \left(v'w'' - v''w'\right)ds\right\}\right] - \frac{1}{2}\left[v'\left(v'v'' + w'w''\right)\right. \\
&\left.+w''\int_0^s \left(v'w'' - v''w'\right)ds\right]' - \frac{1}{2}\alpha\gamma w'\int_0^s \left(v'w'' - v''w'\right)ds,
\end{aligned}
\tag{2}
$$

where $(\,\cdot\,)$ and $(\,')$ denote the derivatives with respect to t^* and s^*, respectively. The asterisks $*$ indicating the dimensionless variables are omitted in Eq. (1), Eq.

(2) and hereinafter. The third term on the left side of Eq. (1) represents the effect of gravity. These equations agree with the equations of previous studies [15] when the sign of the gravitational force is reversed. The governing equation and its boundary conditions in the $x-z$ plane can be written in the same manner. Four non-dimensional parameters are involved in Eqs. (1) and (2) as follows:

$$\alpha = \frac{M}{(m+\rho S)\,\ell}, \quad \beta = \frac{\rho S}{m+\rho S}, \quad V = \sqrt{\frac{\rho S \ell^2}{EI}}U, \quad \gamma = \frac{(m+\rho S)\,g\ell^3}{EI}. \tag{3}$$

α, β, V and γ denote the non-dimensional additional mass, the non-dimensional fluid mass, the non-dimensional flow velocity and the ratio of the gravitational force to the elastic force. In experiments, it is possible to change α and V independently by changing additional mass M and mean flow velocity U.

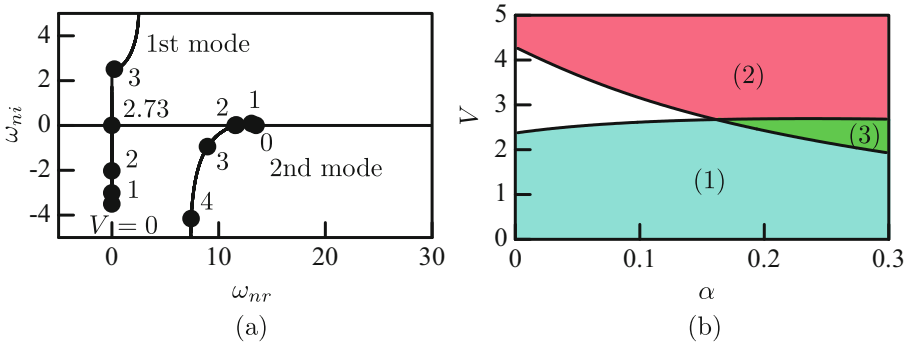


Fig. 2. Linear stability: (a) Eigenvalues $\lambda_j = -\omega_{ji} + i\omega_{jr}$ $(j = 1, 2)$ as a function of V ($\alpha = 0.27$, $\beta = 0.14$, $\gamma = 13$), (b) Stability boundaries of the standing pipe with an additional mass. (1) $\omega_{1i} < 0, \omega_{1r} = 0$, (2) $\omega_{2i} < 0, \omega_{2r} \neq 0$, (3) $\omega_{1i} < 0, \omega_{1r} = 0, \omega_{2i} < 0, \omega_{2r} \neq 0$.

3 Linear Stability

In order to make the experimental plan and determine the parameter regions in which static and dynamic instabilities occur, we first conduct linear stability analyses. By letting $v = \Phi_n(s)e^{\lambda_n t}$ and neglecting the nonlinear terms, we can construct the eigen-value problem. Φ_n is the eigen-function and $\lambda_n = -\omega_{ni} + i\omega_{nr}$ is eigen-value. Figure 2(a) shows the variations $\lambda_j = -\omega_{ji} + i\omega_{jr} (j = 1, 2)$ as a function of V ($\alpha = 0.27$, $\beta = 0.14$, $\gamma = 13$). ω_{ji} and ω_{jr} mean damping ratio and eigen-frequency for jth mode respectively. At $V = 0$, the system is buckled in the first mode under its own weight. The first mode gain the stability at $V = 2.73$. When the flow velocity increases, the system meets dynamical instability in the second mode at V=2.06. Therefore, in case of $2.06 \leqq V \leqq 2.73$, both first and second modes are unstable.

Figure 2(b) shows the effects of α on the linear stability of the standing pipe. The unstable region is divided into three regions: (1) $\omega_{1i} < 0, \omega_{1r} = 0$ (buckling), (2) $\omega_{2i} < 0, \omega_{2r} \neq 0$ (flutter), (3) $\omega_{1i} < 0, \omega_{1r} = 0, \omega_{2i} < 0, \omega_{2r} \neq 0$ (buckling, flutter). In case of large γ, the gravitational force exceeds the restoring force of the pipe. In this situation, even without an additional mass α, the system loses stability by buckling without flow and gains stability when the flow velocity V increases. For large α, at the higher flow velocity, the system also undergoes flutter instability in the second mode. Therefore, in certain parameter region, Fig. 2(b) shows that there is a double degeneracy point.

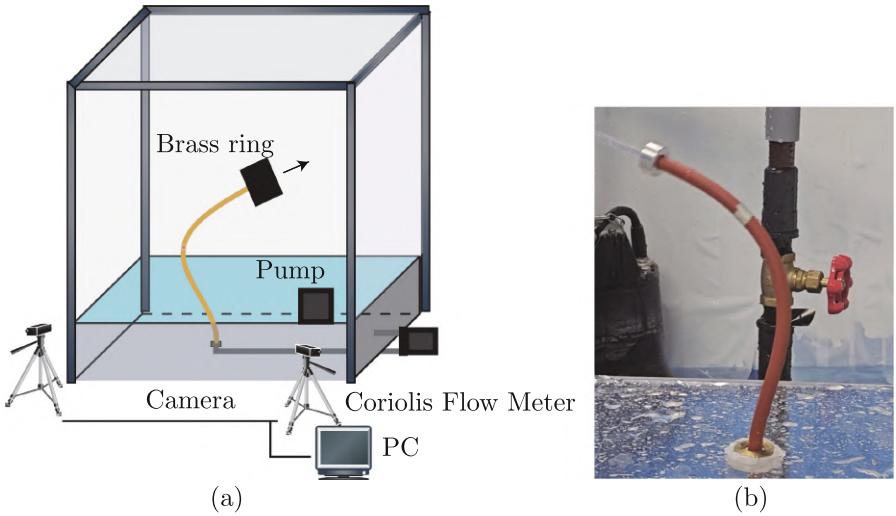


Fig. 3. Experimental apparatus.

4 Experiment

4.1 Experimental Apparatus

We used a silicone rubber pipe with an external diameter of 8.9 mm and an internal diameter of 3.85 mm. The pipe had a circular cross section and had an overall length of 217 mm. Because the initial imperfections of pipes have strong influence on the post-critical nonlinear motions, we have to carefully chose the pipe that is as straight as possible without deformation. The silicone rubber pipe was molded inside two layers of rigid aluminum cylinder. A cross-section of the pipe is almost circular. Flowing fluid was water. Pipe was clamped at the lower end and was fitted with a mass M at the upper free end. The overall length of the pipe and the additional mass are chosen such that the first mode in the system loses the stability by buckling without flow. The parameter values in

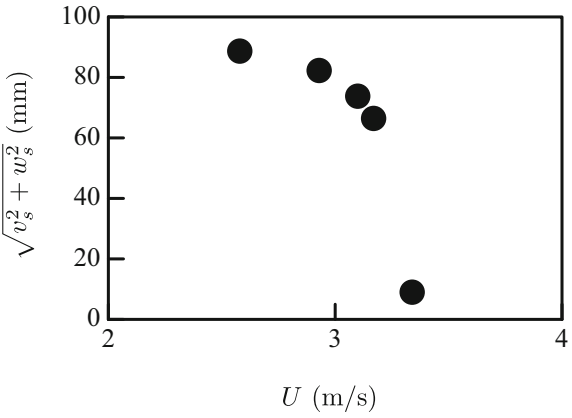


Fig. 4. Static deformation of the pipe as a function of flow velocity U with an additional mass ($M = 2$g).

experiments match the non-dimensional parameter values used in the theoretical analysis shown in Fig. 2.

The schematic experimental apparatus is shown in Fig. 3(a). Figure 3(b) is a photograph of the standing silicone rubber pipe. A cover to prevent water from scattering was installed at a sufficient distance so as not to affect pipe motions. Mean flow velocity in the pipe was measured using a Coriolis flow meter (KEYENCE FD-SS20 A) installed in the flow channel. The image processing system (Dipp-Motion V, Ditect) enabled us to conduct non-contact measurements of pipe motion in 3D space at a rate of 200 times per second. Lateral displacements v and w of the pipe were sensed by the system at $s = 114$ mm. Measurements were performed after a sufficient time had passed after changing the flow velocity. The measurement time is 34 s. Non-planar motions in the neighborhood of double degeneracy points are examined by varying α and V as the control parameters. No qualitatively different phenomena were observed for different initial conditions throughout the experiments.

4.2 Experimental Results

The experiments were conducted by changing the additional mass M from 0g to 6g in 1g increments. The overall length of the pipe was chosen so that it would be buckled under its own weight even if the additional mass was 0g.

First, we describe the experimental results when M is 2g. Without the flow, the pipe was buckled under its own weight. Figure 4 shows the variation of the static deformation of pipe as a function of flow velocity U. The static deformation gradually decreased with the increase of flow velocity. The pipe gains stability when the flow velocity U reaches 3.5 m/s. Even if an initial disturbance was applied to the pipe, the pipe soon returned to its stable straight position. Over $U = 3.8$ m/s, planar oscillation around straight position was observed. This

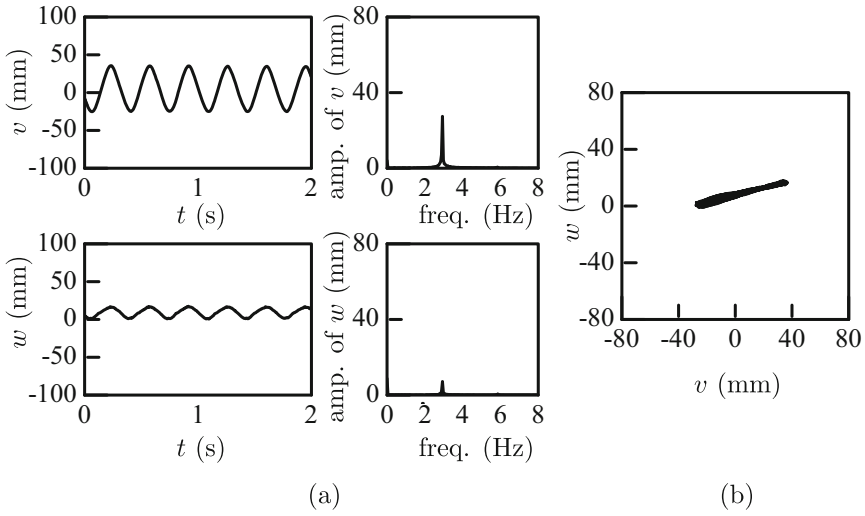


Fig. 5. Planar pipe oscillation around straight position of the pipe ($M = 2$g, $U = 4.3$ m/s): (a) time histories of v and w and their spectrum analyses, (b) pipe motion in $y - z$ plane.

result ($M = 2$g) qualitatively agrees to the results for $M = 0, 1$g. Additional mass $M = 2$g corresponds to non-dimensional parameter value $\alpha = 0.11$.

Figure 5(a) shows the time histories of v and w and their spectrum analyses ($M = 2$g, $U = 4.3$ m/s). The dominant frequency of pipe motion corresponds to the second mode natural frequency which is expected from the linear stability analysis. Figure 5(b) shows the pipe motion in $y - z$ plane. The pipe motion is almost planar oscillation around the straight position.

Next, we describe the experimental results when M is 5g. Figure 6 shows the variation of the pipe motion in $y - z$ plane. The system loses stability under its own weight by buckling below $U = 3.4$ m/s, as shown in Fig. 6(a). The static deformation decreased as the flow velocity increased. However, slightly over $U = 3.4$ m/s, the pipe experienced flutter instability. As shown in Fig. 6(b), the direction of buckling is orthogonal to the plane in which the self-excited oscillation is involved. When the flow velocity is slightly increased, the pipe motion traces a Fig. 8, as shown in Fig. 6(c). The displacement in the direction of buckling contains a frequency component twice the frequency of self-excited oscillation. When U increased further, the pipe motion became complex non-planar motion as shown in Fig. 6(d). For short time intervals, the pipe motion looks like Fig. 8. The center position of this vibration fluctuated with the passage of time, and the motion of the pipe exhibited a complex non-planar motion. This beating oscillation gradually shifted to periodic motion as flow velocity increased. In the plane orthogonal to a direction of the static deformation, the pipe oscillated with a dominant frequency which corresponded to the second mode natural frequency.

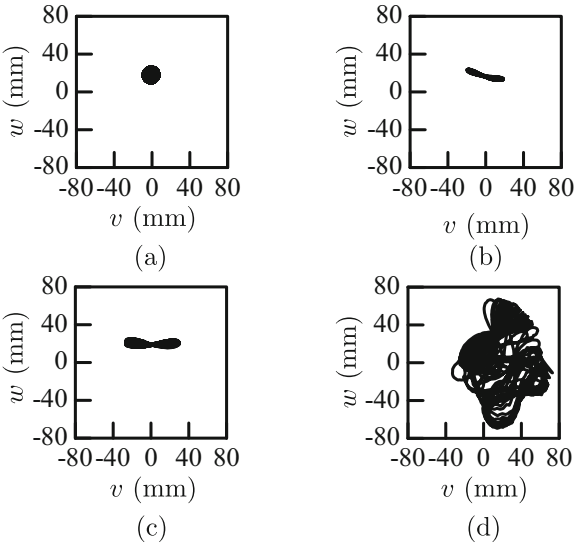


Fig. 6. Pipe motion in $y - z$ plane ($M = 5$g): (a) $U = 3.3$ m/s, (b) $U = 3.4$ m/s, (c) $U = 3.7$ m/s, (d) $U = 4.5$ m/s.

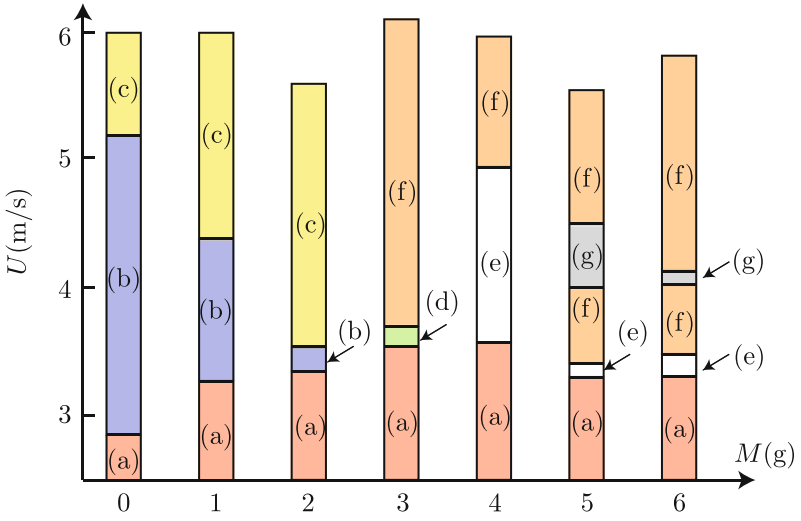


Fig. 7. Effects of U and α on the pipe motion: (a) buckling, (b) straight position, (c) self-excited vibration around straight position of pipe, (d) buckling and self-excited oscillation occur in a same plane, (e) the direction of buckling is orthogonal to the plane in which planar self-excited vibration is involved, (f) pipe motion looks like Fig. 8 as shown in Fig. 6(c) and (g) beating pipe motion as shown in Fig. 6(d).

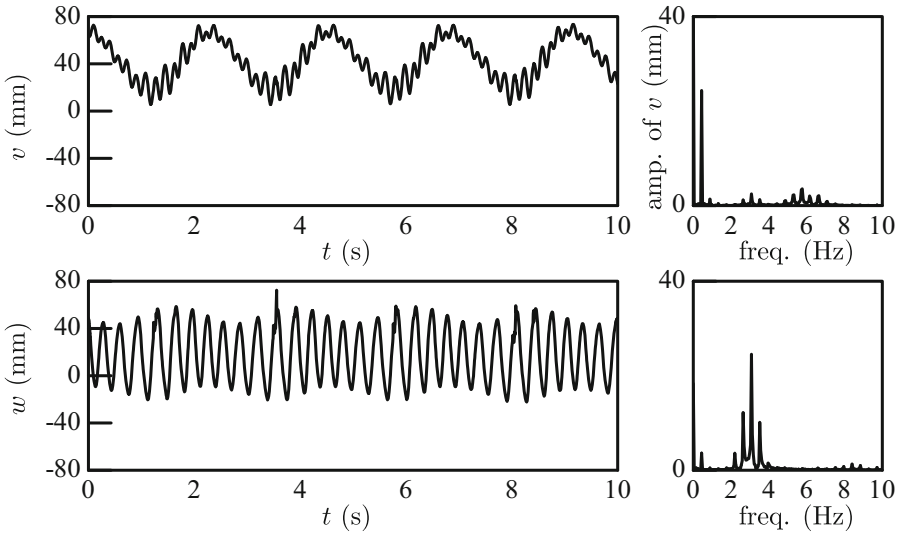


Fig. 8. Time histories of v and w and their spectrum analyses ($M = 6$g, $U = 3.9$ m/s).

In the plane parallel to a direction of the static deformation, the pipe oscillated with a dominant frequency which was twice the second mode natural frequency.

As shown in Fig. 7, we classify the pipe motion into 7 regions: (a) buckling, (b) straight position, (c) self-excited vibration around straight position of pipe, (d) buckling and self-excited oscillation occur in a same plane, (e) the direction of buckling is orthogonal to the plane in which planar self-excited vibration is involved, (f) pipe motion looks like Fig. 8, as shown in Fig. 6(c) and (g) beating pipe motion as shown in Fig. 6(d). Time histories of v and w and their spectrum analyses for beating non-planar motion are shown in Fig. 8. Over $M = 3$g (non-dimensional value $\alpha = 0.17$), we observed some non-planar motions which are caused by static and dynamic instabilities.

5 Conclusion

In this study, we experimentally investigated a standing pipe conveying fluid. We particularly focus on the non-planar pipe motions which can be measured by a non-contact measurement system. When the overall length of the pipe exceeds a certain value, the system experiences the static bifurcation (pitchfork bifurcation). Because this system belongs to non-self-adjoint system, this system can also reveal dynamic instability (Hopf bifurcation). An additional mass on the free end of the pipe is a key parameter which can easily produce double degeneracy (pitchfork and Hopf bifurcation). Flow velocity and additional mass were changed as control parameters.

First, in order to make the experimental plan, we conduct linear stability analyses. As predicted by linear stability analyses, for large additional mass, we

observed post-critical pipe motion which were subjected to buckling and self-excited oscillation. A typical example of such motions is non-planar motion in which the direction of buckling is orthogonal to the plane in which self-excited oscillation is involved. When the flow velocity was slightly increased, this non-planar motion shifted to non-planar motions looks like Fig. 8. In some parameter regions, beating non-planar motions were observed. In this motion, the buckling displacement fluctuated greatly over time. As a spring supported pipe exhibited a variety of non-planar motions near double degeneracy point, a standing pipe also exhibited some interesting non-planar motions in experiments.

When we consider large end mass and high flow velocity, dynamic instability of other modes can also become a problem, and the post-critical pipe motion can become more complex. In order to clarify the modes that contribute to complex non-planar motions and nonlinear modal interactions, we plan to conduct multipoint measurement in the next experiments. A theoretical approach to clarify post-critical non-planar motions is a problem of the future.

References

1. Païdoussis, M.P.: Fluid-structure interactions: slender structures and axial flow. Academic Press (1998)
2. Ibrahim, R.A.: Overview of mechanics of pipes conveying fluids-part I: fundamental studies. ASME J. Press. Vessel Tech. **132**, 1–32 (2010). https://doi.org/10.1115/1.4001271
3. Steindl, A., Troger, H.: Nonlinear three-dimensional oscillations of elastically constrained fluid conveying viscoelastic tubes with perfect and broken O(2)-symmetry. Nonlinear Dyn. **7**, 165–193 (1995). https://doi.org/10.1007/BF00053707
4. Łuczko, J., Czerwiński, A.: Nonlinear three-dimensional dynamics of flexible pipes conveying fluids. J. Fluids Struct. **70**, 235–260 (2017). https://doi.org/10.1016/j.jfluidstructs.2017.02.002
5. Wang, L., Jiang, T.L., Dai, H.L.: Three-dimensional dynamics of supported pipes conveying fluid. Acta. Mech. Sin. **33**(6), 1065–1074 (2017). https://doi.org/10.1007/s10409-017-0718-z
6. Sugiyama, Y., Kawagoe, H., Kishi, T., Nishiyama, S.: Studies on the stability of pipes conveying fluid: (the combined effect of a spring support and a lumped mass). JSME Int. J. Ser. **1**(31), 20–26 (1988). https://doi.org/10.1299/jsmea1988.31.1_20
7. Païdoussis, M.P., Semler, C.: Non-linear dynamics of a fluid-conveying cantilevered pipe with a small mass attached at the free end. Int. J. Non-Linear Mech. **33**, 15–32 (1998). https://doi.org/10.1016/S0020-7462(97)00002-4
8. Jin, J.D.: Stability and chaotic motions of a restrained pipe conveying fluid. J. Sound Vib. **208**, 427–439 (1997). https://doi.org/10.1006/jsvi.1997.1195
9. Païdoussis, M.P., Semler, C.: Nonlinear dynamics of a fluid-conveying cantilevered pipe with an intermediate spring support. J. Fluids Struct. **7**, 269–298 (1993). https://doi.org/10.1006/jfls.1993.1017
10. Ghayesh, M.H., Païdoussis, M.P.: Three-dimensional dynamics of a cantilevered pipe conveying fluid, additionally supported by an intermediate spring array. Int. J. Non-Linear Mech. **45**, 507–524 (2010). https://doi.org/10.1016/j.ijnonlinmec.2010.02.001

11. Ghayesh, M.H., Païdoussis, M.P., Modarres-Sadeghi, Y.: Three-dimensional dynamics of a fluid-conveying cantilevered pipe fitted with an additional spring-support and an end-mass. J. Sound Vib. **330**, 2869–2899 (2011). https://doi.org/10.1016/j.jsv.2010.12.023
12. Modarres-Sadeghi, Y., Païdoussis, M.P.: Chaotic oscillations of long pipes conveying fluid in the presence of a large end-mass. Comput. Struct. **122**, 192–201 (2013). https://doi.org/10.1016/j.compstruc.2013.02.005
13. Li, G.X., Païdoussis, M.P.: Stability, double degeneracy and chaos in cantilevered pipes conveying fluid. Int. J. Non-Linear Mech. **29**, 83–107 (1994). https://doi.org/10.1016/0020-7462(94)90054-X
14. Copeland, G.S., Moon, F.C.: Chaotic flow-induced vibration of a flexible tube with end mass. J. Fluids Struct. **6**, 705–718 (1992). https://doi.org/10.1016/0889-9746(92)90004-M
15. Yamashita, K., Kitaura, K., Nishiyama, N., Yabuno, H.: Non-planar motions due to nonlinear interactions between unstable oscillatory modes in a cantilevered pipe conveying fluid. Mech. Syst. Signal Process. **178**, 109183 (2022). https://doi.org/10.1016/j.ymssp.2022.109183

Nonlinear Dynamics and Control

Numerical Study of Chatter Control in End Milling Process by Workpiece Excitation

Yutaka Nakano(✉), Satoru Sekikawa, and Hiroki Takahara

Tokyo Institute of Technology, 2-12-1, O-okayama, Meguro-ku 152-8552, Tokyo, Japan
nakano.y.aj@m.titech.ac.jp

Abstract. Regenerative chatter causes deterioration of machining accuracy and tool wear. Various studies have been conducted on the suppression of regenerative chatter. One of the active chatter control techniques is the application of spindle speed variation. The variation of the spindle speed during the cutting process can control regenerative chatter, because the variation of the time delay can disturb the regenerative effect. However, the frequency and amplitude of spindle speed variation cannot be extremely high because the impact on spindle life and reliability must be minimized. This work investigates the effect of the workpiece excitation to fluctuate the relative speed between a cutting edge and workpiece at the cutting point on regenerative chatter as an alternative method to the spindle speed variation technique. The time history analysis was conducted to investigate the excitation amplitude and frequency to reduce the regenerative chatter vibration amplitude. It was confirmed increasing the excitation amplitude with increasing spindle speed so that the time delay variation amplitude is equal has the chatter suppression effect even in the high spindle speed range.

Keywords: Self-Excited Vibration · Regenerative Chatter · Variable Time Delay · Vibration Suppression · End Milling Process

1 Introduction

Regenerative chatter caused by the regenerative effect is one of the self-excited vibrations that occur during machining process. Regenerative chatter (hereinafter, referred to as "chatter") in machining process causes a poor surface finish of a workpiece, a tool wear and a machine tool damage. To control chatter, a wide variety of approaches have been proposed [1]. For example, one passive chatter suppression measure is the installation of a tuned mass damper on the tool holder [2, 3]. Through the use of tuned mass dampers, it is possible to increase the lower limit of the critical axial depth of cut at onset of chatter. However, the tuned mass damper must be installed to avoid interfering with the tool path, and the damper tuning parameters must be determined according to the vibration modes associated with chatter. In contrast, the spindle speed variation technique is one of active chatter control techniques [4–6]. Varying the spindle speed can control chatter by variation of the time delay and interference with the regenerating effect. However, the true optimum time delay variation conditions for chatter suppression may not have

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 397–404, 2025.
https://doi.org/10.1007/978-3-031-72794-8_31

been selected because the amplitude and frequency of spindle speed variation cannot be extremely large to minimize their impact on spindle life and reliability. The present study investigates chatter suppression by exciting a workpiece in the tangential direction of a cutting tool to fluctuate the relative speed between a cutting edge and workpiece at the cutting point as an alternative method to the spindle speed variation technique. The time delay variation to achieve the relative velocity variation is derived and investigated the excitation amplitude and frequency to reduce chatter vibration amplitude by conducting time history analysis.

2 Analytical Model and Excitation Parameters

Figure 1 shows a workpiece model fixed to a vibration table which can be vibrated in x and y directions independently. Synchronized control of the tool rotation direction and each excitation direction of the vibration table enables workpiece excitation in the tangential direction of the cutting edge in case of the use of a straight flute end milling tool. Therefore, the time delay can be varied by fluctuating the tangential relative velocity between the cutting edge and the workpiece at the cutting point. To accomplish this, the workpiece must be subjected to excitation displacement in the x and y directions during the cutting of the j-th cutting edge as expressed in the following equation:

$$\begin{aligned} x_{ej}(t) &= A \sin \omega_e t \cos \phi_j(t) \\ y_{ej}(t) &= -A \sin \omega_e t \sin \phi_j(t) \end{aligned} \tag{1}$$

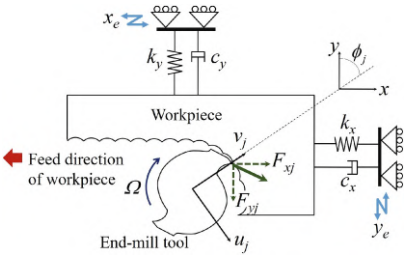


Fig. 1. Workpiece model fixed to active workpiece holder.

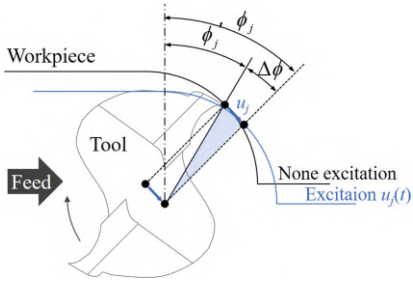


Fig. 2. Relationship between excitation displacement and instantaneous angular immersion.

where A is the excitation amplitude, ω_e is the excitation frequency and $\phi_j(t)$ is the instantaneous angular immersion expressed by the following expression:

$$\phi_j(t) = \Omega t + \frac{2\pi}{N}(j-1), \quad (j = 1, \cdots, N) \tag{2}$$

where, Ω is the angular velocity of the tool and N is the number of the flutes. Note that, the present study assumes that no more than two cutting edges are involved in cutting at the same time during cutting in order to realize the tangential excitation of a cutting edge. When the number of teeth N is two, the excitation displacements of the vibration table in the x and y directions are expressed as follows:

$$\begin{aligned} x_e(t) &= A\gamma(t)\sin\omega_e t\cos\Omega t \\ y_e(t) &= -A\gamma(t)\sin\omega_e t\sin\Omega t \end{aligned} \tag{3}$$

where,

$$\gamma(t) = \begin{cases} 1 & (\mathrm{mod}(t, 2\tau_0) < \tau_0) \\ -1 & (\mathrm{mod}(t, 2\tau_0) \geq \tau_0) \end{cases} \tag{4}$$

where, τ_0 represents the average tooth passing period, which is expressed by the following equation:

$$\tau_0 = \frac{2\pi}{N\Omega} \tag{5}$$

The excitation displacement u_e in the tangential direction of the cutting edge is expressed by the following expression in accordance with Eq. (3).

$$\begin{aligned} u_e(t) &= x_e(t)\cos\Omega t - y_e(t)\sin\Omega t \\ &= A\gamma(t)\sin\omega_e t \end{aligned} \tag{6}$$

The instantaneous angular immersion fluctuation can be calculated by Eq. (7) as shown in Fig. 2 when the excitation displacement is small:

$$\Delta\phi(t) = \frac{2u_e(t)}{d} = \frac{2A\gamma(t)}{d}\sin\omega_e t \tag{7}$$

where d is the tool diameter. From Fig. 2, when the workpiece is positively displacement in the tangential direction of the cutting edge, the instantaneous angular immersion of the j-th cutting edge considering the workpiece excitation decreases as expressed in the following expression:

$$\phi_j'(t) = \phi_j(t) - \Delta\phi(t) \tag{8}$$

The following expression is derived from the relationship that the instantaneous angular immersion during the current cut is equivalent to that during the previous one.

$$\phi_j'(t) = \phi_{j+1}'(t - \tau(t)) \tag{9}$$

where $\tau(t)$ is defined as the time delay variation considering the effect of the workpiece excitation. Substituting Eq. (2), Eq. (5), Eq. (7) and Eq. (8) into Eq. (9) gives:

$$\begin{aligned}\tau(t) &= \tau_0 + \frac{2A\gamma(t)}{d\Omega}\{\sin\omega_e t + \sin\omega_e(t-\tau(t))\} \\ &= \tau_0 + \frac{4A\gamma(t)}{d\Omega}\cos\frac{\omega_e\tau(t)}{2}\sin\omega_e\left(t-\frac{\tau(t)}{2}\right)\end{aligned} \tag{10}$$

Under the assumption that the time delay $\tau(t)$ varies minutely with respect to the average tooth passing period τ_0, $\tau(t)$ is approximately expressed by the following expression:

$$\tau(t) \simeq \tau_0 + \frac{4A\gamma(t)}{\Omega d}\cos\frac{\omega_e\tau_0}{2}\sin\omega_e\left(t-\frac{\tau_0}{2}\right) = \tau_0 + \tau_a\sin\omega_e\left(t-\frac{\tau_0}{2}\right) \tag{11}$$

where,

$$\tau_a = \frac{4A\gamma(t)}{\Omega d}\cos\frac{\omega_e\tau_0}{2} \tag{12}$$

The amplitude of the time delay variation τ_a is zero when the product of the excitation frequency ω_e and the average tooth passing period τ_0 satisfies the following expression, since the argument of the cosine function in Eq. (12) includes the excitation frequency ω_e and the average tooth passing period τ_0.

$$\begin{aligned}&\cos\frac{\omega_e\tau_0}{2} = 0 \;\Rightarrow\; \frac{\omega_e\tau_0}{2} = \frac{\omega_e}{2}\frac{2\pi}{2\Omega} = \frac{2m-1}{2}\pi \quad (m = 1,\cdots,\infty) \\ &\Rightarrow R_f = \frac{\omega_e}{2\Omega} = \frac{2m-1}{2}, \quad (m = 1,\cdots,\infty)\end{aligned} \tag{13}$$

where R_f is the frequency ratio of the excitation frequency ω_e to the tooth passing frequency 2Ω. On the other hand, the condition for maximizing the variation amplitude of the time delay can be obtained from the following expression:

$$\left|\cos\frac{\omega_e\tau_0}{2}\right| = 1 \;\Rightarrow R_f = m, \quad (m = 1,\cdots,\infty) \tag{14}$$

The expression (14) implies that the chatter suppression effect is maximized when the excitation frequency is equal to an integer multiple of the tooth passing frequency.

The equations of motion for the 2 degree of freedom workpiece milling model subject to the tangential workpiece excitation shown in Fig. 1 are expressed as follows:

$$\begin{cases} m_x\ddot{x} + c_x\dot{x} + k_x x = \sum_{j=1}^{N} F_{x,j} \\ m_y\ddot{y} + c_y\dot{y} + k_y y = \sum_{j=1}^{N} F_{y,j} \end{cases} \tag{15}$$

where, $F_{x,j}$ and $F_{y,j}$ are respectively the cutting forces in the x and y directions acting on the workpiece and are expressed as follows:

$$\begin{cases} F_{x,j} = F_{u,j}\cos\phi'_j(t) + F_{v,j}\sin\phi'_j(t) \\ F_{y,j} = -F_{u,j}\sin\phi'_j(t) + F_{v,j}\cos\phi'_j(t) \end{cases} \tag{16}$$

where $F_{u,j}$ and $F_{v,j}$ respectively represent the cutting forces in the tangential and radial directions of the tool acting on the workpiece from the j-th cutting edge. Under the assumption that the cutting force $F_{u,j}$ and $F_{v,j}$ are proportional to the cutting cross sectional area which is the product of the axial depth of cut a and the uncut chip thickness $h\left(\phi_j'(t)\right)$, $F_{u,j}$ and $F_{v,j}$ are expressed by the following expressions:

$$\begin{cases} F_{u,j} = K_{\mathrm{u}} a h(\phi_j'(t))\chi(h(\phi_j'(t))) \\ F_{v,j} = K_{\mathrm{r}} F_{u,j} \end{cases} \tag{17}$$

where, K_{u} and K_{r} are respectively the cutting stiffness in the tangential direction of the cutting edge and the cutting stiffness ratio of the tangential direction to the radial direction. $h(\phi_j'(t))$ is the uncut chip thickness expressed as follows.

$$h\left(\phi_j'(t)\right) = \left[\{f + x(t-\tau(t)) - x(t)\}\sin\phi_j'(t) + \{y(t-\tau(t)) - y(t)\}\cos\phi_j'(t)\right] g\left(\phi_j'(t)\right) \tag{18}$$

where f is the feed rate. $g(\phi_j'(t))$ is the function which evaluates whether the j-th cutting edge is engaged with the workpiece as follows:

$$g(\phi_j'(t)) = \begin{cases} 1, & (\phi_{\mathrm{st}} < \phi_j'(t) < \phi_{\mathrm{ex}}) \\ 0, & (\phi_j'(t) < \phi_{\mathrm{st}}, \phi_j'(t) > \phi_{\mathrm{ex}}) \end{cases} \tag{19}$$

where ϕ_{st} and ϕ_{ex} represent the start and exit cutting angle. The present analysis focusses the up-cut milling process. Therefore, the start and exit cutting angle are defined as $\phi_{\mathrm{st}} = 0$ and $\phi_{\mathrm{ex}} = \cos^{-1}(1 - 2r_d/d)$, where r_{d} is the radial depth of cut. Considering the loss of contact of the tool and the workpiece, $h(\phi_j'(t))$ is zero when the uncut chip thickness $h(\phi_j'(t))$ is negative. $\chi(h(\phi_j'(t)))$ is the function to determine the sign of $h(\phi_j'(t))$ defined as follows:

$$\chi\left(h(\phi_j'(t))\right) = \begin{cases} 1, & (h(\phi_j'(t)) \geq 0) \\ 0, & (h(\phi_j'(t)) < 0) \end{cases} \tag{20}$$

Therefore, the equations of motion Eq. (15) include the nonlinearity.

3 Numerical Simulation

The stability limit of chatter depends on the phase difference of the dynamic displacement between the previous and present tooth periods. The phase difference is expressed as the product of the tooth passing period and the chatter frequency. Chatter control techniques which perform time delay variation, such as the spindle speed variation and the workpiece excitation to vary the time delay, are most effective at the spindle speed located in the local minimum of the stability limit diagram.

Firstly, the chatter stability limit diagram without workpiece excitation was derived by conducting the numerical integration of the equations of motion Eq. (15) using the Runge-Kutta method in order to investigate the spindle speed which is located in the local minimum of the stability limit diagram. The stability limit was determined by identifying

the chatter frequency component of the Fourier spectrum of the workpiece vibration waveform obtained by the numerical integration according to the method proposed in previous work [7]. The accuracy of the time history analysis results was evaluated by comparing the stability limit diagram obtained by stability analysis using the semi-discretization method [8]. Table 1 shows the calculation parameters. Figure 3 shows the chatter stability limit diagrams without the workpiece excitation obtained by the numerical integration (black line) and the semi-discretization method (red circle mark plots). The abscissa represents the dimensionless rotation angular velocity of the tool, and the ordinate represents the dimensionless axial depth of cut. The system becomes unstable when the depth of cut exceeds the stability limit diagram. Therefore, chatter can occur. From Fig. 3, the stability limit diagrams obtained by the numerical integration and the semi-discretization method coincide well.

Table 1. Calculation parameters used in analysis.

