ABSTRACT

Title of dissertation: ESTIMATION AND CONTROL OF
NONLINEAR SYSTEMS: MODEL-BASED
AND MODEL-FREE APPROACHES

Debdipta Goswami
Doctor of Philosophy, 2020

Dissertation directed by: Professor Derek A. Paley
Department of Aerospace Engineering
Department of Electrical and Computer
Engineering
Institute for Systems Research

State estimation and subsequent controller design for a general nonlinear system is an important problem that have been studied over the past decades. Many applications, e.g., atmospheric and oceanic sampling or lift control of an airfoil, display strongly nonlinear dynamics with very high dimensionality. Some of these applications use smaller underwater or aerial sensing platforms with insufficient on-board computation power to use a Monte-Carlo approach of particle filters. Hence, they need a computationally efficient filtering method for state-estimation without a severe penalty on the performance. On the other hand, the difficulty of obtaining a reliable model of the underlying system, e.g., a high-dimensional fluid dynamical environment or vehicle flow in a complex traffic network, calls for the design of a data-driven estimation and controller when abundant measurements are present from a variety of sensors. This dissertation places these problems in two broad categories: model-based and model-free estimation and output feedback.

In the first part of the dissertation, a semi-parametric method with Gaussian mixture model (GMM) is used to approximate the unknown density of states. Then a Kalman filter and its nonlinear variants are employed to propagate and update each Gaussian mode with a Bayesian update rule. The linear observation model permits a Kalman filter covariance update for each Gaussian mode. The estimation error is shown to be stochastically bounded and this is illustrated numerically. The estimate is used in an observer-based feedback control to stabilize a general closed-loop system. A transfer-operator-based approach is then proposed for the motion update for Bayesian filtering of a nonlinear system. A finite-dimensional approximation of the Perron-Frobenius (PF) operator yields a method called constrained Ulam dynamic mode decomposition (CU-DMD). This algorithm is applied for output feedback of a pitching airfoil in unsteady flow.

For the second part, an echo-state network (ESN) based approach equipped with an ensemble Kalman filter is proposed for data-driven estimation of a nonlinear system from a time series. A random reservoir of recurrent neural connections with the echo-state property (ESP) is trained from a time-series data. It is then used as a model-predictor for an ensemble Kalman filter for sparse estimation. The proposed data-driven estimation method is applied to predict the traffic flow from a set of mobility data of the UMD campus. A data-driven model-identification and controller design is also developed for control-affine nonlinear systems that are ubiquitous in several aerospace applications. We seek to find an approximate linear/bilinear representation of these nonlinear systems from data using the extended dynamic mode decomposition algorithm (EDMD) and apply Lie-algebraic methods to analyze the controllability and design a controller. The proposed method utilizes the Koopman canonical transform (KCT) to approximate the dynamics into a bilinear system (Koopman bilinear form) under certain assumptions. The accuracy of this approximation is then analytically justified with the universal approximation property of the Koopman eigenfunctions. The resulting bilinear system is then subjected to controllability analysis using the Myhill semigroup and Lie algebraic structures, and a fixed endpoint optimal controller is designed using the Pontryagin's principle.

ESTIMATION AND CONTROL OF NONLINEAR SYSTEMS: MODEL-BASED AND MODEL-FREE APPROACHES

by

Debdipta Goswami

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2020

Advisory Committee:
Professor Derek A. Paley, Chair/Advisor
Professor Perinkulam S. Krishnaprasad
Professor Nuno M. C. Martins
Professor Eyad H. Abed
Professor Nikhil Chopra

Dedication

আমার দুঃখদিনের রক্তকমলগুলি চয়ন করে মালা গাঁথলাম। যাঁদের পরশে সেই ফুলেরা একটি একটি করে ফুটে উঠত, তাঁদেরই চরণে এই মালা অর্পণ করলাম।

Translation: I have weaved this wreath picking the red roses of my struggling days. I lay this at their feet who made the flowers bloom, one by one, by their gentle touch.

Acknowledgments

“রেখো মা দাসেরে মনে, এ মিনতি করি পদে।
সাধিতে মনের সাধ, ঘটে যদি পরমাদ,
মধুহীন করো না গো তব মনঃকোকনদে।”
- মাইকেল মধুসূদন দত্ত[1]

I have tried to be the bright flame that would draw some eyes, only to burn myself out without even putting out enough warmth. So, the city of College Park, I leave you a charred residue of my soul, and take a little piece of you in the secret yard of my heart. Not many people have the privilege to call more than one place home, but Washington-Dulles airport has ushered me with the same pleasure of homecoming many times that Dumdum airport does. Whenever I look at the Google map, the confluence of Anacostia and Potomac rivers looks like the outstretched arms of a mother calling her children home. Call it a fruit of my ripe imagination, but don’t we call our school “alma mater”? I don’t expect you to remember this insignificant adopted child of yours, but please keep a sanctuary for these broken souls who might continue to take refuge there in the days to come.

---

[1] Michael Madhusudan Dutt, 19th century Bengali poet

Thank you.

Thank you Dr. Paley[2]! It has been a roller-coaster ride for the past five years. Thank you for giving me an invaluable opportunity to work on an extremely interesting project of nonlinear estimation and control. Many a times I have resisted your ideas only to find out later their paramount importance to succeed as a researcher. Your commitment to the academic and professional development of your students is commendable, as is your effort in time-management. You will continue to be my role model as a successful academic who has maintained a strong commitment to his family as well. And thank you for bearing with me when my productivity was quite low.

I am thankful to all the wonderful professors whom I crossed path with during my graduate study here. I sincerely thank Prof. P. S. Krishnaprasad for his close mentorship, instruction, and support. He has been extremely helpful in undoing the complex mathematical knots I sometimes stumble into and didn't mind spending hours to discuss topics related to my research. I am grateful to my dissertation committee (Prof. Krishnaprasad, Prof. Nuno Martins, Prof. Eyad Abed, and Prof. Nikhil Chopra) for all the thoughtful advice and discussions. I am also grateful to Profs. Nuno Martins, Behtash Babadi, Michael Rotkowitz, P. S. Krishnaprasad, André Tits, Wojciech Czaja, John Baras, and Armand Makowski for the great education I received from them. I am largely indebted to Prof. André Tits, with whom I also served as an Editorial Assistant to Automatica Rapid Publications, for his unwavering support to my endeavours. His door was always open for me specially when I was going through a turmoil in my professional and personal life.

---

[2] Dr. Derek A. Paley

I would also like to thank my professors at Jadavpur University, India, where I pursued my undergraduate study. A big thank you note to Prof. Subhankar Ray for introducing me to the universe of nonlinear dynamics which ultimately has proved to be the central theme of my graduate research.

Close collaborations, both internal and external to UMD, enriched the research process for me, strengthening the questions I asked and my methods for answering them. I gratefully acknowledge Emma Thackray from Mount Holyoke College for the simulations in Chapter 3. I am also thankful to Justin Lidard from UMD Aerospace undergraduate program. He did a fantastic work in airfoil-lift maximization problem in Chapter 3. I sincerely wish him a productive career ahead. I also acknowledge the help from Dr. Artur Wolek for processing the traffic data from Numina sensors in Chapter 4.

My colleagues at the Collective Dynamics and Control Laboratory have enriched my graduate life in many ways and deserve a special mention. Frank[3] was my "in-house" mentor until he graduated in 2017, and I was probably the last graduate from the "famous Frank's High-School" as we jokingly called the rich collection of his mentees. Daigo[4] never failed to inspire me with his cool quadcopters and Matlab simulations and always left me amazed with his mastery in visual skills. My interaction with Dr. Feitian Zhang, Brett Barkley, Dr. Brian Free, Dr. Will Craig, Jinseong Lee, Daniel Gomez, Travis Burch, Jordan Boehm, Dr. Will Scott, Dr. Artur Wolek, Sheng Cheng, Curtis Merrill, Animesh Shastry, Rachael Suitor, and Anthony Thompson has been very fruitful. A special shout-out to Will Craig for his enthusiasm in organizing lab happy hours; we haven't had one

---

[3] Dr. Francis D. Lagor Jr. [4] Dr. Daigo Shishika

since he graduated.

It would be an injustice not to acknowledge the helps I received from Melanie, Emily, Maria, and Vivian from the office of ECE Graduate Studies; and Renee, Lavita, and Otto from the Aerospace Graduate office. They kept the bureaucratic process running so that I never had to worry about my pay-cheques.

A big shout-out to all the fellow board-members of Electrical and Computer Engineering Graduate Student Association (ECEGSA) with whom I had the privilege of serving there. All those coffee-hours, hiking trips, and ski-trips were amazing.