Parameters	Values	Parameters	Values
m_x [kg]	0.48	K_u [N/m^2]	8.57×10^8
m_y [kg]	0.23	K_r	0.085
$\omega_x/2\pi$ [Hz]	860	r_d [mm]	1.5
$\omega_y/2\pi$ [Hz]	400	d [mm]	6
ζ_x	0.02	f [mm/rev]	0.05
ζ_y	0.05	N	2

Next, the effect of workpiece excitation on chatter was verified at the rotation angular velocity at the local minimum of the stability limit diagram, where the effect of the time delay is significant. Figure 4 shows the workpiece vibration displacement with and without the workpiece excitation under conditions where chatter occurs at the dimensionless rotation angular velocity of 0.147 and the dimensionless axial depth of cut of

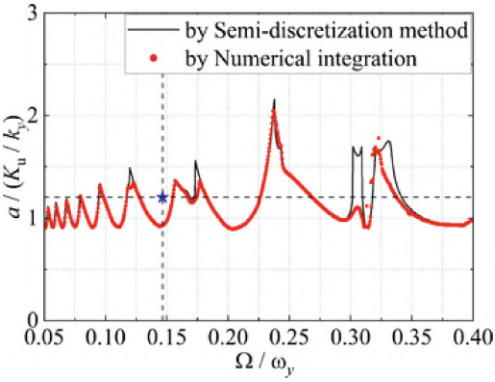
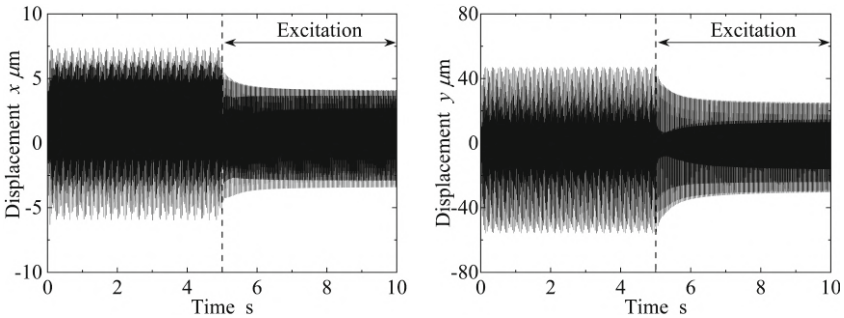


Fig. 3. Chatter stability limit diagrams without workpiece excitation.

1.20 indicated with a blue star plot in Fig. 3. The workpiece excitation was applied after 5 s from the cutting start. The vibration waveforms in Fig. 4 were obtained by numerically integrating Eq. (15). The excitation frequency is 3 times of the tooth passing frequency (excitation frequency ratio R_f of 3.0) and the excitation amplitude A is 300 μm (amplitude of time delay variation τ_a of 545.7 μs). Figure 4 shows that the peak to peak value of the workpiece vibration displacement is reduced by 43.2% when the workpiece excitation is applied.



(a) x direction (b) y direction

Fig. 4. Workpiece vibration displacement with and without excitation.

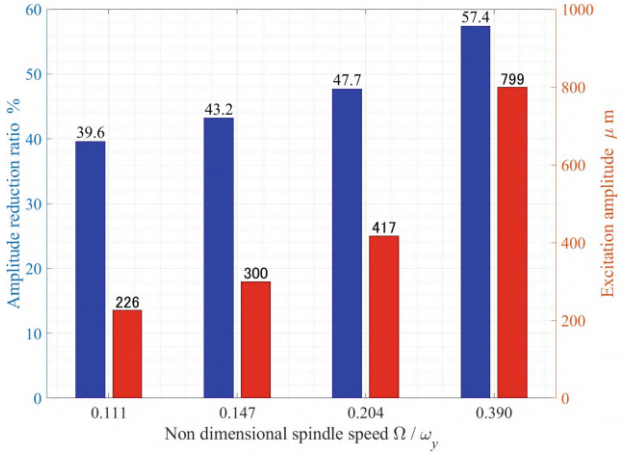


Fig. 5. Relation between excitation amplitude and reduction rate.

Figure 5 shows the relation between the amplitude reduction ratio and the excitation amplitude for different spindle speeds when the excitation frequency ratio is 3.0 and the amplitude of time delay variation τ_a is 545.7 μs. Namely, the excitation amplitude is changed according to the spindle speed. Figure 5 confirms that increasing the excitation amplitude with increasing spindle speed according to Eq. (12) so that the time delay variation amplitude is equal has the chatter suppression effect even in the high spindle speed range.

4 Conclusions

This study is summarized as follows. This paper confirmed the chatter suppression effect by exciting a workpiece in two orthogonal directions to change the relative velocity between a cutting edge and a workpiece at the cutting point. In addition, the excitation amplitude and frequency to effectively reduce the chatter vibration amplitude is derived. It was confirmed increasing the excitation amplitude with increasing spindle speed so that the time delay variation amplitude is equal has the chatter suppression effect even in the high spindle speed range. Therefore, the chatter suppression effect was confirmed at a time delay fluctuation frequency higher than that of the conventional spindle speed variation. In our future work, we will perform experimental verification to validate the proposed method for chatter suppression.

References

1. Munoa, J., et al.: Chatter suppression techniques in metal cutting. CIRP Ann. Manuf. Technol. **65**, 785–808 (2016)
2. Yang, Y., Munoa, J., Altintas, Y.: Optimization of multiple tuned mass dampers to suppress ma-chine tool chatter. Int. J. Mach. Tools Manuf **50**, 834–842 (2010)
3. Nakano, Y., Takahara, H., Kondo, E.: Countermeasure against chatter in end milling operations using multiple dynamic absorbers. J. Sound Vibr. **332**(6), 1626–1638 (2013)
4. Seguy, S., Insperger, T., Arnaud, L., Dessein, G., Peigné, G.: On the stability of high-speed milling with spindle speed variation. Int. J. Adv. Manuf. Technol. **48**(9), 883–895 (2010)
5. Xie, Q., Zhang, Q.: Stability predictions of milling with variable spindle speed using an improved semi-discretization method. Math. Comput. Simul **85**, 78–89 (2012)
6. Qin, X.B., Wan, M., Zhang, W.H., Yang, Y.: Chatter suppression with productivity improvement by scheduling a c^3 continuous feedrate to match spindle speed variation. Mech. Syst. Sig. Process. **188**, 110021 (2023)
7. Nakano, Y., De Silva, A.D., Takahara, H., Akiyama, Y.: Countermeasure against regenerative and forced chatter of flexible workpieces in milling process using a two directional excitation. Trans. Nanjing Univ. Aeronaut. Astronaut. **35**(1), 10–19 (2018)
8. Insperger, T., Stépán, G.: Semi-Discretization for Time-Delay Systems: Stability and Engineering Applications. Springer, New York (2011). https://doi.org/10.1007/978-1-4614-0335-7

Nonlinear Guidance of a Human Driver via an Automated Vehicle

Bence Szaksz[1,2(✉)], Gábor Orosz[3,4], and Gábor Stépán[1,5]

[1] Department of Applied Mechanics, Budapest University of Technology and Economics, Budapest 1111, Hungary
szaksz@mm.bme.hu
[2] MTA-BME Lendület "Momentum" Global Dynamics Research Group, Budapest University of Technology and Economics, Budapest 1111, Hungary
[3] Department of Mechanical Engineering, University of Michigan, Ann Arbor 48109, USA
[4] Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor 48109, USA
[5] HUN-REN–BME Dynamics of Machines Research Group, Budapest University of Technology and Economics, Budapest 1111, Hungary

Abstract. This study investigates the dynamics of a human-driven vehicle (HV) following an automated vehicle (AV) when the latter performs a guidance of the former one. After performing linear stability analysis, the relevant nonlinearity in the range policy of the human driver is taken into account, and the string stability of the system is analysed for different excitation amplitudes of the reference velocity. The results show that applying small positive values for the cruise control gain is advisable, while the backward-looking gain should somewhat be increased when the nonlinear effects are relevant.

Keywords: Autonomous vehicles · human and vehicle interaction · multi-vehicle systems · nonlinear frequency response

1 Introduction

Automated vehicles (AVs) are gaining popularity, but human-driven vehicles (HVs) are expected to continue dominating traffic in the coming years. Still, the presence of AVs has the potential to significantly influence the dynamics of traffic, particularly in terms of reducing the risk of accidents and decreasing energy consumption also for the vehicles behind them [6,7,18].

If there is no vehicle ahead of the AV then it can follow a predefined reference velocity, for which a conventional cruise control is satisfactory. However, in real-world scenarios, this free road condition cannot be assured, therefore, an adaptive cruise control (ACC) is needed. Furthermore, AVs have the potential to get information from connected human-driven vehicles (CHVs), which are located further away in the traffic. In case of connected cruise control (CCC) [20,23],

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 405–416, 2025.
https://doi.org/10.1007/978-3-031-72794-8_32

the CHV is ahead, while in case of connected traffic control (CTC) [18,28], this CHV is behind the AV. Finally, the scenario, when a pair of AVs is connected through traffic, is analyzed in [10].

String stability is a key measure for traffic dynamics, which assesses how the velocity fluctuations of a vehicle affect the velocity fluctuations of the following vehicles [3,22,26]. At the linear level, this can be described by a transfer function between the reference velocity and the velocity of a selected vehicle in the traffic (practically the last vehicle in a chain) [29]. However, transfer functions are used for linear systems, while the traffic dynamics is fundamentally nonlinear. There are various approaches for the calculation of the corresponding nonlinear frequency response functions including the harmonic balance method (HBM) [4,14,16], the Volterra series approach [13,21,24], the method of multiple scales [1,15,19], and the projection of the dynamics onto spectral submanifolds [5,11].

In this paper, the simplest traffic control scenario is considered, when an AV is driving in front of an HV. A reference velocity is provided to the AV which is utilized in its cruise controller. In the meantime, the AV also senses the velocity of the HV behind, which allows the AV to perform a guidance for the human driver. While previous studies have primarily focused on the string stability of linearized systems, here, we take into account the nonlinear range policy of the human driver, and present string stability charts for the nonlinear guidance of the HV.

The paper is organized as follows. Section 2 introduces the simplified car following model together with the corresponding control. In Sect. 3, the linearized system is analysed, which is followed by the investigation of the effect of the nonlinearities in Sect. 4. The nonlinear frequency response function is calculated with the harmonic balance method, and string stability charts are presented for certain large excitation amplitudes in the plane of the control gains of the AV. Finally, Sect. 5 includes some concluding remarks.

2 Modeling

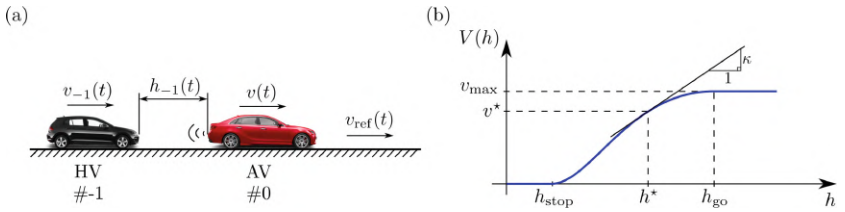


Fig. 1. Panel (a) presents the model of the guidance of a human-driven vehicle, while panel (b) shows the nonlinear range policy of the human driver.

Let us consider the simple car-following scenario when an AV is driving in front of an HV; the distance headway between the two vehicles is $h_{-1}(t)$, while the

velocities of the AV and the HV are denoted by $v(t)$ and $v_{-1}(t)$, respectively (see Fig. 1(a)). A possibly time-varying reference velocity $v_{\mathrm{ref}}(t)$ is provided to the AV that also senses the velocity difference between the two vehicles. Utilizing the translational symmetry of the system, the dynamics can be described by three first order differential equations:

$$\dot{h}_{-1} = v - v_{-1}, \tag{1}$$

$$\dot{v}_{-1} = \alpha(V(h_{-1}) - v_{-1}) + \beta(W(v) - v_{-1}), \tag{2}$$

$$\dot{v} = \hat{\beta}(v_{\mathrm{ref}} - v) + \beta_{-1}(W(v_{-1}) - v). \tag{3}$$

Here, $\hat{\beta}$ and β_{-1} are the cruise control gain and the backward-looking gain of the automated vehicle, respectively; while $W(v) = \min(v, v_{\max})$ is the velocity saturation function. Furthermore, we assume that the human driver is responding to the HV's relative velocity with respect to the AV with a control gain β, and also considers the so-called range policy $V(h_{-1})$ with the control gain α. Note that the saturation function $W(v)$ is non-smooth, which is a coarse approximation of human behavior, but this breakpoint is not reached in our domain of interest.

Figure 1(b) presents the nonlinear range policy function that determines the relation between the distance headway h and the corresponding desired velocity. We assume that if h gets smaller than the critical value h_{stop}, then the human driver intends to stop; while if the vehicle ahead is sufficiently far away, that is, $h > h_{\mathrm{go}}$, then the driver aims to travel with a comfortable $v_{\max}$ velocity. Finally, if the headway takes a value in between h_{stop} and h_{go}, then the driver is responding according to a strictly monotonically increasing function $F(h)$. Thus, the range policy function assumes the form

$$V(h) = \begin{cases} 0, & \text{if } h < h_{\mathrm{stop}}, \\ F(h), & \text{if } h_{\mathrm{stop}} \le h < h_{\mathrm{go}}, \\ v_{\max}, & \text{if } h_{\mathrm{go}} \le h. \end{cases} \tag{4}$$

Here, we approximate $F(h)$ with the third-degree polynomial

$$F(h) = v_{\max} \frac{(3h_{\mathrm{go}} - h_{\mathrm{stop}} - 2h)(h - h_{\mathrm{stop}})^2}{(h_{\mathrm{go}} - h_{\mathrm{stop}})^3}, \tag{5}$$

which introduces the relevant nonlinearity into the system.

This model of the human driver is called optimal velocity model, which is widely used in the literature [9,12,17,25]; moreover, the typical numerical values of the human gains α and β have already been measured and estimated in [2].

3 Linear Analysis

First, let us investigate the linearized dynamics to provide a basis of comparison for nonlinear studies.

In case of constant reference velocity $v_{\mathrm{ref}}(t) = \bar{v}_{\mathrm{ref}} < v_{\max}$, the governing equations (1)-(3) yield the steady state motion $v^{\star}_{-1} = v^{\star} = \bar{v}_{\mathrm{ref}}$, while the inverse of the range policy determines the distance between the two vehicles: $h^{\star}_{-1} = V^{-1}(\bar{v}_{\mathrm{ref}})$. As Fig. 1(b) presents, the slope of the range policy at the steady state is denoted by $\kappa = V'(h^{\star}_{-1})$.

Let us introduce the time varying reference velocity $v_{\mathrm{ref}}(t) = \bar{v}_{\mathrm{ref}} + \tilde{v}_{\mathrm{ref}}(t)$ and the shifted state vector $\mathbf{x} = [\tilde{h}_{-1}\ \tilde{v}_{-1}\ \tilde{v}]^{\mathrm{T}}$ with

$$\tilde{h}_{-1} = h_{-1} - h^{\star}_{-1}, \quad \tilde{v}_{-1} = v_{-1} - \bar{v}_{\mathrm{ref}}, \quad \text{and} \quad \tilde{v} = v - \bar{v}_{\mathrm{ref}}, \tag{6}$$

which leads to the linearized state space model

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\tilde{v}_{\mathrm{ref}}, \tag{7}$$

$$y = \mathbf{C}\mathbf{x}. \tag{8}$$

Here, the velocity of the HV is selected as the output y, and the coefficient matrices assume the form

$$\mathbf{A} = \begin{bmatrix} 0 & -1 & 1 \\ \alpha\kappa & -(\alpha+\beta) & \beta \\ 0 & \beta_{-1} & -(\hat{\beta}+\beta_{-1}) \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ \hat{\beta} \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 0\ 1\ 0 \end{bmatrix}. \tag{9}$$

The stability of the linearized system can be analysed in two ways. On one hand, the system is called plant stable if it is stable at constant reference velocity $v_{\mathrm{ref}} = \bar{v}_{\mathrm{ref}}$; on the other hand, the system is called string stable if the oscillations in the input $\tilde{v}_{\mathrm{ref}}(t)$ do not lead to increased oscillations in the output y.

3.1 Plant Stability

The plant stability is determined by the roots of the characteristic function $D(s) = \det(s\mathbf{I} - \mathbf{A})$ which leads to the characteristic equation

$$s^3 + (\alpha + \beta + \hat{\beta} + \beta_{-1})s^2 + (\alpha\hat{\beta} + \alpha\beta_{-1} + \beta\hat{\beta} + \alpha\kappa)s + \alpha\kappa\hat{\beta} = 0. \tag{10}$$

The system is stable if and only if all the characteristic roots have negative real parts. Because of Rouche's theorem, the location of the roots is a continuous function of the control gains. Thus, varying the parameters, the system may loose its stability in two ways: either a root crosses the imaginary axis through the origin of the complex plane leading to static loss of stability (steady-state bifurcation); or a pair of complex conjugate roots crosses the imaginary axis with nonzero imaginary part, which leads to dynamic loss of stability (Hopf bifurcation).

Thus, the static stability boundary can be determined by substituting $s = 0$ into Eq. (10), which yields

$$\alpha\kappa\hat{\beta} = 0. \tag{11}$$

Similarly, the dynamic stability boundary can be calculated by substituting $s = \mathrm{j}\Omega$ into the characteristic equation, where Ω is the angular frequency, with

which the system starts oscillating when it loses stability. This leads to a complex equation, the real and imaginary parts of which have to be 0 at the same time. From the system of the corresponding two scalar algebraic equations, the critical control gains of the AV can be expressed as the function of Ω:

$$\beta_{-1}(\Omega) = -\frac{\Omega^4 + ((\alpha+\beta)^2 - 2\alpha\kappa)\Omega^2 + \alpha^2\kappa^2}{\beta\Omega^2 + \alpha^2\kappa}, \tag{12}$$

$$\hat{\beta}(\Omega) = \frac{\Omega^4 + \alpha(\alpha+\beta-\kappa)\Omega^2}{\beta\Omega^2 + \alpha^2\kappa}. \tag{13}$$

Figure 2 presents a stability chart in the plane of the AV control gains β_{-1} and $\hat{\beta}$, while the control gains of the human driver are fixed (see [2]). The static and dynamic stability boundaries are denoted by dashed and solid red lines, respectively; both the dark and the light gray regions are plant stable. This means that the cruise control gain $\hat{\beta}$ has to be positive, while the backward-looking gain β_{-1} may also be tuned to negative values.

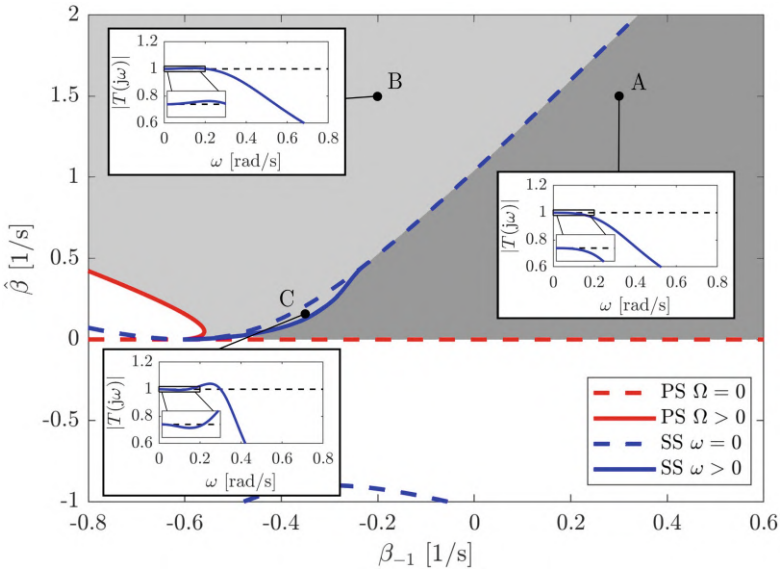


Fig. 2. Linear stability chart ($\alpha = 0.3\ \mathrm{s}^{-1}$, $\beta = 0.4\ \mathrm{s}^{-1}$, $\kappa = 0.6\ \mathrm{s}^{-1}$). Both the light gray and the dark gray regions are plant stable, while the dark gray region is also string stable. The inlets display typical frequency response curves at the corresponding control gain combinations.

3.2 String Stability

The string stability is defined by the transfer function between the input and the output of the system, which are now the reference velocity and the velocity

of the HV. The corresponding transfer function takes the form

$$T(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} =: \frac{N(s)}{D(s)}, \tag{14}$$

where $D(s)$ is the same characteristic function as it was used in the plant stability calculations, while the numerator assumes the form

$$N(s) = s\beta\hat{\beta} + \alpha\kappa\hat{\beta}. \tag{15}$$

The system is called string stable if

$$|T(\mathrm{j}\omega)| < 1, \text{ for all } \omega > 0. \tag{16}$$

Again, two types of string instability may occur: in case of the string instability at $\omega = 0$, the norm of the transfer function goes above 1 immediately after $\omega = 0$; while in case of the string instability at an $\omega > 0$, the norm of the transfer function first decreases but it has a peak later, at which it exceeds 1.

As it is derived in [27], the $\omega = 0$ string instability boundary is located at

$$\hat{\beta} = -\beta_{-1}\frac{\alpha + \beta - \kappa}{\alpha + 2\beta - 2\kappa} \pm \frac{\sqrt{\beta_{-1}^2(\alpha + \beta - \kappa)^2 - \alpha(\alpha + 2\beta - 2\kappa)(\beta_{-1} + \kappa)^2}}{\alpha + 2\beta - 2\kappa}. \tag{17}$$

The closed-form expression of the $\omega > 0$ string instability boundary has a lengthy expression that is not manageable in the present format here.

In Fig. 2, the dashed and solid blue curves correspond to the $\omega = 0$ and to the $\omega > 0$ string stability boundaries, respectively. Therefore, the dark gray region is both plant and string stable. Furthermore, the inlets present some typical transfer functions at three different control gain configurations. At point A, the system is string stable and (16) is satisfied. At point B, the norm of the transfer function goes beyond 1 immediately after $\omega = 0$. Finally, point C is close to an $\omega > 0$ string stability boundary, that is, $|T(\mathrm{j}\omega)|$ first decreases then it exceeds 1 later, as shown in the inlet.

4 Effect of Nonlinearities

Taking into account the relevant nonlinearities of the range policy modifies both the plant dynamics and the string stability. Assuming constant reference velocities, stable or unstable self-excited oscillations will occur close to the dynamic stability boundaries, which originate from super- or subcritical Hopf bifurcations. Considering the nonlinear string stability, the corresponding nonlinear frequency response function (FRF) is a function of the perturbation amplitude and includes not only the basic harmonics but also higher ones. Moreover, in the case of asymmetric nonlinearites, it also contains a shift term. The actual paper focuses on the string stability and the corresponding nonlinear FRFs of the car following model.

Let us assume that the reference velocity is subjected to a sinusoidal perturbation $\tilde{v}_{\mathrm{ref}}(t) = \varepsilon \cos(\omega t)$ leading to $v_{\mathrm{ref}}(t) = \bar{v}_{\mathrm{ref}} + \varepsilon \cos(\omega t)$. In this case, the equation of motion will take the form of the nonlinear non-autonomous differential equation

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{g}(\mathbf{x}) + \mathbf{B}\varepsilon \cos(\omega t), \tag{18}$$

where $\mathbf{g} = \mathbf{g}_2 + \mathbf{g}_3$ contains the second- and third-degree nonlinearities, respectively, which now assume the form

$$\mathbf{g}_2(\boldsymbol{\Phi}, \boldsymbol{\Psi}) = \begin{bmatrix} 0 \\ \frac{\alpha}{2} V''(h^\star_{-1}) \Phi_1 \Psi_1 \\ 0 \end{bmatrix}, \quad \mathbf{g}_3(\boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}) = \begin{bmatrix} 0 \\ \frac{\alpha}{6} V'''(h^\star_{-1}) \Phi_1 \Psi_1 \Lambda_1 \\ 0 \end{bmatrix}. \tag{19}$$

Subscript 1 denotes the first element of the 3-dimensional vectors $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$, or $\boldsymbol{\Lambda}$.

The nonlinear equation of motion (18) can not be solved in closed form, however, there are various approaches to approximate the solution analytically. In this paper, we apply the harmonic balance method [4,14,16].

4.1 Harmonic Balance

Since it is easier to work with exponential functions, let us substitute $\cos(\omega t)$ with the Euler's formula in (18):

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{g}(\mathbf{x}) + \frac{1}{2}\mathbf{B}\varepsilon \left(\mathrm{e}^{\mathrm{j}\omega t} + \mathrm{e}^{-\mathrm{j}\omega t}\right). \tag{20}$$

In what follows, assume that the linear unexcited system is stable. In the case of small excitation amplitudes, it can be assumed that the $\mathcal{O}(\varepsilon)$ terms of the solution do not depend on the nonlinearities, that is, these form the solution of the excited linearized system. So, the solution is approximated as

$$\mathbf{x}(t) = \varepsilon \left(\mathbf{a}_{10}\mathrm{e}^{\mathrm{j}\omega t} + \mathbf{a}_{01}\mathrm{e}^{-\mathrm{j}\omega t}\right) + \mathcal{O}(\varepsilon^2). \tag{21}$$

Substituting (21) into the equation of motion (20), the coefficients of $\varepsilon \mathrm{e}^{\mathrm{j}\omega t}$ can be balanced:

$$\mathrm{j}\omega \mathbf{a}_{10} = \mathbf{A}\mathbf{a}_{10} + \frac{1}{2}\mathbf{B}, \tag{22}$$

which results in the solution

$$\mathbf{a}_{10} = \frac{1}{2}\left(\mathrm{j}\omega \mathbf{I} - \mathbf{A}\right)^{-1}\mathbf{B}. \tag{23}$$

Looking for the coefficients of $\varepsilon \mathrm{e}^{-\mathrm{j}\omega t}$ yields that $\mathbf{a}_{01} = \bar{\mathbf{a}}_{10}$, where overbar denotes the complex conjugate. This is consistent with the requirement that the solution $x(t)$ has to be a real function of time. The higher-order terms can be calculated successively by substituting the actual approximation of $x(t)$ in the

nonlinear equation of motion (20). The second- and third-order nonlinearities yield that the particular solution assumes the form

$$\begin{aligned}\mathbf{x}(t) = &\varepsilon\left(\mathbf{a}_{10}\mathrm{e}^{\mathrm{j}\omega t} + \mathbf{a}_{01}\mathrm{e}^{-\mathrm{j}\omega t}\right) + \varepsilon^2\left(\mathbf{a}_{20}\mathrm{e}^{2\mathrm{j}\omega t} + \mathbf{a}_{11} + \mathbf{a}_{02}\mathrm{e}^{-2\mathrm{j}\omega t}\right) \\ &+ \varepsilon^3\left(\mathbf{a}_{30}\mathrm{e}^{3\mathrm{j}\omega t} + \mathbf{a}_{21}\mathrm{e}^{\mathrm{j}\omega t} + \mathbf{a}_{12}\mathrm{e}^{-\mathrm{j}\omega t} + \mathbf{a}_{03}\mathrm{e}^{-3\mathrm{j}\omega t}\right) + \ldots.\end{aligned} \tag{24}$$

Note, that $\mathrm{e}^{j\omega t}$ terms appear not only in the coefficients of ε, but also in the coefficients of ε^3 and all higher-order terms with odd order. Thus, applying the harmonic balance method, the coefficients of $\varepsilon^2\mathrm{e}^{2\mathrm{j}\omega t}$ yield that

$$\mathbf{a}_{20} = \left(2\mathrm{j}\omega\mathbf{I} - \mathbf{A}\right)^{-1}\mathbf{g}_2(\mathbf{a}_{10}, \mathbf{a}_{10}), \tag{25}$$

while, again, $\mathbf{a}_{02} = \bar{\mathbf{a}}_{20}$.

As it can be expected from the presence of the second-degree nonlinearities, there will be a shift in the solution. This corresponds to the $\varepsilon^2\mathrm{e}^0$ terms, which yields

$$\mathbf{a}_{11} = -2\mathbf{A}^{-1}\mathbf{g}_2(\mathbf{a}_{10}, \mathbf{a}_{01}). \tag{26}$$

One can also calculate the $\mathbf{a}_{30}$ and $\mathbf{a}_{21}$ vectors from the balance of the coefficients of $\varepsilon^3\mathrm{e}^{3\mathrm{j}\omega t}$ and $\varepsilon^3\mathrm{e}^{\mathrm{j}\omega t}$. These take the form

$$\mathbf{a}_{30} = \left(3\mathrm{j}\omega\mathbf{I} - \mathbf{A}\right)^{-1}\left(2\mathbf{g}_2(\mathbf{a}_{10}, \mathbf{a}_{20}) + \mathbf{g}_3(\mathbf{a}_{10}, \mathbf{a}_{10}, \mathbf{a}_{10})\right), \tag{27}$$

$$\mathbf{a}_{21} = \left(\mathrm{j}\omega\mathbf{I} - \mathbf{A}\right)^{-1}\left(2\mathbf{g}_2(\mathbf{a}_{01}, \mathbf{a}_{20}) + 2\mathbf{g}_2(\mathbf{a}_{10}, \mathbf{a}_{11}) + 3\mathbf{g}_3(\mathbf{a}_{10}, \mathbf{a}_{10}, \mathbf{a}_{01})\right); \tag{28}$$

and finally, $\mathbf{a}_{03} = \bar{\mathbf{a}}_{30}$ and $\mathbf{a}_{12} = \bar{\mathbf{a}}_{21}$.

Note that neglecting the nonlinearity, the response will only be determined by $\mathbf{a}_{10} = \bar{\mathbf{a}}_{01}$; so, Eq. (23) together with Eq. (8) yield the transfer function (14).

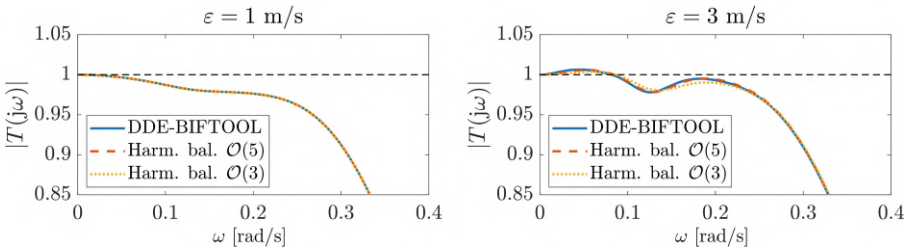


Fig. 3. Comparison of the nonlinear frequency response $M(\omega)$ calculated with 3rd and 5th order harmonic balance method and with the DDE-BIFTOOL package for excitation amplitudes $\varepsilon = 1$ m/s and $\varepsilon = 3$ m/s ($\alpha = 0.3$ s^{-1}, $\beta = 0.4$ s^{-1}, $\beta_{-1} = -0.3$ s^{-1}, $\hat{\beta} = 0.18$ s^{-1}, $v_{\max} = 30$ m/s, $\bar{v}_{\mathrm{ref}} = 26.55$ m/s, $h_{\mathrm{stop}} = 5$ m, $h_{\mathrm{go}} = 55$ m).

Let $M(\omega)$ denote the nonlinear magnification between the amplitude of oscillation of the reference velocity and that of the HV. Figure 3 compares this nonlinear magnification calculated with harmonic balance method up to 3rd and 5th order with the results of the DDE-BIFTOOL package in MATLAB [8] (see

details in the Appendix). As it can be observed, the frequency response depends on the amplitude of the excitation. If $\varepsilon = 1$ m/s, then the three curves run exactly on each other and the response is string stable; while in case of $\varepsilon = 3$ m/s, there is a slight difference between the responses. In this case, the 5^{th} order harmonic balance method approximates the continuation-based solution better, but all the three results show string unstable property. This comparison demonstrates the accuracy of the harmonic balance method, which is computationally more efficient than the continuation based solvers, making it a good candidate for performing multi-dimensional parameter analysis.

Figure 4 presents stability charts in the plane of the control gains of the AV, which are colored according to the maximum of the norm of the nonlinear frequency response. The linear string stability limit is recovered for small excitation amplitudes (cf. Fig. 2), while the string stable domain shrinks for small and extends for large values of $\hat{\beta}$, respectively, as the excitation amplitude is increased. Note, that the large values of $\hat{\beta}$ might lead to unrealistic large accelerations of the AV. Panel (c) of Fig. 4 shows the change of the transfer functions for the gains $\beta_{-1} = -0.3\ \text{s}^{-1}$ and $\hat{\beta} = 0.18\ \text{s}^{-1}$ as the excitation amplitude increases; this specific case was analysed in Fig. 3. Here, it can be observed in more detail how the nonlinear system loses its string stability as the excitation amplitude increases. Note that these figures are created using the 5^{th} order harmonic balance method, which still does not take into account either the saturation of the range policy function or the velocity saturation $W(v)$. However, in case of the string stable and the slightly string unstable solutions, these saturation scenarios do not occur.

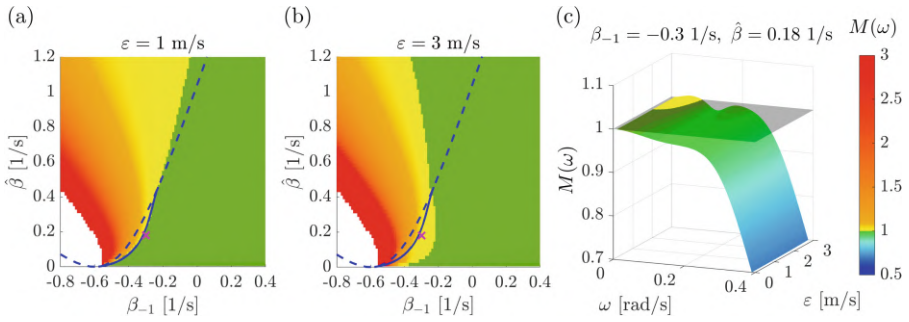


Fig. 4. Panels (a) and (b) present stability charts in case of different excitation amplitudes colored according to the maximum of the norm of the frequency response function ($\alpha = 0.3\ \text{s}^{-1}$, $\beta = 0.4\ \text{s}^{-1}$, $v_{\max} = 30$ m/s, $\bar{v}_{\text{ref}} = 26.55$ m/s, $h_{\text{stop}} = 5$ m, $h_{\text{go}} = 55$ m). The green region is plant and string stable, while the blue dashed and continuous lines refer to linear string stability boundaries. Panel (c) shows the change of the frequency response curves as the excitation amplitude increases (for the parameter combination $\beta_{-1} = -0.3\ \text{s}^{-1}$, $\hat{\beta} = 0.18\ \text{s}^{-1}$, marked with magenta crosses in the stability charts).

5 Conclusion

In this study, we investigated the dynamics of an HV following an AV equipped with a backward-looking control. The results of the linear analysis showed that it is advisable to apply small positive values for the cruise control gain, while the backward-looking gain can take both positive and negative values. However, taking into account the nonlinearities in the range policy function, we found that the backward looking gain should be increased. Measurements are to be performed in the future to tune the model of the human driver and verify the theoretical results. This will enable us to design control systems for automated vehicles and guide the dynamics of the traffic flow in mixed traffic scenarios.

Appendix

The DDEBIFTOOL MATLAB package is a continuation based solver primarily developed for the analysis of delay differential equations but it can also work with autonomous systems without delay [8]. Although, our system is non-autonomous, it can be transformed to an autonomous ODE by extending it with a two dimensional subsystem, the periodic solution of which is $[y(t),\ z(t)]^{\mathrm{T}} = [\sin(\omega t + \gamma),\ \cos(\omega t + \gamma)]^{\mathrm{T}}$. The extended ODE takes the form

$$\begin{bmatrix} \dot{\mathbf{x}} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \varepsilon\mathbf{B} \\ \mathbf{0} & 1 & \omega \\ \mathbf{0} & -\omega & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ y \\ z \end{bmatrix} + \begin{bmatrix} \mathbf{g}(\mathbf{x}) \\ -y(y^2+z^2) \\ -z(y^2+z^2) \end{bmatrix}, \tag{29}$$

the periodic solution of which is the same as that of the original system (18).

References

1. Ahmadi, H., Foroutan, K.: Nonlinear primary resonance of spiral stiffened functionally graded cylindrical shells with damping force using the method of multiple scales. Thin-Walled Struct. **135**, 33–44 (2019)
2. Avedisov, S.S., Bansal, G., Orosz, G.: Impacts of connected automated vehicles on freeway traffic patterns at different penetration levels. IEEE Trans. Intell. Transp. Syst. **23**(5), 4305–4318 (2022)
3. Besselink, B., Johansson, K.H.: String stability and a delay-based spacing policy for vehicle platoons subject to disturbances. IEEE Trans. Autom. Control **62**(9), 4376–4391 (2017)
4. Blahoš, J., Vizzaccaro, A., Salles, L., El Haddad, F.: Parallel harmonic balance method for analysis of nonlinear dynamical systems. In: Turbo Expo: Power for Land, Sea, and Air, vol. 84232. American Society of Mechanical Engineers (2020). V011T30A028
5. Breunung, T., Haller, G.: Explicit backbone curves from spectral submanifolds of forced-damped nonlinear mechanical systems. Proc. Roy. Soc. A: Math. Phys. Eng. Sci. **474**(2213), 20180083 (2018)
6. Caveney, D.: Cooperative vehicular safety applications. IEEE Control Syst. Mag. **30**(4), 38–53 (2010)

7. Chan, E., Gilhead, P., Jelinek, P., Krejci, P., Robinson, T.: Cooperative control of SARTRE automated platoon vehicles. In: Proceedings of the 19th ITS World Congress, Vienna (2012)
8. Engelborghs, K., Luzyanina, T., Samaey, G., et al.: DDE-BIFTOOL: a Matlab package for bifurcation analysis of delay differential equations. TW Report **305**(1) (2000)
9. Gisolo, C.M., Delle Monache, M.L., Ferrante, F., Frasca, P.: Nonlinear analysis of stability and safety of optimal velocity model vehicle groups on ring roads. IEEE Trans. Intell. Transp. Syst. **23**(11), 20628–20635 (2022)
10. Guo, S., Orosz, G., Molnar, T.G.: Connected cruise and traffic control for pairs of connected automated vehicles. IEEE Trans. Intell. Transp. Syst. **24**(11), 12648–12658 (2023)
11. Haller, G., Ponsioen, S.: Nonlinear normal modes and spectral submanifolds: existence, uniqueness and use in model reduction. Nonl. Dyn. **86**, 1493–1534 (2016)
12. Jin, Y., Meng, J.: Dynamical analysis of an optimal velocity model with time delayed feedback control. Commun. Nonl. Sci. Numer. Simulat. **90**, 105333 (2020)
13. Jones, J.P., Yaser, K.: Computation of the MIMO Volterra frequency response functions of nonlinear systems. Mech. Syst. Signal Process. **134**, 106323 (2019)
14. Krack, M., Gross, J.: Harmonic Balance for Nonlinear Vibration Problems, vol. 1. Springer (2019)
15. Lelkes, J., Kalmár-Nagy, T.: Bifurcation analysis of a forced delay equation for machine tool vibrations. Nonl. Dyn. **98**(4), 2961–2974 (2019). https://doi.org/10.1007/s11071-019-04984-w
16. Marathe, A., Chatterjee, A.: Wave attenuation in nonlinear periodic structures using harmonic balance and multiple scales. J. Sound Vib. **289**(4–5), 871–888 (2006)
17. Martinovich, K., Kiss, A.K.: Nonlinear effects of saturation in the car-following model. Nonl. Dyn. **111**(3), 2555–2569 (2023)
18. Molnár, T.G., Hopka, M., Upadhyay, D., Van Nieuwstadt, M., Orosz, G.: Virtual rings on highways: traffic control by connected automated vehicles. In: AI-enabled Technologies for Autonomous and Connected Vehicles, pp. 441–479. Springer (2023)
19. Nayfeh, A.H.: Resolving controversies in the application of the method of multiple scales and the generalized method of averaging. Nonl. Dyn. **40**, 61–102 (2005)
20. Orosz, G.: Connected cruise control: modelling, delay effects, and nonlinear behaviour. Veh. Syst. Dyn. **54**(8), 1147–1176 (2016)
21. Peng, Z., Lang, Z., Billings, S., Tomlinson, G.: Comparisons between harmonic balance and nonlinear output frequency response function in nonlinear system analysis. J. Sound Vib. **311**(1–2), 56–73 (2008)
22. Ploeg, J., van de Wouw, N., Nijmeijer, H.: L_p string stability of cascaded systems: application to vehicle platooning. IEEE Trans. Control Syst. Technol. **22**(2), 786–793 (2014)
23. Qin, W.B., Gomez, M.M., Orosz, G.: Stability and frequency response under stochastic communication delays with applications to connected cruise control design. IEEE Trans. Intell. Transp. Syst. **18**(2), 388–403 (2017)
24. Rugh, W.J.: Nonlinear System Theory. Johns Hopkins University Press, Baltimore (1981)
25. Shen, M., He, C.R., Molnar, T.G., Bell, A.H., Orosz, G.: Energy-efficient connected cruise control with lean penetration of connected vehicles. IEEE Trans. Intell. Transp. Syst. **24**(4), 4320–4332 (2023)

26. Swaroop, D., Hedrick, J.K.: String stability of interconnected systems. IEEE Trans. Autom. Control **41**(3), 349–357 (1996)
27. Szaksz, B., Orosz, G., Stepan, G.: Guided control of a human driver via an automated vehicle. IFAC-PapersOnLine **56**(2), 899–904 (2023)
28. Wang, J., Zheng, Y., Chen, C., Xu, Q., Li, K.: Leading cruise control in mixed traffic flow: System modeling, controllability, and string stability. IEEE Trans. Intell. Transp. Syst. (2021)
29. Zhang, L., Orosz, G.: Motif-based design for connected vehicle systems in presence of heterogeneous connectivity structures and time delays. IEEE Trans. Intell. Transp. Syst. **17**(6), 1638–1651 (2016)

Performance Evaluation of Model Predictive Control for Active Suspension Systems Employing an Inerter with Road Preview

Wanshu Wang[1(✉)] and Takehiko Asai[2]

[1] Graduate School of Science and Technology, University of Tsukuba, Tsukuba, Japan
wangwans18@gmail.com
[2] Faculty of Engineering Information and Systems, University of Tsukuba, Tsukuba, Japan
asai@kz.tsukuba.ac.jp

Abstract. Advancements in sensor technology have enabled the measurement of road profile information ahead of the vehicle. In this context, model predictive control (MPC) emerges as a viable control strategy for implementing active suspension systems, especially when equipped with road preview sensors. This preference is attributed to the ability of the MPC approach to make real-time predictions about the future behavior of the system and to accommodate constraints during the optimization process. Additionally, it is worth noting that the effectiveness of employing an inerter in passive suspension systems has been well-demonstrated in the previous research. Therefore, this study proposes an inerter-based active suspension controlled by the MPC scheme with the road preview information. To assess the efficacy of the proposed strategy, numerical simulations are conducted using a quarter car model. By comparing to the passive control and the existing MPC approaches, the advantages of the MPC with preview for performance improvement of the suspension were validated.

Keywords: Inerter · Active suspension · Model predictive control · Road preview · Vehicle dynamics

1 Introduction

Over the past few decades, inerter, a component renowned for producing amplified inertial forces proportional to the relative acceleration across its two terminals, has garnered significant attention for passive vehicle suspension systems. Following the preliminary optimization study of the potential superiority of the inerter in the passive suspensions [1], subsequent studies have shed light on performance limitations related to suspension deflection for certain inerter-based configurations [2].

Recent studies incorporated various optimal control strategies into active and semi-active inerter-based suspensions to enhance dynamic performance across diverse operational contexts. Notably, the H_2 optimal control approach can be transformed into an optimal feedback system utilizing linear state feedback control, rendering it highly

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 417–425, 2025.
https://doi.org/10.1007/978-3-031-72794-8_33

amenable for practical engineering implementation [3]. A subclass of H_2 optimal control, the linear quadratic regulator (LQR) control strategy, is particularly prominent due to its straightforward and tangible physical interpretation [4]. Furthermore, due to the constraints on suspension working space, generator input current and its rate, there has been a growing interest in model predictive control (MPC) because it enables the explicit incorporation of constraints, distinguishing it from the LQR controller [5]. Additionally, recent advancements in road preview sensors have paved the way for their seamless integration with active suspensions utilizing MPC control strategies. This integration holds the potential to significantly enhance the dynamic performance of active suspension systems [6].

The principal contribution of this study centers on the incorporation of the MPC approach, coupled with the road preview capability, into an active suspension system integrated with an inerter. By replacing the conventional fixed damper with a motor capable of generating variable electromagnetic force, the current of the motor can be effectively controlled to improve the dynamic performance, thus enhancing ride comfort and stability of the active suspension system.

2 Modeling

2.1 Quarter Car Model

In this paper, a quarter car (QC) model is considered for the analysis and control of the suspension vibration performance. Figure 1 shows a typical QC model in which an inerter and a motor are installed between the sprung mass and the unsprung mass. The dynamic equation can be developed as Eq. (1)

$$\begin{aligned} &m_s\ddot{z}_s + c(\dot{z}_s - \dot{z}_u) + k_1(z_s - z_u) - K_e u + b(\ddot{z}_s - \ddot{z}_u) = 0 \\ &m_u\ddot{z}_u - c(\dot{z}_s - \dot{z}_u) - k_1(z_s - z_u) + K_e u - b(\ddot{z}_s - \ddot{z}_u) + k_t(z_u - z_{r0}) = 0 \end{aligned} \tag{1}$$

where m_s and m_u are the sprung and unsprung masses; k_1 and c are the vertical suspension stiffness and residual damping of the passive components; b is the inertance of the inerter; k_t is the tire stiffness; z_s and z_u are displacements of sprung and unsprung masses; and z_{r0} is the vertical displacement of the tire contact patch. Note that the term $K_e u$ represents the electromagnetic force generated by the motor [7], where K_e is a constant associated with the back-EMF of the generator, and u is the controllable current of the motor.

Then Eq. (1) can be converted into an equivalent continuous-time state-space representation:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}u(t) + \mathbf{G}z_{r0}(t) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{D}u(t) + \mathbf{H}z_{r0}(t) \end{aligned} \tag{2}$$

where the state and output vectors are respectively defined as $\mathbf{x} = \left[z_s \; z_u \; \dot{z}_s \; \dot{z}_u \right]^T$ and $\mathbf{y} = \left[\ddot{z}_s \; k_t(z_u - z_r) \right]^T$; $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are the system matrices; $\mathbf{G}$ and $\mathbf{H}$ are the road disturbance matrices; and t is time.

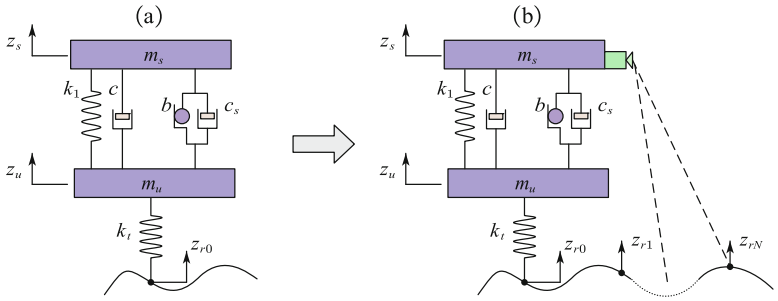


Fig. 1. QC model: (a) Typical layout, including motor and inerter in parallel, (b) Augmented layout, including the road profile model and preview capability.

2.2 Augmented State-Space Representation with Road Preview

To implement the preview control, a preview sensor is mounted to scan the road conditions ahead of the vehicle, as depicted in Fig. 1. Dynamics of the road disturbance information can be described by a shift register [8], which is expressed in discrete-time form as

$$\begin{bmatrix} z_{r0}(k+1) \\ z_{r1}(k+1) \\ z_{r2}(k+1) \\ \vdots \\ z_{rN-1}(k+1) \\ z_{rN}(k+1) \end{bmatrix} = \begin{bmatrix} 0\,1\,0\,0\cdots 0 \\ 0\,0\,1\,0\cdots 0 \\ 0\,0\,0\,1\cdots 0 \\ \vdots\,\vdots\,\vdots\,\vdots\,\ddots\,\vdots \\ 0\,0\,0\,0\cdots 1 \\ 0\,0\,0\,0\cdots 0 \end{bmatrix} \begin{bmatrix} z_{r0}(k) \\ z_{r1}(k) \\ z_{r2}(k) \\ \vdots \\ z_{rN-1}(k) \\ z_{rN}(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} y_r(k) \tag{3}$$

where k indicates the current time step. Equation (3) can be simplified as

$$\hat{z}_r(k+1) = \begin{bmatrix} 0 & \mathbf{A}_{rd} \end{bmatrix} \hat{z}_r(k) + \mathbf{E}_{rd} y_r(k) \tag{4}$$

where $\hat{z}_r = \begin{bmatrix} z_{r0} \cdots z_{rN} \end{bmatrix}^T$ is the vector of the road states, i.e., the road profile heights ahead of the tire, which consists of N points spaced according to the time step T_s; $\mathbf{A}_{rd}$ is the shift model matrix; $y_r(k) = z_{rN}(k+1)$ is the road disturbance input; and $\mathbf{E}_{rd}$ is the road disturbance matrix. Accordingly, to integrate with the MPC control scheme, the continuous-time state-space form should be discretized with a sampling time applying the zero-order hold method. Thus Eq. (2) is converted into

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}_d \mathbf{x}(k) + \mathbf{B}_d u(k) + \mathbf{G}_d z_{r0}(k) \\ \mathbf{y}(k) &= \mathbf{C}_d \mathbf{x}(k) + \mathbf{D}_d u(k) + \mathbf{H}_d z_{r0}(k) \end{aligned} \tag{5}$$

where $\mathbf{A}_d, \mathbf{B}_d, \mathbf{C}_d, \mathbf{D}_d, \mathbf{G}_d, \mathbf{H}_d$ are the discrete system matrices.

Substituting Eq. (4) into Eq. (5), the augmented discrete-time form can be obtained as

$$\begin{aligned} \mathbf{x}(k+1) &= \begin{bmatrix} \mathbf{A}_d & 0 \\ 0 & \mathbf{A}_{rd} \end{bmatrix} \mathbf{x}(k) + \begin{bmatrix} \mathbf{B}_d \\ 0 \end{bmatrix} u(k) + \begin{bmatrix} \mathbf{G}_d \\ 0 \end{bmatrix} z_{r0}(k) + \begin{bmatrix} 0 \\ \mathbf{E}_{rd} \end{bmatrix} y_r(k) \\ \mathbf{y}(k) &= \begin{bmatrix} \mathbf{C}_d & 0 \end{bmatrix} \mathbf{x}(k) + \mathbf{H}_d z_{r0}(k) + \mathbf{D}_d u(k) \end{aligned} \tag{6}$$

which can be simplified to

$$\begin{aligned}\mathbf{x}(k+1) &= \mathbf{A}_{sd}\mathbf{x}(k) + \mathbf{B}_{sd}u(k) + \mathbf{G}_{sd}z_{r0}(k) + \mathbf{E}_{sd}y_r(k)\\ \mathbf{y}(k) &= \mathbf{C}_{sd}\mathbf{x}(k) + \mathbf{D}_d u(k) + \mathbf{H}_d z_{r0}(k)\end{aligned} \tag{7}$$

2.3 Model Predictive Controller Formulation

The predicted output $\mathbf{Y}$ consists of the state vector $\mathbf{x}(k)$, the controlled input $\mathbf{U}$ over control horizon, and the road profile $\tilde{z}_{r0}$ over the prediction horizon, which is given as

$$\mathbf{Y} = \mathbf{F}\mathbf{x}(k) + \Phi\mathbf{U} + \Theta\tilde{z}_{r0} \tag{8}$$

where $\mathbf{F}$, Φ, and Θ are resulting matrices derived from Eq. (7). Then the optimization problem for the model predictive controller is to minimize the cost function:

$$\min_{\mathbf{U}} (\mathbf{Y}^T\mathbf{Q}\mathbf{Y} + \mathbf{U}^T\mathbf{R}\mathbf{U}) \tag{9}$$

where $\mathbf{Q}$ and $\mathbf{R}$ are weighting matrices that are both positive definite. For the QC model, $\mathbf{Q}$ is chosen to minimize vehicle body acceleration and dynamic tire load. $\mathbf{R}$ focuses on the control input current of the motor, thereby reflecting the energy demand. Considering the constraints of Eq. (8) and the generator input current and its rate, then the optimization problem can be converted to an equivalent quadratic programming format as follows:

$$\begin{aligned}&\min_{\mathbf{U}}\left(\frac{1}{2}\mathbf{U}^T\mathbf{S}\mathbf{U} + \mathbf{f}^T\mathbf{U}\right)\\ &\text{s.t. } \mathbf{P}\mathbf{U} \le \mathbf{M}_1 + \mathbf{M}_2\mathbf{x}(k)\end{aligned} \tag{10}$$

where $\mathbf{S}$, $\mathbf{f}$, $\mathbf{P}$, $\mathbf{M}_1$, and $\mathbf{M}_2$ are appropriate constant matrices relevant to the quadratic programming problem. The basic task in quadratic programming involves determining the minimum of a positive definite quadratic function that is subject to certain constraints. In this study, the optimization approach enables the calculation of the optimal current for each time step within the preview horizon. Based on the electromechanical characteristics of the motor, the desired damping force can be generated to minimize the vehicle body acceleration and dynamic tire.

3 Numerical Simulation

To illustrate and compare the dynamic performance of the inerter-based suspension among the passive control, MPC controller without preview, and MPC with preview, numerical simulations are carried out with appropriate parameter values. Table 1 gives the parameters of the QC model used in this study. Moreover, the sampling time is set to 10 ms and the preview time for the proposed MPC approach is set to 90 ms. The weighting matrices $\mathbf{Q}$ and $\mathbf{R}$ are well-tuned to obtain good results for the MPC with and without preview over different road profiles.

For the comprehensive numerical analysis, in this paper, two typical road profiles are respectively utilized as disturbance inputs: a bump road based on a simple harmonic function to simulate a car running through an obstacle like a speed bump, and random road based on the white noise to simulate passing through more general road surfaces. The bump and random road profiles can be respectively described as the following equations:

$$\dot{z}_{bump}(t) = \frac{A}{2}(1 - \cos\frac{2\pi V(t - t_0)}{L}),\ t_0 \le t \le \frac{L}{V} \tag{11}$$

$$\dot{z}_{random}(t) = -\alpha V z_{random}(t) + w(t) \tag{12}$$

where $z_{bump}(t)$ and $z_{random}(t)$ are the displacements of the bump and random road inputs respectively, A is the bump height, L is the bump length, V represents the car speed, t_0 is the starting time of the bump, t is the simulation time length, α is the road roughness coefficient and set as 0.127, and $w(t)$ is the Gaussian white noise. Figure 2 depicts the bump road input for the car driving at constant speed 40 km/h and random road input at 60 km/h, respectively.

Table 1. Parameters selected for numerical simulations.

Parameters	Values
Sprung mass m_s	320 kg
Unsprung mass m_u	45 kg
Inertance b	5 kg
Main spring stiffness k_1	14 kN/m
Tire stiffness k_t	190 kN/m
Inherent damping coefficient c	100 Ns/m
Electromechanical coupling K_e	100 Vs/m

To evaluate the ride comfort and stability, vehicle body acceleration and dynamic tire load are selected as performance indicators. Comparing the peak values in Fig. 3, a considerable improvement can be observed when using the existing MPC approach, with a reduction of 66.5% in vehicle body acceleration and 53.7% in dynamic tire load, compared to the passive case. However, the utilization of the MPC with road preview yields even further reductions of 49.5% and 53.2%. Figure 4 gives the peak values of the evaluating performance at different speeds. In Fig. 4, it is found that for the vehicle body acceleration and dynamic tire load, the peak values of the proposed MPC with road preview are decreased by 21.5% to 49.5%, compared to the MPC without preview. Most notably, as illustrated in Fig. 4, the peak values of the suspension with the MPC with preview remained consistently low across the broad speed range. In fact, when contrasted with the MPC without preview, the peak values are significantly decreased by 47.8% to 83.1%.

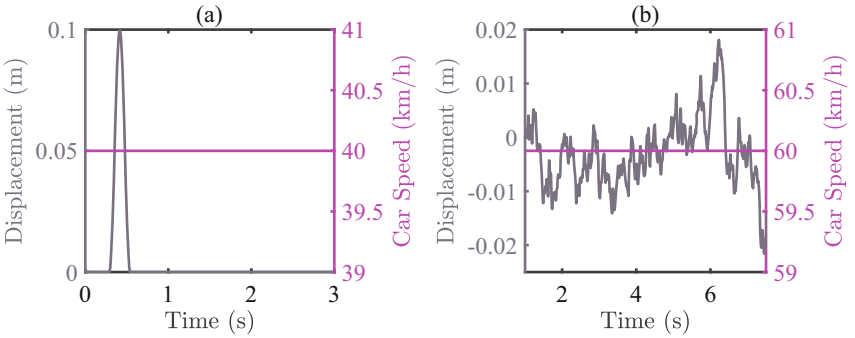


Fig. 2. Road profiles for simulations: (a) Bump road at constant car speed 40 km/h, (b) Random road at constant car speed 60 km/h.

Comparisons involving the random road profile yield the similar outcomes. Given the nature of the random road, root mean square (RMS) values are considered more appropriate measures for comparison. As shown in Fig. 5, the RMS values of the vehicle body acceleration and dynamic tire load of the suspension utilizing the MPC with preview exhibit substantial reductions, with decreases of 37.5% and 33.9%, respectively, in comparison to the MPC without preview. Figure 6 presents the comparisons of the RMS values at different car speeds. Similar to the results from the bump road, Fig. 6 illustrates that the RMS values of the vehicle body acceleration are impressively decreased by 57.8% to 58.4% compared to the passive control, and 21.5% to 49.5% compared to the MPC without preview. However, it is worth noting that Fig. 6 indicates that the existing MPC without preview does not have a significant advantage over the passive control, especially at lower car speeds, in terms of the dynamic tire load. In contrast, the proposed approach, the MPC with road preview, consistently delivers much better results. When compared to both passive control and MPC without preview, the RMS values of the

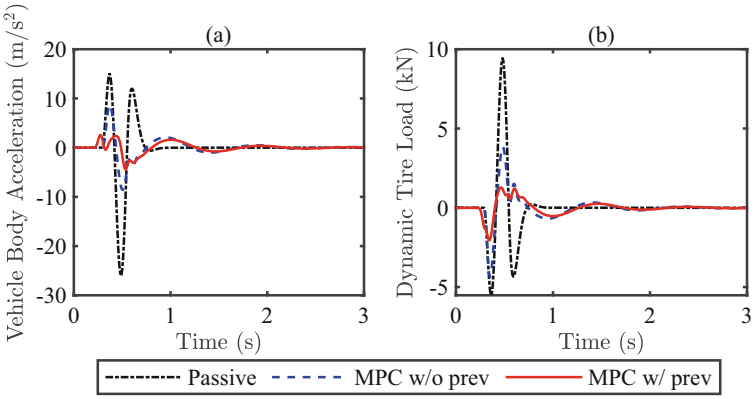


Fig. 3. Response time histories of (a) the vehicle body acceleration and (b) dynamic tire load on the bump road at car speed 40 km/h.

MPC with road preview demonstrate a steady decrease ranging from 34.4% to 35.7%, highlighting its reliability and effectiveness.

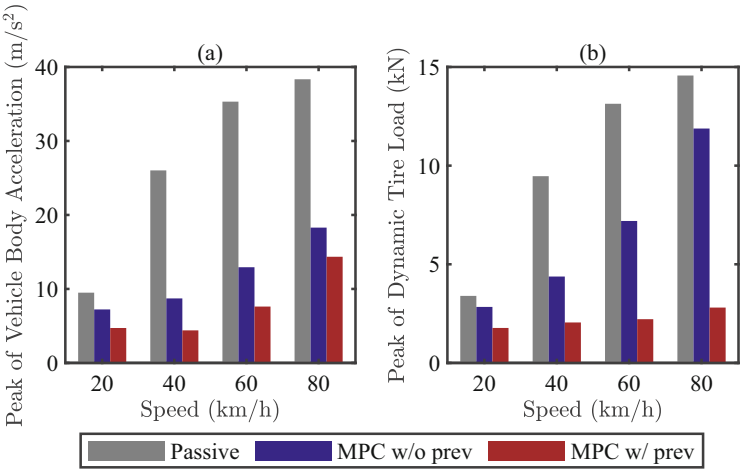


Fig. 4. Peak values of (a) vehicle body acceleration and (b) dynamic tire load under different speeds on the bump road.

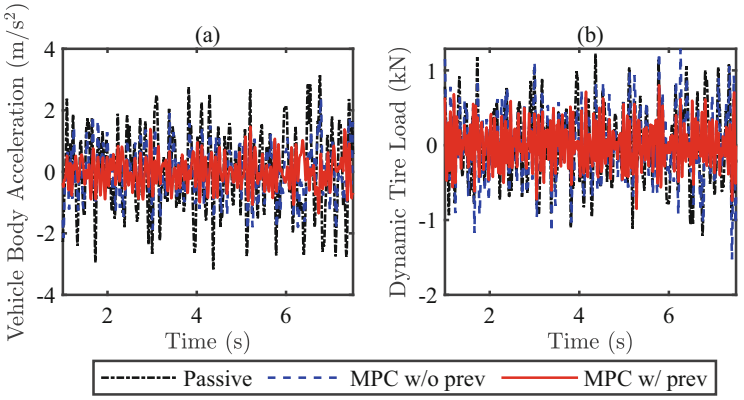


Fig. 5. Response time histories of (a) the vehicle body acceleration and (b) dynamic tire load on the random road at car speed 60 km/h.

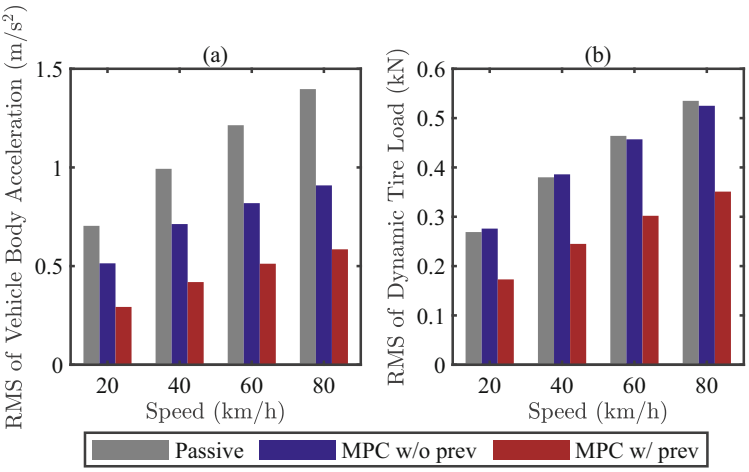


Fig. 6. RMS values of (a) vehicle body acceleration and (b) dynamic tire load under different speeds on the random road.

4 Conclusions

This paper proposed an improved control scheme using MPC with road preview method for an active inerter-based suspension. The simulation results demonstrated the superiority of the proposed design over other two approaches: for both bump and random road profiles, the MPC with road preview can effectively enhance the damping performance of the suspension system. For future work, the control parameters will be further optimized and numerical analysis in the frequency domain will also be carried out. Furthermore, to evaluate more performance indicators such as the lateral stability, it is also of great interest to study the full-car suspension model.

References

1. Smith, M.C., Wang, F.C.: Performance benefits in passive vehicle suspensions employing inerters. Veh. Syst. Dyn. **42**(4), 235–257 (2004)
2. Hu, Y., Chen, M.Z., Shu, Z.: Passive vehicle suspensions employing inerters with multiple performance requirements. J. Sound Vibr. **333**(8), 2212–2225 (2014)
3. Daraghmeh, A., Qatanani, N.: Numerical error bound of optimal control for homogeneous linear systems. Arch. Control Sci. **29**(2), 323–337 (2019)
4. Wang, Y., Li, H.X., Meng, H.D., Wang, Y.: Dynamic characteristics of underframe semi-active inerter-based suspended device for high-speed train based on LQR control. Bull. Pol. Acad. Sci. Tech. Sci. **70**(4), 141722 (2022)
5. Canale, M., Milanese, M., Novara, C.: Semi-active suspension control using "fast" model-predictive techniques. IEEE Trans. Control Syst. Technol. **14**(6), 1034–1046 (2006)
6. Göhrle, C., Schindler, A., Wagner, A., Sawodny, O.: Model predictive control of semi-active and active suspension systems with available road preview. In: 2013 European Control Conference (ECC), pp. 1499–1504. IEEE, Switzerland (2013)

7. Haraguchi, R., Asai, T.: Enhanced power absorption of a point absorber wave energy converter using a tuned inertial mass. Energy **202**, 117740 (2020)
8. Theunissen, J., et al.: Regionless explicit model predictive control of active suspension systems with preview. IEEE Trans. Industr. Electron. **67**(6), 4877–4888 (2019)

Experimental Study of Vibration Control Using Autoparametric Vibration Absorber

Yuma Kawasumi[1(✉)], Kenji Yasui[1], and Hiroshi Yabuno[2]

[1] Okumura Corporation Technical Research Institute, 387 Osuna, Tsukuba City, Ibaraki Prefecture, Japan
yuma.kamagatani@okumuragumi.jp

[2] University of Tsukuba, 1-1-1 Tennodai, Tsukuba City 305-8577, Ibaraki Prefecture, Japan

Abstract. A tuned mass damper, also known as a dynamic damper, is often used for passive vibration control of a structure. The damper incorporates an additional mass that is tuned to the natural frequency of the structure in question. This type of vibration control device has a significant damping effect at the natural frequency of the structure but is subject to the resonance of the new vibration system created by the additional mass. With the goal of eliminating this unwanted resonance, we have been pursuing research on a new vibration control device using parametric excitation characteristics. We developed the vibration control device consisting of a pendulum connected to a torsion bar spring and confirmed its effect through vibration experiments and response analysis.

Keywords: parametric excitation · tuned mass damper · free vibration experiment · excitation experiment · response analysis

1 Introduction

Tuned mass dampers are used in various applications to reduce structural vibration [1, 2]. The damper controls the vibration of a structure by identifying the natural frequency of the structure, tuning an additional mass to that natural frequency, and allowing the mass to oscillate largely to absorb the vibration energy. Examples of a one-degree-of-freedom system to be added include the springs and weights used to control the vertical vibration of large slabs [3] and the pendulums used to reduce the wind-induced vibration of high-rise buildings [4–6]. Figure 1 shows, as the characteristics of a tuned mass damper, the relationship between the ratio of the external excitation frequency p (frequency ratio) to the natural frequency of the main system ω_m and the response magnification factor obtained.

Shibata et al. have undertaken research on a vibration control device based on parametric excitation characteristics with the goal of preventing this unwanted resonance [7]. Parametric excitation is an oscillatory phenomenon caused by cyclic changes in the parameters of a vibration system which are normally constants [8]. For example, the mass, spring constant, and other parameters of a spring-mass system that are normally not vary periodically.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 426–437, 2025.
https://doi.org/10.1007/978-3-031-72794-8_34

We thus pursued research on a new vibration control device using such parametric excitation characteristics. In this study, we developed a vibration control device incorporating the theory of parametric excitation and performed vibration experiments with the device installed on a simplified structure model. We created a model by joining the vibration control device to the vibration model and subjected it to response analysis.

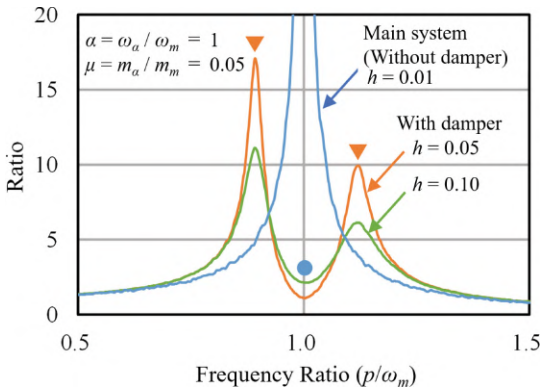


Fig. 1. Response magnification factor of the main system with a tuned mass damper attached. The natural frequency ratio $a = 1$, mass ratio $\mu = 0.05$, and damping constants $h = 0.05$ and $h = 0.10$ are added to a vibration model with a damping constant $h = 0.01$ (the main system). The chart indicates a significant damping effect (●) at the natural frequency of the main system (ω_m), but also indicates two peaks of the unwanted resonance (▼) of the newly created two-mass spring system due to the addition of the single-mass spring system.

2 Experimental Setup and Vibration Control Device Based on Parametric Excitation

2.1 Dynamic Model of Device and Main System

Figure 2 gives an overview of the vibration control device based on parametric excitation.

This device is a vertically vibrating main system with pendulum attached. The natural frequency of the pendulum is tuned to 1/2 (f) of the natural frequency of the vertical vibration of the main system ($2f$). The vertical vibration of the main system is equivalent to the fluctuation in the gravitational acceleration (g) acting on the pendulum with twice the natural frequency of the pendulum, and it excites the pendulum due to the parametric excitation characteristics [9].

As the pendulum completes one horizontal oscillation, the main system reciprocates twice vertically. This means that the vibration of the main system can be reduced by the large oscillation of the pendulum tuned to the vertical natural frequency of the main system.

Since the natural frequency of the pendulum is half that of the main system and their vibration directions are perpendicular to each other, the pendulum will not generate

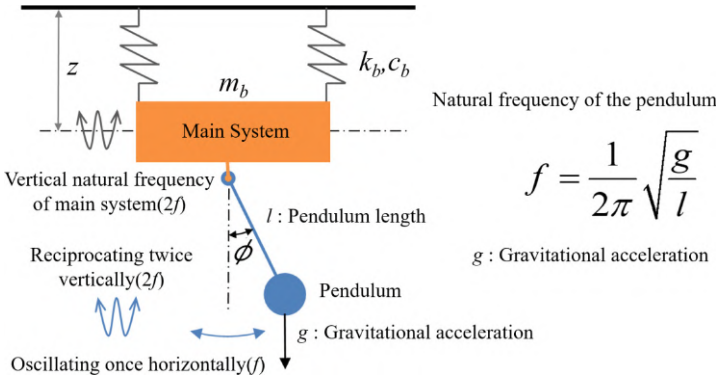


Fig. 2. Main system where this device is installed

vibrational coupling; nor will it create a new vibration system. The pendulum operates only within the frequency band in which the pendulum is excited, and the greater the amplitude of the pendulum, the larger the reduction of the response. This effect results even if the pendulum mass is small; in fact, for damping, the smaller the better. This major feature of this vibration control device distinguishes it from conventional tuned mass dampers.

However, the vertical vibration of the main system will not act in the direction required to generate oscillation of the pendulum because the pendulum is initially positioned right under the main system. To cause the pendulum to oscillate, an external force must be applied.

2.2 Experimental Setup

To confirm the effect of the vibration control device, we performed an excitation experiment by preparing a steel simple beam supported on both ends by small pins (which had a small effect on damping) to increase the vibration of the main system, mounting the vibration control device at the center of the beam, and deploying a piezo actuator. The piezo actuator and the exciting mass were mounted 70 mm from the center, as close to the beam center as possible.

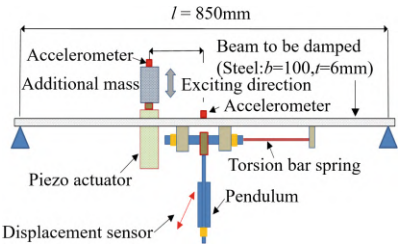


Fig. 3. Overview of the experimental setup

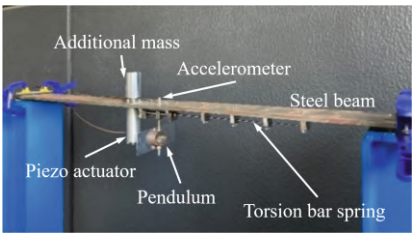


Fig. 4. Photo1 Appearance of experimental setup

2.3 Vibration Control Device

Figure 3 and Photo 1 present the overview of the experimental setup.

We coupled a torsion bar spring to the pendulum to allow the natural frequency of the vibration control device to be selected over a wide range. The natural frequency of the vibration control device is determined by the pendulum length, pendulum mass, and the torsional rigidity of the torsion bar spring. To tune the natural frequency of the vibration control device to half that of the beam, we made the pendulum length adjustable.

The mass of the pendulum, the movable part of the vibration control device, was 0.065 kg, and the total mass of the beam to be damped and the additional mass was approximately 4.9 kg. The mass ratio was 1/75. Table 1 gives the specifications of the vibration control device and Table 2 gives the mass of each elements of the experimental setup.

Table 1. Specifications of the vibration control device

Name	Item	Unit	Measurement
Torsion bar spring	Diameter	m	0.002
	Length	m	0.250
	Torsional rigidity	N•m^2	0.462
Pendulum moving part	Moment of inertia	kg•m^2	7.726×10^{-5}
	Pendulum length	m	0.025-0.060
	Mass	kg	0.065

Table 2. Mass of each elements of the experimental setup

Name	Mass (kg)
Beam	3.978
Piezo actuator	0.205
Additional mass	0.395
Device (other than moving parts)	0.224
Device moving parts	0.065

2.4 The Piezo Actuator

To excite the experimental setup, we connected the 0.395 kg additional mass with the piezo actuator as an exciter along the vertical axis. The piezo actuator excites vibration with controlled displacement. It excites the beam vertically with a sine wave of the desired frequency and displacement using the function generator.

The controllable response displacement ranges from 0 μm to 80 μm, based on a control signal between 0 and 5 V. We applied an offset voltage of 2.5 V and performed excitation at ±2.0 V with a sine wave.

2.5 Measurement Conditions

We installed two accelerometers at the center of the beam and on the additional mass connected to the piezo actuator to confirm the damping effect, and installed a displacement sensor along the horizontal axis of the pendulum to determine the movement of the vibration control device. To measure acceleration, we used a charge mode accelerometer (mass: 2 g, sensitivity: 0.3 pC/mm/s^2, frequency range: *fc* to 10 kHz, max. acceleration: 50,000 m/s^2). For displacement, we used a non-contact thru-beam laser displacement sensor (measurement width: 30 mm; resolution: 15 μm).

The swing angle of the pendulum is defined as horizontal displacement divided by the pendulum length within the range in which $sin\theta \approx \theta$ holds true.

Measurements were continued for more than 100 s to achieve a frequency resolution of 1/100 Hz. The sampling frequency was 655.36 Hz; the number of measurement points was 65,536.

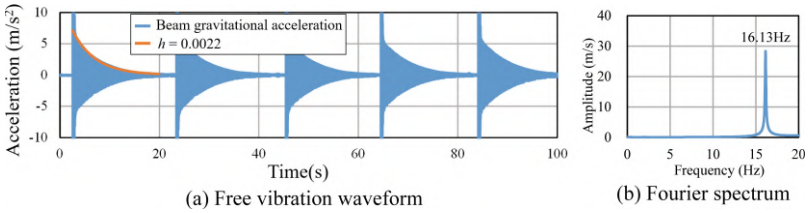
3 Vibration Experiment Result

3.1 Tuning by Free Vibration Experiment

We performed a free vibration experiment to determine the natural frequency of the beam to be damped along the vertical axis. We struck the center of the beam with an impact hammer while the pendulum was fixed and measured the acceleration. A piezo actuator and an additional mass as well as the vibration control device were attached to the beam. During the 100-s measurement, we struck the beam five times.