Well, now it is time to give thanks to all my friends and family in College Park, DC, Blacksburg, and Calcutta. And it is a long list. So, leave if you will, else buckle up! Thank you Udit Da[5] and Dipankar Da[6], my former and only roommates ever, to help this socially awkward guy to come out of his cocoon. 3404 #14 will always stay in my heart (specially the dinner table discussions). Chiranjib[7], lend me your grit in scholarly research, will you? Keep publishing, keep patenting. I sincerely hope that we can go on that cross-country trip one day. Debi Di[8], thank you for being a sister from another mum, for providing the emotional cushion needed during the periods of isolation and frustration. Babua[9] and Soumojit[10], the late night discussions on Fridays were real stress-busters. Thanks Da and Di[11], for letting me crash at your place whenever I visited Princeton, and Da, specially for all the technical discussions on those untimely phone-calls. Thank you, UMD Counseling Center and Dr. Allison Asarch, for your professional support when I went into clinical depression. Without Dr. Asarch's sessions, I would have left the programme long ago.

---

[5] Dr. Udit Halder [6] Dr. Dipankar Maity [7] Chiranjib Saha [8] Dr. Debarati Kundu [9] Prithviraj Dhar [10] Soumojit Das [11] Dr. Biswadip Dey and Shawon Sarkar Dey

And thank you, Minna, you are the only permamnent officemate I ever got in these five years. It was good to know that there is another living thing, even if not vertebrate, breathing in the Manufacturing Building during the all-nighters. For those who do not know Minna, she is the cutest possible Chilean rose tarantula who spends most of her time sleeping in her terrarium.

Thanks everyone in the Anne Arundel Radio Club (W3VPR) for introducing me the wonderful world of Ham radio. I will keep listening for the club call-sign in the coming field days if I stay in this country. Otherwise we can always catch up during a solar peak from any part of the world provided I can put up a decent antenna. Until then, 73 AC3EH[12].

Thank you Ma (Mum)[13] and Baba (Dad)[14] for your unwavering support throughout the long road of graduate school. You are so much a part of who I am. I thank each of them for always being there for me, letting me talk through my highs and lows, and for sharing in my successes and failures. A special thanks to Ma for gifting me the UK trip in my final year of the graduate school even after the huge screw-up at Belgium. I am missing you here during my graduation. Please forgive my tantrums over those Skype calls. Love you!

As I am writing this dissertation, the dark shadow of the COVID-19 pandemic is all over the world. Thank you all the public heath professionals (including my dad) and medical researchers who are fighting at the forefront. Thanks everyone who, in the midst of a pandemic, are risking their lives to keep the essential services up and running.

---

[12] My FCC call-sign [13] Kumkum Goswami [14] Dr. Prabir K. Goswami

Without them I would be in no position to complete my dissertation.

I would like to acknowledge financial support from the Air Force Office of Scientific Research (AFOSR FA95501810137) and the National Science Foundation (NSF CMMI1362837) for all the projects discussed herein.

It is impossible to remember all, and I apologize to those I've inadvertently left out. And finally, the deepest and sincerest gratitude to Rabindranath Tagore[15], whose creations resonate with every emotions of mine and continue to push me through the dark days; and to J. K. Rowling, because the original Bloomsbury books[16] stayed, and will stay, on my night-stand "Until the very end".

---

[15] Rabindranath Tagore, FRAS, Indian polymath and the first Nobel laureate from Asia

[16] Harry Potter series, originally published by Bloomsbury Publishing PLC, London

Table of Contents

List of Tables

List of Figures

List of Nomenclatures

	Uppercase Greek letters
Γ	Non-recurrent surface (Chapter 5)
$\Phi_f(t,\cdot)$	Flow map of $\dot{\mathbf{x}} = f(\mathbf{x})$
∇	Gradient operator

	Lowercase Greek letters
α	Angle of attack (Chapter 3), leakage rate (Chapter 4)
λ	Eigenvalue of a matrix or a linear operator
$\rho(\cdot)$	Probability density function (Chapter 3)
τ_1, τ_2	Flow settling times (Chapter 3)
$\phi(\cdot)$	Koopman/PF eigenfunctions (Chapters 3, 5)
$\psi(\cdot)$	Activation function for the neural network (Chapter 4)
χ_B	Characteristic (Indicator) function of set B (Chapter 3)

	Uppercase Roman letters
C_L	Coefficient of lift Chapter 3)
$G(n,p)$	Erdös-Rényi graph with n vertices and edge-probability p (Chapter 4)
$\mathrm{GL}(n,\mathbb{R})$	General linear group (group of all $n \times n$ invertible real-valued matrices)
$H(t,\cdot,\cdot,\cdot)$	Pre-Hamiltonian (Chapter 5)
$\mathbf{H}(\mathbf{x})$	Set of dictionary functions (Chapter 5)
$\mathbf{I}$	Identity matrix
K	Kalman gain
$L_{\mathbf{f}}$	Lie-derivative with respect to the vector-field $\mathbf{f}$
L^p	Space of p^{th}-integrable functions, $p \in [1,\infty)$
L^{inf}	Space of essentially bounded measurable functions
M	Mixture complexity (Chapter 2)
P	Covariance matrix
P_τ	Markov approximation of $\mathcal{P}^\tau$ (Chapter 3)
S^1	Circle group
$\mathbf{V}$	Measurement noise (Chapter 2)
$\mathbf{W}$	Process noise (Chapter 2)
W	Reservoir weight matrix (Chapter 4)
W_{in}	Input weight matrix to an ESN (Chapter 4)
W_{out}	Readout matrix from an ESN (Chapter 4)

	Lowercase Roman letters
$\det(\cdot)$	Determinant of a square matrix
dw_t	Wiener increment (Chapter 3)
$Im(\cdot)$	Imaginary component operator
k_1, k_2	Control gains (Chapter 3)
$\ker(A)$	Null-space (kernel) of matrix A
m	Lebesgue measure
$Re(\cdot)$	Real component operator
$\mathbf{r}$	Reference signal (Chapter 2), reservoir states (Chapter 4)
T	Koopman canonical transform (KCT) operator (Chapter 5)
$\mathrm{tr}(\cdot)$	Trace of a square matrix
$\mathbf{u}(\cdot)$, $u(\cdot)$	Control input
w_j	Weight of j^{th} mixture component (Chapter 2)
$\mathbf{x}$, x	State variable
$\overline{\mathbf{x}}$	Mean of the stochastic states/ensemble of the state vectors
$\mathbf{y}$, y	Output/measurement variable
$\mathbf{z}$	Transformed state variable (Chapter 5)

	Calligraphic, blackboard, and decorative letters
$\mathcal{B}(\mathbb{X})$	Borel sigma algebra over $\mathbb{X}$
$\mathbb{C}$	Set of complex numbers
$\mathbb{E}(X)$	Expectation of random variable X
$\mathcal{H}(t,\cdot,\cdot)$	Control Hamiltonian
$\mathcal{K}^{\tau}$	Koopman operator semigroup for time τ
$\mathcal{N}(\mathbf{x};\overline{\mathbf{x}},P)$	Probability of obtaining state $\mathbf{x}$ from a Gaussian distribution with mean $\overline{\mathbf{x}}$ and covariance P (Chapter 2)
$\mathbb{P}$	Probability measure
$\mathcal{P}^{\tau}$	Perron-Frobenius operator semigroup for time τ (Chapter 3)
$\mathbb{R}$	Set of real numbers
$\mathbb{X}$	State-space/manifold

	List of abbreviations
BIC	Bayesian information criterion
CU-DMD	Constrained Ulam dynamic mode decomposition
DMD	Dynamic mode decomposition
DMZ	Duncan-Mortensen-Zakai (equation)

EDMD	Extended dynamic mode decomposition
EKF	Extended Kalman filter
EM	Expectation maximization
EnKF/S	Ensemble Kalman filter/smoother
ESN	Echo-state network
ESP	Echo-state property
GK	Goman-Khrabrov (model)
GMM	Gaussian mixture model
GMM-EKF	Gaussian mixture model-extended Kalman filter
GMM-KF	Gaussian mixture model-Kalman filter
GMM-UKF	Gaussian mixture model-unscented Kalman filter
i.i.d.	Independent and identically distributed
KCT	Koopman canonical transform
KF	Kalman filter
MAV	Micro-air vehicle
ODE	Ordinary differential equation
PDE	Partial differential equation
PDF	Probability density function
PF	Perron-Frobenius (operator)
PMF	Probability mass function
RC	Reservoir computing
RNN	Recurrent neural network
SDE	Stochastic differential equation
UKF	Unscented Kalman filter
UT	Unscented transform

Chapter 1

Introduction

Estimation and control of high-dimensional nonlinear systems is of paramount importance in many control systems application. In many of the aerospace applications including atmospheric and oceanic sampling, nonlinear filtering, even with non-Gaussian driving noise, has become more important in recent days. And most of them use smaller aerial or underwater vehicles that have serious restrictions regarding to on-board computational power. Hence the need for approximate nonlinear filtering without extensive use of Monte-Carlo simulations on-board arises. Furthermore, it is sometimes difficult to obtain a reliable model of the underlying system, e.g., a high-dimensional fluid dynamical environment or vehicle flow in a complex traffic network. But with the abundant availability of cheap sensors, it is easier to obtain a lot of spatiotemporal sampling data for these cases, thereby requiring a model training using the available data, and its use in the subsequent estimation and controller design.