To tune the natural frequency of the pendulum to half that of the beam, we performed a free vibration experiment with the pendulum (Fig. 4). After moving and releasing the pendulum, we measured its displacement. Figure 5(a) to (d) show the results of the experiment conducted after the pendulum length was adjusted to tune the frequency. The damping constants of the pendulum and the beam were 0.014 and 0.0024, respectively. The natural frequencies of the beam and pendulum were 16.13 Hz and 8.06 Hz, respectively, which confirmed that the natural frequency of the pendulum was tuned to about half that of the beam.

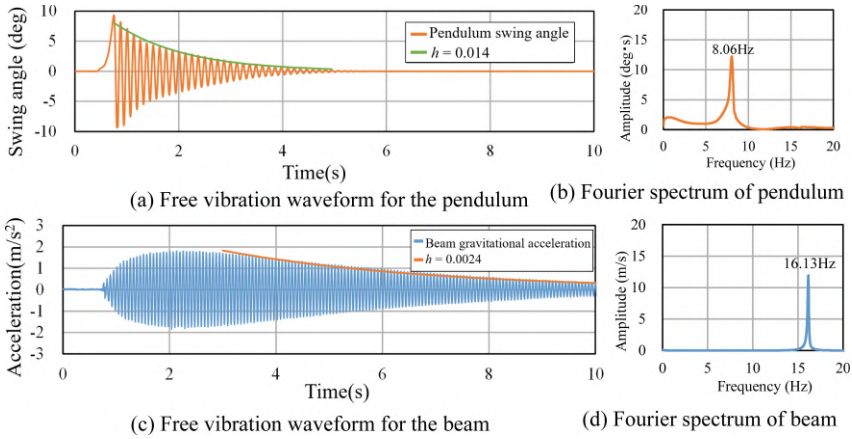
Based on the free vibration experiment results with the pendulum, we evaluated the damping constants of the pendulum and the beam using the logarithmic decrement method.



(a) Free vibration waveform (b) Fourier spectrum

Fig. 5. Free vibration experimental results for the beam

The beam vibrated for more than 10 s each time it was struck. The natural frequency was 16.13 Hz. From the first free vibration waveform, we evaluated the damping constant using the logarithmic decrement method. Figure 4 also shows the fitting result. The damping constant h was 0.0022.



(a) Free vibration waveform for the pendulum (b) Fourier spectrum of pendulum

(c) Free vibration waveform for the beam (d) Fourier spectrum of beam

Fig. 6. Free vibration experiment results for the pendulum

When we moved the pendulum and allowed it to oscillate freely, the resulting oscillation acted on the beam as an excitation force, exciting vibration in the beam. After the pendulum began to oscillate, the vibration of the beam gradually increased. As the pendulum oscillation weakened, the beam vibration decreased. The pendulum oscillation ceased in approximately five seconds, whereas the beam vibration continued for more than 10 s.

3.2 Examination of Damping Effect by Excitation Experiment

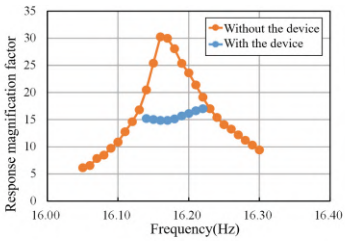
To confirm the damping effects of the vibration control device, we performed an excitation experiment using a piezo actuator. Figure 6 shows the results of the excitation experiment.

Considering the natural frequency of the beam obtained by the free vibration experiment, we performed excitation in 0.01 Hz increments from 16.05 Hz to 16.30 Hz.

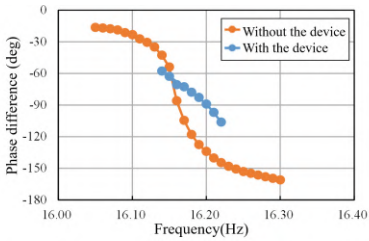
When at a complete standstill, this device cannot be activated by vertical vibration alone, therefore we applied a slight external force to the pendulum.

To evaluate the damping effect, the difference between the response acceleration of the beam and the response acceleration of the additional mass of the piezoelectric actuator was used as the input acceleration to the beam, and the response magnification factor and phase difference of the response acceleration of the beam to the input acceleration were calculated. For the calculations, we applied the Fourier spectrum of the acceleration waveform at each frequency.

We conducted experiments at several frequencies and confirmed that two peaks of the unwanted resonance did not occur. This device produces its desired effect through resonance, which only comes about in the natural frequency band. This means it will not generate adverse effects, such as the response of the main system with the vibration control device exceeding the response without the device.



(a) Excitation frequency and response magnification factor



(b) Excitation frequency and phase difference between response acceleration and input acceleration

Fig. 7. The results of the excitation experiment: (a) The response magnification factor; (b) The phases difference both with and without the device

Without the device, the response magnification factor was greatest at 30.5 at 16.16 Hz. The response magnification factor is not symmetrical with respect to the natural frequency. The resonance curve exhibits weak nonlinearity, rising steeply at frequencies lower than the natural frequency and sloping gently at higher frequencies. With the device, the response magnification factor at the natural frequency is approximately 15, almost half that without damping. The response magnification factor is smallest at the natural frequency, increasing as the excitation frequency diverges away from the natural frequency.

3.3 Damping Effect at Resonance Peaks

We examined the damping effect for the resonant frequency of 16.16 Hz while varying the input signal level to the piezo actuator ($\pm$1.75 V [0.24 m/s^2], $\pm$ 2.0 V [0.29 m/s^2] and $\pm$ 2.25 V [0.34 m/s^2]). By changing the input signal based on $\pm$2.0 V, we confirmed that the attenuation effect was independent of the input signal. The results are shown in Fig. 7 and Fig. 8.

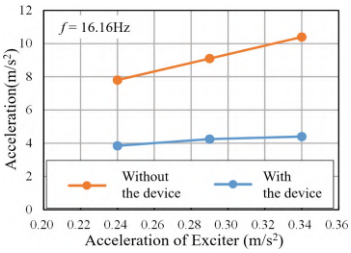


Fig. 8. The relationship between the input acceleration and the response acceleration of the beam

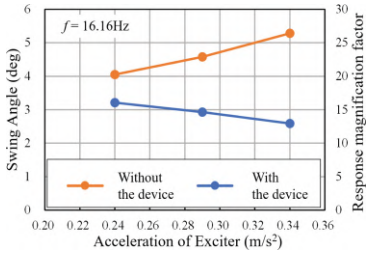


Fig. 9. The relationships between the input acceleration and the swing angle of the pendulum and between the input acceleration and the response magnification factor with the device

Without the device, the response acceleration increases as the acceleration of the additional mass increases. With the device, an increase in the response acceleration is much smaller than that in the input acceleration.

As the input acceleration increases, the swing angle of the pendulum increases, producing a greater damping effect and reducing the response magnification factor of the beam (main system).

4 Modelling of Vibration Experiment

4.1 Modelling of Beam

This section examines the vibration experiment result using response analysis. We modeled the beam used for the vibration experiment as shown in Fig. 9. The eigenvalue analysis of this model showed that the first natural frequency was 16.20 Hz, which almost equaled the natural frequency obtained from the resonance curve (16.16 Hz).

Next, we represented the beam by an equivalent spring-mass model to simplify the response analysis.

The kinetic energy of the beam T is expressed by Eq. (1), where the equivalent mass of the beam is m_b and the vertical displacement at the center is z.

$$T = \frac{1}{2} m_b \dot{z}^2 \tag{1}$$

As shown in Fig. 10, when a simple beam with a uniform cross section vibrates, the deformation is concentrated in the center of the beam. Therefore, static vertical displacement u at a distance x from the sup-porting point is expressed by Eq. (2).

$$u = z\left(\frac{3x}{l} - \frac{4x^3}{l^3}\right) 0 \le x \le \frac{l}{2} \tag{2}$$

The kinetic energy of the beam T is expressed by Eq. (3), where the cross-sectional area of the beam is A, and the density is ρ.

$$\begin{aligned} T &= \frac{1}{2}(m_p + m_d)\dot{z}^2 + \frac{1}{2}\int_0^{l/2} 2\rho A\dot{u}^2 dx \\ &= \frac{1}{2}(m_p + m_d + \frac{17}{35}\rho Al)\dot{z}^2 \end{aligned} \tag{3}$$

Based on Eqs. (1) and (3), the equivalent mass of the beam m_b is obtained with the Eq. (4).

$$m_b = m_p + m_d + \frac{17}{35}\rho Al \tag{4}$$

The equivalent mass of the beam calculated using Eq. (4) is 2.828 kg. Since $k_b = 48EI/l^3$, the spring constant of the simple beam is 2.854×10^4 kgs^{-2}. This gives a natural frequency for the equivalent spring-mass model of 15.99 Hz. This is slightly less than the natural frequency obtained in the experiment. We believe the spring constant was greater than that given by the analysis model because the pin supporting the beam of the experimental setup is firmer than the ideal pin assumed in the analysis model.

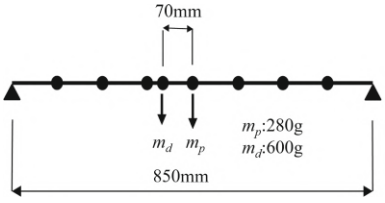


Fig. 10. Beam model for vibration experiment

We divided the beam into nine elements considering bending and shear and assumed that both ends were super acted on the mass point ported by pins. The mass of the vibration control device m_p and the mass of the piezo actuator and additional mass m_d.

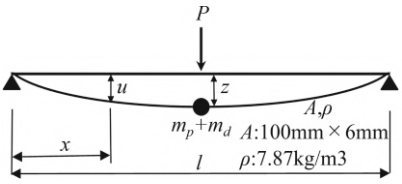


Fig. 11. How to calculate equivalent mass

The input signal y^* and output acceleration to the piezo actuator used as an exciter will be modeled as a second-order lag system as shown below, and will be incorporated into the equation of motion. Time constants of piezo actuator are as shown in Table 3.

$$y^* = T_1\ddot{y} + T_2\dot{y} + y \tag{5}$$

Table 3. Time constants of piezo actuator

Name	Time constant
T_1	2.0×10^{-6}
T_2	5.0×10^{-4}

5 Equations of Motion of Vibration Control Device

The beam and the vibration control device attached to the beam are modeled as a mass spring system. The Eqs. (6) to (8) show the equations of motion.

$$\begin{aligned}&(m_b + m_p + m_d)\ddot{z} + c_b\dot{z} + k_b z \\ &+m_p l\ddot{\phi}\sin\phi + m_p\dot{\phi}^2\cos\phi \\ &-\frac{m_d T_2}{T_1}\dot{y} - \frac{m_d}{T_1}y + \frac{m_d}{T_1}y^* = 0\end{aligned} \tag{6}$$

$$I_p\ddot{\phi} + m_p l\ddot{z}\sin\phi + c_p\dot{\phi} + m_p gl\sin\phi + k_\phi\phi = 0 \tag{7}$$

$$\ddot{y} + \frac{T_2}{T_1}\dot{y} + \frac{1}{T_1}y - \frac{1}{T_1}y^* = 0 \tag{8}$$

Equation (6) is the equation of motion along the vertical axis of the beam modeled as a mass point. The Eq. (7) is in the rotational direction of the pendulum of the vibration control device. The Eq. (8) is along the vertical axis of the piezo actuator.

When $m_p l\ddot{\phi}\sin\phi$ of the Eq. (6) and $m_p l\ddot{z}\sin\phi$ of the Eq. (7) interact with each other and the vertical vibration of the beam is input to the pendulum of the vibration control device, the vibration of the beam is decreased.

In response analysis, using the Newmark-beta method for integral calculus, the state quantities are obtained with the Newton-Raphson method for every integral step. The state quantities will be $x = (z, \phi, y, \dot{z}, \dot{\phi}, \dot{y}, \ddot{z}, \ddot{\phi}, \ddot{y})$ (Fig. 11).

6 Response Analysis Result

Figure 12 shows the relationship between the frequency ratio to natural frequency and the response magnification factor of both the analysis result and the experiment result.

The response magnification factor at the natural frequency without device is approximately 30, almost equal to the experiment result, which suggests that the beam is adequately modeled. The response magnification factor with device is around 15 in the operating range of the pendulum and thus the model almost reproduces the experiment result including the damper operating range. This indicates that the vibration control device is appropriately modeled.

However, the analysis result shows almost symmetrical curves about the natural frequency. To represent the weak nonlinearity exhibited by the experiment result, nonlinearity must be considered in modeling (Fig. 13).

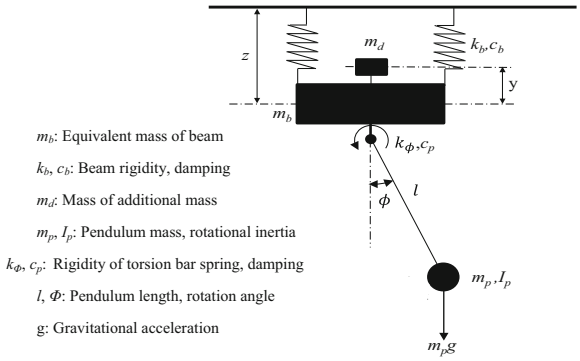


Fig. 12. Modelling of beam and vibration control device

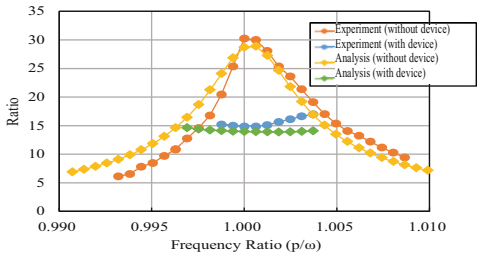


Fig. 13. The relationship between the frequency ratio to natural frequency and the response magnification factor of both the analysis result and the experiment result

7 Conclusions and Afterword

We confirmed the damping effect of a newly developed vibration control device based on the theory of parametric excitation. We have confirmed the following through vibration model experiments using the vibration control device consisting of a pendulum connected with a torsion bar spring and the response analysis.

i. The vibration control device with 1/75 of the total mass reduces the response of the beam by approximately 1/2 of that without newly generating resonance.
ii. A response analysis using appropriate models of the beam and the vibration control device can evaluate the damping effect with high accuracy.

This device, despite the extremely small mass ratio, produces a large damping effect, which increases as the amplitude increases, without causing a parasitic resonance response. We will continue to improve this device with the goal of making it suitable for practical applications.

References

1. Rana, R., Soong, T.T.: Parametric study and simplified design of tuned mass dampers, design of tuned mass dampers. Eng. Struct. **20**(3), 193–204 (1998)

2. Soto, M.G., Adeli, H.: Tuned mass dampers. Arch. Comput. Methods Eng. **20**, 419–431 (2013)
3. Yaginuma, K., Inadome, K.: Development of tuned mass damper for slab vibration. Okumura Corporat. Tech. Res. Annu. Rep. **34**, 141–146 (2008). in Japan
4. Yasui, K., Kawai, N., Ohtsuka, S.: Verification test and response observation of a building with tuned roller pendulum damper system. In: 2nd World Conference on Structural Control, Kyoto (1998)
5. Carlot, J.: Effects of a tuned mass damper on wind-induced motions in tall buildings. Massachusetts Inst. Technol. Archiv. (2012)
6. Webster, A.C., Vaicaitis, R.: Application of tuned mass dampers to control vibrations of composite floor systems. Engineering **1**, 116–124 (1992). AISC
7. Shibata, A., Ohishi, S., Yabuno, H.: Passive method for controlling the nonlinear characteristics in a parametrically excited hinged-hinged beam by the addition of a linear spring. J. Sound Vib. **350**, 111–122 (2015). in Japan
8. Haxton, R.S., Barr, A.D.S.: The autoparametric vibration absorber. J. Eng. Ind. **94**(1), 119–125 (1972)
9. Brzeskia, P., Perlikowskia, P., Yanchukb, S., Kapitaniaka, T.: The dynamics of the pendulum suspended on the forced Duffing oscillator. J. Sound Vib. (2012)

Nonlinear Dynamics in Practical Systems

Numerical Bifurcation Analysis of a Two-DOF Shaft-System Coupled by a Universal Joint with Clearance

Junaid Ali(✉), Gregory Shaver, and Anil K. Bajaj

Ray W. Herrick Laboratories, School of Mechanical Engineering, Purdue University, West Lafayette, IN 47907, USA
ali181@purdue.edu

Abstract. The presence of clearances is inevitable in practical mechanisms. Components are designed to maintain optimal clearances between mating parts, enabling relative motion. There is extensive literature available on U-Joints from manufacturing tolerances perspective. In this work the emphasis is on clearances specifically existing between a yoke and trunnion, which affects the input/output motion characteristics. More specifically, the focus is on bifurcation analysis of systems with clearances between the yoke and crosspiece. To the best of authors knowledge literature is limited in this area. So, this effort will shed light on non-linear characteristics of motions in U-Joints. Clearances in U-Joint are categorized into axial and rotational type. Here we chose to explore the bifurcation characteristics associated with pure rotational type clearances and the effect of U-Joint inclination angle on the dynamical behavior of clearance between the yoke and the crosspiece of U-Joint. Several period-doubling route to chaos within a range of clearance sizes are observed. Other interesting bifurcations such as interior, boundary and basin-boundary crisis are also observed as a function of clearances.

Keywords: Universal joint · Crisis · Clearances · Period-doubling · Rotating Systems

1 Introduction

Universal joints (U-joints) play a crucial role in the transmission of torque and motion between non-collinear shafts in a driveline. Due to their inherent kinematics, U-joints introduce non-linearity into the system dynamics. The driveshafts interconnected with U-Joints at non-zero operating angles produce second-order torsional accelerations due to inherent non-linearities. Parametric instabilities caused by operating angle and phasing of the U-Joints have been studied in the literature. Asokanthan & Meehan [1] obtained conditions for parametric instability in a 2-DOF shaft system using the method of averaging and computationally demonstrated the quasi-periodic route to chaos for the full non-linear model utilizing maximal Lyapunov exponent.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 441–456, 2025.
https://doi.org/10.1007/978-3-031-72794-8_35

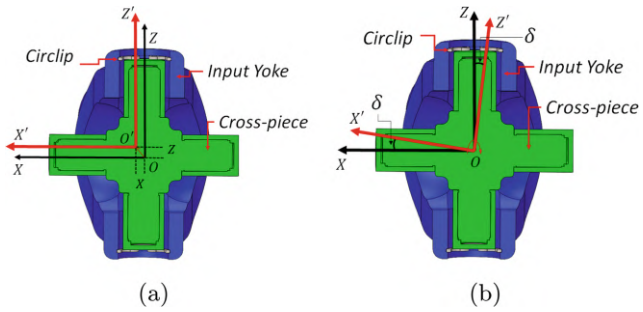


Fig. 1. Exaggerated 3-D representation of (a) Axial clearance. True center of rotation of trunion is O and shifted axis of rotation is O' where (X, Z) represents the units shifted in XZ-plane, (b) Rotational clearance (δ) in U-Joint.

Bulut and Parlar [2] studied the linear time-dependent dynamics of a 2-DOF shaft system interconnected with a U-Joint using the monodromy matrix method. Strutt diagrams were generated utilizing the Floquet theory to determine regions of parametric instabilities. Combination-type instabilities were observed first using this approach. SoltanRezaee et al. [3] performed a similar study but on a multi-U-Joint driveline using the monodromy matrix method and identified regions of sub-harmonic and combination type resonances. They concluded that U-Joint angle at the input end has a significant effect on the driveline torsional signature. Also, combination-type instabilities were found to be present for small joint angles as well which was not the case with single joint driveline. Most other literature has mainly focused on developing equations of motion and studying the parametric dynamical behavior of U-Joints without clearances. However, in practical mechanical systems, clearances between components can arise not only from manufacturing tolerances but also from wear and tear, such as brinelling and spalling. These clearances pose additional challenges in the dynamics of mechanical systems.

Clearances are inherent aspects of mechanical components due to limitations in manufacturing. Components are designed to maintain optimal clearances between mating parts, enabling relative motion. In the design of universal joints, clearances are intentionally incorporated to facilitate the relative motion.

Hummel and Chassapis [4] categorized these clearances in universal joints in two distinct ways. Firstly, there is clearance allowing bi-directional axial motion of the cross-piece concerning the yokes along the axis of cross-piece rotation, assuming no axial motion of the cross-piece along the input yoke's axis of rotation (see Fig. 1(a)). Secondly, there is rotational clearance, which pertains to the gap between the bearing cap and the yoke. The pure rotational clearance results in lead/lag behavior in the input-output relationship of the coupling (see Fig. 1(b)). Existing literature on clearances in U-Joints focuses on the design optimization of the joint from a manufacturing perspective. Relationships to prevent binding between the mechanism's components are established for ideal joints and joints

with clearances [5]. The primary objective of these studies was to propose design improvements in the manufacturing processes of universal joints to prevent yoke interference during motion, without delving into the dynamic analysis of U-joints with clearances.

In present study, clearances in a single universal joint are modeled as a three-body problem, involving the input yoke, trunnion, and output yoke, with additional four degrees of freedom. Among these four degrees of freedom, two account for the bi-axial motion of the trunnion relative to both yokes, while the other two describe the rotational clearance effect relative to the input and output yokes. Investigating the dynamic behavior of the trunnion in the presence of these clearances when it is perturbed from ideal input-output flow patterns becomes an intriguing prospect.

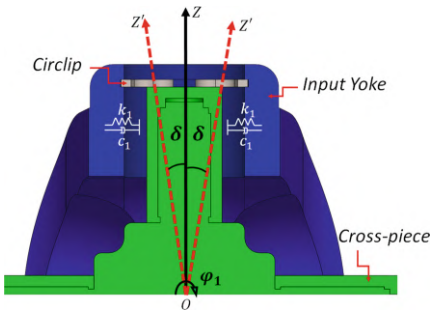


Fig. 2. Cross-section view of input yoke and rigid cross-piece with rotational clearance (δ).

There is currently no research that examines the dynamic characteristics of these axial and rotational clearances for U-Joints. To begin with, we choose to focus on studying effects of pure rotational clearances first assuming that axial clearances are absent. Additionally, we assume that rotational clearances only exist between the input yoke and the cross-piece, simplifying the relationship between the input and output of the U-joint. The contact between the bearing cap and the yoke for the rotational clearances (see Fig. 2) can be modeled as linear piece-wise spring force as a function of input angular displacement [6,7]. This assumption holds true when clearances exist between the input yoke and the trunnion, and the inertia of the trunnion is assumed negligible compared to the inertia of the input and output shafts [8]. Furthermore, in scenarios where clearances are not the primary focus of the study, these simplifications are invaluable for conducting comprehensive vehicle-level driveline simulations, where universal joints are just one component among many in complex assemblies. The paper is structured into several sections, with Sect. 1 serving as the Introduction, where the modeling assumptions are made. Section 2 is devoted to Mathematical Modeling, Sect. 3 covers Numerical Integration, and Sect. 4 is dedicated to the examination of non-linear bifurcation analysis. The final section

summarizes the study's findings and closes with a discussion of potential avenues for future research.

2 Mathematical Modeling

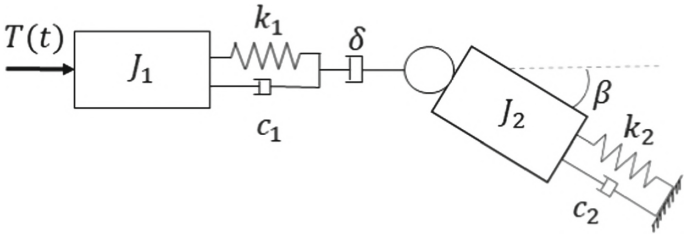


Fig. 3. Schematic of a two degree-of-freedom shaft system interconnected with a U-joint having rotational clearance δ between input yoke and rigid cross-piece.

Consider a shaft system interconnected through a U-Joint misaligned with angle β (see Fig. 3). The kinematic relationship (see Eq. 1) between input and output can be written as [9],

$$\varphi_1 = \arctan(\tan(\varphi_2) \cdot \cos(\beta)) \tag{1}$$

where φ_1 and φ_2 are the input and the output angular displacements of the U-Joint. The presence of a trunnion in a universal joint introduces a holonomic constraint on the system's dynamic behavior. This constraint imposes restrictions on the permissible motions of the joint. When one yoke undergoes rotation, the motion of the other yoke becomes complex due to the imposed constraint. Equation 2 captures the relationship and limitations imposed on the system's motion,

$$\dot{\varphi}_1 = \frac{\cos(\beta)}{1 - (\sin(\beta))^2 \cdot (\sin(\theta_2))^2} \dot{\varphi}_2 \equiv \eta(t)\dot{\varphi}_2 \tag{2}$$

where $\dot{\varphi}_2 = \dot{\psi}_2 = \dot{\theta}_2$ and $\dot{\psi}_1 = \dot{\varphi}_1 + \dot{\theta}_1$. Here, ψ_1 and ψ_2 are the torsional displacements of the inertia J_1 and J_2 of the input and output shaft. Let the torsional stiffnesses of the input and output shaft be k_1 and k_2. The input and output shaft damping coefficients be c_1 and c_2. Let θ_1 and θ_2 be the relative torsional displacements; their first time-derivatives are the relative angular velocities of springs k_1 and k_2. The clearance between the input yoke and the output yoke are represented by δ (see Fig. 1 (b)). The system is driven by a sinusoidal torque input $T(t) = T_0 \sin(\Omega t)$ with forcing frequency Ω on one end whilst the other end is fixed. Newton-Euler approach is used to derive the equations of motion for the system under investigation. For the free body diagram (FBD) of the system, see Fig. 4, which illustrates the forces and moments acting on the shafts

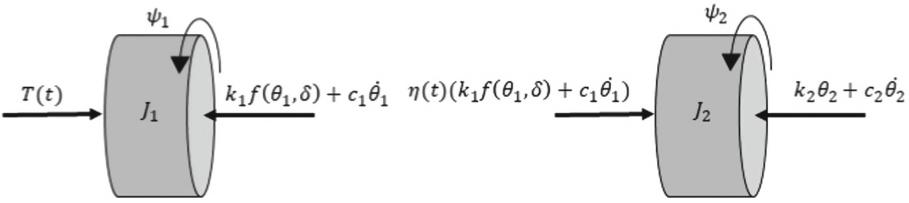


Fig. 4. Free body diagram of the dynamic system.

and U-Joint. The equations of motion of the system are then given by equations below:

$$J_1(\ddot{\theta}_1(t) + \eta(t)\ddot{\theta}_2(t) + \dot{\eta}(t)\dot{\theta}_2(t)) + k_1 f(\theta_1, \delta) + c_1\dot{\theta}_1(t) = T(t) \tag{3}$$

$$J_2(\ddot{\theta}_2(t)) - \eta(t)(k_1 f(\theta_1, \delta) + c_1\dot{\theta}_1(t)) + k_2\theta_2(t) + c_2\dot{\theta}_2(t) = 0 \tag{4}$$

The non-linearity resulting from the clearance is modeled as a piece-wise function $f(\theta_1, \delta)$ defined below. This means that the behavior here of the system changes depending on the position θ_1 of the clearance δ within the range of motion [6,7].

$$f(\theta_1, \delta) = \begin{cases} \theta_1 - \delta & \text{if } \theta_1 > \delta \\ 0 & \text{if } -\delta \leq \theta_1 \leq \delta \\ \theta_1 + \delta & \text{if } \theta_1 < -\delta \end{cases} \tag{5}$$

3 Numerical Simulation

Table 1. Numerical simulation parameters for the system in Fig. 4.

Parameter	Value
Input shaft inertia $J_1(kg.m^2)$	0.001
Output shaft inertia $J_2(kg.m^2)$	0.001
Input shaft stiffness $k_1(Nm/rad)$	10
Output shaft stiffness $k_2(Nm/rad)$	10
Damping ratio (ζ)	0.01
T_0 (Nm)	1
Ω (rad/sec)	20
β	5°

The non-linear equations of motion (3 and 4) are numerically integrated using the Runge-Kutta 45 routine in MATLAB. The system parameters chosen

for the investigation are shown in Table 1. A comparison of the steady-state time response plots of the non-linear equations of motion reveals distinct differences between the systems with and without clearance. The system with clearance exhibits a higher level of dynamic behavior compared to the system without clearance (see Fig. 5(a) to (d)).

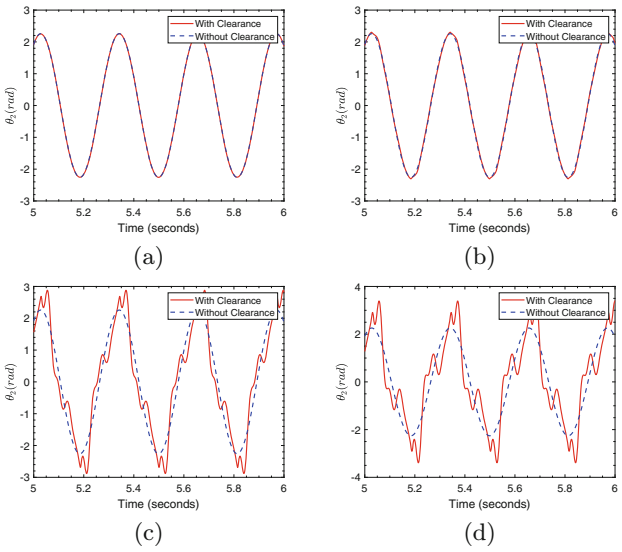


Fig. 5. Steady-state time response with and without clearance effect: (a) $\delta = 0$ rad, (b) $\delta = 0.0005$ rad, (c) $\delta = 0.005$ rad, (d) $\delta = 0.008$ rad.

As the clearance increases, the complexity of the dynamic response becomes more pronounced. This can be clearly observed from the plots in Fig. 5(a) to (d), where larger clearance results in more intricate oscillation patterns. To understand the effect of clearance size on the dynamics of the system, forward parameter sweeps bifurcation diagrams are generated for various U-Joint angles. Moreover, for selected points on bifurcation diagrams, phase portraits and Poincaré maps in $(\dot{\theta}_i, \theta_i)$ plane are shown. In the absence of external input, the nonlinear kinematical relationship between the input and output angular velocity of a U-Joint periodically oscillates in nature resulting in a π-periodic variation in the coefficients of the dynamic system [2]. Here, the Poincaré maps are constructed by selecting a period value of $T = \frac{2\pi}{\Omega}$, and then the plots are created by considering 5000 periods of the system. To quantify chaos, the largest Lyapunov exponent is calculated from the time series response of the perturbed trajectories for the forward sweep of clearance parameter δ [10].

4 Results and Discussion

In this section, numerical bifurcation diagrams are presented for various U-Joint angles, encompassing a wide spectrum of rotational clearance values ranging from 0 to 0.2 rad. The deliberate inclusion of a large clearance value in the bifurcation analysis is intended to provide a thorough exploration of the system's dynamical behavior under a worst-case scenario. Moreover, supplementary analyses for smaller clearance values are undertaken, utilizing tools such as phase portraits, the largest Lyapunov exponent, and Poincaré maps. These investigations are executed across different clearance values, specifically chosen to closely mirror realistic conditions and enhance the practical relevance of the findings.

Table 2. Numerical simulation parameters for the bifurcation diagram shown in Fig. 6(d).

Parameter	Value
Input shaft inertia $J_1(kg.m^2)$	0.0015
Output shaft inertia $J_2(kg.m^2)$	0.0015
Input shaft stiffness $k_1(Nm/rad)$	20
Output shaft stiffness $k_2(Nm/rad)$	20
Damping ratio (ζ)	0.01
T_0 (Nm)	1
Ω (rad/sec)	25
β	5°

4.1 Numerical Bifurcation Analysis

By investigating the corresponding values of the parameters δ and β we observe distinct patterns of period-doubling cascades and the emergence of stabilized periodic states. In the bifurcation diagram depicted in Fig. 6(a), it can be observed that the system showcases distinct patterns of behavior. Initially, for values of δ within the range of 0 to 0.0509, the system displays predictable periodic motion characterized as period-1 (1T). At $\delta = 0.0509$, a notable transition occurs in the system's attractor, leading to subsequent bifurcations. Specifically, as δ continues to increase between 0.05091 and 0.0850, the system undergoes a series of bifurcations, namely 2T, 4T, and 8T forming a Feigenbaum cascade. As δ surpasses the threshold of 0.0850, the system's behavior becomes increasingly complex, exhibiting chaotic dynamics. The bifurcation diagram depicted in Fig. 6(d) illustrates the system's behavior based on the parameters outlined in Table 2. The transition of attractor from periodic to chaotic states, accompanied by the emergence of multi-periodic windows appears to be a universal phenomenon of the system.

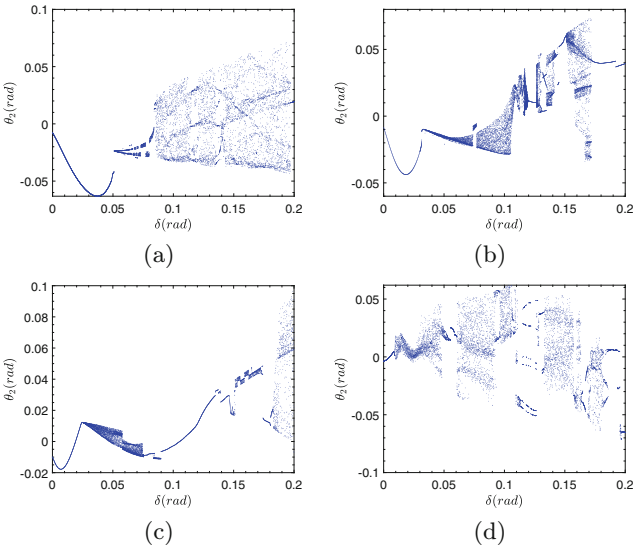


Fig. 6. Bifurcation Diagram as function of clearances for various inclination angles: (a) $\beta = 5°$, (b) $\beta = 15°$, (c) $\beta = 25°$, (d) Bifurcation diagram for Table 2 parameters ($\beta = 5°$).

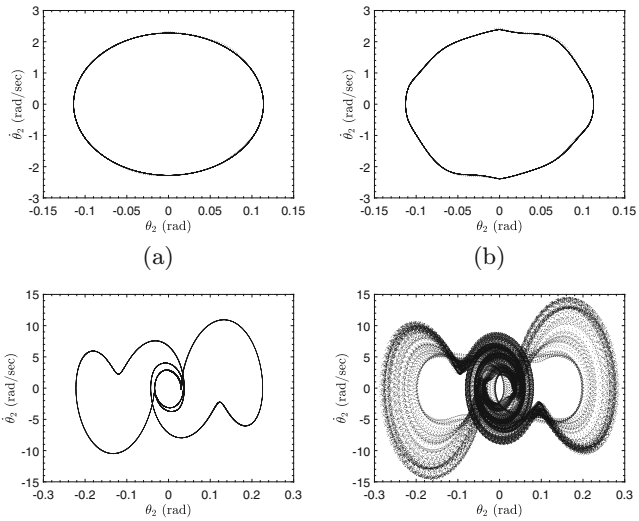


Fig. 7. Phase Portraits ($\beta = 5°$) as a function of clearance: (a) $\delta = 0$ rad, (b) $\delta = 0.0005$ rad, (c) $\delta = 0.055$ rad, (d) $\delta = 0.1$ rad.

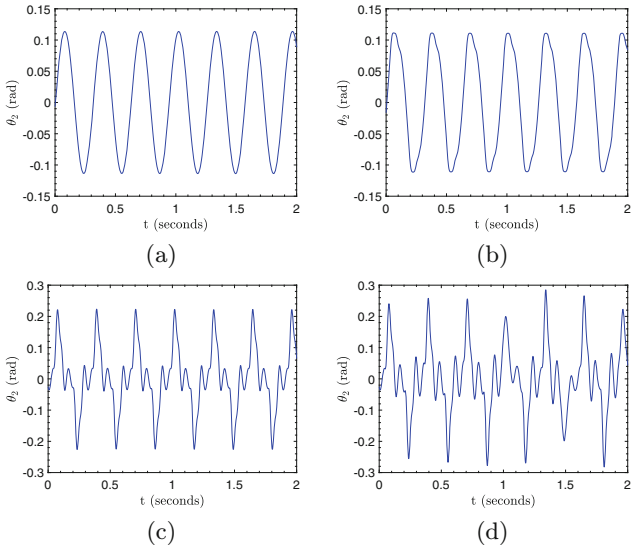


Fig. 8. Time Response corresponding to Fig. 7 Phase Portraits ($\beta = 5^\circ$) as a function of clearance: (a) $\delta = 0$ rad , (b) $\delta = 0.0005$ rad, (c) $\delta = 0.055$ rad, (d) $\delta = 0.1$ rad.

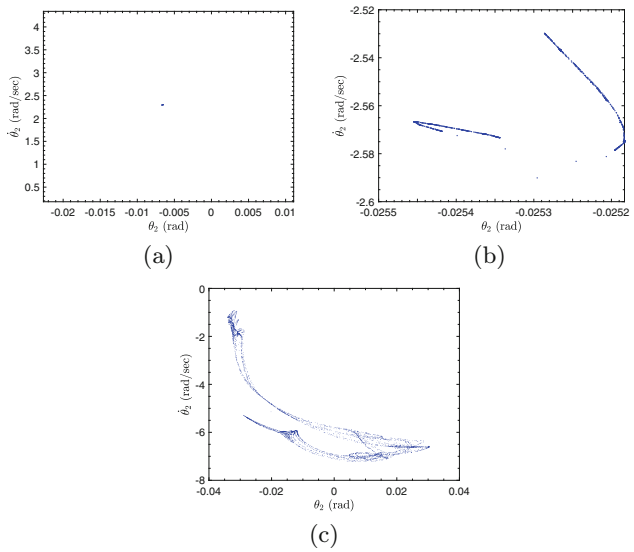


Fig. 9. Poincaré Maps corresponding to phase portraits shown in Fig. 7 for ($\beta = 5^\circ$): (a) $\delta = 0$ rad, (b) $\delta = 0.055$ rad, (c) $\delta = 0.1$ rad.

The phase portraits depicted in Fig. 7(a)–(d) and their corresponding time responses shown in Fig. 8(a)–(d) provide valuable insights into the influence of clearance size δ on the dynamic behavior of the system. Remarkably, when no clearance ($\delta = 0rad$) is present the motion trajectories exhibit periodicity. However, upon introducing a sufficiently large clearance with $\delta = 0.055$ rad (see Fig. 7(c)), a complex motion pattern emerges, indicating a period-doubling route. Moreover, for significantly large values of $\delta = 0.1$, the motion becomes chaotic. The Poincaré maps illustrated in Fig. 9(a)–(c) corroborate these findings, revealing periodic motion for $\delta = 0$ rad, multi-periodic behavior for $\delta = 0.055$ rad, and chaotic motion for $\delta = 0.1$.

For $\beta = 15°$, revealing a similar pattern of period-doubling cascades for $0.0 \leq \delta \leq 0.1$ (see Fig. 6(b)). Further investigation of the system with $\beta = 25°$ (see Fig. 6(c)) showed a familiar pattern of period-doubling cascades for $0.0 \leq \delta \leq 0.055$.

The Poincaré maps shown in Fig. 10(a)–(c) illustrate the attractors for various values of δ which is consistent with the results obtained through bifurcation analysis for $\beta = 15°$. Similarly for $\beta = 25°$ in Fig. 11(a)–(c), illustrate different motion characteristics of the system ranging from periodic to chaotic and then to multi-periodic. Additionally, the examination of the largest Lyapunov exponent in relation to clearance provides complementary insights into the system dynamics as shown in Fig. 12(a)–(c). As the U-Joint angle increases, the bifurcation point arises at lower clearances i.e., $(\beta, \delta) = \{(5°, 0.0509), (15°, 0.030), (25°, 0.024)\}$. Furthermore, as δ increases, the Lyapunov exponent becomes positive, indicating a divergence of nearby trajectories. This positive exponent signifies the potential presence of chaos or unpredictable behavior in the system, highlighting the critical role of δ as a controlling factor.

4.2 Route Out of Chaos

The bifurcation diagram for $\beta = 5°$ reveals intriguing insights into the behavior for specific ranges of the clearances δ. In Fig. 13(a), within the ranges of $0.093 \leq \delta \leq 0.094$ and $0.095 \leq \delta \leq 0.097$, the system exhibits multi-periodic motion; the multi-periodic non-chaotic saddle cycles become a single chaotic attractor upon collision with chaotic saddle when δ is increased. The negative Lyapunov exponents advocate for the non-chaotic nature of periodic islands as shown in Fig. 14(a). The extended bifurcation diagram (see Fig. 13(b)) demonstrates that for $\delta \geq 0.197$, the system's motion becomes periodic, while before this threshold, the motion exhibits chaotic dynamics. This transition from chaos to stability is further illustrated in the figure plotting the largest Lyapunov exponent against δ. The Lyapunov exponent becomes positive for the chaotic region, but changes sign at $\delta \geq 0.197$, indicating the emergence of a stable 1T periodic solution. Interestingly, one can observe that the stable periodic attractor is a subset of the chaotic attractor and the emergence of the stable periodic attractor is sudden which relates to a special type of bifurcation behavior called interior crisis [11]. Figure 15(a) and (b) depict the Poincaré sections of points within the chaotic attractor and stable periodic attractor regimes, respectively,

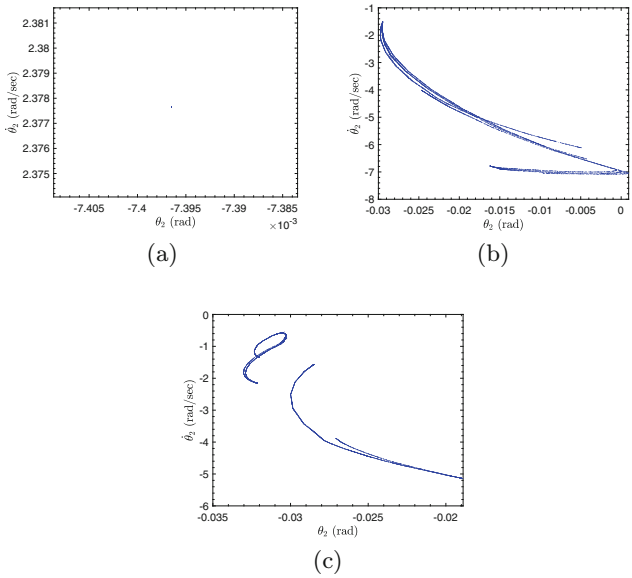


Fig. 10. Poincaré Maps for system with various rotational clearances ($\beta = 15°$): (a) $\delta = 0.0$ rad, (b) $\delta = 0.10$ rad, (c) $\delta = 0.142$ rad.

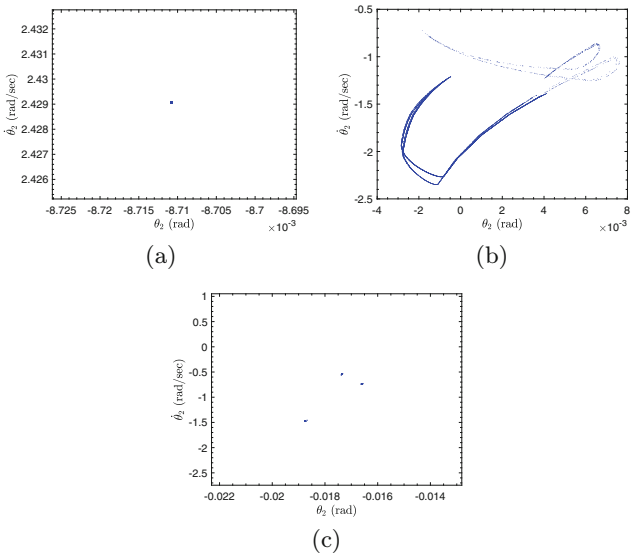


Fig. 11. Poincaré Maps ($\beta = 25°$) for various clearances: (a) $\delta = 0.0$ rad, (b) $\delta = 0.05$ rad, (c) $\delta = 0.139$ rad.

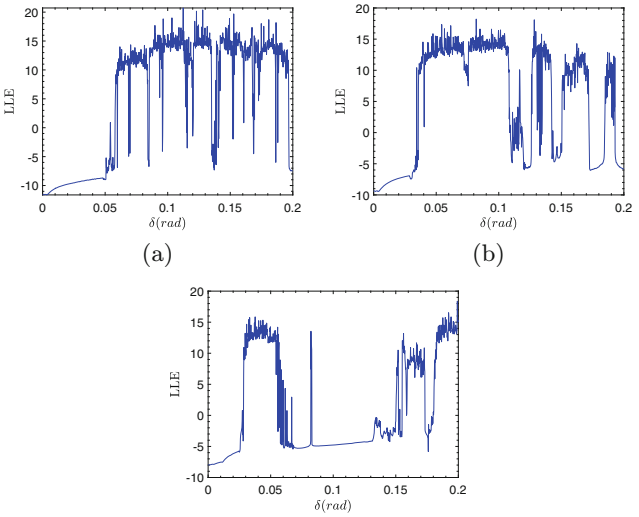


Fig. 12. Largest Lyapunov exponents as a function of clearances: (a) $\beta = 5°$, (b) $\beta = 15°$, (c) $\beta = 25°$.

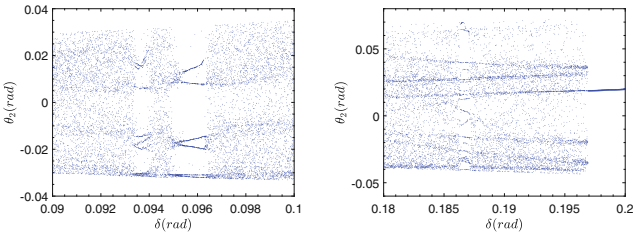


Fig. 13. Bifurcation diagrams ($\beta = 5°$) as a function of clearances: (a) Windows of Periodic behaviour within chaos, (b) System transition from chaos to periodic behaviour at large clearances.

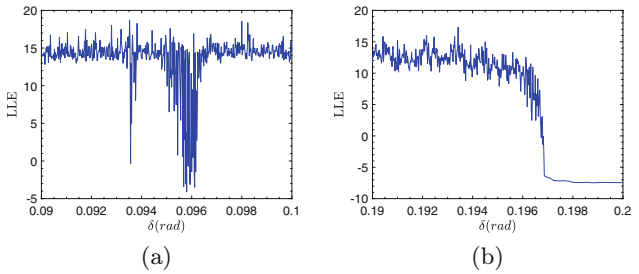


Fig. 14. Largest Lyapunov exponents as a function of clearances: (a) Windows of Periodic behaviour within chaotic regions, (b) System transition from chaos to periodic behaviour at large clearances.

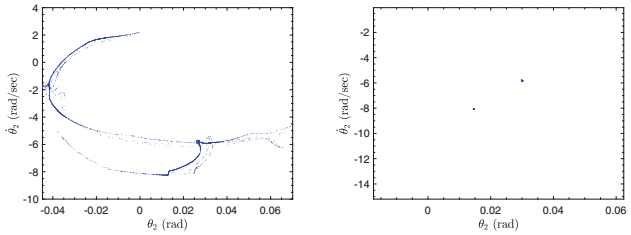


Fig. 15. Poincaré Maps ($\beta = 5°$) for various clearances: (a) Folded annular structure at $\delta = 0.196$ rad, (b) Periodic motion at $\delta = 0.20$ rad.

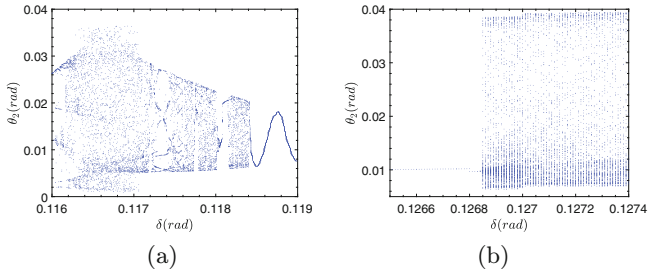


Fig. 16. Bifurcation Diagram ($\beta = 15°$) as a function of clearances: (a) Windows of periodicty & Boundary Crisis, (b) Explosive bifurcation.

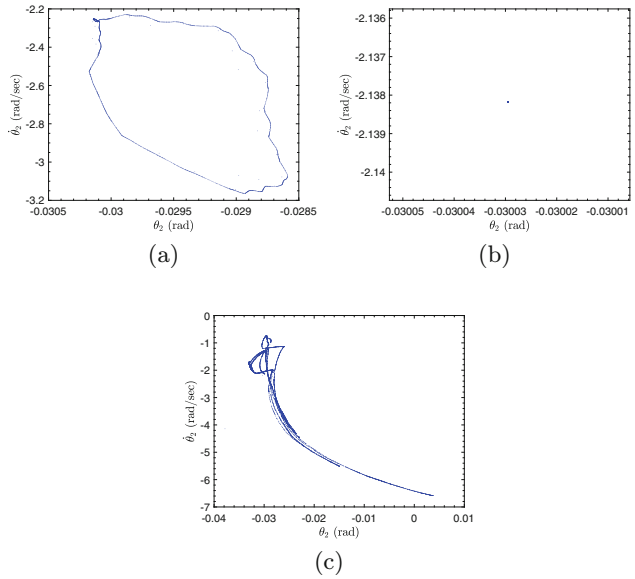


Fig. 17. Poincaré Maps ($\beta = 15°$) as a function of clearances: (a) Closed loop structure at $\delta = 0.118$ rad, (b) Periodic motion at $\delta = 0.124$ rad, (c) Chaotic motion at $\delta = 0.129$ rad.

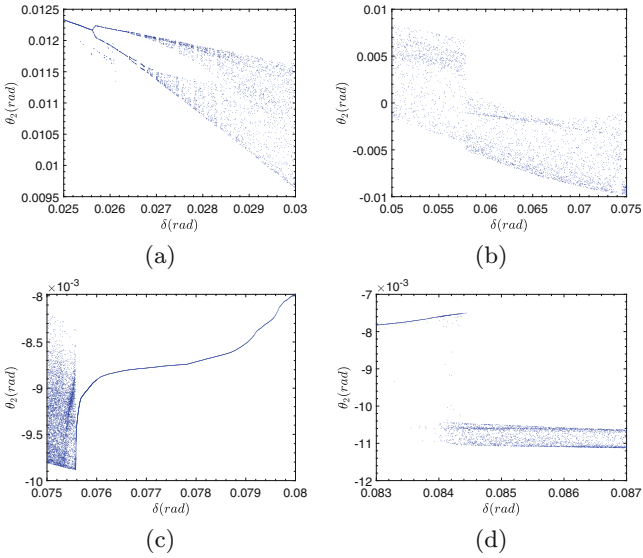


Fig. 18. Bifurcation Diagram ($\beta = 25^{\circ}$) as a function of clearances: (a) Period Doubling Cascades, (b) Chaos -to- Chaos Transition of Attractor, (c) Chaotic attractor to periodic cycle transition , (d) Basin Boundary Crisis.

at clearance values of $\delta = 0.196$ rad and $\delta = 0.20$ rad. These Poincaré sections provide visual insights into the strange attractor's dynamic behavior and distinct characteristics. The stable periodic attractor can be easily observed as a subset of the chaotic attractor in the Poincaré mapping.

The bifurcation diagram depicted in Fig. 6(b) reveals an intriguing pattern within a particular parameter range, specifically $0.116 \leq \delta \leq 0.128$ as shown in Fig. 16. Within the range of $0.1170 \leq \delta \leq 0.1175$, the chaotic attractor has a 6T periodic saddle cycle which collides with the chaotic saddle and enlarges into a chaotic attractor again. Moreover, with further increase in δ the chaotic attractor implodes into stable periodic motion at $\delta = 0.118$ which explodes again into chaotic attractor upon collision with unstable saddle [12]. This explosion and implosion cycle continues until $\delta = 0.127$. Between $0.1268 \geq \delta \geq 0.1270$ the stable periodic attractor collides with unstable chaotic saddle and explodes into a distant attractor, the stable attractor being the subset of the chaotic attractor makes it a case of interior crisis [11]. It can be quantified through largest Lyapunov exponent analysis shown in Fig. 12(b) where Lyapunov exponent becomes positive as it enters the chaotic domain. The corresponding Poincaré sections are shown in Fig. 17(a) and (c), it can be observed that for $\delta = 0.118$ the Poincaré map shows closed loop structure as shown in Fig. 17(a) which is a 2D representation of a higher-dimensional torus structure whereas for $\delta = 0.124$ the motion characteristic is periodic Fig. 17(b), and when $\delta = 0.129$ the attractor becomes chaotic and the corresponding Poincaré section shows structure as shown in Fig. 17(c), which is consistent with bifurcation diagram.

For $\beta = 25°$ and $0.083 < \delta < 0.087$ (see Fig. 18(d)) the stable periodic and chaotic attractor co-exists for a small range of values, which is evidence of basin boundary crisis; small changes in initial conditions can cause trajectories to switch between the two attractors. As δ is increasing, the non-chaotic limit cycle approaches the unstable saddle cycle, and for a critical value they both coalesce and annihilate each other. A new attractor is born which is chaotic henceforth.

5 Conclusion

In this paper, we investigated the effect of rotational clearances in the U-joints and the impact of the U-joint inclination angle on the rotational clearance dynamics. The rotational clearance has been modeled as a discontinuity in the flow which strictly depends on the angular position of the input shaft. We assumed the effect of axial clearances to be zero and only focused our analysis on rotational clearances between the input yoke and rigid cross-piece.

Numerical simulations of the dynamic system with various rotational clearance values (δ) showed that the response becomes larger with an increase in rotational clearance. Bifurcation analysis describes the period-doubling cascades that emerge after a stable periodic motion for a specific clearance value. This specific value varies with the U-joint angle, the Feigenbaum cascade begins for smaller clearance values for larger U-joint angles. More complex dynamic behavior can be observed for sufficiently larger clearance values, as quantified by largest Lyapunov exponents, the nature of motion becomes chaotic. Moreover, for $\beta = 5°$ it is observed that for a specific clearance range the chaotic attractor suddenly merges into a stable periodic attractor where the stable periodic attractor is a subset of the unstable chaotic attractor, such type of bifurcations is termed as "Boundary Crisis" where the chaotic attractor ceases to exist and gets destroyed with motion becoming a stable periodic cycle. Our analysis reveals a dynamic sequence for $\beta = 15°$ where the 6T multi-period attractor collides with and enlarges the chaotic attractor as δ increases. Subsequently, the motion implodes into a stable periodic attractor, which explosively bifurcates, transitioning between consecutive explosions and implosion. This behavior constitutes an interior crisis bifurcation, signifying the attractor's sudden shifts to distant orbits after colliding with the chaotic attractor. Co-existing attractors were observed for a small range of δ, which is evidence of basin boundary crisis: the non-chaotic limit cycle approaches the unstable saddle cycle, and for a critical value they both coalesce and annihilate each other. A new attractor is born which could be chaotic or non-chaotic depending on the initial conditions.

Parametric studies of shaft systems incorporating U-Joint couplings have yielded valuable insights in past. It has been observed that, in the absence of rotational clearance, the system exhibits a quasi-periodic route to chaos [1]. However, with the introduction of rotational clearance, the system transitions to a period-doubling route leading to chaos. To summarize, the analysis conducted in this study primarily concentrated on the effect of rotational clearances on the

driveline torsionals with a single joint. Further analysis is needed to study the effect of pure axial clearances and the effect of axial and rotational clearances on the overall dynamics of the driveline.

Acknowledgement. We gratefully acknowledge the financial support provided by Allison Transmissions Inc., which greatly contributed to the successful completion of this research.

Conflict of Interest. The authors declare no conflict of interest.

References

1. Asokanthan, S.F., Meehan, P.A.: Non-linear vibration of a torsional system driven by a Hooke's joint. J. Sound Vib. **233**(2), 297–310 (2000)
2. Bulut, G., Parlar, Z.: Dynamic stability of a shaft system connected through a Hooke's joint. Mech. Mach. Theory **46**(11), 1689–1695 (2011)
3. SoltanRezaee, M., Ghazavi, M.-R., Najafi, A.: Parametric resonances for torsional vibration of excited rotating machineries with nonconstant velocity joints. J. Vib. Control **24**(15), 3262–3277 (2018)
4. Hummel, S.R., Chassapis, C.: Configuration design and optimization of universal joints with manufacturing tolerances. Mech. Mach. Theory **35**(3), 463–476 (2000)
5. Hummel, S.R., Chassapis, C.: Configuration design and optimization of universal joints. Mech. Mach. Theory **33**(5), 479–490 (1998)
6. Luo, G., Ma, L., Lv, X.: Dynamic analysis and suppressing chaotic impacts of a two-degree-of-freedom oscillator with a clearance. Nonlinear Anal. Real World Appl. **10**(2), 756–778 (2009)
7. Wang, Yu., Li, F.-M.: Nonlinear dynamics modelling and analysis of two rods connected by a joint with clearance. Appl. Math. Mech. **39**(9), 2518–2527 (2015)
8. Šalinić, S., Vranić, A., Nešić, N.D., Tomović, A.M.: On the torque transmission by a Cardan-Hooke joint. FME Trans. **45**(1), 117–121 (2017)
9. Rattan, S.S.: Theory of Machines. McGraw-Hill Education, India (2014)
10. Wolf, A., Swift, J.B., Swinney, H.L., Vastano, J.A.: Determining Lyapunov exponents from a time series. Physica D: Nonl. Phenom. **16**(3), 285–317 (1985)
11. Thompson, J.M.T., Stewart, H.B., Ueda, Y.: Safe, explosive, and dangerous bifurcations in dissipative dynamical systems. Phys. Rev. E **49**(2), 1019–1027 (1994)
12. Grebogi, C., Ott, E., Yorke, J.A.: Crises, sudden changes in chaotic attractors, and transient chaos. Physica D: Nonl. Phenom. **7**(1–3), 181–200 (1983)

Soft Robotics Powered by Rod Theory

Hiromi Mochiyama(✉)

University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki 305-8573, Japan
motiyama@iit.tsukuba.ac.jp
https://www.frlab.iit.tsukuba.ac.jp

Abstract. This paper presents a few illustrative examples of soft robots that can achieve controlled behaviors even in their inherent hyper-under-actuated natures. The rod theory plays an important role in the achievements considering the perspectives of robotics. Here, three types of soft robots, hyper-flexible manipulators, snap motors, and ostrich-inspired arms are presented with appropriate references.

Keywords: soft robotics · rod theory

1 Introduction

Soft robotics is a research field that attempts to identify new robotic functions by maximizing the remarkable mechanical, electrical-electronic, and chemical properties of soft materials [1]. The rod theory considers an exact modeling of largely-deformable rod and is a fundamental tool for addressing nonlinear problems of elasticity [2]. Because largely-deformable rods are typical objects in soft robotics, the rod theory can be adopted for soft robots.

Consider a rod with length L and mass M. Without loss of generality, it is assumed that the mass distribution of a rod is uniform along the direction of its length. The largely-deformable rod can be modeled as the limit of a serial chain of rigid bodies connected with passive three-dimensional rotational joints as the number of degrees of freedom increases to the infinity [3]. The discretized kinematics of the rod [1] with the partition number n can be expressed by

$$\boldsymbol{F}_i \;= \boldsymbol{F}_{i-1}\boldsymbol{R}_{\mathrm{J}}(\boldsymbol{\theta}_i), \quad \boldsymbol{F}_0 = \boldsymbol{F}_{\mathrm{b}} \tag{1}$$

$$\tilde{\boldsymbol{p}}_i = \tilde{\boldsymbol{p}}_{i-1} + \frac{1}{n}\boldsymbol{F}_i\boldsymbol{e}_{\mathrm{x}}, \quad \tilde{\boldsymbol{p}}_0 = \frac{\boldsymbol{p}_{\mathrm{b}}}{L} \tag{2}$$

where $i \in \{0, \cdots, n\}$ is the index of joints or rigid bodies (links), $\boldsymbol{F}_i \in \mathrm{SO}(3)$ is the orientation matrix of the i-th link, $\tilde{\boldsymbol{p}}_{i-1} \in \Re^3$ is the normalized position vector of the i-th joint, $\boldsymbol{\theta}_i \in C^3$ is the angles of the i-the joint, $\boldsymbol{R}_{\mathrm{J}}(\boldsymbol{\theta}_i) \in \mathrm{SO}(3)$ is the rotational action of the i-the joint which depends on $\boldsymbol{\theta}_i \in C^3$, $\boldsymbol{e}_{\mathrm{x}} := [1\ 0\ 0]^T$

[1] This discretized rod kinemtic is considered as an extended version of the Frenet-Serret formulas for spatial curves. See [4] for the geometric treatment.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 457–470, 2025.
https://doi.org/10.1007/978-3-031-72794-8_36

is the unit vector in the x direction, $\boldsymbol{F}_{\mathrm{b}} \in \mathrm{SO}(3)$ is the orientation matrix of the base, $\boldsymbol{p}_{\mathrm{b}} \in \Re^3$ is the position vector of the base origin to which one end of the rod is fixed. The actual position vector of each joint before normalization, $\boldsymbol{p}_i \in \Re^3$, can be obtained by multiplying its normalized counterpart by L, i.e., $\boldsymbol{p}_i = L\tilde{\boldsymbol{p}}_i$.

The dynamics of the discretized rod can be expressed with respect to the i-th joint moment by

$$\mathcal{I}_i + \mathcal{G}_i + \mathcal{K}_i = \mathcal{E}_i + \mathcal{U}_i \tag{3}$$

where $\mathcal{I}_i$ is the inertial moment, $\mathcal{G}_i$ is the gravitational moment, $\mathcal{K}_i$ is the viscoelastic moment, $\mathcal{E}_i$ is the external moment from a surrounding environment, and $\mathcal{U}_i$ is the control moment. Each moment term has been normalized and non-dimensionalized and can be represented as follows.

First, the inertial moment $\mathcal{I}_i$ can be expressed in the following summation form thanks to the Newton-Euler formulation of the dynamics of a serial chain mechanism [5,6]:

$$\mathcal{I}_i = \sum_{j=i}^{n} \left\{ (\tilde{\boldsymbol{r}}_j - \tilde{\boldsymbol{p}}_{i-1}) \times \Delta\tilde{\boldsymbol{f}}_j + \Delta\tilde{\boldsymbol{m}}_j \right\} \tag{4}$$

where $\Delta\tilde{\boldsymbol{f}}_i$ and $\Delta\tilde{\boldsymbol{m}}_i \in \Re^3$ are the nondimensional inertial force and moment of the i-th link, respectively, and $\tilde{\boldsymbol{r}}_i \in \Re^3$ denotes the normalized position vector of the mass center of the i-th link. The inertial force and moment can be written by

$$\Delta\tilde{\boldsymbol{f}}_i = \frac{1}{n}\frac{d^2\tilde{\boldsymbol{r}}_i}{d\tilde{t}^2} \tag{5}$$

$$\Delta\tilde{\boldsymbol{m}}_i = \tilde{\boldsymbol{I}}_i \frac{d\tilde{\boldsymbol{\omega}}_i}{d\tilde{t}} + [\tilde{\boldsymbol{\omega}}_i \times]\, \tilde{\boldsymbol{I}}_i \tilde{\boldsymbol{\omega}}_i \tag{6}$$

where $\tilde{t} \in \Re$ is the time, $\tilde{\boldsymbol{I}}_i$ is the inertia tensor, $\tilde{\boldsymbol{\omega}}_i \in \Re^3$ is the angular velocity vector of the i-th link, and $[\cdot\times] : \Re^3 \to \mathrm{so}(3)$ denotes an operator that expands a three dimensional vector into a corresponding three-dimensional skew-symmetric matrix. The relationships between the dimensional values and the non-dimensionalized counterparts (which are denoted by $\tilde{(\cdot)}$) are summarized as follows:

$$\boldsymbol{r}_i = L\tilde{\boldsymbol{r}}_i \tag{7}$$

$$t = T_{\mathrm{r}}\hat{t} \tag{8}$$

$$\boldsymbol{I}_i = ML^2\tilde{\boldsymbol{I}}_i \tag{9}$$

$$\boldsymbol{\omega}_i = T_{\mathrm{r}}\tilde{\boldsymbol{\omega}}_i \tag{10}$$

where $T_{\mathrm{r}} := \sqrt{\frac{L}{g}}$ is the representative time, and g is the gravitational acceleration constant.

Second, the gravitational moment $\mathcal{G}_i$ is expressed by

$$\mathcal{G}_i = \sum_{j=i}^{n} \left\{ (\tilde{\boldsymbol{r}}_j - \tilde{\boldsymbol{p}}_{i-1}) \times \left(-\frac{1}{n} \boldsymbol{e}_{\mathrm{z}} \right) \right\} \tag{11}$$

where $\boldsymbol{e}_{\mathrm{z}} := [0\ 0\ 1]^T$ is the unit vector in the z direction.

Third, the visco-elastic moment $\mathcal{K}_i$ can be represented by

$$\mathcal{K}_i = \boldsymbol{A}_i(\boldsymbol{\theta}_i) \left(\tilde{\boldsymbol{K}}_{\mathrm{d}} \boldsymbol{\theta}_i + \tilde{\boldsymbol{D}}_{\mathrm{d}} \frac{d\boldsymbol{\theta}_i}{d\tilde{t}} \right) \tag{12}$$

where $\tilde{\boldsymbol{K}}_{\mathrm{d}}$ is a three dimensional diagonal matrix whose diagonal elements are the rotational spring constants of an elastic joint around the joint axes, $\tilde{\boldsymbol{D}}_{\mathrm{d}}$ is a three dimensional diagonal matrix whose diagonal elements are the rotational viscous friction coefficient of an elastic joint around the joint axes, $\boldsymbol{A}_i : C^3 \rightarrow \Re^{3\times 3}$ is the joint axes matrix which can be obtained by arranging the three unit vectors of the i-th joint axes. [2]. Nondimensional $\tilde{\boldsymbol{K}}_{\mathrm{d}}$ and $\tilde{\boldsymbol{D}}_{\mathrm{d}}$ can be restored to their dimensional values $\boldsymbol{K}_{\mathrm{d}}$ and $\boldsymbol{D}_{\mathrm{d}}$ by

$$\boldsymbol{K}_{\mathrm{d}} = MgL\tilde{\boldsymbol{K}}_{\mathrm{d}} \tag{13}$$

$$\boldsymbol{D}_{\mathrm{d}} = \frac{ML^2}{T_{\mathrm{r}}} \tilde{\boldsymbol{D}}_{\mathrm{d}} \tag{14}$$

The stiffness of the elastic rod $\boldsymbol{K}_{\mathrm{d}}$ can be determined by its material properties (the Young's modulus and Poisson's ratio), the geometric parameters (the rod length, the shape of the cross section along the backbone), and the partition number [7] while the viscosity $\boldsymbol{D}_{\mathrm{d}}$ should be identified experimentally.

Fourth, the external moment $\mathcal{E}_i$ in the right-hand side depends on the situation. One of the most typical case is to have an external force and moment at the tip of a rod. In this case, this term can be written by

$$\mathcal{E}_i = (\tilde{\boldsymbol{p}}_n - \tilde{\boldsymbol{p}}_{i-1}) \times \tilde{\boldsymbol{f}}_{\mathrm{e}} + \tilde{\boldsymbol{m}}_{\mathrm{e}} \tag{15}$$

where $\tilde{\boldsymbol{f}}_{\mathrm{e}}$ and $\tilde{\boldsymbol{m}}_{\mathrm{e}} \in \Re^3$ are the external force ad moment at the tip of a rod. The dimensional external force and moment at the tip can be reconstructed from their nondimensional counterparts by

$$\boldsymbol{f}_{\mathrm{e}} = Mg\tilde{\boldsymbol{f}}_{\mathrm{e}} \tag{16}$$

$$\boldsymbol{m}_{\mathrm{e}} = MgL\tilde{\boldsymbol{m}}_{\mathrm{e}} \tag{17}$$

Note that $\boldsymbol{f}_{\mathrm{e}}$ and $\boldsymbol{m}_{\mathrm{e}}$ are usually unknown, and must be measured by some sensor.

Finally, the control moment $\mathcal{U}_i$ applied on the i-th joint depends on a system under consideration. If we can drive all the joints using sufficiently high-power actuators, which means no restriction on $\mathcal{U}_i$, the rod can form any shape, such

[2] An explicit representation of the joint axes matrix can be found in [1].

that it can be moved on any target curve [8,9], as well as achieve enveloping whole-arm compliance [10]. However, the full actuation to all the degrees of freedom is infeasible, and some actuation strategies have to be considered, which implies that soft robotics is always faced with hyper-underactuated problems. Here, illustrative examples of soft robots that solve hyper-underactuated problems to some extent are introduced with the adoption of the rod theory modified for soft robotics.

The contribution of the paper is to illustrate the cases that reveal how hyper-many-degree-of-freedom dynamics can yield functions of soft robots, which may leverage to generate some ideas for new directions for the theory of dynamical systems. Here, the rod theory with an appropriate nondimensinalization treatment plays an important role to model the dynamics of soft robots successfully.

2 Hyper-Flexible Manipulator

A hyper-flexible manipulator (HFM) is a continuum robot with a one-dimensional mechanism such as a cable, rope, and string [11]. In this case, $\mathcal{K}_i$ is negligibly small in comparison with other terms. Then, the system dynamics can be reduced to

$$\mathcal{I}_i + \mathcal{G}_i = \mathcal{U}_i \tag{18}$$

where we suppose no external moment at the tip from the surrounding environment.

One of the possibilities of control actuation for an HFM is to move the base where one end of an HFM is attached. Figure 1 shows an HFM system where a hyper-flexible rod is attached to a slider moving in the x-direction. The position vector of the connecting point is $\boldsymbol{p}_0$. In the left drawing, the slider stays at the origin, and the hyper-flexible rod is at rest under gravity. Therefore, its shape is a vertical straight line on the z-axis starting from the origin. In the right drawing, the slider moves along the x-axis with an acceleration u which causes a deformation of the hyper-flexible rod. This control action can be expressed by

$$\mathcal{U}_i = -\tilde{u} \sum_{j=i}^{n} \left\{ (\tilde{\boldsymbol{r}}_j - \tilde{\boldsymbol{p}}_{i-1}) \times \left(\frac{1}{n} \boldsymbol{e}_{\mathrm{x}} \right) \right\} \tag{19}$$

where $\tilde{u} \in \Re$ is the nondimensional base acceleration along the x direction as the control input of this system. The actual control acceleration, u, can be represented by $u = g\tilde{u}$. In this actuation case, owing to the port-controlled Hamiltonian structure of this rod system, it is verified that the horizontal base acceleration as a control input, $\tilde{u}$, is energetically dual with $\tilde{\xi}$ which is the horizontal velocity of the mass center of the hyper-flexible body [12], i.e.,

$$\tilde{\xi} = \frac{1}{n} T_{\mathrm{r}} \boldsymbol{e}_{\mathrm{x}}{}^{T} \sum_{j=1}^{n} \frac{d\tilde{\boldsymbol{r}}_j}{d\tilde{t}} \tag{20}$$

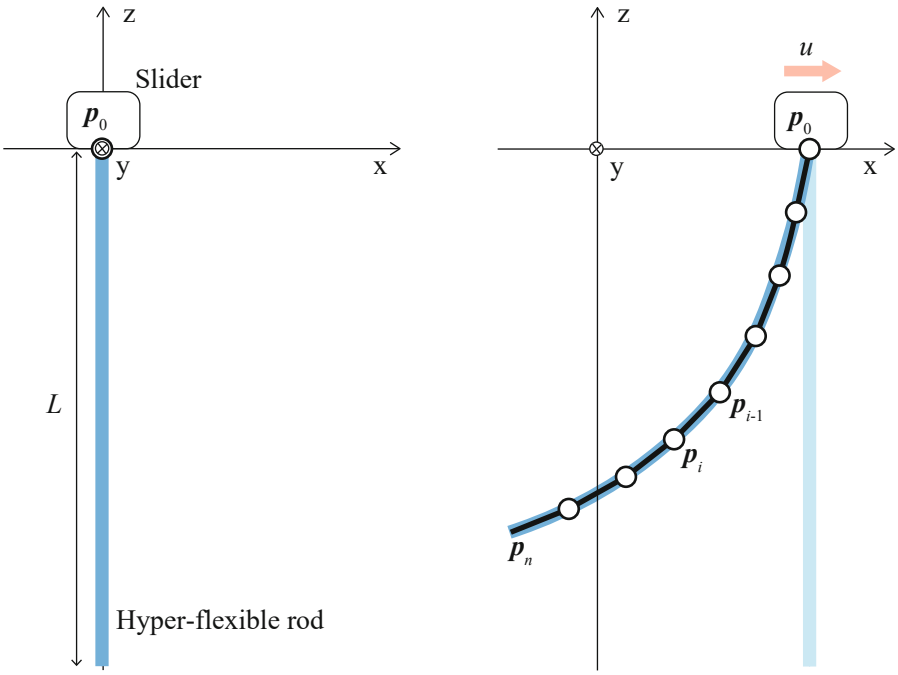


Fig. 1. A hyper-flexible manipulator (HFM) with a horizontally moving slider. (Left) A hyper-flexilbe manipulator is at rest. The shape of the hyper-flexible rod is a vertical straight line. (Right) When the slider moves in the x direction with an acceleration, the hyper-flexible rod deforms. The serial chain of rigid bodies connected with free joints is also drawn. Free joints are drawn as white circles, while rigid bodies are expressed by black solid lines. In this case, $n = 8$.

This means that we can attenuate the swing of a hyper-flexible rod by horizontally moving the upper end of the rod in the direction where the center of gravity of the rod moves horizontally. Based on this system property, we can achieve damping injection by the feedback $\tilde{u} = -\phi(\tilde{\xi})$ where $\phi(\cdot)$ is a monotonically increasing function. The effect of this damping injection was verified not only by simulation but also by experiments. See [12,13] for more detail. Moreover, by inversely using this property, we can also achieve impact manipulation using an HFM where we can apply an impact force to a given target point [14].

Another hyper-flexible possibility is the chameleon-like shooting manipulation system for acting on a distant target object [15]. In this manipulation, the end effector of an HFM is catapulted to the target. A hyper-flexible body is connected to the end effector and used to constrain the end effector's motion. In this case, the large mass at the tip end should be considered explicitly, and $\mathcal{U}_i$ should be differently modeled from the previous case such that an impulsive force is applied at the tip end mass.

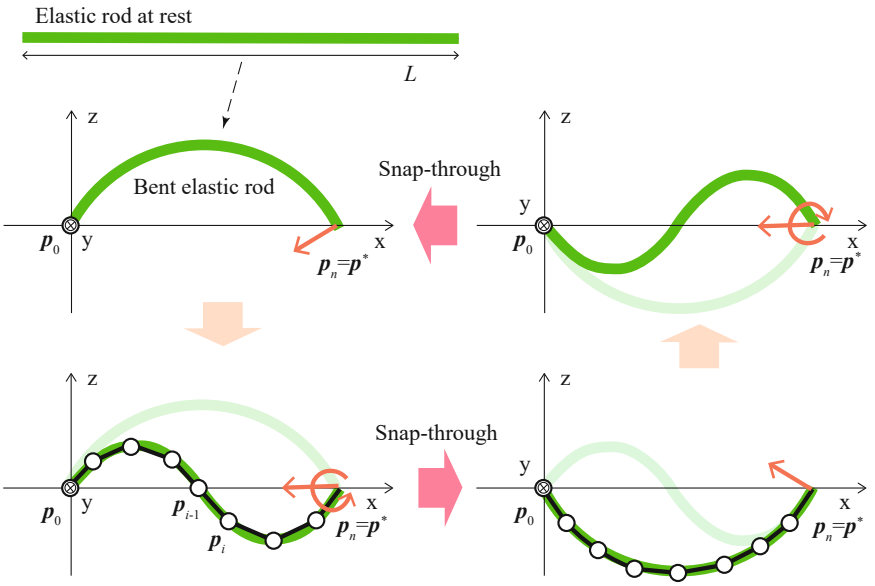


Fig. 2. Impulse force generation process of a typical snap motor. (Top) The original shape of an elastic rod is a straight line. (Middle-Left) One end of the elastic rod is attached to a free joint at the origin. The other end is constrained s.t. $\boldsymbol{p}_n = \boldsymbol{p}^*$. The rod is C-shaped. (Bottom-Left) By applying an appropriate force and moment, the rod deforms gradually. Eventually, the rod becomes mirrored S-shaped. (Bottom-Right) The rod changes its shape rapidly to a simple C-shape due to a snap-through buckling. (Middle-Right) By applying an appropriate force and moment, the rod deforms, and the rod becomes S-shaped. A snap-through buckling occurs again, and the rod return to the original C-shape shown in the Middle-Left drawing. This process can be repeated.