For a linear system with Gaussian process and sensor noise and with Gaussian initial density, the optimal estimator is given by the celebrated Kalman filter [3], whereas for a nonlinear system with non-Gaussian process noise, optimal filtering is infinite dimensional. The optimal estimator for the nonlinear system involves solving Duncan-Mortensen-Zakai (DMZ) [4] stochastic Partial Differential Equation (PDE) for continuous measurement or recursive Bayesian filtering by solving Fokker Planck PDE for discrete measurement case. However both the approaches require solution of the PDEs and hence no finite-dimensional filtering is possible in general. Although there are certain cases where finite dimensional solutions for this problem have been discovered [5], [6], nonlinear filtering is mostly facilitated by several approximate techniques like extended Kalman filter (EKF), unscented Kalman filter (UKF), projection filtering [7, 8, 9], and particle filters [10]. The EKF and UKF aim to make the filtering problem finite dimensional using a finite number of parameters and relying on a fully parametric and, in most cases, Gaussian process noise model, which does not guarantee convergence for the nonlinear system. The latter techniques use Monte-Carlo simulations of particles in the state space and update their weights to approximate the probability density function of the states. But they require large amount of computing power and suffer from the issue of degeneracy. The latter issue can be mitigated by resampling, but that, in turn, increase computational burden even more.

In many aerospace applications, nonlinear filtering, even with non-Gaussian driving noise, has become more important in recent days. And most of them use smaller aerial or underwater vehicles that have serious restrictions regarding to on-board computational power. Hence the need of approximate nonlinear filtering without extensive use of Monte-Carlo simulations on-board arises. Furthermore, the necessity to design observer-based feedback controller using the estimate poses a real challenge. On the other hand, with the evolution of cheap sensors providing abundant data and an unprecedented increase in offline computational power, we have moved forward with the great hope of controlling more complicated systems where a reliable model is often hard to find. Sometimes those models are very high-dimensional and hence difficult to tame.

This dissertation aims to solve the aforementioned problems in two successive steps:

1. The first part proposes two new techniques for model-based estimation and output feedback for nonlinear systems. Both of them are designed to deal with non-Gaussian prior and process noise profiles.

    - First, the Gaussian mixture model (GMM) as a semi-parametric approach to approximate the probability density and Kalman filter based motion and measurement updates are used for nonlinear filtering with non-Gaussian process noise. The stochastic error bounds of such filters are analytically proved and numerically illustrated. The ultimate stochastic boundedness of the output feedback control using such filter is also analyzed.
    - Next, the Perron-Frobenius (PF) operator, a linear transfer operator that describes the evolution of probability density in a dynamical system, is recalled. A finite-dimensional approximation of the Perron-Frobenius operator is employed with a Bayesian measurement assimilation technique for nonlinear estimation. The proposed method is applied for the output feedback control of a pitching airfoil in order to maximize lift.

2. In the second part, we shift our focus to the estimation and control of unmodeled or partially modeled systems. Here we rely on the richness of available data for prediction and control of a system.
    - A framework for data-driven estimation of complex systems from time-series data is developed using an echo-state network (ESN), a special type of recurrent neural network (RNN) and it is applied to estimate the daily cycles of traffic congestion from a real set of mobility data.
    - Lastly, a data-driven control framework for a control-affine nonlinear system is developed using the spectra of Koopman operator , the dual of the PF operator, when the drift vector field is unknown.

1.1 Survey of the relevant literature

Due to the wide breadth of topics covered in this dissertation, this survey only focuses on the most pertinent works known to the author at the time of writing. Additional references appear elsewhere in the text of the dissertation as appropriate.

1.1.1 Nonlinear estimation and feedback control using system models

1.1.1.1 Gaussian mixture model based estimation

For a nonlinear system with non-Gaussian process noise, the optimal estimator is represented by a recursive Bayesian filter that converts the prior probability density function

(PDF) into a posterior PDF using a likelihood function of the observation. However, in many cases, the prior PDF is not explicitly known, and has to be approximated from an ensemble realization. The most common approach is to use a particle filter [11] with an effective method of re-sampling to tackle highly non-Gaussian distributions. But sampling and re-sampling become computationally intractable for a high-dimensional system. Hence, a semi-parametric approach may be necessary.

Alspach and Sorenson [12] proved that any sufficiently smooth probability density function is approximated arbitrarily closely by an ensemble realization of a weighted sum of Gaussian PDFs. This technique, called the Gaussian Mixture Model (GMM), gives an approximate way to explicitly calculate the *posteriori* density of the states of a stochastic nonlinear system, even in the presence of non-Gaussian process noise.

There are several approaches for time and measurement updates using GMM. The GMM-DO filter of Sondergaard and Lermusiaux [13] uses GMM equipped with Monte-Carlo data-fitting based on the Expectation Maximization (EM) algorithm [14] and the Bayesian information criterion (BIC) [15]. GMM represents the PDF using a weighted sum of Gaussian PDFs, each of which is updated using individual Kalman filters if the measurement model is linear and the measurement noise profile is Gaussian. However, [13] requires re-sampling at each time instant at which an update is produced, thus increasing the computational burden.

For high dimensional systems, [16], [17] have developed an implicit formulation of ensemble Kalman filters and smoothers (EnKF/S). This approach operates directly on the ensemble members. Reference [18] provides an approach of clustering the forecast ensembles around centroids obtained using a clustering algorithm, e.g., K-means. The Kalman gain for each cluster is defined locally using an empirical covariance matrix. Reference [19] uses EM and BIC to approximate the forecast distribution using GMM; the update takes place by projecting it on a posterior Gaussian distribution. Reference [16] removes the constraint of a Gaussian posterior and [17] introduces balanced sampling to determine the ensemble member for update. Reference [20] develops a mixture-model EnKF that uses an exact update equation without requiring explicit knowledge of the mixture moments.

Many sensors used in atmospheric and oceanic sampling have a linear (or linearized) observation model and an additive Gaussian measurement noise profile [13]. Taking these assumptions into account, GMM along with Kalman filter updates proves to be an efficient estimator, even for non-Gaussian process noise. For nonlinear systems, the Kalman update can be replaced by the extended Kalman filter, which gives the prediction up to first-order precision [21]. The unscented Kalman filter [22], [23] uses the unscented transform (UT) to transport the Gaussian covariance up to third-order precision with significantly less computational effort compared to Monte-Carlo covariance propagation in each step.

1.1.1.2 Estimation using operator approximation

The operator-theoretic approach to dynamical systems deals with the evolution of measurable maps under the system dynamics. The two main candidates of this approach are the Koopman operator and its dual, the Perron-Frobenius operator. While the Koopman operator is useful in studying observables, the Perron-Frobenius (PF) operator acts on the space of densities. Hence the PF operator is important when dealing with uncertainties in the system, especially when the likelihood of the state is given in the form of a probability density function under a suitable absolutely continuous probability measure. Formally, the PF operator is an infinite-dimensional operator operating on the space of L^1 functions.

The PF operator is used extensively to analyze the global behavior of dynamical systems, especially for fluid dynamics [24], and to estimate almost-invariant sets with efficient toolboxes like GAIO [25]. The PF operator, being able to transport density in a dynamical system, is important in recursive estimation problems. The approximated PF operator can be used to transport density with less computational effort than solving a partial differential equation. Hence an accurate and efficient approximation technique of the PF operator is needed. As the PF operator operates on infinite-dimensional spaces, it is customary to project it onto the span of suitable basis functions to approximate their dominant actions in finite dimensions. This projection is usually accomplished by Galerkin methods using a weak approximation of the operand functions [26], [27].

One technique to approximate the PF operator is Ulam's method [28], where the chosen basis functions are the characteristic functions of grids spanning the state space. The PF operator is reduced to a Markov state transition matrix and approximated by one-pass Monte Carlo simulation of a large number of initial conditions. However, this method requires the time step of the simulation to be sufficiently long [28] to avoid degeneracy. This limitation poses a difficulty when using the approximated PF operator for the estimation, because precise density transport over small time intervals is necessary to accommodate frequent observations. To enable PF approximation for a short time step, a multi-pass approach is necessary.

Multi-pass data-driven approaches, e.g., dynamic mode decomposition (DMD), are ubiquitous for the approximation of the Koopman operator, the adjoint of the PF operator. Rowley et al. [29] describes the relationship between DMD and the Koopman operator; Schmid et al. [30] demonstrates the use of DMD on experimental fluid flow data; and Mohr et al. [31] extends mean ergodic theorem using Laplace averages to construct Koopman eigenfunctions. The extended DMD (EDMD) [32, 33] uses a time series of the observable data to approximate the eigenvalues of the Koopman operator on a finite-dimensional basis set chosen from a dictionary of appropriate functions.

EDMD solves an unconstrained least-squares problem to approximate the Koopman operator in the form of a matrix. Huang et al. [34] proposes a structured version of EDMD that takes the positivity of the Koopman operator into account to generate a more accurate estimate of the Koopman eigenfunctions. To replicate this approach for the PF operator, two challenges need to be addressed. Unlike Koopman, the PF operator is Markov, and hence preserves the measure. Moreover, the PF operator operates on densities, which implies it is not directly observable, so it must be approximated from a time series of Monte Carlo simulations.

1.1.2 Data-driven estimation of nonlinear system without a system model

With the availability of cheap sensors and high computational power, prediction and estimation techniques are being developed for systems when a reliable model is not available.

1.1.2.1 Prediction using an echo-state network

Recently developed machine learning techniques have been useful for solving a wide variety of problems, e.g., classification, speech recognition [35], and board games [36]. Recurrent neural networks (RNN) have been particularly useful for model-free prediction of dynamical systems. For example, an echo-state network (ESN) [37] can model chaotic systems with great effect [2, 38]. The ESN adopts an input-output neural network with a randomly generated recurrent reservoir. Linear regression determines the output weights. It has been shown that an ESN with fading memory can universally model nonlinear dynamics [39, 40]. However, these prediction techniques assume no measurements are available after training and rely instead on a free-running neural network to predict the dynamical system. But in many practical cases, a stream of sensor measurements, even if sparse and/or noisy, may be available. Such applications include fluid flow over an airfoil, atmospheric dynamics, and traffic network data. After training the neural network, estimates of the quantities of interest may be obtained with or without subsequent measurements.

Neural-network predictors do not use a dynamic model. Instead, they utilize the measurement and state data for training, and then run freely from an initial condition to predict future states. This prediction requires a reasonably accurate initial condition and does not incorporate any subsequent measurements. Some ESN-based methods [2] have been developed to utilize measurements for predicting unmeasured variables in the testing phase when all of the states are used during the training phase. These methods, however, feed the measurements directly into the ESN and rely on the ESN's structure to assimilate them for prediction of the unmeasured states. These methods also rely only on the current measurement, rather than the history of measurements, and do not take measurement noise into account.

1.1.2.2 Koopman operator based approach

The Koopman operator's action on an observable function is to describe its evolution along the trajectory of the original system. Being a linear operator, it can be used for spectral analysis. Koopman eigenfunctions are directly related to the geometry of the system dynamics, e.g., periodic partitions in an ergodic system [41]. A nonlinear flow can be characterized by its dominant Koopman modes [29]. The Koopman operator can be approximated from data, without knowledge of the system, by algorithms like extended dynamic mode decomposition [42] and Krylov-subspace methods [29]. An operator theoretic approach helps to transform a nonlinear system into a linear or bilinear one and enables the design of a controller that exploits the structure of the transformed system.

Operator-theoretic methods embed finite-dimensional dynamics in an infinite-dimensional function space in which functions evolve under a linear operator. Spectral properties of the Koopman operator are well analyzed, see, e.g., [41]. The Koopman operator offers effective methods to characterize a nonlinear system in terms of stability [43] and linearization [44]. The global stability properties of fixed points and limit cycles can also be characterized by Koopman eigenfunctions [43]. Furthermore, the Koopman operator underlies numerical algorithms like dynamic Mode decomposition [29, 30].

The application of Koopman methods to actuated systems has proven to be difficult, because the Koopman operator changes its spectral properties with the actuation signals. Proctor et al. [45] introduces a method to incorporate the control input in a Koopman framework by re-defining the Koopman operator with two arguments. Williams et al. [42] bridge the gap between the analysis and simulation by providing a method to determine the spectral properties of the Koopman operator of the underlying unforced system from the data of the actuated dynamics. Surana et al. [46] propose an observer design for a discrete-time nonlinear system using Koopman eigenfunctions. This framework is extended in [47] with the introduction of the Koopman canonical transform (KCT) for a continuous-time control-affine nonlinear system. This approach deals only with the Koopman operator related to the unactuated part of the dynamics and relies on Lipschitz continuity of the control vector fields to design a high-gain observer.

1.2 Contributions to the state-of-the-art

This dissertation provides research contributions in the general areas of model-based and model-free estimation and control of nonlinear systems. The main results of this dissertation have been published or submitted for publication in archival journals [48, 49, 50, 51]. Earlier research results related to this dissertation appeared in conference proceedings

[52, 53]. Some additional discussion and results appear in this dissertation that have not appeared elsewhere. Unless otherwise stated, the following claims of contribution were performed together with colleagues in the aforementioned publications.

We formulate a Gaussian mixture model based Kalman filter (extended/unscented Kalman filter for nonlinear systems) with a stochastic boundedness guarantee and an output feedback control design using the estimate. We derive an improved approximation technique for the Perron-Frobenius (PF) operator to use it for the motion update for nonlinear estimation. This method is applied in output feedback of a pitching airfoil to maximize its lift performance.

We also derive a framework for data-driven estimation of complex systems using an echo-state network (ESN), a special kind of recurrent neural network equipped with an ensemble Kalman filter. This method is used to improve the estimation accuracy over time for a chaotic nonlinear system and to predict daily cycles of traffic congestion from a real set of mobility data. The model-free estimation algorithm developed here has wide applications for estimation of complex dynamics when a reliable model is not available.

Finally, a control framework for a control-affine nonlinear system is designed from data using the Koopman canonical transform (KCT) [47]. The nonlinear system is approximated as a bilinear one and the convergence of such approximation is proved. An optimal control framework is then developed using Pontryagin's principle.

1.3 Outline of the dissertation

The organization of the dissertation is as follows. Part I includes chapters 2 and 3 that discuss the model-based estimation and output feedback strategies. Chapter 2 describes the Gaussian mixture models (GMM) as a semi-parametric approach to approximate the probability density and Kalman filter based motion and measurement updates are used to propagate each of the Gaussian modes. Alspach and Sorenson [12] proved that any sufficiently smooth probability density function is approximated arbitrarily closely by an ensemble realization of a weighted sum of Gaussian PDFs, thereby providing an approximate way to explicitly calculate the *posteriori* density of the states of a stochastic nonlinear system, even in the presence of non-Gaussian process noise. Stochastic boundedness of the estimation-errors from GMM-based filters and the corresponding output feedback control are analyzed and numerically illustrated.

Chapter 3 proposes a novel finite dimensional approximation technique for the Perron-Frobenius (PF) operator for motion update in a Bayesian nonlinear filter. Application of the proposed method for the ouput feedback control of a pitching airfoil shows its potential for nonlinear estimation and control.

In part II, we explore data-driven methods for estimation and output feedback for nonlinear systems. Part II includes chapters 4 and 5. Chapter 4 derives an echo-state network (ESN) based estimation framework for data-driven estimation of nonlinear systems. A case-study with a real set of mobility data for the traffic flow prediction is also presented.

Chapter 5 provides a data-driven control framework for partially modeled control-affine nonlinear systems with unknown drift vector field. A control-affine system is approximated as a bilinear one through its Koopman spectra and the convergence of such approximation is proved. This chapter also includes controllability results and an optimal-control design for the approximated bilinear system using the Pontryagin's principle.

Chapter 6 summarizes the primary contributions of of this dissertation and and discusses directions for future research.

Part I

Model-based estimation and control of nonlinear systems

Chapter 2

Non-Gaussian estimation and dynamic output feedback using the Gaussian mixture model Kalman filter

For a nonlinear system with non-Gaussian process noise, the optimal estimator is represented by a recursive Bayesian filter that converts the prior probability density function (PDF) into a posterior PDF using a likelihood function of the observation. However, in many cases, the prior PDF is not explicitly known, and has to be approximated from an ensemble realization. The most common approach is to use a particle filter [11] with an effective method of re-sampling to tackle highly non-Gaussian distributions. But sampling and re-sampling become computationally intractable for a high-dimensional system. In this chapter we look into the Gaussian Mixture Model (GMM) as a semi-parametric approach to facilitate the density propagation for the recursive Bayesian estimation process.

Alspach and Sorenson [12] proved that any sufficiently smooth PDF can be approximated arbitrarily closely by an ensemble realization of a weighted sum of Gaussian PDFs. This technique, called the Gaussian mixture model (GMM), gives an approximate way to explicitly calculate the *posteriori* density of the states of a stochastic nonlinear system, even in the presence of non-Gaussian process noise.

GMM represents the PDF using a weighted sum of Gaussian PDFs, each of which is updated using individual Kalman filters if the measurement model is linear and the measurement noise profile is Gaussian. This chapter utilizes the Expectation Maximization (EM) algorithm, and Bayesian Information Criteria (BIC) to determine the Gaussian mode parameters and the total number of such modes. Since many sensors used in atmospheric and oceanic sampling have a linear (or linearized) observation model and additive Gaussian measurement noise profile [13], we have assumed a linear observation profile in the rest of the chapter. Taking these assumptions into account, GMM along with Kalman filter updates proves to be an efficient estimator, even for non-Gaussian process noise. For nonlinear systems, the Kalman update can be replaced by the extended Kalman filter (EKF), which gives the prediction up to first-order precision [21]. The unscented Kalman filter (UKF) [22], [23] uses the unscented Transform to transport the Gaussian covariance up to third-order precision with significantly less computational effort compared to Monte-Carlo covariance propagation in each step.