3 Snap Motor

The snap motor is a simple impulsive force generator typically comprising a loop mechanism of a largely-deformable elastic rod, an actuator, and a sufficiently stiff frame [16,17]. In comparison with other conventional continuum-elastic mechanisms in soft robotics, such as a wire-driven elastic rod [18] or parallel extensible balloons [19], the distinguished feature of the snap motor is to utilize a bistable mechanism and generate snap-though buckling of a closed elastic rod intentionally via actuation. Figure 2 shows a typical mechanism of impulsive force generation in a snap motor: 0) Assume that an elastic rod is straight if no force is applied to it as shown in the top drawing. 1) Then, the elastic rod is attached so that one end is connected to a free joint at the origin while the other end is constrained at $\boldsymbol{p}^*$ on the x axis. The shape of the rod is bent as shown in the middle-left drawing. 2) By applying an appropriate force and moment at the tip end with keeping the tip position s.t. $\boldsymbol{p}_n = \boldsymbol{p}^*$, the elastic rod deforms. Eventually, the rod forms a mirrored S-shape as shown

in the bottom-left drawing. 3) The elastic rod rapidly changes its shape to a simple C-shape as shown in the bottom-right drawing. This transition is called snap-through buckling. 4) Again, by applying an appropriate force and moment at the tip end with keeping the tip position s.t. $\boldsymbol{p}_n = \boldsymbol{p}^*$, the elastic rod deforms. Eventually, the rod forms an S-shape as shown in the middle-right drawing. 5) The snap-through buckling occurs again. The elastic rod return to a simple C-shape as shown in the middle-left drawing. The process from 1) to 4) can be repeated.

Quasi-static shape transition simulation using the shooting method [26] allows us to estimate the released energy of a snap motor, which is useful in designing a better snap motor as an impulse force generator [27].

In quasi-static analysis of the snap motor, we ignore the inertia, gravity and external moments as well as the viscous term as follows:

$$\mathcal{K}_i = \mathcal{U}_i \tag{21}$$

where $\frac{d\boldsymbol{\theta}_i}{d\tilde{t}} = \boldsymbol{0}$. We also assume that the positions of the both ends of the rod are fixed, but an orientation of the one end of the rod is rotated. This control structure can be expressed by

$$\mathcal{U}_i = (\tilde{\boldsymbol{p}}_n - \tilde{\boldsymbol{p}}_{i-1}) \times \tilde{\boldsymbol{u}}_{\mathrm{f}} + \tilde{\boldsymbol{u}}_{\mathrm{m}} + \tilde{\boldsymbol{u}}_{\mathrm{a}} \tag{22}$$

where $\tilde{\boldsymbol{u}}_{\mathrm{f}}$ and $\tilde{\boldsymbol{u}}_{\mathrm{m}} \in \Re^3$ are the controlled force and moment which are adjusted to hold the following geometric constraints on the tip position and orientation of the elastic rod:

$$\tilde{\boldsymbol{p}}_n = \frac{\boldsymbol{p}^*}{L} \tag{23}$$

$$\boldsymbol{F}_n = \boldsymbol{F}^* \tag{24}$$

where $\boldsymbol{p}^* \in \Re^3$ and $\boldsymbol{F}^* \in \mathrm{SO}(3)$ are the constrained position and orientation at the tip of the elastic rod which should be properly selected so that snap-through buckling occurs. Usually, $\tilde{\boldsymbol{u}}_{\mathrm{f}}$ and $\tilde{\boldsymbol{u}}_{\mathrm{m}}$ are "structurally" generated by the loop mechanism of the snap motor. Moreover, $\tilde{\boldsymbol{u}}_{\mathrm{a}} \in \Re^3$ is the additional controlled moment for rotating the tip orientation of the elastic rod back and forth repeatedly to obtain repeated snap-through buckling. Therefore, the tip orientation of the rod (24) can be replaced with

$$\boldsymbol{F}_n = \boldsymbol{F}^* \boldsymbol{R}(\boldsymbol{e}_{\mathrm{z}}, \theta(\tilde{t})) \tag{25}$$

where $\boldsymbol{R}(\boldsymbol{a}, \theta) : U \times C \to \mathrm{SO}(3)$ denotes a rotational action around axis $\boldsymbol{a}$ with a rotational angle θ, and U is a set of unit vectors. Here $\theta(\tilde{t}) : \Re \to C$ is a time function that repeatedly cycles back and forth between the minimum $\theta_{\min}$ and the maximum $\theta_{\max}$. Typically, $\tilde{\boldsymbol{u}}_{\mathrm{a}}$ is generated by a rotational motor in the loop mechanism, and $\theta_{\min}$ and $\theta_{\max}$ are properly selected such that the snap-through angles exist between these values.

For efficient quasi-static shape transition simulation of the snap motor, it is preferable to understand the system properties deeply [28–32]. In order to

do this, the base wrench expression is useful. The nondimensional force and moment applied from the elastic rod to the base, $\tilde{\boldsymbol{f}}_{\mathrm{b}}, \tilde{\boldsymbol{m}}_{\mathrm{b}} \in \Re^3$, are related to the controlled force and moments by

$$\tilde{\boldsymbol{f}}_{\mathrm{b}} = \tilde{\boldsymbol{u}}_{\mathrm{f}}, \tag{26}$$

$$\tilde{\boldsymbol{m}}_{\mathrm{b}} = (\tilde{\boldsymbol{p}}_n - \tilde{\boldsymbol{p}}_0) \times \tilde{\boldsymbol{u}}_{\mathrm{f}} + \tilde{\boldsymbol{u}}_{\mathrm{m}} + \tilde{\boldsymbol{u}}_{\mathrm{a}}. \tag{27}$$

Here, the relationships between the actual dimensional values and the corresponding nondimentional ones are summarized as follows:

$$\boldsymbol{u}_{\mathrm{f}} = Mg\tilde{\boldsymbol{u}}_{\mathrm{f}} \tag{28}$$

$$\boldsymbol{u}_{\mathrm{m}} = MgL\tilde{\boldsymbol{u}}_{\mathrm{m}} \tag{29}$$

$$\boldsymbol{u}_{\mathrm{a}} = MgL\tilde{\boldsymbol{u}}_{\mathrm{a}} \tag{30}$$

$$\tilde{\boldsymbol{f}}_{\mathrm{b}} = Mg\boldsymbol{f}_{\mathrm{b}} \tag{31}$$

$$\tilde{\boldsymbol{m}}_{\mathrm{b}} = MgL\boldsymbol{m}_{\mathrm{b}} \tag{32}$$

We can obtain the following moment balance expressions

$$\boldsymbol{A}_i(\boldsymbol{\theta}_i)\tilde{\boldsymbol{K}}_{\mathrm{d}}\boldsymbol{\theta}_i = (\tilde{\boldsymbol{p}}_0 - \tilde{\boldsymbol{p}}_{i-1}) \times \tilde{\boldsymbol{f}}_{\mathrm{b}} + \tilde{\boldsymbol{m}}_{\mathrm{b}} \tag{33}$$

The following approximation is useful for efficient computation:

$$\boldsymbol{\theta}_i \approx \tilde{\boldsymbol{K}}_{\mathrm{d}}^{-1}\boldsymbol{A}_i(\boldsymbol{\theta}_i)^T \left\{(\tilde{\boldsymbol{p}}_0 - \tilde{\boldsymbol{p}}_{i-1}) \times \tilde{\boldsymbol{f}}_{\mathrm{b}} + \tilde{\boldsymbol{m}}_{\mathrm{b}}\right\} \tag{34}$$

which is valid when the partition number n is sufficiently large[3] If the base wrench can be measured by a six-axes force sensor, it is possible to estimate the shape of the rod in real time, which is significantly beneficial for robotic applications [28,29].

With these snap motors, we can develop a compact jumping mechanism. The 14-gram and business-card-size snap motor designed for generating unidirectional impulsive forces can quickly climb up a normal stair by repeated jumps [20]. The snap motor can also be adopted for a jumping mechanism of a wheeled robot that can jump while rolling [21]. This robot is equipped with an arched snap motor where three-dimensional shape transitions of the mechanism should be considered. The snap motor is sufficiently light; hence, it can be utilized as the perturbator of a wearable device to estimate human joint impedance [22,23]. A pulsatile pump can also be fabricated by a snap motor [24]. Here, a snap motor was used to expand and flatten a soft bag, to generate a pulsated flow. Recently, owing to its compact nature, light-weight, and low power consumption, the snap motor is mounted on flightless drones moving on walls for automated hammering inspections [25].

[3] The partition number can be determined by experiment from the view point of model accuracy as shown in [28]. In [28], the relationship between the estimation error at the tip end of an elastic rod and the partition number is shown as a graph. From this graph, it can be seen that the model errors decrease rapidly as the numbers of the partition increase.

The impulse generation capability of a snap motor can be evaluated by an area of a time graph of measured impulse forces such as by a load cell. A generated impulse of a snap motor is strongly correlated to the difference in stored energies between the rod shapes before and after snap-through buckling. Therefore, one of the typical ways to change a generated impulse of a snap motor is to move the constrained tip position $\boldsymbol{p}^*$ in order to change the snap-through shape. To do this, we have to pay a cost of some additional actuator. The stored energies difference before and after snap-through buckling can be estimated by the quasi-static simulation efficiently based on the rod theory [1]. Analysis of dynamics of snap-through buckling can be found in [33–35].

4 Ostrich-Inspired Arm

There is a historical fact that good ideas of bio-inspiration have led to the progress of robotics. Such examples include long-armed monkeys (as brachiation animals of) [36–39], snakes[40–42], and earthworms[43]. Here, we focus on ostriches[4]. An ostrich-inspired arm is a wire-driven mechanism with many degrees of freedom. The important feature of this manipulator is that the number of actuation is significantly smaller than the number of kinematic degrees of freedom, and it has to move against the gravity. Therefore, an ostrich-inspired manipulator is inherently hyper-underactuated, and the control problem is similar to a multi-dimensional version of an inverted pendulum. In this case, $\mathcal{K}_i$ and $\mathcal{E}_i$ can be negligible. The dynamics of an ostrich-inspired arm can be described by

$$\mathcal{I}_i + \mathcal{G}_i = \mathcal{U}_i \tag{35}$$

For the first step of analysing this complex dynamical system, the condition of the equilibrium is sought here. Assume that $\mathcal{U}_i$ is static. In a static case, $\mathcal{I}_i = \mathbf{0}$. Then, the static balance of an ostrich-inspired arm is expressed by

$$\mathcal{G}_i = \mathcal{U}_i \tag{36}$$

The prototype of the manipulator is a serial chain of 18 rigid links connected with rotation joints moving in a sagittal plane under gravity and driven through two asymmetric antagonistic wire systems connected to two levers that are directly operated by a human operator functioning as the controller [5] [45,46]. This wire-driven control mechanism can be represented by

$$\mathcal{U}_i = (\tilde{\boldsymbol{p}}_n - \tilde{\boldsymbol{p}}_{i-1}) \times \tilde{\boldsymbol{f}}_{\mathrm{W}} + \tilde{\boldsymbol{m}}_{\mathrm{W}} \tag{37}$$

where $\tilde{\boldsymbol{f}}_{\mathrm{W}}$ and $\tilde{\boldsymbol{m}}_{\mathrm{W}} \in \Re^3$ are the nondimensional force and moment generated by the wires at the rod tip and expressed by

$$\tilde{\boldsymbol{f}}_{\mathrm{W}} = -(\tilde{u}_{\mathrm{D}} + \tilde{u}_{\mathrm{V}})\, \boldsymbol{F}_n \boldsymbol{e}_{\mathrm{x}} \tag{38}$$

$$\tilde{\boldsymbol{m}}_{\mathrm{W}} = (\tilde{u}_{\mathrm{D}} \tilde{\boldsymbol{r}}_{\mathrm{D}} + \tilde{u}_{\mathrm{V}} \tilde{\boldsymbol{r}}_{\mathrm{V}}) \times (-\boldsymbol{F}_n \boldsymbol{e}_{\mathrm{x}}) \tag{39}$$

[4] See [45] for the detailed reasons for why ostriches were selected as model animals.

[5] The wire-driven actuation by a human operator to seek the feasibility of the mechanism was inspired by the running robot research [44].

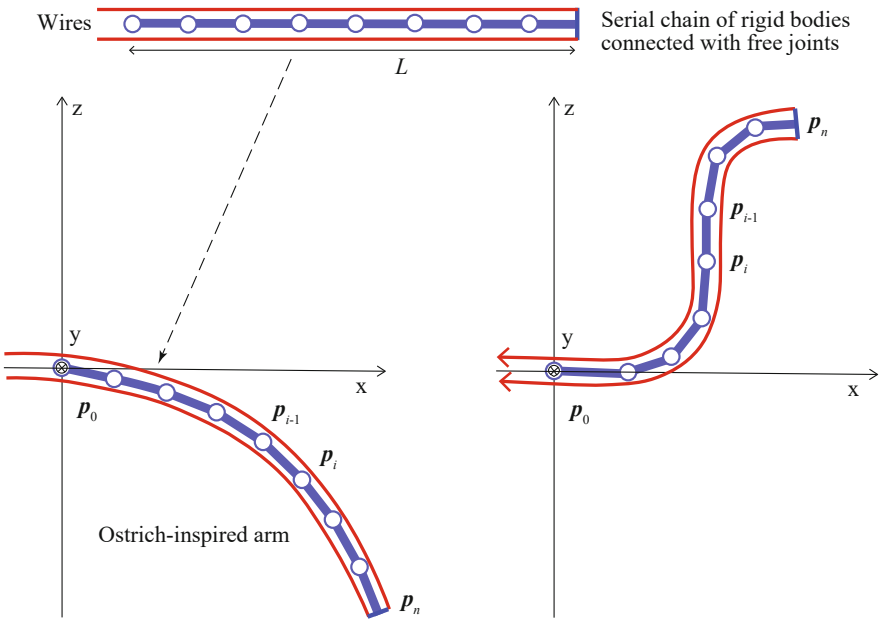


Fig. 3. Movement of an ostrich-inspired arm. (Top) An ostrich-inspired arm is typically composed of a serial chain of rigid bodies connected with free joints with joint angle limitations, and with two antagonistic wires. (Bottom-Left) When one end of the serial chain is horizontally attached at the origin without wire tensions, the arm forms an arc. (Bottom-Right) By applying wire tensions appropriately, the head of the arm can reach an upper position in spite of its hyper-underactuated nature.

where $\tilde{u}_{\mathrm{D}}$ and $\tilde{u}_{\mathrm{V}} \in [0\ \infty)$ are the tensions of the dorsal and ventral wires, respectively, and $\tilde{\boldsymbol{r}}_{\mathrm{D}}$ and $\tilde{\boldsymbol{r}}_{\mathrm{V}} \in \Re^3$ are the directional vectors of the moment arms for the dorsal and ventral wires, respectively. The relationships between the dimensional values and the corresponding nondimentional ones are summarized as follows:

$$u_{\mathrm{D}} = Mg\tilde{u}_{\mathrm{D}} \tag{40}$$
$$u_{\mathrm{V}} = Mg\tilde{u}_{\mathrm{V}} \tag{41}$$
$$\boldsymbol{r}_{\mathrm{D}} = L\tilde{\boldsymbol{r}}_{\mathrm{D}} \tag{42}$$
$$\boldsymbol{r}_{\mathrm{V}} = L\tilde{\boldsymbol{r}}_{\mathrm{D}} \tag{43}$$

Figure 3 shows the movement of an ostrich-inspired arm by pulling dorsal and ventral wires. An ostrich-inspired arm is not made of continuum bodies, but is composed of a serial chain of rigid bodies connected with free joints with two wires as shown in the top drawing. One end of each wire is fixed to the terminal rigid body with the moment arm $\boldsymbol{r}_{\mathrm{D}}$ or $\boldsymbol{r}_{\mathrm{V}}$. Each wire is constrained so as to move along the arm body. The base end of the serial chain is horizontally attached at the origin. The range of each free joint is limited. Therefore, initially, the

arm forms an arc as shown in the bottom-left drawing. By applying appropriate tensions to the wires, the ostrich-inspired arm can form the shape where the head of the arm reaches an upper position as shown in the bottom-right drawing.

The obtained experimental results demonstrate that, although the mechanism is significantly underactuated, a human operator can control this manipulator, such that its tip can reach several positions, including an upper position against gravity, thus indicating the potential of ostrich-inspired manipulators. A more advanced version of ostrich-inspired arms can be found in [47]. To analyze the dynamics of the ostrich-inspired arm, a general model of a wire-driven continuum manipulator will be beneficial [48].

5 Conclusion

Here, I present illustrative examples of soft robots where the rod theory is successfully utilized with appropriate robotics considerations. The future work includes extension of a more general theory handling largely-deromable plates, shells, and three-dimensional bodies, considering other conventional types of soft robots, e.g. soft pneumatic actuators (i.e., balloons), and analysis of normalized and nondimensionalized dynamics of elastic rods for system theory of soft robots across different length/time scales [49].

Acknowledgement. This study was supported in part by JSPS KAKENHI Grant-in-Aid for Scientific Research on Innovative Areas "Science of Soft Robot" Project under Grant JP18H05466, and for Scientific Research (B) JP20H02106.

References

1. Suzumori, K., Fukuda, K., Niiyama, R., Nakajima, K. (eds.): The Science of Soft Robots: Design, Materials and Information Processing. Springer (2023)
2. Antman, S.: Nonlinear Problems of Elasticity, 2nd edn. Springer (2004)
3. Mochiyama, H.: Model Validation of discretized spatial closed elastica. In: Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 5216–5223 (2016)
4. O'Neill, B.: Elementary Differential Geometry, Revised 2nd edn. Academic Press (2006)
5. Murray, R.M., Li, Z., Sastry, S.S.: A Mathematical Introduction to Robotic Manipulation. CRC Press (1994)
6. de Wit, C.C., Siciliano, B., Bastin, G. (eds.): Theory of Robot Control. Springer, London (1996). https://doi.org/10.1007/978-1-4471-1501-4_9
7. Timoshenko, S.P., Goodier, J.N.: Theory of Elasticity, 3rd edn. McGraw-Hill (1970)
8. Mochiyama, H., Shimemura, E., Kobayashi, H.: Shape control of manipulators with hyper degrees of freedom. Int. J. Robot. Res. **18**(6), 584–600 (1999)
9. Chirikjian, G.S., Burdick, J.W.: A modal approach to hyper-redundant manipulator kinematics. IEEE Trans. Robot. Autom. **10–3**, 343–354 (1994)
10. Mochiyama, H.: Whole-arm impedance of a serial-chain manipulator. In: Proceedings of the 2001 IEEE International Conference on Robotics and Automation, pp. 2223–2228 (2001)

11. Mochiyama, H., Suzuki, T.: Kinematics and Dynamics of a cable-like hyper-flexible manipulator. In: Proceedings of the 2003 IEEE International Conference on Robotics and Automation, pp. 3672–3677 (2003)
12. Mochiyama, H., Fujimoto, H.: Robotic manipulation of a hyper-flexible body. In: Proceedings of the 16th IFAC World Congress, 6p. (2005)
13. Mochiyama, H., Fujimoto, H.: Damping manipulation of a hyper-flexible string-like robot. In: Preprints of the IFAC 3rd Workshop on Lagrangian and Hamiltonian Methods for Nonlinear Control, pp. 221–226 (2006)
14. Mochiyama, H., Fujimoto, H.: Impact manipulation by a hyper-flexible robotic manipulator. In: Proceedings of the 2007 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, 6p. (2007)
15. Hatakeyama, T., Mochiyama, H.: Shooting manipulation inspired by Chameleon. IEEE/ASME Trans. Mechatron. **18**(2), 527–535 (2013)
16. Mochiyama, H., Watari, M., Fujimoto, H.: A robotic catapult based on the closed elastica and its application to robotic tasks. In: Proceedings of the: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1508–1513 (2007)
17. Yamada, A., Mochiyama, H., Fujimoto, H.: Kinematics and statics of robotic catapults based on the closed elastica. In: Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3993–3998 (2007)
18. Hirose, S., Kado, T., Umetani, Y.: Tensor actuated elastic manipulator. In: Proceedings of the Sixth World Congress on Theory of Mechanisms, pp. 978–981 (1983)
19. Suzumori, K., Iikura, S., Tanaka, H.: Applying a flexible microactuator to robotic mechanisms. IEEE Control Syst. Magaz. **12**(1), 21–27 (1992)
20. Tsuda, T., Mochiyama, H., Fujimoto, H.: Quick stair-climbing using snap-through buckling of closed elastica. In: Proceedings of the 2012 International Symposium on Micro-NanoMechatronics and Human Science, pp. 368–373 (2012)
21. Misu, K., Yoshii, A., Mochiyama, H.: A compact wheeled robot that can jump while rolling. In: Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 7507–7512 (2018)
22. Yagi, K., Suzuki, K., Mochiyama, H.: Human joint impedance estimation with a new wearable device utilizing snap-through buckling of closed-elastica. IEEE Robot. Automat. Lett. **3**(3), 1506–1513 (2018)
23. Endo, Y., Yagi, K., Mori, Y., Takei, T., Mochiyama, H.: Tele-snap: a joint impedance estimation system using snap motor and openPose for remote rehabilitation diagnosis. Adv. Robot. **37–8**, 528–539 (2023)
24. Arakawa, K., Giorgio-Serchi, F., Mochiyama, H.: Snap pump: a snap-through mechanism for a pulsatile pump. IEEE Robot. Automat. Lett. **6**(2), 803–810 (2021)
25. Nishimura, Y., Takahashi, S., Mochiyama, H., Yamaguchi, T.: Automated hammering inspection system with multi-copter type mobile robot for concrete structures. IEEE Robot. Automat. Lett. **7**(4), 9993–10000 (2022)
26. Keller, H.B.: Numerical Methods for Two-Point Boundary Value Problems. Blaisdell Publishing (1968)
27. Mochiyama, H., Kinoshita, A., Takasu, R.: Impulse force generator based on snap-through buckling of robotic closed elastica: analysis by quasi-static shape transition simulation. In: Proceedings of the 2013 IEEE/RSJ IROS, pp. 4583–4589 (2013)
28. Takano, R., Mochiyama, H., Takesue, N.: Real-time shape estimation of Kirchhoff elastic rod based on force/torque sensor. In: Proceedings of the: IEEE International Conference on Robotics and Automation, pp. 2508–2515 (2017)
29. Nakagawa, N., Mochiyama, H.: Real-time shape estimation of an elastic rod using a robot manipulator equipped with a sense of force. In: Proceedings of the:

IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 8067–8073 (2018)
30. Bretl, T., McCarthy, Z.: Quasi-static manipulation of a kirchhoff elastic rod based on a geometric analysis of equilibrium configurations. Int. J. Robot. Res. **33**(1), 48–68 (2014)
31. Rucker, D.C., Webster III, R.J.: Statics and dynamics of continuum robots with general tendon routing and external loading. IEEE Trans. Robot. **27**(6), 1033–1044 (2011)
32. Till, J., Bryson, C.E., Chung, S., Orekhov, A., Rucker, D.C.: Efficient computation of multiple coupled cosserat rod models for real-time simulation and control of parallel continuum manipulators. In: Proceedings of ICRA2015 (2015)
33. Armanini, C., Dal Corso, F., Misseroni, D., Bigoni, D.: From the elastica compass to the elastica catapult: an essay on the mechanics of soft robot arm. Proc. Roy. Soc. A **473**, 20160870 (2017)
34. Cazzolli, A., Dal Corso, F.: Snapping of elastic strips with controlled ends. Int. J. Solids Struct. **162**, 285–303 (2019). https://doi.org/10.1016/j.ijsolstr.2018.12.005
35. Cazzolli, A., Misseroni, D., Dal Corso, F.: Elastica catastrophe machine: theory, design and experiments. J. Mech. Phys. Solids **136**, 103735 (2020)
36. Fukuda, T., Hosokai, H., Kondo, Y.: Brachiation type of mobile robot. In: Fifth International Conference on Advanced Robotics Robots in Unstructured Environments, vol. 2, pp. 915–920 (1991). https://doi.org/10.1109/ICAR.1991.240556
37. Saito, F., Fukuda, T., Arai, F.: Swing and locomotion control for a two-link brachiation robot. IEEE Control Syst. Mag. **14**(1), 5–12 (1994). https://doi.org/10.1109/37.257888
38. Kajima, H., Hasegawa, Y., Doi, M., Fukuda, T.: Energy-based swing-back control for continuous brachiation of a multilocomotion robot. Int. J. Intell. Syst. **21**(9), 1025–1043 (2006)
39. A Study of the Brachiation Type of Mobile Robot, History of Robotics Research and Development of Japan. https://robogaku.jp/en/history/locomotion/L-1990-1.html
40. Hirose, S.: Biologically Inspired Robots: Snake-Like Locomotors and Manipulators. Oxford University Press (1993)
41. Mori, M., Hirose, S.: Development of active cord mechanism ACM-R3 with agile 3D mobility. In: Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 3, pp. 1552–1557 (2001)
42. Mori, M., Hirose, S.: Three-dimensional serpentine motion and lateral rolling by active cord mechanism ACM-R3. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 1, pp. 829–834 (2002)
43. Nakamura, T., Iwanaga, T.: Locomotion strategy for a peristaltic crawling robot in a 2-dimensional space. In: Proceedings of the 2008 IEEE International Conference on Robotics and Automation (ICRA), pp. 238–243 (2008)
44. Nagoya Institute of Technology. Robot goes running. https://www.youtube.com/watch?v=_swh12QvHRU
45. Mochiyama, H., Gunji, M., Niiyama, R.: Ostrich-inspired soft robotics: a flexible bipedal manipulator for aggressive physical interaction. J. Robot. Mechatron. **34**(2), 212–218 (2022)
46. Misu, K., et al.: Robostrich arm: wire-driven high-DOF underactuated manipulator. J. Robot. Mechatron. **34**(2), 328–338 (2022)
47. Nakano, K., et al.: "RobOstrich" manipulator: a novel mechanical design and control based on the anatomy and behavior of an ostrich neck. IEEE Robot. Automat. Lett. **8**(5), 3062–3069 (2023). https://doi.org/10.1109/LRA.2023.3265301

48. Hsiao, K., Mochiyama, H.: A Wire-driven continuum manipulator model without assuming shape curvature constancy. In: Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 436–433 (2017)
49. Mochiyama, H.: The elastic rod approach toward system theory for soft robotics. In: Preprints of the 21st IFAC World Congress 2020, vol. 53, no. 2, pp. 9175–9180 (2020)



Dynamic Characteristics of a Long-Period, Cable-Stayed Bridge Infrastructure Subjected to Seismic Waveforms

Kakeru Sagara(✉) and Gaku Shoji

University of Tsukuba, Tsukuba, Japan
s2320858@u.tsukuba.ac.jp, gshoji@kz.tsukuba.ac.jp

Abstract. For a three-dimensional nonlinear numerical model of a structural network system consisting of linear elements and nonlinear ones considering elasto-plastic behaviours, phenomenological and geometrical nonlinearities, the coupling vibration mechanism between the different modes of the system was clarified when the stiffness of the dominant structural parameter of the system was varied. The interaction of the dynamic responses of the linear and nonlinear elements to extreme non-stationary disturbances on the structural network system was clarified for the aforementioned changes in the structural parameter. The cable-stayed bridge infrastructure is taken as a representative example of the structural network system, and the axial stiffness of the elastic restraint cables is taken as the dominant structural parameter for the coupling vibration characteristics of the system. The coupled modes of the first swing mode of the girder and the first laterally symmetric bending mode were observed as the axial stiffness of the elastic restraint cables varied. Therefore, changes in the responses were observed in the main structural members of the main girder and cables, and the interactions of nonlinear responses of structural members such as the foundation and the piers were observed.

Keywords: Structural network system · Nonlinear structural elements · Coupling vibration · Nonlinear response interactions

1 Introduction

In structural network systems composed of linear and nonlinear elements, changes in the stiffness of the dominant parameters of the system can cause the coupling vibration between the different modes, altering the vibration characteristics of the entire system and leading to interactions in the dynamic response between linear and nonlinear elements under the action of non-stationary disturbances. It is very important to clarify the nonlinear response of long-period structural network systems to extreme non-stationary excitations such as long-period seismic excitations. A typical example of a structural network composed of linear and nonlinear elements is a cable-stayed bridge.

In this study, the nonlinear vibration of the entire structural system is discussed from the viewpoint of coupling of modes of the above-mentioned multi-degree-of-freedom system. As this theoretical background, we will give an example of a previous study on

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 471–485, 2025.
https://doi.org/10.1007/978-3-031-72794-8_37

the formulation of nonlinear vibration modes. Mikhlin [1] discussed two formulations of nonlinear normal vibrational modes in conservative and near-conservative systems.

The target structural system in this study is modeled with various structural elements such as springs and viscous damping, and nonlinear behavior can occur due to internal resonance between these elements and the overall system. As a theoretical existing study on the internal resonance of such systems, Flosi et al. [2] considered a periodic chain of mass-in-mass cells with pure cubic nonlinearity.

In a previous study from the perspective of response localization due to the coupling of multiple modes in a system formed from nonlinear elements similar to those in this study, Harata et al. [3] employed modal analysis to investigate the localization phenomena in an N-pendulum array subjected to a vertical, harmonic excitation. As for the control of response localization by the coupling of those modes, Kimura and Hikihara [4] discussed the control of intrinsic localized mode in a coupled cantilever array.

This study focuses on the first swing mode of the girder in the target system, which is a typical example of self-excited vibration, as variations in structural parameters may change the vibration mode and cause coupled vibration. As a case study of the theoretical and experimental analysis of the control of such self-excited vibration, Yang and Chao [5] investigated the vibration energy transfer behaviour of forced self-excited systems. Nishijima and Yabuno [6] proposed a method to produce the self-excited oscillations with the first or second mode and both modes in a two-degree-of-freedom spring-mass-damper system.

Furthermore, in this study, vibration control devices such as stoppers and vane dampers installed in structural systems are modeled as nonlinear structural elements and their responses are also focused on. Many previous studies have been conducted on the effects of such nonlinear vibration control devices on the dynamic response of multi-degree-of-freedom oscillators and on theoretical considerations for their control. As an example, Kawasumi et al. [7] considered an autoparametric vibration absorber nonlinearly coupled with the main structure and developed the actual device using a pendulum connected with a torsion bar spring. Also, Zhang and Yabuno [8] revealed the optimum length of a vibration absorber by experiments using a simple cantilever-type autoparametric vibration absorber.

From the academic background above, this study discusses the coupling vibration between different modes of the system when the stiffness of the dominant structural parameter of the system is varied for a three-dimensional nonlinear numerical model of a structural network system composed of linear and nonlinear elements. The interaction of the dynamic response of the linear and nonlinear elements under extreme non-stationary disturbances to the structural network system is clarified for the aforementioned changes in the structural parameter. The cable-stayed bridge is taken as a representative example of a structural network system, and the axial stiffness of elastic restraint cables is taken as the structural parameter that dominates the coupling vibration characteristics of the system. Eigenmode diagrams were analysed from the first to sixth-order modes for a 50% reduction in the axial stiffness of an elastic restraint cable, and the coupling vibration mechanism between the different modes of the system was clarified.

2 Equation of Motion and Modeling of the Damping Matrix

Herein, the non-stationary excitations in this analysis are given by multivariate inputs at each foundation, where the x, y and z components are influenced by observation waves in the longitudinal (LG), transverse (TR) and vertical (UD) directions, respectively. The equation of motion for the distributed-mass non-proportional damping system model is derived as follows.

The absolute displacement matrix of ground Y is an input:

$$\mathrm{X}_1 = \mathrm{X}_0 + \mathrm{X}, \tag{1}$$

where X_1 is the absolute displacement matrix of the structure, X_0 is the absolute displacement matrix of the structure when excited by Y and X is the relative displacement matrix between the structure and the ground. The relationship between the external force P_1 and absolute displacement u_1 of the structure and the relationship between the reaction force P_Y and absolute displacement u_Y of the input are given by the following equation:

$$\begin{Bmatrix} \mathrm{P}_1 \\ \mathrm{P}_\mathrm{Y} \end{Bmatrix} = \begin{bmatrix} \mathrm{K}_{11}\ \mathrm{K}_{12} \\ \mathrm{K}_{21}\ \mathrm{K}_{22} \end{bmatrix} \begin{Bmatrix} \mathrm{u}_1 \\ \mathrm{u}_\mathrm{Y} \end{Bmatrix}, \tag{2}$$

where K_{11} is the summation of the stiffness of the structure and the spring stiffness between the structure and the ground, K_{12} and K_{21} are the spring stiffnesses between the structure and the ground and K_{22} is the summation of the stiffness outside the ground and the spring stiffness between the structure and the ground. Given the external force applied to the structure is $P_1 = 0$, and the absolute displacement of the ground is $u_Y = Y$, then:

$$\mathrm{u}_1 = \mathrm{X}_0 = \left[-\mathrm{K}_{11}^{-1}\mathrm{K}_{12}\right]\mathrm{Y} = \mathrm{rY}, \tag{3}$$

where X_0 is the displacement of the structure when only the displacement of the ground Y is input, and r is called the influence vector.

Now, the equation for the equilibrium of forces during vibration is given as follows:

$$\begin{Bmatrix} \mathrm{M}_1\ddot{\mathrm{X}}_1 \\ \mathrm{P}_\mathrm{Y} \end{Bmatrix} + \begin{bmatrix} \mathrm{C}_{11}\ \mathrm{C}_{12} \\ \mathrm{C}_{21}\ \mathrm{C}_{22} \end{bmatrix} \begin{Bmatrix} \dot{\mathrm{X}}_1 \\ \dot{\mathrm{Y}} \end{Bmatrix} + \begin{bmatrix} \mathrm{K}_{11}\ \mathrm{K}_{12} \\ \mathrm{K}_{21}\ \mathrm{K}_{22} \end{bmatrix} \begin{Bmatrix} \mathrm{X}_1 \\ \mathrm{Y} \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}, \tag{4}$$

where M_1 is the mass matrix of the structure, C_{11} is the summation of the structure damping and the damping between the structure and the ground, C_{12} and C_{21} are the damping between the structure and the ground, and C_{22} is the summation of the damping outside the ground and the damping between the structure and the ground. Given the absolute displacement of the ground Y, the equation for the equilibrium force of the structure is given by Eq. (5):

$$\mathrm{M}_1\ddot{\mathrm{X}}_1 + \mathrm{C}_{11}\dot{\mathrm{X}}_1 + \mathrm{K}_{11}\mathrm{X}_1 = -\mathrm{C}_{12}\dot{\mathrm{Y}} - \mathrm{K}_{12}\mathrm{Y}. \tag{5}$$

Substituting Eq. (3) into Eq. (1) and then substituting Eq. (1) into Eq. (5) yields the following equation:

$$\mathrm{M}_1\ddot{\mathrm{X}}_1 + \mathrm{C}_{11}\dot{\mathrm{X}} + \mathrm{K}_{11}\mathrm{X} = -\mathrm{M}_1\mathrm{r}\ddot{\mathrm{Y}} - \mathrm{C}_{11}\mathrm{r}\dot{\mathrm{Y}} - \mathrm{K}_{11}\mathrm{r}\mathrm{Y} - \mathrm{C}_{12}\dot{\mathrm{Y}} - \mathrm{K}_{12}\mathrm{Y}. \tag{6}$$

Assuming that stiffness K and damping C are proportional, we can obtain Eq. (7a, 7b):

$$C_{11}r\dot{Y} = -C_{11}K_{11}^{-1}K_{12}\dot{Y} = -C_{11}C_{11}^{-1}C_{12}\dot{Y} = -C_{12}\dot{Y} \tag{7a}$$

$$K_{11}rY = -K_{11}K_{11}^{-1}K_{12}Y = -K_{12}Y. \tag{7b}$$

Finally, substituting Eq. (7a, 7b) into Eq. (6) yields Eq. (8):

$$M_1\ddot{X} + C_{11}\dot{X} + K_{11}X = -M_1 r\ddot{Y}. \tag{8}$$

where $\ddot{Y}$ is the input ground dynamic acceleration. Before solving the equation of motion (8), the balancing of the stress field due to the own weight of the structural elements of the system is taken into account.

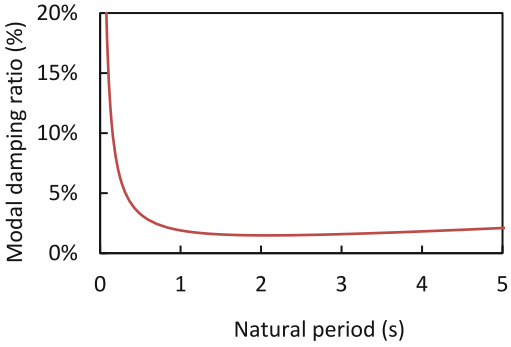


Fig. 1. The modal damping constant for Rayleigh damping

The Newmark β method, with $\beta = 0.25$, was used for the numerical integration to solve the equation of motion in (8), with an integration interval of 0.004 s. The damping matrix in the equations of motion was assumed to be Rayleigh damping.