This chapter uses the GMM framework to capture the non-Gaussian noise profile, and uses the Kalman filter and its variants (EKF and UKF) to propagate the parameters of the GMM forward in time. This approach does not require re-sampling whenever an observation is obtained, thus saving computational cost. This chapter also proposes the notion of *bounded with probability* to provide an analytical justification of GMM-KF, GMM-EKF, and GMM-UKF.

The goal of this chapter is to provide a theoretical guarantee of stochastic boundedness of the estimation error while using GMM-KF/GMM-EKF/GMM-UKF and a proof of the ultimate boundedness of the state in a nonlinear system using Lyapunov's method for observer-based feedback.

2.1 The Gaussian mixture model

The Gaussian mixture model (GMM) provides a semi-parametric approach to quantify the prior distribution from a set of ensemble realizations [54]. It uses a weighted sum of Gaussian distributions to approximate a non-Gaussian prior. The primary advantage is that each of the Gaussian modes is updated using the Kalman filter equations, augmented with a weight analysis.

The Gaussian mixture model Kalman filter (GMM-KF) has three principal components [13]: the Gaussian mixture model, the expectation maximization (EM) algorithm and the Bayesian information criterion . Let w_j, $j = 1, \ldots, M$, be scalar weights such that $\sum_{j=1}^{M} w_j = 1$. Let $\overline{\mathbf{x}}_j$ and P_j be the mean vector and covariance matrix respectively for a multivariate Gaussian $\mathcal{N}\left(X; \overline{X}_j, P_j\right)$, $j = 1, \ldots, M$. The weighted sum of the M Gaussian densities [13]

$$p_X\left(\mathbf{x}; \{(w^j, \overline{\mathbf{x}}^j, P^j)\}_{j=1}^{M}\right) = \sum_{j=1}^{M} w^j \mathcal{N}\left(\mathbf{x}; \overline{\mathbf{x}}^j, P^j\right) \tag{2.1}$$

is a valid PDF known as a Gaussian mixture that integrates to unity and has an analytical representation. Through the selection of the weights, means, covariances, and number of mixture components, (2.1) can represent highly non-Gaussian distributions with vanishing error [12].

The EM algorithm [14] is an iterative procedure for estimating the parameters of a target distribution that maximize the probability of obtaining a given set of realizations $\mathbf{x} = \{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$. This procedure involves the maximization of a log-likelihood function given by the log of the product of the parametric distribution with the likelihoods of each of the ensemble members: i.e., maximizing the logarithm of the likelihood function $p_{\mathbf{X}}(\mathbf{x};\theta) = \prod_{i=1}^{N} p_{\mathbf{X}_i}(\mathbf{x}_i;\theta)$ with respect to θ, where θ is the parameter vector and N is the number of ensemble members. Given the set of ensemble realizations, $\mathbf{x} = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$, and initial parameter estimates, $\theta^{(0)} = \{w^1_{(0)}, \cdots, w^n_{(0)}, \overline{\mathbf{x}}^1_{(0)}, \cdots, \overline{\mathbf{x}}^n_{(0)}, P^1_{(0)}, \cdots, P^n_{(0)}\}$, The EM algorithm for a GMM [13] is given in Algorithm 1.

Step (2.2) of the EM algorithm calculates the probability of mixture component j having generated realization $\mathbf{x}_i$ based on the present parameter estimates, across all possible pairs of realizations and components. Step (2.3) updates the parameter values in accordance with their weighted averages across all realizations. The convergence of EM algorithm is proved by first showing that the likelihood function increases in each iteration and is bounded above, and then using the monotone-convergence theorem [55].

To choose the optimal mixture complexity, we maximize

$$p_{\mathbf{X}|\Theta}(\mathbf{x}|\theta; M) = \prod_{i=1}^{N} p_{\mathbf{X}_i|\Theta}(\mathbf{x}_i|\theta; M),$$

Algorithm 1: Expectation Maximization for GMM

```
input  : Ensemble $\mathbf{x} = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$
output: $\theta = \{w^1, \cdots, w^n, \bar{\mathbf{x}}^1, \cdots, \bar{\mathbf{x}}^n, P^1, \cdots, P^n\}$
Initialize: $\theta^{(0)} = \{w^1_{(0)}, \cdots, w^n_{(0)}, \bar{\mathbf{x}}^1_{(0)}, \cdots, \bar{\mathbf{x}}^n_{(0)}, P^1_{(0)}, \cdots, P^n_{(0)}\}$
repeat
    for $i \in \{1, \cdots, N\}$ and for $j \in \{1, \cdots, M\}$ do
        $$\tau^j_k(\mathbf{x}_i) = \frac{w^j_{(k)} \mathcal{N}\left(\mathbf{x}_i; \bar{\mathbf{x}}^j_{(k)}, P^j_{(k)}\right)}{\sum_{m=1}^{M} w^m \mathcal{N}(\mathbf{x}_i; \bar{\mathbf{x}}^m_{(k)}, P^m_{(k)})}. \qquad (2.2)$$
    end
    for $j \in \{1, \cdots, M\}$ and $N^j_k = \sum_{i=1}^{N} \tau^j_k(\mathbf{x}_i)$ do
        $$\begin{aligned} w^j_{k+1} &= \frac{N^j_k}{N} \\ \bar{\mathbf{x}}^j_{k+1} &= \frac{1}{N^j} \sum_{i=1}^{N} \tau^j_k(\mathbf{x}_i)\mathbf{x}_i \\ P^j_{k+1} &= \frac{1}{N^j_k} \sum_{i=1}^{N} \tau^j_k(\mathbf{x}_i)(\mathbf{x}_i - \bar{\mathbf{x}}^j_{k+1})(\mathbf{x}_i - \bar{\mathbf{x}}^j_{k+1})^T. \qquad (2.3) \end{aligned}$$
    end
until convergence
```

where $p_{\mathbf{X}|\Theta}(\mathbf{x}|\theta, M)$ is the distribution of the ensemble conditioned on the parameters θ. The Bayesian information criterion (BIC) is [13]

$$\text{BIC} = \min_{M} \left[-2 \sum_{i=1}^{N} \log p_{\mathbf{X}_i|\Theta}(\mathbf{x}_i|\theta_{\text{ML}}; M) + K \log N \right], \tag{2.4}$$

where K is the number of parameters in the model, θ_{ML} is the maximum likelihood set of parameters (e.g., produced by the EM algorithm). For a multivariate Gaussian mixture, $K = M\left(2d + (d(d-1))/2 + 1\right)$ is the number of free parameters, where d is the dimension of the state vector. Note that BIC has two components: the first component evaluates the goodness-of-fit for the model of complexity M and the second component is a penalty on the overall model complexity [13].

For many data-assimilation applications [13], the observation-operator C linearly extracts the measurement from the state vector, i.e.,

$$\mathbf{y} = C\mathbf{x} + \mathbf{V}, \text{where } \mathbf{V} \sim \mathcal{N}(\mathbf{0}, K_V), \tag{2.5}$$

where $\mathbf{V}$ is a zero-mean Gaussian measurement noise with covariance K_V. For a (single) Gaussian forecast PDF, Gaussian measurement noise, and a linear observation operator, the Kalman filter analysis equations represent the optimal approach to Bayesian assimilation of a measurement. In the case of a mixture of Gaussians, the Kalman analysis equations may be augmented as follows with a weight analysis equation to yield the proper application of Bayes' rule for each component in the mixture [13].

The Bayesian update of a Gaussian mixture prior with a Gaussian observation model yields a Gaussian mixture posterior [13]. For prior GMM

$$p_{\mathbf{X}}\left(\mathbf{x}\right) = \sum_{j=1}^{M} w^j \mathcal{N}\left(\mathbf{x}; \overline{\mathbf{x}}^j, P^j\right),$$

and a Gaussian observation model $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}; C\mathbf{x}, R)$, the posterior PDF is [13]

$$p_{\mathbf{X}|\mathbf{Y}}\left(\mathbf{x}|\mathbf{y}\right) = \sum_{j=1}^{M} \hat{w}^j \mathcal{N}\left(\mathbf{x}; \hat{\mathbf{x}}^j, \hat{P}^j\right), \tag{2.6}$$

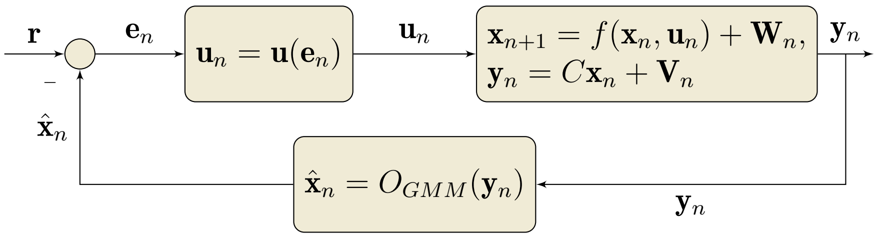


Figure 2.1: Block diagram of a discrete time closed-loop dynamic output feedback control system with a Gaussian mixture model filter O_{GMM}

where

$$\begin{aligned}
\hat{\mathbf{x}}^j &= \overline{\mathbf{x}}^j + K^j(\mathbf{y} - C\overline{\mathbf{x}}^j) \\
K^j &= P^j C^T (C P^j C^T + K_V)^{-1} \\
\hat{P}^j &= (I - K^j C) P^j \\
\hat{w}^j &= \frac{w^j \mathcal{N}(\mathbf{y}; C\overline{\mathbf{x}}^j, CP^jC^T + K_V)}{\sum_{m=1}^{M} w^m \mathcal{N}(\mathbf{y}; C\overline{\mathbf{x}}^m, CP^mC^T + K_V)}.
\end{aligned} \tag{2.7}$$

Combining the weighted Gaussians produces the posterior-mean estimate.