The damping ratio for the upper parts of the main girder and tower was set to 0.02, for the piers to 0.05 and for the foundation-ground spring to 0.2. The modal damping ratio was calculated based on the strain-energy proportional damping approach and the damping matrix was modelled. Herein, the constants α_0 and α_1 were defined according to Eq. (9).

$$\begin{Bmatrix} \alpha_0 \\ \alpha_1 \end{Bmatrix} = \frac{2\omega_1\omega_2}{\omega_2^2 - \omega_1^2} \begin{bmatrix} \omega_2 & -\omega_1 \\ -\frac{1}{\omega_2} & \frac{1}{\omega_1} \end{bmatrix} \begin{Bmatrix} \xi_1 \\ \xi_2 \end{Bmatrix} \tag{9}$$

where (ω_1, ω_2) and (ξ_1, ξ_2) are the natural angular frequencies and damping ratios for setting Rayleigh damping. (ω_1, ω_2) and (ξ_1, ξ_2) are selected from the two dominant vibration modes of the system, thereby constructing the matrix of Rayleigh damping. From this, the natural angular frequencies and modal damping ratios for the 1st-order mode, which represents the first symmetric bending mode, and the 11th-order mode,

Table 1. Modal analysis results

	Natural vibration mode	Natural period (s)
1	1st symmetric bending mode	4.811
2	1st lateral symmetric bending mode	3.925
3	1st asymmetric bending mode	3.455
4	1st swing mode of a girder	2.831
5	1st torsional mode	2.001
6	2nd symmetric bending mode	1.953
7	2nd asymmetric bending mode	1.658
8	1st lateral asymmetric bending mode	1.542
9	3rd symmetric bending mode	1.470
10	1st bending mode of P2	1.431
11	1st bending mode of P3	1.359

which represents the first bending mode of P3 main tower, were adopted for the modeling, based on the results of the eigenvalue analysis presented below Table 1. The modal damping ratio obtained by the Rayleigh damping modeling described above varies with the natural period as shown in Fig. 1.

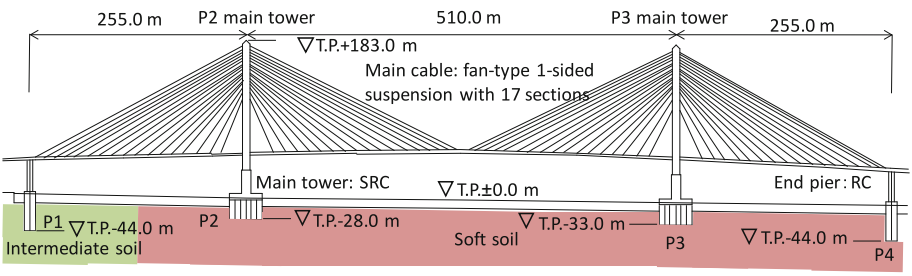


Fig. 2. Structural network system consisting of linear and nonlinear elements

3 Analysed System and Analysis Model

As a case study of a structural network system consisting of linear and nonlinear elements, the Tsurumi Tsubasa Bridge, a cable-stayed bridge shown in Fig. 2, was taken up. A three-dimensional nonlinear frame model of the bridge which is equipped with various vibration control devices was constructed based on the findings of Yamamoto et al. [9], the design data of the Metropolitan Expressway Company [10] and the findings of Utsunomiya and Shoji [11] as shown in Fig. 3.

As a linear element, the main tower, main girder and end piers were modelled using linear beam elements. The steel main cable was modelled as a truss element with six different cross sections transmitting only axial forces in the tensile direction. Since the main cable has a sag due to its own weight, geometric nonlinearity must be taken into account, and the axial stiffness of the 34 cables per main tower was modelled using Ernst's modified elastic modulus E_t [12] shown in Eq. (10).

$$E_t = \frac{E_c}{1+\frac{(\gamma_s L_H)^2}{12\sigma_d^3}E_C} \tag{10}$$

where E_c is Young's modulus of the steel, γ_s is the unit volume weight of the cable, L_H is the horizontal projected length of the cable and σ_d is the stress due to dead load on the cable. The initial tension of the main cable was introduced at 1.35×10^6 to 8.99×10^6 N, referring to the literature [10]. Figure 4 shows the distribution of the modified Young's modulus for each cable obtained by Eq. (10). The cable number is No. 1 on the end pier side and No. 34 on the centre span side. Pendel, vertical and horizontal bearings were modelled as truss and spring elements, with sufficiently large cross-sectional areas and fixed directional spring constants. The elastic restraint cables are regularly located locally in the system and are dominant for the coupling vibration characteristics of the system, and they were modelled as linear truss elements, with a maximum initial tension of 6.4×10^6 N. Four cables (approximately 117 m each) were installed per the main towers.

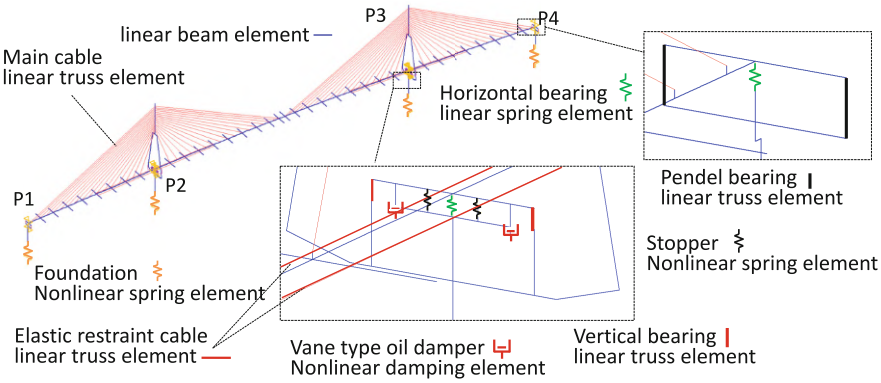


Fig. 3. The analysis model

Next, nonlinear elements are explained. The piers from the main tower transverse beam to the top of the foundation were modelled using nonlinear beam elements. The relationship between the bending moment (M) and curvature (φ) of the piers at a main tower was modelled using the Takeda [13] model. As shown in Fig. 5(a), it considers the fracture process of core concrete restrained by steel bars and the fourth-order slope, i.e., material nonlinearity. The model between the foundation and ground was a Sway-Rocking model, which considered translational and rotational springs and their coupling in the direction and perpendicular to the bridge axis, based on the literature [9]

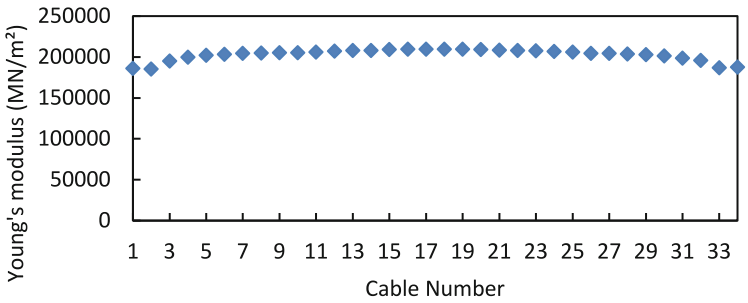


Fig. 4. Modified Young's modulus for each cable

and Metropolitan Expressway Company [10]. As shown in Fig. 5(b), the nonlinearity was similar to a nonlinear beam element and the load-displacement and moment-rotation angles were considered up to the fourth-order gradient, which expresses the different material nonlinearities. As shown in Fig. 5(c), the vane damper was modelled using a nonlinear viscous element and the nonlinearity was modelled as bilinear which took into account the nonlinearity of the viscosity. As shown in Fig. 5(d), the stopper was modelled using a nonlinear spring element that started to act at 0.45 m intervals and the nonlinearity was modelled using a trilinear element representing phenomenological nonlinearity like pounding between the elements. The displacement dependence of tension and compression was symmetrical.

The boundary conditions were fixed in all directions, with spring fixed points between the foundation and the ground. These were fixed in the vertical direction and around the vertical axis at the spring attachment points among the base of the piers, foundation and ground. The other nodes were free.

4 Characteristics of Nonlinear Response Interactions

The waveforms shown in Fig. 6, estimated from those of the Great Kanto earthquake of 1923 by finite difference method, were applied to the subject bridges at multiple points. Strong ground motions caused by megathrust earthquakes have a long-period component of more than 2 s and a long duration. The ground motion in this analysis was given by the multi-variant input at each foundation, as described earlier.

4.1 Changes in Natural Vibration Characteristics

The dominant parameter with respect to the coupling vibration characteristics of a three-dimensional nonlinear frame model of a cable-stayed bridge, consisting of nonlinear and linear elements, is the axial stiffness of the elastic restraint cables which are the linear structural elements.

Figure 7 shows the load-displacement relationship when the axial stiffness varies by decreasing the cross-sectional area of the elastic restraint cable. The decreasing rate denotes a minus percentage value like −50% and varies from −0% (standard model) to −100%. Figure 8 shows the variation in the natural period of the system with respect to

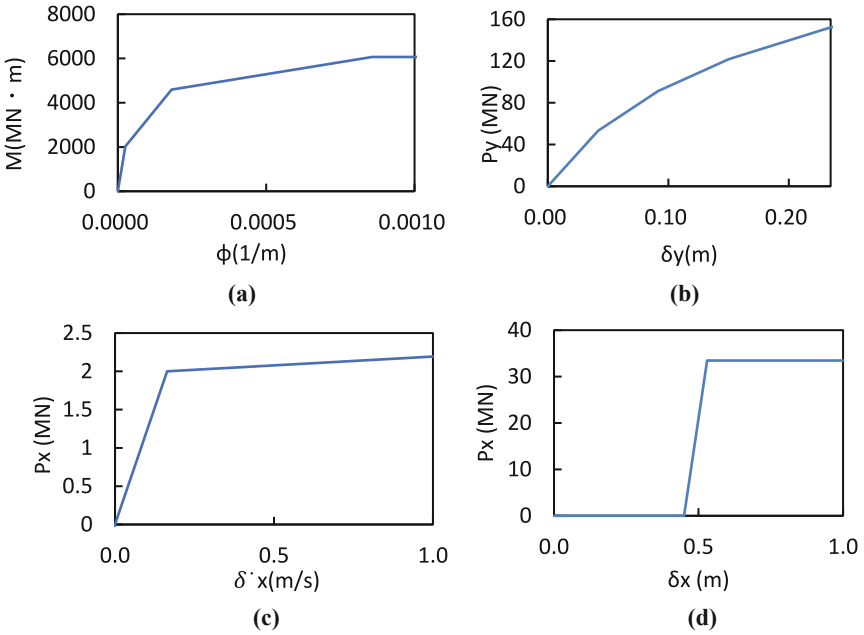


Fig. 5. Nonlinear characteristics of nonlinear elements. **(a)** M−φ relationship around the element axis at the base of the main tower P3 pier. **(b)** Py−δy relationship of the P4 position foundation. **(c)** Vane damper Px−$\dot{\delta}$ x relationship. **(d)** Stopper Px−δx relationship.

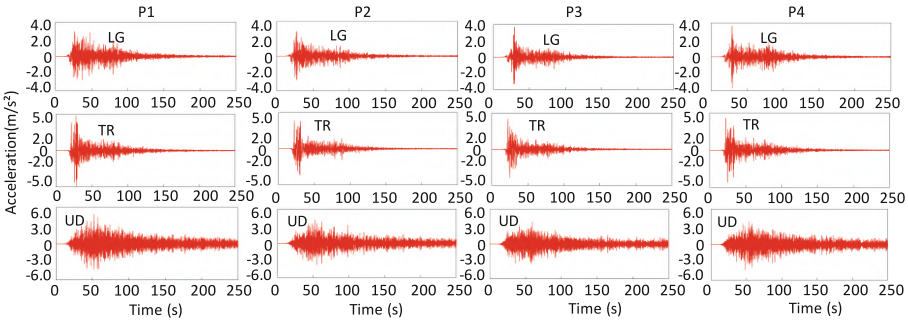


Fig. 6. Input acceleration on an engineering basis

the variation in the axial stiffness of the elastic restraint cable. Figure 7 shows that the response displacement of the elastic restraint cable increases from 0.17 m to 0.20 m as the axial stiffness of the cable decreases. In the −50% case, the response of the elastic restraint cable is reduced compared to the other two cases. Figure 8 shows that from −40% to −50% axial stiffness of the elastic restraint cable, the first swing mode of the girder approaches and slightly exceeds the first laterally symmetric bending mode. From −50% to −70%, the first swing mode of the girder approaches the first symmetric bending mode. When the axial stiffness is further reduced (corresponding to −100%),

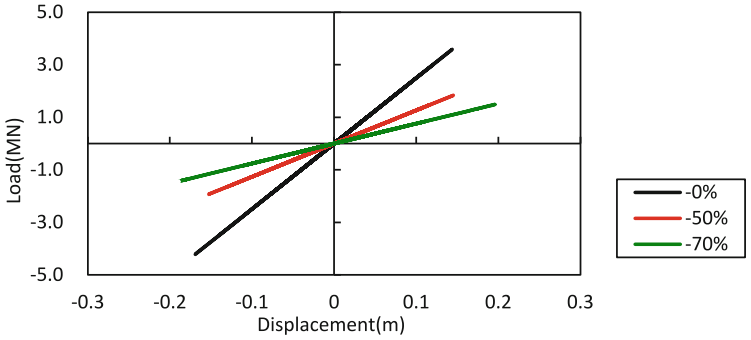


Fig. 7. Variation in the load-displacement relationship of the elastic restraint cables

however, without elastically restraint cables, the eigen-period of the first swing mode of the girder increases further, reaching 14.237 s, making the first swing mode of the girder the overall first-order mode. This indicates that the first swing mode of the girder is substantially sensitive to changes in the axial stiffness of the elastic restraint cable.

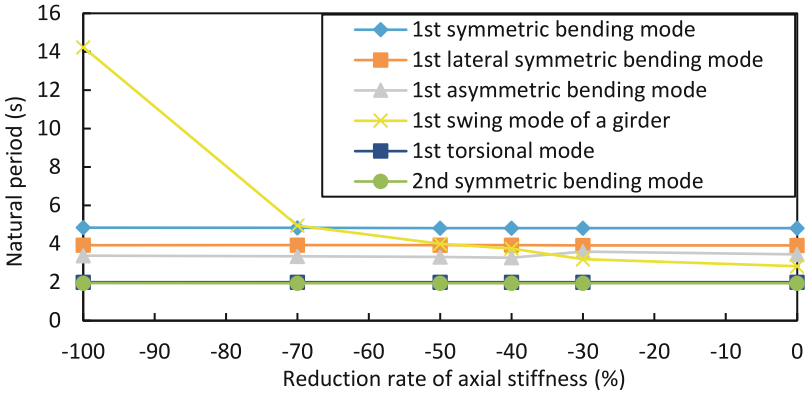


Fig. 8. Rate of decrease in the axial stiffness and natural period of the elastic restraint cable

4.2 Interaction of Major Structural Components Within a Structural Network System

Figure 9 shows the time histories of the response accelerations at the structural elements which are the main structural elements in a structural network system modelled by linear beam elements when the axial stiffness of the elastic restraint cable is varied. Herein, the responses of the centre span deck and the P3–P4 deck were considered when the coupling of the above modes occurred.

On the centre span deck, the response in the LG direction was maximum in the −50% axial stiffness case, 1.12 times that of the standard model. This is due to the

change in the action of the vane damper, which is discussed later. The response in the UD direction increases as the axial stiffness of the elastic restraint cable decreased. The magnitude of the increase was greatest from −0% to 50% axial stiffness, which was approximately 1.12 times greater. This change in response was because the amplitude in the UD direction of the first swing mode of the girder at −50% axial stiffness was larger than that of the first swing mode of the girder in the standard model due to the coupling of first swing mode of the girder and first lateral symmetric bending mode.

In the deck from P3–P4, the response in the LG direction tended to decrease as the axial stiffness of the elastic restraint cable decreased. In the UD direction, the degree of increase in response acceleration is greater from −0% to −50% and −70% axial stiffness. On the LG direction, differences are observed from the centre span deck.

Figure 10 shows the relative displacement between the two nodes of each truss element for the main cable connecting the top of the P3 main tower to the centre span (inside cable) and the main cable connecting the P3 main tower to the P4 end (outside cable). The relative displacements change as the axial stiffness of the elastic restraint cables decreases. Their absolute values are basically small and are not at a level where the geometric nonlinearity of the main cables is noticeable. Then, the relative displacement between the two nodes of the main cables, both inside and outside, has a maximum in the LG direction with the −50% axial stiffness. The response increases in the UD direction when the axial stiffness of the elastic restraint cable decreases. The inside cable is more related to the response of the centre span deck, while the outside cable is more related to the response of the deck from P3–P4, with dynamic interaction of the structural responses as described above.

4.3 Responses of Structural Elements with Different Material Nonlinearities

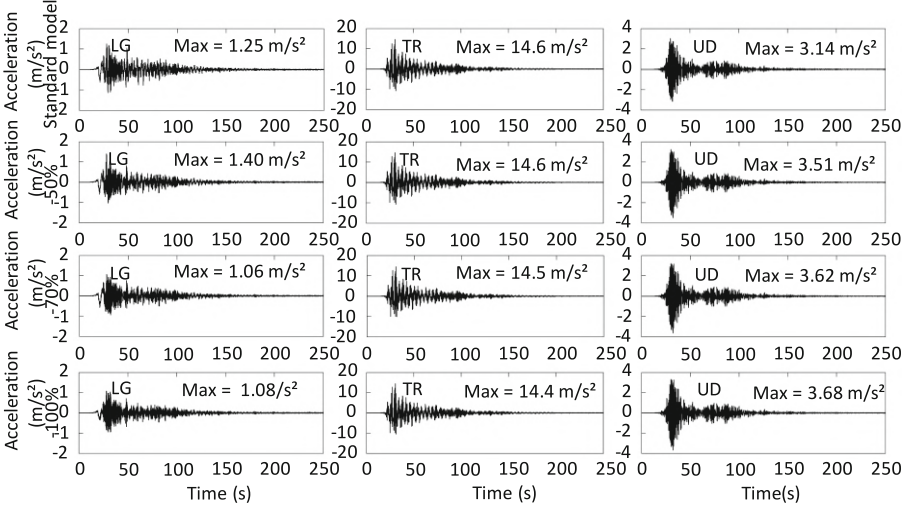
As the behavior of structural elements with different material nonlinearities, Fig. 11 shows the $M-\theta$ relationships for the P3 main tower pier and $N-\delta$ and $M-\theta$ relationships for the P3 foundation-ground spring. In the beam elements of the piers, the circumferences of the x-, y- and z-axes correspond to the UD, TR and LG axes, respectively.

The history of the P3 main tower pier, a nonlinear beam element, shows significant changes in the response around the TR and UD axes due to changes in the axial stiffness of the elastic restraint cables. Moments are smaller in the −50% axial stiffness case than the other three cases.

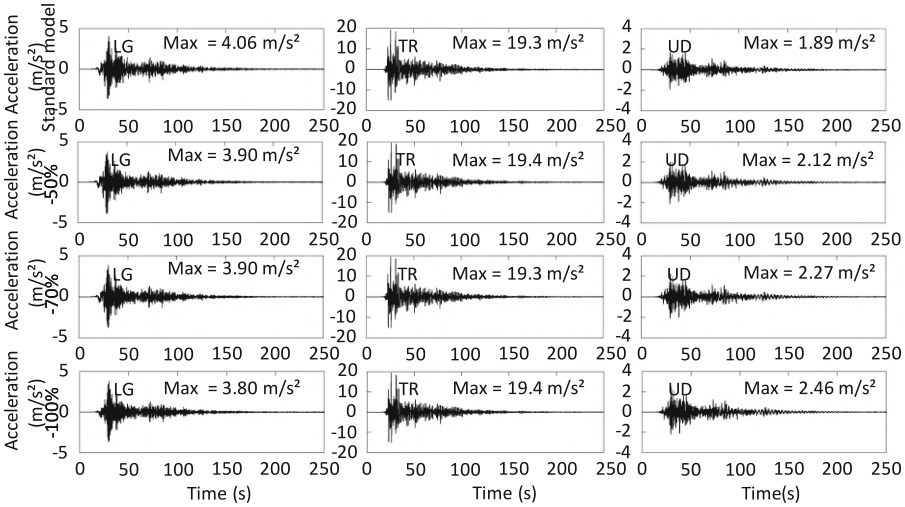
In addition, in the foundation-to-ground spring at position P3, the element x-, y- and z-axes correspond to the LG, TR and UD axes, respectively. At position P3, regarding the foundation-to-ground spring history, the change in load in the LG direction and change in moments around the LG axis are remarkable. Although moments in the TR direction and around the TR axis are nonlinearized, the change in response owing to the changes in the axial stiffness of the elastic restraint cable is minimal.

4.4 Interaction of Vibration Control Device Responses

As the behavior of structural elements with nonlinear viscosity and phenomenological nonlinearity, Fig. 12 shows the damping force-velocity relationship in the LG direction



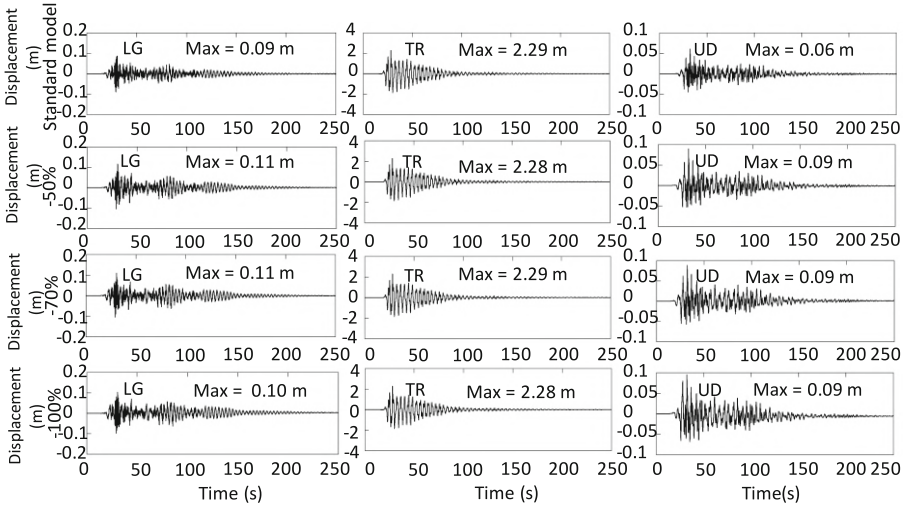
(a) Centre span deck



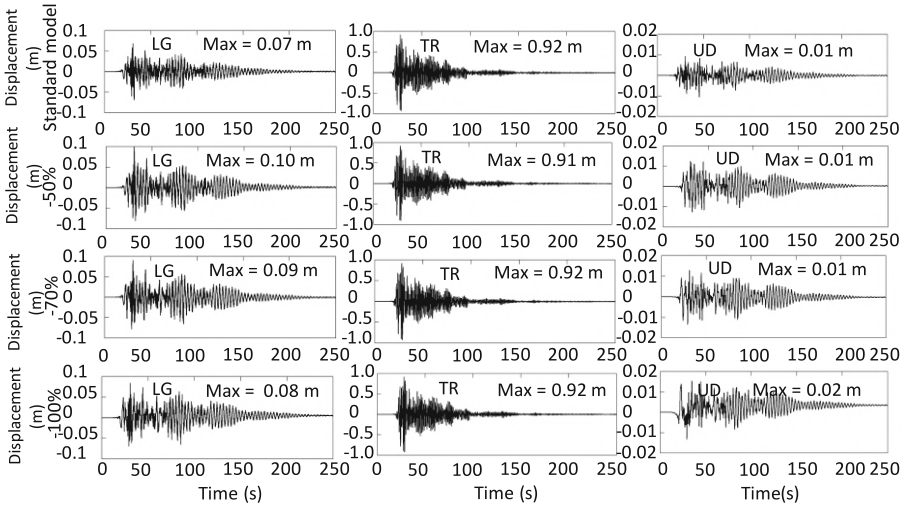
(b) Deck from P3-P4

Fig. 9. Sensitivity of the response acceleration time histories to changes in the axial stiffness of the elastic restraint cables

for the vane damper and the load-displacement relationship in the LG direction for the stopper. The vane oil damper with an axial stiffness of −50% has a smaller damping force than the other three cases. This is owing to the suppression of the amplitude in the LG direction in the first swing mode of the girder, resulting in a large response in the LG direction at the centre span deck. As the axial stiffness of the elastic restraint cable decreases, the displacement of the stopper increases. However, in none of the cases it reaches 0.45 m, where the stopper acted. This response change occurred because the



(a) Inside cable

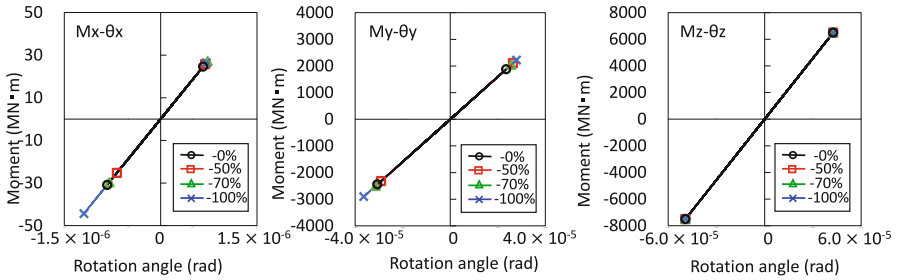


(b) Outside cable

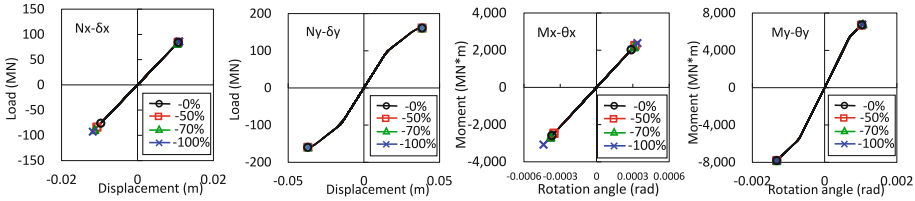
Fig. 10. The relative displacement between the two nodes of the main cable

stopper restrained the displacement of the elastic restraint cable, corresponding to the displacement of the elastic restraint cable.

To completely demonstrate the effectiveness of these vibration control devices, verifying the amplitude reduction theory is required, which is examined by Kawasumi et al. [7] and Zhang and Yabuno [8].

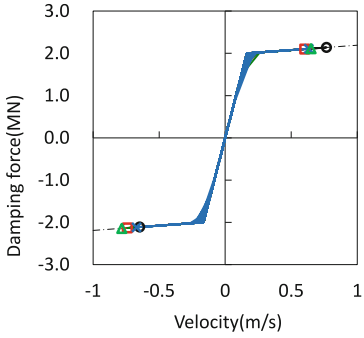


(a) M-θ relationship of the P3 main tower pier

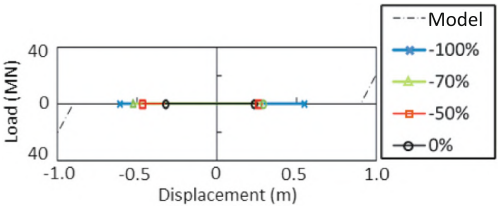


(b) Response of the P3 foundation-subsurface spring position

Fig. 11. Response of the piers and foundation-to-ground springs



(a) Response of a vane damper



(b) Load-displacement relationship of a stopper

Fig. 12. Response of vibration control devices

5 Conclusion

This study revealed the coupling vibration of different modes in a three-dimensional nonlinear numerical model of a structural network system composed of linear and nonlinear elements when the dominant structural parameter is varied. The associated interactions of the nonlinear structural elements and the resulting sensitivity of the main linear structural elements of the system to extreme non-stationary excitations were also examined.

When the axial stiffness of the elastic restraint cable which is the dominant structural parameter for the coupling vibration characteristics of the system was −50%, the first swing mode of a girder approached the first lateral symmetric bending mode. This

resulted in a decreased amplitude in the LG direction and an increased amplitude in the UD direction of the first swing mode of the girder. The response around the TR axis of the P3 main tower pier increased. Response in the LG direction and around the LG axis of the foundation-to-ground springs, which also have different material nonlinearities, increased. The damping force of the vane damper, which is a nonlinear element with nonlinear viscosity, decreased. The maximum displacement of the stopper, which is a nonlinear element with phenomenological nonlinearity like pounding between the elements, increased. These interactions increase the response in the UD direction on the P3–P4 deck and in the LG and UD directions on the centre span deck. The inside cable is more related to the response of the centre span deck, while the outside cable is more related to the response of the deck from P3–P4.

Acknowledgements. This work was supported by JSPS KAKENHI Grant Number 22H01569. The authors would like to thank MARUZEN-YUSHODO Co., Ltd. (https://kw.maruzen.co.jp/kousei-honyaku/) for the English language editing.

References

1. Mikhlin, Y.V.: Two formulations of nonlinear normal vibration modes and their applications. In: Wiercigroch, M., Rega, G. (eds.) IUTAM Symposium on Nonlinear Dynamics for Advanced Technologies and Engineering Design. IUTAMBOOK, vol. 32, pp. 31–45. Springer Dordrecht (2012). https://doi.org/10.1007/978-94-007-5742-4_3
2. Flosi, J., Savadkoohi, A., Lamarque, C.: Inter-modal interactions in a chain of mass-in-mass non-linear oscillators. In: IUTAM Symposium on Nonlinear Dynamics for Design of Mechanical Systems Across Different Length/Time Scales, pp. 230–231 (2023)
3. Harata, Y., Ikeda, T., Terada, H.: Modal analysis for localization phenomena of parametric resonance in pendulum arrays subjected to vertical excitation. In: IUTAM Symposium on Nonlinear Dynamics for Design of Mechanical Systems Across Different Length/Time Scales, pp. 262–263 (2023)
4. Kimura, M., Hikihara, T: Control of intrinsic localized mode in coupled cantilever array. In: Wiercigroch, M., Rega, G. (eds.) IUTAM Symposium on Nonlinear Dynamics for Advanced Technologies and Engineering Design. IUTAMBOOK, vol. 32, pp. 359–374. Springer, Dordrecht (2012). https://doi.org/10.1007/978-94-007-5742-4_28
5. Yang, J., Chao, C.: Vibration energy flow behaviour of self-excited dynamic systems. In: IUTAM Symposium on Nonlinear Dynamics for Design of Mechanical Systems Across Different Length/Time Scales, pp. 186–187 (2023)
6. Nishijima, M., Yabuno, H.: Selective self-excited oscillations in a two-degree-of-freedom spring-mass-damper system in two modes by using band-pass filter. In: IUTAM Symposium on Nonlinear Dynamics for Design of Mechanical Systems Across Different Length/Time Scales, pp. 190–191 (2023)
7. Kawasumi, Y., Yasui, K., Yabuno, H.: Experimental study of vibration control using autoparametric vibration absorber. In: IUTAM Symposium on Nonlinear Dynamics for Design of Mechanical Systems Across Different Length/Time Scales, pp. 172–173 (2023)
8. Zhang, C., Yabuno, H.: Influence of viscous damping in autoparametric vibration absorber on its performance at the anti-resonance point. In: IUTAM Symposium on Nonlinear Dynamics for Design of Mechanical Systems Across Different Length/Time Scales, pp. 174–175 (2023)

9. Yamamoto, Y., Fujino, Y., Yabe, M.: Comparison of seismic records with seismic response of long-span cable-supported bridges identified using a dynamic analytical model. J. JSCE A **65**(3), 738–757 (2009)
10. Metropolitan Expressway Public Corporation: Construction historical document on Tsurumi Tsubasa Bridge, pp. 5-100–5-111 (1994)
11. Utsunomiya, H., Shoji, G.: Seismic response characteristics of a long-period and long-span bridge based on recorded waveforms. In: 17th World Conference on Earthquake Engineering (17WCEE), Sendai, Japan, Paper No. C001667 (2020)
12. Ernst, H.: The modulus of elasticity of cable taking into account of catenary action. Der Bauingenieur **40**(2), 52–55 (1965)
13. Takeda, T., Sozen, M.A., Nielsen, N.N.: Reinforced concrete response to simulated earthquakes. J. Struct. Div. **96**(12), 2557–2573 (1970)



Feasibility Studies of Using Vibration Sensor to Monitor Tool Wear Progress During Finishing Step of Hardened Steel Machining

Jonny Herwan(✉), German Herrera-Granados, Ichiro Ogura, Yoshiyuki Furukawa, Takashi Misaka, and Hitoshi Komoto

Industrial Cyber-Physical Systems Research Center, AIST, Tsukuba, Japan
jonny.herwan@aist.go.jp

Abstract. In conventional mold and die manufacturing, the machining process is conducted in the annealed state (lower hardness of steel) followed by heat treatment, electric discharge machining (EDM), and finishing by surface grinding or polishing. Using high-speed machining (HSM), it is possible to perform the machining in the hardened state, eliminating the heat treatment and EDM at the middle stage, and even perform a better surface quality, thus reducing the effort for grinding or polishing. However, in HSM, tool wear is a serious issue and requires a robust monitoring system to avoid defective products. Furthermore, there is a challenge in tool monitoring for this application where the depth of cut is very small with a high cutting speed, thus difficult to capture the cutting process through a sensor compared to conventional machining. This work is addressed to conduct feasibility studies of using vibration sensors to monitor tool wear progress. Five positions of sensors are investigated: close to the workpiece, Y-axis slideways, spindle housing, and two direct motors for workpiece rotation. The signals are extracted into several features; maximum, root mean square (RMS), mean, standard deviation, kurtosis, skewness (in the time domain), and peak spectral amplitude (in the frequency domain). The best feature is selected using Pearson correlation analysis, then plotted together with the tool flank wear length data to confirm the capability of tool wear monitoring. The results showed that the tool wear progress can be monitored using the RMS value of signals from a sensor embedded close to the workpiece.

Keywords: High-speed machining · Hardened steel · Vibration sensor

1 Introduction

High-speed machining (HSM) of hardened steel promises a more efficient manufacturing process such as in mold and die manufacturing. The HSM technique eliminates the heat treatment that is required after conventional machining (machining in the annealed state of steel). Furthermore, HSM also provides a better surface roughness, thus eliminating or significantly reducing the polishing effort [1].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 486–495, 2025.
https://doi.org/10.1007/978-3-031-72794-8_38

Tool wear becomes a serious issue during HSM of hardened steel and requires robust control and monitoring. Tool wear monitoring could be done by direct methods, such as by investigating the cutting edge using a high-speed camera, or by indirect methods, usually conducted using several types of sensors e.g.: dynamometer (force sensor), acoustics emission sensor (AE), accelerometer (vibration sensor), and electric current sensor. From the previous experience, the dynamometer showed the best performance on detecting tool wear, but it is difficult to use it to retrofit the machine tool due to its size and the requirement of assembling it under the workpiece, representing a potential problem for the motion of the machine tool itself, as shown in Fig. 1 (a). On the other hand, embedded sensors such as AE sensors or accelerometers can be easily fixed in other parts of the machine tool allowing a higher flexibility for the monitoring. In our previous experiment, the AE sensor showed a better capability compared to the accelerometer. However, the sampling rate of the AE sensor represents a challenge due to its higher sensitivity (during the cutting process it can reach up to 800 kHz compared to the 12.5 kHz of the accelerometer), hence it requires a large size of data storage and a more complex data processing system. Meanwhile, the electric current sensor showed a low capability to capture the progress of tool wear. Therefore, in this study, the capability of the accelerometer on detecting the tool wear progress from several positions is investigated.

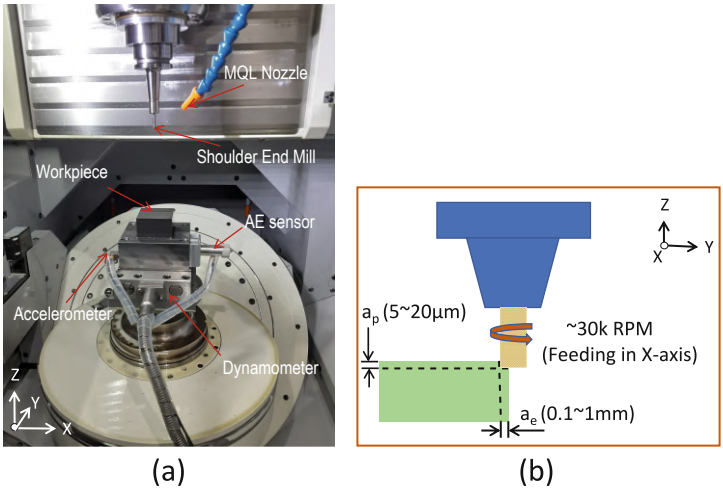


Fig. 1. (a) Common sensors set-up in a laboratory environment, (b) Actual condition of finishing mold manufacturing using HSM

2 Literature Review

In the previous work, an accelerometer has been successfully used in monitoring of turning process [2–5] including high-speed turning [3–5]. In these studies, accelerometers have also proven to have capability for tool wear monitoring in positions outside

the cutting area. However, the tool monitoring during the finishing step of hardened steel machining for mold and die using HSM is much more difficult since it is usually conducted with a very low depth of cut (around 20 microns) while the spindle rotation could reach 30,000 rpm as illustrated in Fig. 1(b). In such conditions, the vibration of the spindle overwhelms the vibration created by the cutting process.

Based on the latest reviews about tool wear monitoring during the milling process [6, 7], only a few works focus on the machining of hardened steel with a maximum depth of cut of 0.1 mm [8–11]. Three of them [8–10] are focused on micro milling processes where AE sensors and dynamometers are used for the monitoring. Meanwhile, another paper [11] used indexable and solid ball end mills (each type with diameters of 8 and 12 mm), and a dynamometer was also used to monitor the tool wear. The use of an accelerometer has not been much explored yet, and thus become the focus of this work.