The state estimate thus obtained may be used for feedback stabilization as shown in Fig. 2.1. In this model, $\mathbf{x}_n$ and $\mathbf{y}_n$ are the state and observation vectors, respectively, at the n^{th} time step. The process is inherently noisy with an additive process noise $\mathbf{W}_n$ and the observation is corrupted by the measurement noise $\mathbf{V}_n$. The state-estimate $\hat{\mathbf{x}}_n$ is obtained using a GMM-based filter (denoted O_{GMM}) and this estimate is used to derive the output feedback control signal $\mathbf{u}_n = K(\mathbf{r} - \hat{\mathbf{x}}_n)$. The objective of the output feedback is to effectively stabilize the system so that the output follows the desired value $\mathbf{r}$.

2.2 Linear system with non-Gaussian noise

To analytically investigate the performance of the GMM-KF for non-Gaussian process noise, we start with a linear system. For theoretical tractability, we calculate the initial Gaussian modes and their parameters with a sufficiently large mixture complexity M, and then propagate forward in time including updates with observations. In the sequel, all norms used in this chapter will be 2-norms or induced 2-norms on Euclidean space. The eigenvalue of the matrix A with largest absolute value is denoted by $|\lambda_{max}(A)|$. The absolute value sign is dropped for symmetric matrices, and the argument is dropped when the context is clear.

Consider a linear system

$$\mathbf{x}_{n+1} = A\mathbf{x}_n + \mathbf{W}_n \text{ and } \mathbf{y}_n = C\mathbf{x}_n + \mathbf{V}_n, n \geq 1, \tag{2.8}$$

where $\mathbf{x}_n \in \mathbb{R}^d$, $\mathbf{y}_n \in \mathbb{R}^q$ and $\mathbf{W}_n, \mathbf{V}_n, n \geq 1$, are zero mean i.i.d. Gaussian random vectors with $\text{cov}(\mathbf{V}_n) = K_V$ and $\text{cov}(\mathbf{W}_n) = K_W$. The best estimate of the state is provided by the Kalman filter: i.e.,

$$
\begin{aligned}
\overline{\mathbf{x}}_n &= A\hat{\mathbf{x}}_{n-1} \\
\hat{\mathbf{x}}_n &= \overline{\mathbf{x}}_n + K_n(\mathbf{y}_n - C\overline{\mathbf{x}}_n) \\
K_n &= S_n C^T (C S_n C^T + K_V)^{-1} \\
S_n &= A\Sigma_{n-1}A^T + K_W \\
\Sigma_n &= (I - K_n C)S_n,
\end{aligned}
\tag{2.9}
$$

where $S_n = \text{cov}(\mathbf{x}_n - \overline{\mathbf{x}}_n)$ and $\Sigma_n = \text{cov}(\mathbf{x}_n - \hat{\mathbf{x}}_n)$. The error $\tilde{\mathbf{x}}_n$ for the Kalman filter is given by $\tilde{\mathbf{x}}_n = (I - K_n C)\left(A\tilde{\mathbf{x}}_{n-1} + \mathbf{W}_{n-1}\right) - K_n \mathbf{V}_n$. The convergence of the error covariance for the Gaussian process noise is given by the following well-known result.

Lemma 2.1. [56]: Let $K_W = QQ^T$. Suppose that (A, Q) is reachable and (A, C) is observable. Then $\Sigma_n \to \Sigma, K_n \to K, S_n \to S$ as $n \to \infty$. The limiting matrices are the only solutions of the equations $\Sigma = (I - KC)S$, $K = SC^T(CSC^T + K_V)^{-1}$, and $S = A\Sigma A^T + K_W$.

To include non-Gaussian process noise, consider the system in Eq. (2.8) with $\mathbf{W}_n$ being non-Gaussian with zero mean and having finite second moment and $\mathbf{V}_n$ is i.i.d. We model the prior using GMM and propagate using Kalman-like updates. The GMM-KF is, for

$j = 1, \cdots, M$,

$$\begin{aligned}
\overline{\mathbf{x}}_n^j &= A\hat{\mathbf{x}}_{n-1}^j \\
\hat{\mathbf{x}}_n^j &= \overline{\mathbf{x}}_n^j + K_n^j(\mathbf{y}_n - C\overline{\mathbf{x}}_n^j) \\
K_n^j &= S_n^j C^T (C S_n^j C^T + K_V)^{-1} \\
S_n^j &= A\Sigma_{n-1}^j A^T + K_W \\
\Sigma_n^j &= (I - K_n^j C) S_n^j \\
w_n^j &= \frac{w_{n-1}^j \mathcal{N}(\mathbf{y}_n; C\overline{\mathbf{x}}_n^j, C S_n^j C^T + K_V)}{\sum_{m=1}^M w_{n-1}^m \mathcal{N}(\mathbf{y}_n; C\overline{\mathbf{x}}_n^m, C S_n^m C^T + K_V)}
\end{aligned} \tag{2.10}$$

$$\hat{\mathbf{x}}_n = \sum_{j=1}^M w_n^j \hat{\mathbf{x}}_n^j, \tag{2.11}$$

where $\sum_{j=1}^M w_n^j = 1 \ \forall\, n$, $S_n^j = \mathrm{cov}(\mathbf{x}_n - \overline{\mathbf{x}}_n^j)$ and $\Sigma_n^j = \mathrm{cov}(\mathbf{x}_n - \hat{\mathbf{x}}_n^j)$. Assume that the initial values of the parameters have been set by the EM algorithm and BIC.

Usually covariance analysis is the best technique to describe a filter's performance. But due to analytical intractability created by several random variables involved in GMM-KF, we introduce the notion of *bounded with probability*.

Definition 2.1. A random vector $\mathbf{X}$ is said to be *bounded by b with probability P* if $\mathbb{P}[\|\mathbf{X}\| \leq b] = P$. Furthermore $\mathbf{X}$ is said to be *bounded by b with probability at least P* if $\mathbb{P}[\|\mathbf{X}\| \leq b] \geq P$.

Remark 2.1. This definition of a stochastic bound gives a practical way to quantify the performance of a nonlinear filter dealing with non-Gaussian noise when moments are difficult to be computed. Higher probability usually means a more relaxed bound. A similar notion of *practical stability* is discussed in [57].

To show that $\|\mathbf{x}_n - \hat{\mathbf{x}}_n\|$ is bounded with a specified probability, we proceed to the error analysis of the estimator, where the error $\tilde{\mathbf{x}}_n \triangleq \mathbf{x}_n - \hat{\mathbf{x}}_n$ is

$$\tilde{\mathbf{x}}_n = \sum_{j=1}^{M} w_n^j \tilde{\mathbf{x}}_n^j = \sum_{j=1}^{M} w_n^j (\mathbf{x}_n - \hat{\mathbf{x}}_n^j)\,. \tag{2.12}$$

The following two lemmas describe linear algebraic results related to the concept of bounded with probability.

Lemma 2.2. For a given $Q = Q^T > 0$, $\left\|\mathbf{x}\exp(-\mathbf{x}^T Q\mathbf{x})\right\|$, a function of $\mathbf{x} \in \mathbb{R}^d$, is bounded.

Proof.

$$\begin{aligned} \left\|\mathbf{x}\exp(-\mathbf{x}^T Q\mathbf{x})\right\| &\leq \|\mathbf{x}\|\,|\exp(-\mathbf{x}^T Q\mathbf{x})| \\ &\leq \|\mathbf{x}\|\,|\exp(-\lambda_{min}(Q)\,\|\mathbf{x}\|^2)|, \\ &\quad \text{since } Q > 0, \lambda_{min}(Q) > 0, \\ &\leq \frac{1}{\sqrt{2e\lambda_{min}(Q)}}, \end{aligned} \tag{2.13}$$

from single-variable calculus, where e is the base of the natural logarithm. The last step in the inequality is verified by calculating the maximum value of $r\exp(-\lambda_{min}(Q)r^2)$ with $r \geq 0$ by setting the first derivative to zero. □

Lemma 2.3. For all $\mathbf{x} \in \mathbb{R}^d$ and for a random vector $\mathbf{R} \in \mathbb{R}^d$, $\left\|(\mathbf{x} + \mathbf{R})\exp(-\mathbf{x}^T Q\mathbf{x})\right\|$ with $Q > 0$ and symmetric is bounded by $\dfrac{1}{\sqrt{2e\lambda_{min}(Q)}} + b$ with probability at least P if $\mathbf{R}$ is bounded by b with probability P.

Proof. From Lemma 2.2,

$$\left\|(\mathbf{x}+\mathbf{R})\exp(-\mathbf{x}^T Q\mathbf{x})\right\| \leq \frac{1}{\sqrt{2e\lambda_{min}(Q)}} + \|\mathbf{R}\|\,, \tag{2.14}$$

since $|\exp(-\mathbf{x}^T Q\mathbf{x})| \leq |\exp(-\lambda_{min}(Q)\,\|\mathbf{x}\|^2)| \leq 1$. Therefore, $\mathbb{P}[\|\mathbf{R}\| \leq b] = P$, which implies

$$\mathbb{P}\left[\frac{1}{\sqrt{2e\lambda_{min}(Q)}} + \|\mathbf{R}\| \leq \frac{1}{\sqrt{2e\lambda_{min}(Q)}} + b\right] = P,$$

and, hence, from Eq. (2.14),

$$\mathbb{P}\left[\left\|(\mathbf{x}+\mathbf{R})\exp(-\mathbf{x}^T Q\mathbf{x})\right\| \leq \frac{1}{\sqrt{2e\lambda_{min}(Q)}} + b\right] \geq P.$$

□

Lemmas 2.2 and 2.3 are useful in establishing error bounds of the GMM-KF for a linear non-Gaussian system. The primary error bound result is given next.

Theorem 2.1. Consider the system (2.8) with possibly non-Gaussian noise $\mathbf{W}_n$. Let $K_{\mathbf{W}} = QQ^T$ where (A, Q) is reachable and (A, C) is observable. If C has full column rank and $\mathbf{V}_n$ is bounded by b_V with probability P, then $\|\tilde{\mathbf{x}}_n\| = \|\mathbf{x}_n - \hat{\mathbf{x}}_n\|$, where $\hat{\mathbf{x}}_n$ is obtained from GMM-KF (2.10), is bounded by b with probability P, where b is an increasing function of b_V.

Proof. From Lemma 2.1, $\Sigma^j = (I - KC)S^j$, $K^j = S^j C^T(CS^j C^T + K_V)^{-1}$, and $S^j = A\Sigma^j A^T + K_W, \forall\, j = 1, \cdots, M$, are the bounded limits of Σ^j_n, K^j_n, and S^j_n as $n \to \infty$. Let

$R_n^j = CS_n^jC^T + K_V$, d be the dimensionality of $\mathbf{x}_n$, and

$$\beta_n = \sum_{m=1}^{M} w_{n-1}^m \mathcal{N}(\mathbf{y}_n; C\overline{\mathbf{x}}_n^m, CS_n^mC^T + K_V)$$

be a finite normalization factor. From the Kalman update equation (2.10), $\tilde{\mathbf{x}}_n^j = (I - K_n^jC)\,\tilde{r}_n^j - K_n^j\mathbf{V}_n$, where $\tilde{r}_n^j \triangleq A\tilde{\mathbf{x}}_{n-1}^j + \mathbf{W}_{n-1}$. Using this result along with Eq. (2.10), Eq. (2.12) can be expanded as

$$\begin{aligned}
\tilde{\mathbf{x}}_n &= \sum_{j=1}^{M} w_n^j \tilde{\mathbf{x}}_n^j \\
&= \frac{1}{\beta_n}\sum_{j=1}^{M} w_{n-1}^j \mathcal{N}(\mathbf{y}_n; C\overline{\mathbf{x}}_n^j, CS_n^jC^T + K_V)\tilde{\mathbf{x}}_n^j \\
&= \frac{1}{\beta_n}\sum_{j=1}^{M} w_{n-1}^j \mathcal{N}(C(A\mathbf{x}_{n-1} + \mathbf{W}_{n-1}) + \mathbf{V}_n; CA\hat{\mathbf{x}}_{n-1}^j, R_n^j)\tilde{\mathbf{x}}_n^j \\
&= \frac{1}{\beta_n}\sum_{j=1}^{M} w_{n-1}^j \frac{1}{\sqrt{\det(2\pi R_n^j)}} e^{-\frac{1}{2}\left(C\tilde{r}_n^j + \mathbf{V}_n\right)^T (R_n^j)^{-1}\left(C\tilde{r}_n^j + \mathbf{V}_n\right)} \left(\left(I - K_n^jC\right)\tilde{r}_n^j - K_n^j\mathbf{V}_n\right). \\
&= \left[\frac{1}{\beta_n}\sum_{j=1}^{M} w_{n-1}^j \frac{1}{\sqrt{\det(2\pi R_n^j)}} e^{-\frac{1}{2}\left(C\tilde{r}_n^j + \mathbf{V}_n\right)^T (R_n^j)^{-1}\left(C\tilde{r}_n^j + \mathbf{V}_n\right)} \tilde{r}_n^j\right] \\
&\quad - \left[\frac{1}{\beta_n}\sum_{j=1}^{M} w_{n-1}^j \frac{1}{\sqrt{\det(2\pi R_n^j)}} \times e^{-\frac{1}{2}\left(C\tilde{r}_n^j + \mathbf{V}_n\right)^T (R_n^j)^{-1}\left(C\tilde{r}_n^j + \mathbf{V}_n\right)} \right. \\
&\quad \left. \times K_n^j\left(C\tilde{r}_n^j + \mathbf{V}_n\right)\right]. \qquad (2.15)
\end{aligned}$$

From Lemma 2.1, $K_n^j \to K^j$ and $R_n^j \to R^j$ as $n \to \infty$ with finite-norm limit and $(R_n^j)^{-1}$ is positive definite. By Lemma 2.2, each term in the second summation of the Eq. (2.15) is upper bounded by

$$b_j = \frac{\|K^j\|}{\beta_n \sqrt{\det(2\pi R^j)\lambda_{min}((R^j)^{-1})e}}. \tag{2.16}$$

To deal with the first term in (2.15), let $q_n^j = C\tilde{r}_n^j + \mathbf{V}_n$, which in turn yields $\tilde{r}_n^j = (C^T C)^{-1} C^T (q_n^j - \mathbf{V}_n)$, since C has full column rank. Replacing $\tilde{r}_n^j$ by q_n^j in the first term of Eq. (2.15) yields

$$\frac{1}{\beta_n} \sum_{j=1}^{M} w_{n-1}^j \frac{1}{\sqrt{\det(2\pi R_n^j)}} \times e^{-\frac{1}{2}(q_n^j)^T (R_n^j)^{-1} q_n^j} (C^T C)^{-1} C^T (q_n^j - \mathbf{V}_n).$$

Each term of the summation is bounded by $\tilde{b}_j$ with probability at least P if $\mathbf{V}_n$ is bounded by b_V with probability P by Lemma 2.3. The bound $\tilde{b}_j$ is

$$\tilde{b}_j = \frac{1}{\sqrt{\det(2\pi R^j)}\beta_n \|C\|} \left(\frac{1}{\sqrt{e\lambda_{min}((R^j)^{-1})}} + b_V \right), \tag{2.17}$$

as $n \to \infty$. Since $\sum_{j=1}^{M} w_{n-1}^j = 1$, the error $\tilde{\mathbf{x}}_n$ is bounded by b with probability at least P, where $b = \max_j (b_j + \tilde{b}_j)$. □

Remark 2.2. The condition of C having full column rank is restrictive, but it ensures an analytical bound in the presence of non-Gaussian process noise. In practice, the filter has been seen to work well as long as (A, C) is observable.

Remark 2.3. The bound presented in Theorem 2.1 analytically justifies the use of the GMM-KF framework for non-Gaussian estimation. Although this bound may be calculated for a linear system, it can be quite conservative with respect to the actual error. When this framework is extended to nonlinear systems in the next section, it might not be possible to always calculate the bound explicitly.

Remark 2.4. There might be a possibility of imbalance in weights w_j if the GMM-KF runs for a long time. That scenario can possibly be avoided by recalculating the GMM-KF parameters from an ensemble realization with the help of EM and BIC when some weights get too low. Theorem 2.1 is valid in each time span between such re-computations.

Next, we utilize the GMM-KF estimate inside the feedback loop depicted in Fig. 2.1.

Theorem 2.2. Consider the system

$$\begin{aligned} \mathbf{x}_{n+1} &= A\mathbf{x}_n + B\mathbf{u}_n + \mathbf{W}_n, \\ \mathbf{y}_n &= C\mathbf{x}_n + \mathbf{V}_n,\ n \geq 1 \end{aligned}$$

where $\mathbf{W}_n$ is a zero mean i.i.d. possibly non-Gaussian process with finite second moment, and $\mathbf{V}_n$ is i.i.d. Gaussian measurement noise. The dynamic output feedback control signal $\mathbf{u}_n = -K\hat{\mathbf{x}}_n$ is chosen such that the eigenvalue with the maximum magnitude satisfies $|\lambda_{max}(A - BK)| < 1$, where $\hat{\mathbf{x}}_n$ is from GMM-KF. Then the state remains bounded at every time instant with probability P if $\mathbf{V}_n$ and $\mathbf{W}_n$ are bounded with probability P for all $n \geq 0$.