Additionally, the monitoring procedure can be divided into three steps: sensor selection and data acquisition, signals processing and features extraction, and finally decision making about classifying worn tool or predicting remaining useful life (RUL) of the tool. In the step of signal processing and feature extraction, time domain, frequency, and time-frequency domain (wavelet) are usually used. Meanwhile, the decision-making process is usually conducted using deep learning or machine learning process.

3 Research Methodology

3.1 Experimental Set-up

To investigate the performance of the accelerometer to monitor tool wear progress from different sensor positions, five sensor locations were selected as shown in Fig. 2. Two sensors were used during the experiment (Triaxial CCLD accelerometers, 10 mV/g from Bruer & Kjaer), one sensor was always located under the workpiece (location No. 1) as a reference, meanwhile, another sensor was placed alternately from location No. 2 to No. 5. The sensor positions from No. 1 to No. 5 are under the workpiece, at the Y-axis slideway, at the spindle, under the C-axis table, and at B-axis table, respectively.

The experiment was conducted using a 5-axis machining center Makino D200Z, with a maximum spindle speed of 30,000 rpm. The workpiece material was SKD11 (cold work tool steel) with a hardness of 62 HRC. A coated carbide endmill with a diameter of 4 mm and corner radius of 0.1 mm from Moldino (EPDRF4040-12-01-TH) was used with the following cutting conditions: 70 m/min cutting speed (v) which equal to 5,570 rpm, 0.03 mm/tooth feed rate (f), 0.2 mm radial depth of cut (ae) and 0.02 mm axial depth of cut (ap). The workpiece cutting area was 50 × 50 mm and divided into four sections with a size of 12.5 × 50 mm. The tool wear length and workpiece surface roughness were measured after finishing each section which equaled to 64 cutting cycles/section. Note that the 64 cutting cycles are produced by the CAM software considering the tool corner radius of 0.1 mm. The tool flank wear length was measured using a portable optical microscope (Anity 3R Systems with 200x magnifications).

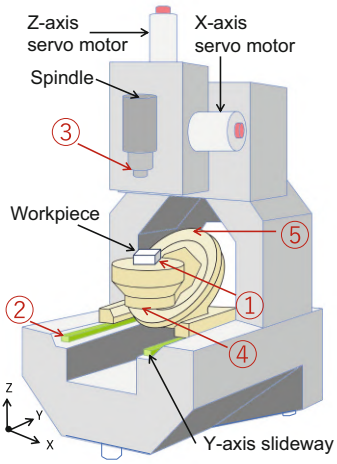


Fig. 2. Selected accelerometer positions for tool wear monitoring

3.2 Signals Recording and Processing

The accelerometers from Bruel & Kjaer type 4520-001 with a sensitivity of 1.02 mV/ms^2 were used in this work. The data was recorded using the National Instrument data acquisition module type NI-9230 with a sampling rate of 12.8 kHz. LabVIEW software was used for the signal recording and processing. The signals were extracted into the following features: maximum value, root means square, mean, standard deviation, skewness, and kurtosis (all in the time domain, following Eqs. 1–6, respectively). Meanwhile, peak spectral amplitude was also extracted as a feature in the frequency domain.

$$x_{max} = \max(x_i) \tag{1}$$

$$x_{rms} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \tag{2}$$

$$x_{mean}(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3}$$

$$x_{std.dev}(\sigma) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{4}$$

$$x_{skew} = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma}\right)^3 \tag{5}$$

$$x_{kurt} = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma}\right)^4 \tag{6}$$

The best feature was then selected using Pearson correlation (Eq. 7).

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{7}$$

where r is the correlation coefficient, x_i are the values of the x-variable in a sample. Meanwhile, $\bar{x}$, $\bar{y}$ are the mean value of the values of the x-variable and y-variable, respectively. The closer the r values to 1 or -1 the stronger the correlation between the two variables. In this case, the signal feature was one variable, and the tool flank wear length was the other variable.

Although there is a trend to use deep learning or machine learning for decision-making, in this work we limited the study to verifying the correlation of signal features with the progressive tool flank wear using Pearson correlation and visually checking it by plotting the best feature vs. tool flank wear. In our opinion, this step is the critical point before applying deep or machine learning where the process is in a black box.

3.3 Experimental Procedure

The experiment was conducted following the steps shown in Fig. 3. The tool wear length measurement was conducted every completing each cutting section (50 × 12.5 mm). If the tool flank wear length was under 0.15 mm, the experiment was repeated. The maximum value of tool flank wear length of 0.15 mm was used considering tool geometry where the flank face length is similar to this value.

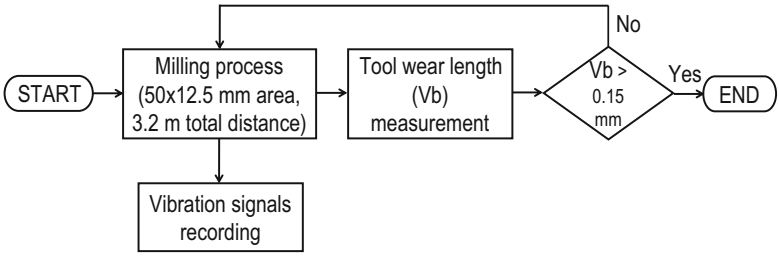


Fig. 3. Experimental flow

4 Results and Discussion

The progressive tool flank wear length measured using an optical microscope is exhibited in Fig. 4. It seems almost no wear appears till a cutting distance of 12.8 m. After this point, the wear amount gradually increased, and the maximum value of tool flank wear length was reached after 51.2 m cutting distance which means after milling the workpiece cutting area (50 × 50 mm) four times.

The respective vibration signals from an accelerometer under the workpiece after 25.6 m and 51.2 m cutting distance are shown in Fig. 5. The left figures show the signals in the time domain, and it is clearly exhibited that amplitude of signals is increased with the increasing tool flank wear. Similarly, in the right figures (signals in frequency domain), the peak spectral amplitude of signals is also increased with the increasing wear amount.

Furthermore, the above-mentioned features are extracted from the signals, the best signal features are investigated using Pearson correlation analysis (Eq. 7), and the results are shown in Fig. 6. From the sensor under the workpiece (Fig. 6a), the analysis indicates that the RMS, the standard deviation, and the peak spectral amplitude are the best features with similar values in all axes (0.98). Meanwhile, from the sensor at the Y-axis slideway (Fig. 6b), the strongest correlation is the maximum value, with Pearson correlation coefficient of 0.91, followed by the mean value (−0.7).

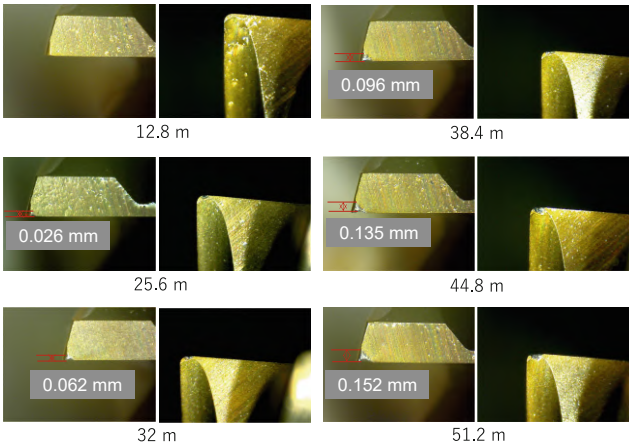


Fig. 4. Tool flank wear lengths with respective cutting distances (top and front view)

To verify the actual correlation, the best features suggested by the Pearson correlation analysis are plotted together with the progressive tool wear length in Fig. 7. The RMS values and standard deviation of time domain signals from the sensor under the workpiece are proved to have a strong correlation with the progressive tool wear as shown in Fig. 7(a). Meanwhile, the best features of the signals from the sensor at Y-axis slideway do not show a good correlation with the progressive tool wear length as depicted in Fig. 7(b), the max-feature in Y-axis and mean-feature in X-axis were selected to represent the highest coefficients from different signal features. As mentioned earlier the visualization of suggested best features vs. progressive tool wear is an important step and was proved that although the Pearson correlation coefficient is relatively high (0.91), the visualization check shows that the actual correlation is still weak.

The performance of the accelerometer at the other three positions (No. 3–5 in Fig. 2) is exhibited in Fig. 8. Instead of evaluating the best features, the investigation was conducted in a basic analysis of whether the signals can recognize the actual cutting process and the air cutting. In position No. 1 (under the workpiece), the three cutting cycles can be easily identified in comparison with the non-cutting movements. However, on the other four signal sets, where three cutting cycles must be identifiable, the accelerometer cannot detect the actual cutting process. At the spindle (position No. 3), the vibration amplitude is much higher than the one under the workpiece which means that spindle vibration overwhelms the cutting vibration. Furthermore, in locations No. 2 and 5, the

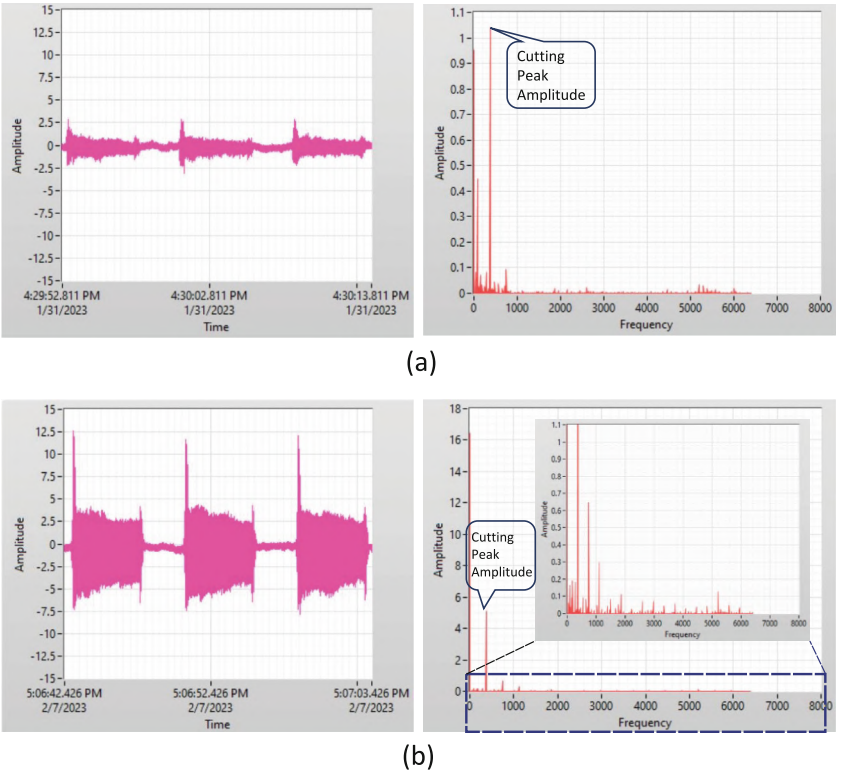


Fig. 5. Signals in the Time and Frequency domain at a cutting distance (a) 25.6 m (b) 51.2 m

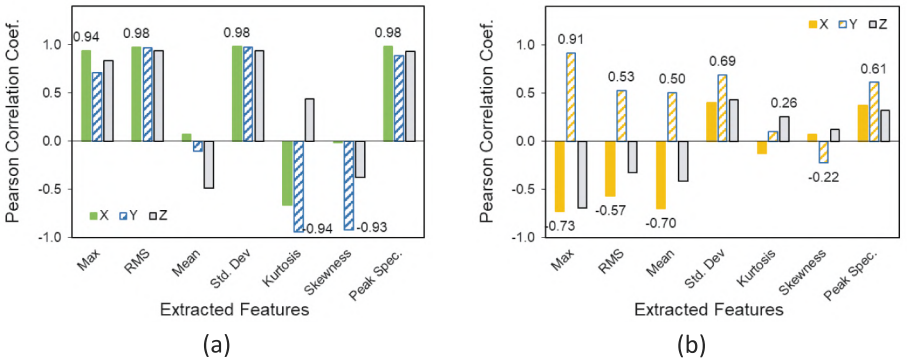


Fig. 6. Pearson correlation analysis of signal features from the accelerometer (a) Under the workpiece (b) at Y-axis slideway.

vibration signals in the y-direction shows a bit different shape but are still not strong enough to capture the progress of tool wear amount.

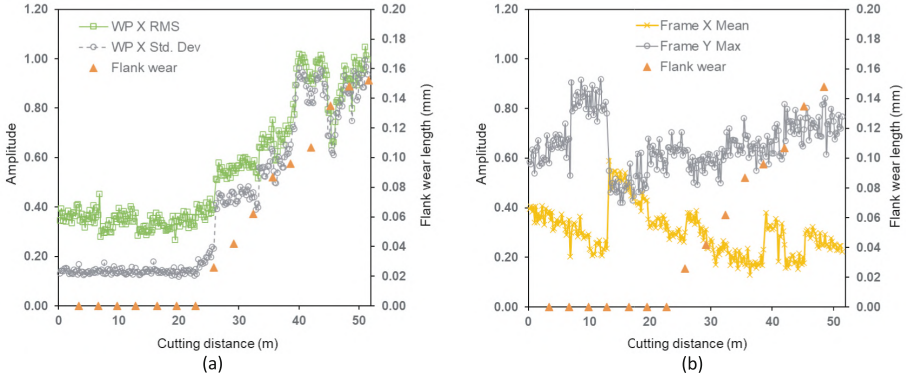


Fig. 7. The progress of tool wear length vs. best signals features with the sensor at (a) Workpiece (b) Y-axis slideway

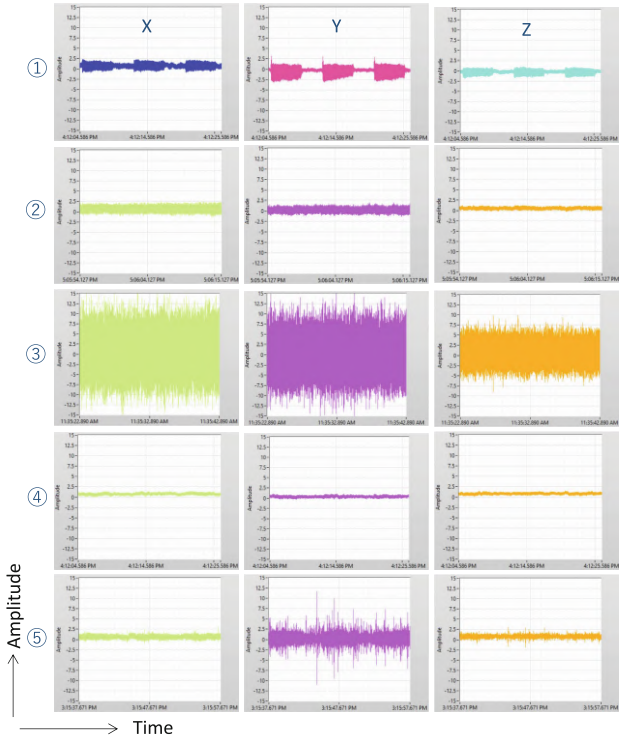


Fig. 8. Vibration signals from different positions, the number indicates the respective position.

5 Conclusion

The objective of this work is to investigate the capability of an accelerometer to be used in tool wear monitoring of high-speed machining of hardened steel for mold and die fabrication. The accelerometer was selected for this study due to showing some advantages compared to other sensors, such as simpler data processing and the flexibility of installation in different parts of the machine tool, compared to the acoustics emission and the dynamometer, respectively. The following points can be concluded from this work:

1. The tool wear progress can be monitored using the RMS value of signals from an accelerometer installed close to the workpiece
2. For this typical application, the signals of the cutting process cannot be well captured using an accelerometer installed at the other four investigated positions (Y-axis slideway, spindle, under the C-axis table, and at B-axis table)

As further works, since the feasible location is only under the workpiece, it is recommended to investigate the performance of a wireless accelerometer to overcome the issue of sensor wire damage due to high-speed movement or rotation. Additionally, it is also suggested to evaluate the performance of acoustic emission sensors at the five positions done in this study for the accelerometer. The issue of large data storage requirements could be solved by saving only the best signal features, instead of saving the original signal data.

References

1. Krajnik, P., Kopac, J.: Modern machining of die and mold tools. J. Mater. Process. Technol. **157–158**, 543–552 (2004)
2. Herwan, J., Kano, S., Oleg, R., Sawada, H., Watanabe, M.: Comparing vibration sensor positions in CNC turning for a feasible application in smart manufacturing system. Int. J. Autom. Technol. **12**(3), 282–289 (2018)
3. Herwan, J., Kano, S., Ryabov, O., Sawada, H., Kasashima, N., Misaka, T.: Predicting surface roughness of dry cut Grey cast iron based on cutting parameters and vibration signals from different sensor positions in CNC turning. Int. J. Autom. Technol. **14**(2), 217–228 (2020)
4. Herwan, J., Kano, S., Ryabov, O., Sawada, H., Kasashima, N., Misaka, T.: Retrofitting old CNC turning with an accelerometer at a remote location towards Industry 4.0. Manuf. Lett. **21**, 56–59 (2019)
5. Herwan, J., Misaka, T., Furukawa, Y., Ogura, I., Komoto, H.: A proposal for improving production efficiency of existing machining line through a hybrid monitoring and optimisation process. Int. J. Prod. Res. **61**(16), 5392–5410 (2023)
6. Zhou, Y., Xue, W.: Review of tool condition monitoring methods in milling processes. Int. J. Adv. Manuf. Technol. **96**(5–8), 2509–2523 (2018). https://doi.org/10.1007/s00170-018-1768-5
7. Mohanraj, T., Shankar, S., Rajasekar, R., Sakthivel, N.R., Pramanik, A.: Tool condition monitoring techniques in milling process—A review. J. Mater. Res. Technol. **9**(1), 1032–1042 (2020)
8. Jemielniak, K., Arrazola, P.J.: Application of AE and cutting force signals in tool condition monitoring in micro-milling. CIRP J. Manuf. Sci. Technol. **1**, 97–102 (2008)

9. Ren, Q., Balazinski, M., Baron, L., Jemielniak, K., Botez, R., Achiche, S.: Type-2 fuzzy tool condition monitoring system based on acoustic emission in micromilling. Inf. Sci. **255**, 121–134 (2014)
10. Li, W., Liu, T.: Time varying and condition adaptive hidden Markov model for tool wear state estimation and remaining useful life prediction in micro-milling. Mech. Syst. Sig. Process. **131**, 689–702 (2019)
11. Aguiar, M.M., Diniz, A.E., Pederiva, R.: Correlating surface roughness, tool wear and tool vibration in the milling process of hardened steel using long slender tools. Int. J. Mach. Tools Manuf. **68**, 1–10 (2013)

Monitoring System for Laser Welding and Emitting Energy Distribution Model to Explain Nonlinear Relationship

Ichiro Ogura(✉)

National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan
ogura.i@aist.go.jp

Abstract. This study describes about the optical monitoring system for laser welding to detect in process welding point, condition, and welding mechanism by using simple apparatus. Two photo diodes and additional optical devices were attached to the end effector of welding robot which holds laser torch. A series of monitoring experiments at several welding positions have been performed and the results showed nonlinear relationship between distance from groove and differential ratio of two photo diodes. A newly model for laser welding was proposed to describe experimental results. This model assumes that emitting light has some energy distribution, and it is changed with detection ratio decided by surface condition. The results of simulation suggested that proposed model has validity to explain this mechanism.

Keywords: Laser welding · Reflecting light · Monitoring · Position · Sensor

1 Introduction

Laser welding has been utilized in industry for half a century. After that, many studies were dedicated to improve, depict or control welding process by using monitoring-technologies [1]. Using multiple sensors is one of effective method to grasp the welding process [2, 3]. Several research and developments of apparatus has already been conducted on methods to confirm the irradiation position and welding state during welding using additional irradiation equipment and camera or sensors [4, 5]. But in this method, it is necessary to add an expensive apparatus unit to the laser head. On the other hand, there seem to be few studies to use same type sensors to monitor laser welding except for optical cameras [6]. Using dual or multiple photo diodes has a possibility to detect laser welding point, condition, and welding mechanism by using simple apparatus with low cost. In this study, simple monitoring apparatus for laser welding was produced and a series of experiments was performed to detect the welding positions. Also new emitting energy distribution model was proposed to explain experimental results.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 496–501, 2025.
https://doi.org/10.1007/978-3-031-72794-8_39

2 Experimental Set-up

Figure 1 shows experimental set-up of optical monitoring system for laser welding. TRUMPF TruDisk2002, fiber laser system which has 1030 nm of wavelength and 60 W to 2000 W of power was used for laser welding light source. The nominal laser spot is 0.3 mm. FANUC M-10iD12 as welding robot holds a laser torch at end effector and moves it to weld materials. Dual photo diodes named PD-R and PD-L were set at a near place of laser head to detect emitting light and reflecting light from welding point. There are supply ports for filler wire and an assisting gas on side position of laser torch. To avoid interference with these, the photodiode was placed so that it could be observed diagonally in front of the laser spot. This arrangement has spatial advantage while this system will be used in practically. Figure 2 shows a configuration diagram of monitoring system. The photo detector is ThorLabs DET50B/M, and the lens tube system holds lenses and an aperture to focus emitting light on the photo diode. A laser line filter with CWL = 1064 nm also set in the lens tube system to select wavelength of incident light. As transmission wavelength is longer than the one of laser light source, the photo diode detects mainly emitting light from plume in this study.

A series of experiments were conducted for laser welding monitoring. 1000 W laser power of continuous wave was selected for experiments. Two flat section SUS304 stainless steel plates with 80 mm square and 3 mm thickness were put on the baseplate, then simple butt joint welding was performed. These plates were put along the straight fixed jigs on the welding baseplate to guarantee same groove position while experiments. The robot moves laser torch along y-direction in 65 mm with scanning-speed of 10 mm/s. At each 0.5 mm different x position, similar experiment was performed and outputs of two photo diode were collected by data acquisition system with 1 ms sampling. Typical welding results are shown in right side of Fig. 1. Welding was performed to the back side of the plate and welding defects could not be detected by eye inspection, then it is thought that the welding speed and laser power were sufficient. Average width of welding marks was 4 mm approximately.

3 Experimental Results and Discussion for Nonlinear Relationship

Before having experiments, outputs of two photo diodes were calibrated by monitoring emitting light from blank specimen. As same as monitoring experiments, the robot moves the laser torch in 10 mm/s scanning for blank stainless plate and average value of each photo diode outputs were compared. The ratio between two photo diodes was 1.11 in this case. The reason of difference is thought that mechanical setting error of monitoring system and original setting tilt of the welding plate against x-y plane. Each photo diodes were fixed at laser torch and robot moves them in same direction, and welded plates were stable while experiments. Therefore, these setting errors were able to correct just by multiplying this ratio. The outputs of photodiodes in experiments after this were corrected by this ratio.

Typical outputs of two photo diodes while laser welding is shown in Fig. 3. In this figure, welding displacement from center position of groove was $s = 0.63$ mm from picture image analysis of actual welded specimen while setting position was 0.5 mm.

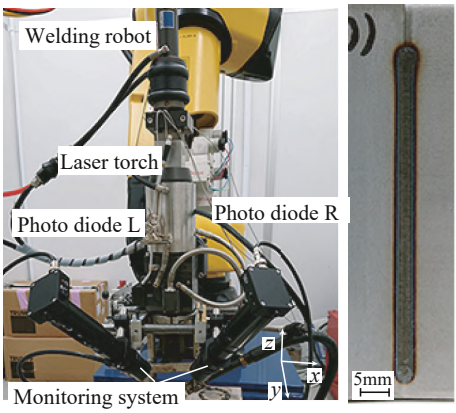


Fig. 1. Experimental set-up and typical welding result

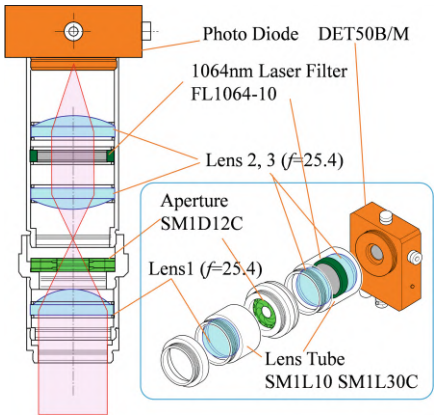


Fig. 2. Monitoring system with photo diode

Time at 0 s is start time of laser light emission. Because of the robot stopped at start position while 1 s to get stable welding laser power, the outputs became higher than residual welding. After moving laser torch, the outputs fell to a stable state gradually until laser stopped the emission. In this experiment, the outputs from 2 s to 7 s were used to get average value and calculation as $R(s) = \frac{PD_L(s)-PD_R(s)}{PD_L(s)+PD_R(s)} \times 100$ gave differential ratio between two photo diodes outputs at $s = s$. In Fig. 3, a differential ratio between two photo diodes outputs in each time is also plotted. $R(s)$ is same as an average value of different ratio while 2 s to 7 s.

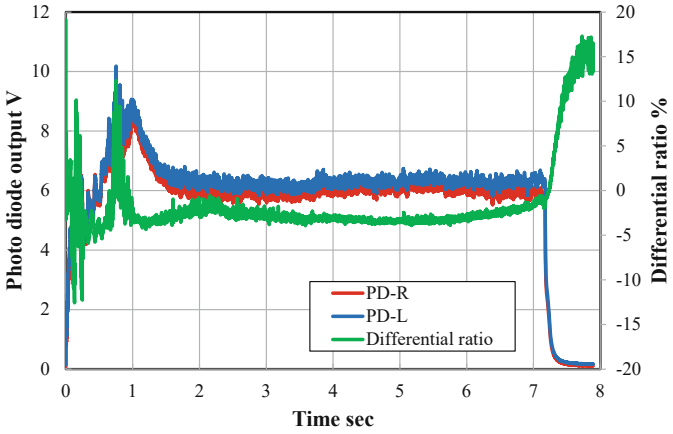


Fig. 3. Output of photo diodes

Figure 4 shows variation of the differential ratio in each distance from the groove. This relationship varies to nonlinearity, which has 1) the increment area, 2) the dead zone, 3) the increment area again and 4) the decrement area. The main purpose of

determining the relationship between laser spot position and output is to confirm that the welding position is suitable while welding. Using this relationship, an in-process correction by robot also becomes possible. A linear relationship between laser spot position and output is reliable to detect or control welding position from grove because of the position is simply determined by the proportionality constant. Although, just fitting curve for Fig. 4 like output will be useful for position detection and control, in order to understand the nonlinear phenomenon and the relationship between welding point and output even for different groove shapes or materials, it is meaningful to attempt to denote this phenomenon through modeling.

To explain these results, an emitting energy distribution model for laser welding is shown in Fig. 5. In this model, some suppositions are set as follows: 1) Photo diode detects the emission along parallel light with each photo diode direction. 2) Welding light emission from blank surface is "Top-hat" energy distribution $I(x)$, which has saturated area with Gaussian function and this distribution is changed with multiplied detection ratio $J(x)$ decided by surface condition. 3) Because of the edge of welded specimen, the detection ratio of PD-L and PD-R are denoted as:

$$J_{L,R}(x, s) = p_0 \pm \frac{\Delta p(x - s)\sqrt{e}}{w} \exp\left(-\frac{(x - s)^2}{2w^2}\right) \tag{1}$$

and it is continuous while welding process. In this equation, p_0 means a nominal detection ratio for each photo diodes while welding for flat blank surface like point A in Fig. 5. s is center position of groove on x-axis. Because of the inclined surface near groove, the detection ratio will be higher when a normal vector of inclined surface is almost parallel to the photodiode direction like point B and one will be lower at point C in Fig. 5 each other. This relationship will be symmetric at PD-L. Δp denotes amount of these varies from nominal ratio p_0 and w is distance of local maximum and minimum position from the center position of groove on x-axis. To represent the change in detection ratio around the groove, as shown in the lower part of Fig. 5, the Eq. (1) is derived from the first derivative of Gaussian function.

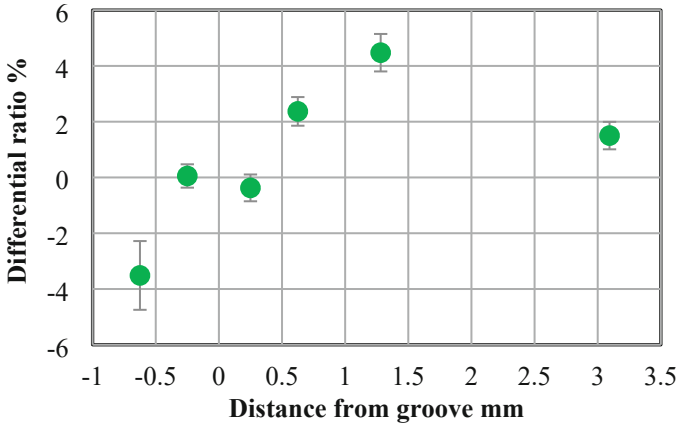


Fig. 4. Differential ratio of dual photodiode

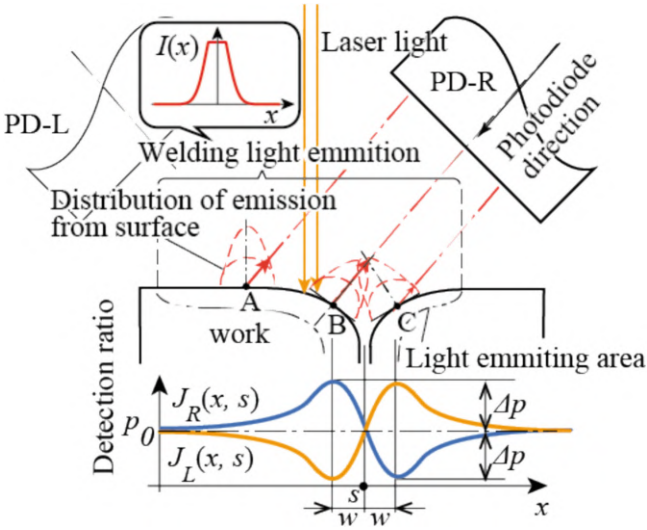


Fig. 5. Emitting energy distribution model for laser welding

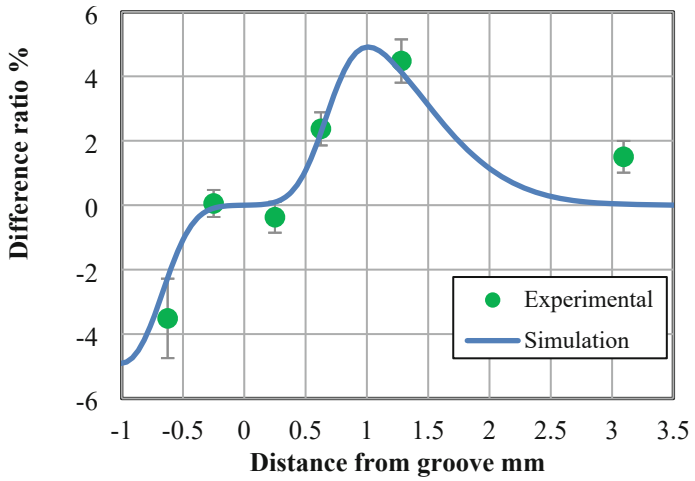


Fig. 6. Results of experiment vs. simulation

Therefore, the outputs of PD-L and PD-R in groove $s = s$ are denoted as:

$$PD_{L,R}(s) = \int_{-\infty}^{\infty} I(x) J_{L,R}(x, s) \mathrm{dx} \tag{2}$$

Figure 6 shows results of experiment and simulation calculated from above model. In this simulation, $p_0 = 0.5$, $\Delta p = 0.3$, $w = 0.2$ of Eq. (1) were set. $p_0 = 0.5$ and $\Delta p = 0.3$ are tentative detection ratio that describe a typical flat surface and the maximum and minimum deviations from it. According to a picture image analysis of Fig. 1, distance

of groove was 0.2 mm. Since the values of w can be supposed as approximately same order to this amount, $w = 0.2$ mm was chosen here. About $I(x)$ function, width of $\pm 3\sigma$ for Gaussian function was 4.5 mm as welding marks was 4 mm. More than 80% of maximum amount in Gaussian function was saturated to realize top hat energy distribution. The maximum amount was 1.0 because any coefficient value of $I(x)$ is canceled while calculating differential ratio. Although discussion is necessary about these parameters setting, this simulation shows well corresponding to experimental results and suggests that the model having validity to explain this mechanism.

4 Conclusion

This study is summarized as follows:

1. Optical monitoring system for laser welding with two photo diodes was produced. It intends to detect in process welding point, condition, and welding mechanism by using simple apparatus.
2. Monitoring experiments were conducted in each distance from groove. The results show non-linearity, which has the increment area, the dead zone, the increment area again, and the decrement area.
3. To explain non-linearity, an emitting model for laser welding was proposed. Although this model has several suppositions and uncertain parameters, the simulation results well corresponded to experimental results.

References

1. Cai, W., Zhuang, J., Jiang, P., Cao, L., Mi, G., Zhou, Q.: Application of sensing techniques and artificial intelligence-based methods to laser welding real-time monitoring: a critical review of recent literature. J. Manuf. Syst. **57**, 1–8 (2020)
2. Zhang, Y., You, D., Gao, X., Katayama, S.: Online monitoring of welding status based on a DBN model during laser welding. Engineering **5**, 671–678 (2019)
3. You, D., Gao, X., Katayama, S.: Multiple-optics sensing of high-brightness disk laser welding process. NDT&E Int. **60**, 32–39 (2013)
4. Dorsh, F.: Process Sensor Systems for Laser Beam Welding Enabling and assuring reliable production. Laser Technik J. **3**, 24–28 (2012)
5. IPG Photonics Homepage. https://www.ipgphotonics.com/en/products/beam-delivery/process-monitoring/process-monitoring/ldd-700. Accessed 01 Oct 2023
6. Tenner, F., Berg, B., Brock, C., Klämpfl, F., Shmidt, F., Experimental approach for quantification of fluid dynamics in laser metal welding. J. Laser Appl. **27**, S29003-1–S29003-7 (2015)

Author Index

A
Abe, Masajiro 221
Ali, Junaid 441
Araki, Kaito 154
Asai, Takehiko 417

B
Bajaj, Anil K. 441
Balachandran, Balakumar 355
Benedetti, Kaio C. B. 125
Birs, Isabela 165
Brunetti, Matteo 75

C
Chao, Cui 99

D
De Keyser, Robin 165

F
Fidlin, Alexander 58, 231
Fukushige, Shinya 281
Furukawa, Yoshiyuki 486

G
Genda, Attila 58
Gendelman, Oleg V. 43, 58
Gonçalves, Paulo B. 125
Gunji, Mamoru 88

H
Hamajima, Tomotaka 335
Hanasaki, Itsuo 265
Herrera-Granados, German 486
Herwan, Jonny 486
Hirose, Kiyoshi 343
Hojo, Masahiro 281
Horiguchi, Hironori 370
Hosaka, Toshihiko 281

I
Inoue, Tsuyoshi 370

K
Kato, Yuzuru 141
Kawaguchi, Natsuki 184
Kawasumi, Yuma 426
Kloda, Lukasz 75
Koike, Sekiya 323, 343
Komoto, Hitoshi 486
Kondo, Shinsuke 199
Kubo, Gai 281
Kubota, Shuya 154
Kurahashi, Takahiko 221
Kuroda, Masaharu 184

L
Lenci, Stefano 125
Li, Xiao 312
Liu, Yuhao 99

M
Matsubara, Hiroharu 154
Matsuda, Akihiro 312, 335
Matsuda, Tetsuya 281, 293
Melnyk, Danyil 75
Meng, Yiming 111
Mircheski, Petar 245
Misaka, Takashi 486
Mitura, Andrzej 75
Mochiyama, Hiromi 457
Mori, Eiichiro 281
Moriguchi, Yuki 184
Morita, Naoki 281
Muresan, Cristina 165

N
Nakano, Yutaka 397
Nakao, Hiroya 141, 245, 255, 305
Namura, Norihisa 255

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
H. Yabuno et al. (Eds.): IUTAM 2023, IUTAM Bookseries 43, pp. 503–504, 2025.
https://doi.org/10.1007/978-3-031-72794-8

O
Ogura, Ichiro 486, 496
Ohyama, Keigo B. 323
Okawa, Hayate 370
Ono, Yusuke 88
Orosz, Gábor 405

P
Pandey, Manoj 355
Peng, Linyu 88
Perkowski, Nicolas 111

R
Raja, Buana Mahkota 383
Rega, Giuseppe 3, 125
Romeo, Francesco 75

S
Sagara, Kakeru 471
Sekikawa, Satoru 397
Seto, Shota 210
Shaver, Gregory 441
Shimura, Yukinobu 293
Shoji, Gaku 471
Sri Namachchivaya, N. 111
Stépán, Gábor 405
Szaksz, Bence 405

T
Taga, Keisuke 305
Takahara, Hiroki 397
Takata, Shohei 141
Takata, Soichiro 154
Takayasu, Akitoshi 199, 210
Takeda, Osamu 343
Tominaga, Tempei 323
Tsujikawa, Chiharu 281

U
Uematsu, Rinri 343

W
Wang, Wanshu 417
Wani, Khawar Zamman 355
Warminski, Jerzy 75
Watanabe, Naoko 154
Willmott, Alexander P. 323, 343

Y
Yabuno, Hiroshi 426
Yagasaki, Kazuyuki 21
Yamashita, Kiyotaka 383
Yamazaki, Yoshihiro 305
Yang, Jian 99
Yasui, Kenji 426
Yokota, Kazuya 221
Yoshikawa, Nobuhiro 281
Yüzbasioglu, Tunc 231

Z
Zhu, Jinjie 245