Proof. By algebraic expansion we get

$$\begin{aligned}
\mathbf{x}_{n+1} &= A\mathbf{x}_n - BK\hat{\mathbf{x}}_n + \mathbf{W}_n \\
&= (A - BK)\mathbf{x}_n + BK(\mathbf{x}_n - \hat{\mathbf{x}}_n) + \mathbf{W}_n \\
&= (A - BK)^n\mathbf{x}_0 \qquad (2.18) \\
&\quad +(A - BK)^{n-1}(BK(\mathbf{x}_1 - \hat{\mathbf{x}}_1) + \mathbf{W}_0) \\
&\quad + \cdots + BK(\mathbf{x}_n - \hat{\mathbf{x}}_n) + \mathbf{W}_n
\end{aligned}$$

Hence,

$$\begin{aligned}
\|\mathbf{x}_{n+1}\| &\le \|(A - BK)^n\| \, \|\mathbf{x}_0\| \\
&\quad + \|(A - BK)^{n-1}\| \, \|BK\| \, (\|\mathbf{x}_1 - \hat{\mathbf{x}}_1\| \qquad (2.19) \\
&\quad + \|\mathbf{W}_0\|) + \cdots + \|BK\| \, \|\mathbf{x}_n - \hat{\mathbf{x}}_n\| + \|\mathbf{W}_n\| \, .
\end{aligned}$$

Now it can be shown [58] that if $|\lambda_{max}(A - BK)| < 1$, then $\sum_{k=1}^{\infty} \left\|(A - BK)^k\right\| \le \beta$ for some finite β. Hence the right-hand side converges with probability P if the process noise is bounded with probability P, because $\|\mathbf{x} - \hat{\mathbf{x}}_n\|$ is bounded with probability at least P from Theorem 2.1. □

Remark 2.5. The assumptions that $\mathbf{W}_n$ and $\mathbf{V}_n$ are bounded with certain probability is valid in most cases, because the probability that the noise reaches a very high value is vanishingly small.

The covariance of the process noise might be computed *a priori* in GMM-KF if the statistics of the process noise is known, or by Monte-Carlo simulation. Otherwise it might be used as a tuning parameter to tune the performance of the filter.

2.3 Extension to a nonlinear system

The extension of the GMM-KF framework to a nonlinear system without re-sampling at each observation step requires a method to propagate the Gaussian modes forward in time. To propagate GMM parameters in the nonlinear setting, we employ two widely used extensions of the Kalman filter for nonlinear systems: the extended Kalman filter (EKF) [21] and the unscented Kalman filter [23]. We retain the linear observation model with Gaussian measurement noise. The system looks like

$$\mathbf{x}_{n+1} = f(\mathbf{x}_n) + \mathbf{W}_n \text{ and } \mathbf{y}_n = C\mathbf{x}_n + \mathbf{V}_n, n \geq 1, \tag{2.20}$$

where $\mathbf{W}_n$ is zero mean possibly non-Gaussian additive noise with finite second moment and $\mathbf{V}_n \sim \mathcal{N}(\mathbf{0}, K_V)$.

2.3.1 Gaussian mixture model extended Kalman filter

Using EKF for propagation of means and covariances, the GMM framework becomes, for $j = 1,\cdots, M$,

$$\begin{aligned}
\overline{\mathbf{x}}_n^j &= f(\hat{\mathbf{x}}_{n-1}^j) \\
\hat{\mathbf{x}}_n^j &= \overline{\mathbf{x}}_n^j + K_n^j(\mathbf{y}_n - C\overline{\mathbf{x}}_n^j) \\
K_n^j &= S_n^j C^T (C S_n^j C^T + K_V)^{-1} \\
S_n^j &= A(\hat{\mathbf{x}}_{n-1})\Sigma_{n-1}^j A(\hat{\mathbf{x}}_{n-1})^T + K_W
\end{aligned} \tag{2.21}$$

$$\begin{aligned}\Sigma_n^j &= (I - K_n^j C) S_n^j \\ w_n^j &= \frac{w_{n-1}^j \mathcal{N}(\mathbf{y}_n; C\overline{\mathbf{x}}_n^j, CS_n^j C^T + K_V)}{\sum_{m=1}^M w_{n-1}^m \mathcal{N}(\mathbf{y}_n; C\overline{\mathbf{x}}_n^m, CS_n^m C^T + K_V)} \\ \hat{\mathbf{x}}_n &= \sum_{j=1}^M w_n^j \hat{\mathbf{x}}_n^j, \end{aligned} \tag{2.22}$$

where $\sum_{j=1}^{M} w_n^j = 1 \; \forall n$, $S_n^j = \mathrm{cov}(\mathbf{x}_n - \overline{\mathbf{x}}_n^j)$, $\Sigma_n^j = \mathrm{cov}(\mathbf{x}_n - \hat{\mathbf{x}}_n^j)$ and $A(\mathbf{x}) = J_f(\mathbf{x})$ is the Jacobian of f evaluated at $\mathbf{x}$. The initial values of the parameters have been set by the EM algorithm and BIC as in the linear case. The error dynamics are

$$\begin{aligned}\tilde{\mathbf{x}}_n &= \sum_{j=1}^M w_n^j \tilde{\mathbf{x}}_n^j \\ &= \frac{1}{\beta_n} \sum_{j=1}^M w_{n-1}^j \frac{1}{\sqrt{\det(2\pi R_n^j)}} e^{-\frac{1}{2}(C\tilde{\mathbf{f}}_n^j + \mathbf{V}_n)^T (R_n^j)^{-1} (C\tilde{\mathbf{f}}_n^j + \mathbf{V}_n)} \\ &\quad \times \left((I - K_n^j C)\tilde{\mathbf{f}}_n^j - K_n^j \mathbf{V}_n \right), \end{aligned} \tag{2.23}$$

where $\tilde{\mathbf{f}}_n^j = f(\mathbf{x}_{n-1}) - f(\hat{\mathbf{x}}_{n-1}^j) + \mathbf{W}_{n-1}$ and $R_n^j = CS_n^j C^T + K_V$. To have $\|\tilde{\mathbf{x}}_n\|$ bounded in the nonlinear case, we need stricter assumptions than before, because R_n^j and S_n^j no longer converge to finite-norm matrices when $n \to \infty$.

Assumption 2.1. $\alpha_j I \leq \Sigma_n^j < \beta_j I$, where $\alpha_j, \beta_j \geq 0$, for all $n \geq 0$ and $j = 1, \cdots, M$.

Remark 2.6. Assumption 2.1 dictates that the covariances of each Gaussian mode remains bounded through the time span over which the filter runs. In a general nonlinear system this might not be satisfied, but it can be enforced by resampling from the ensemble and recalculating the GMM parameters if any covariance violates a pre-determined threshold, i.e., if $\Sigma_n^j > \beta_j I$ for a pre-determined β_j for any $j \in 1, \ldots, M$.

Theorem 2.3. Consider the system (2.20) with possibly non-Gaussian noise $\mathbf{W}_n$. If Assumption 2.1 is satisfied, C has full column rank, and $\mathbf{V}_n$ is bounded by b_V with probability P, then $\|\tilde{\mathbf{x}}_n\| = \|\mathbf{x}_n - \hat{\mathbf{x}}_n\|$ is bounded by $b_{\tilde{x}}(P)$ with probability P, where $b_{\tilde{x}}(P)$ is an increasing function of b_V.

Proof. Under Assumption 2.1, S_n^j and R_n^j, being the positive-definiteness-preserving bilinear transformations of Σ_{n-1}^j and S_n^j (from Eq. (2.21)), respectively, are always bounded above and below by positive-definite matrices. Hence $K_n^j C = S_n^j C^T (R_n^j)^{-1} C$ is also positive definite and bounded above and below. Now, if C has full column rank and $\mathbf{V}_n$ is bounded by b_V with probability P, proceeding in exactly same way as in the proof of Theorem 2.1 (with $\tilde{r}_n^j$ replaced by $\tilde{\mathbf{f}}_n^j$), $\tilde{\mathbf{x}}_n$ is also bounded by some bound $b_{\tilde{x}}(P)$ with probability P if C has full column rank and $\mathbf{V}_n$ is bounded with probability P. $\square$

2.3.2 Gaussian mixture model unscented Kalman filter

Despite being a state-of-the-art nonlinear filter, EKF is plagued with a truncation problem due to its first-order approximation, which might lead to large errors in the posterior. An alternate method is to use the Unscented Transform (UT) with a set of *sigma points* such that their empirical mean and covariances are the same as that of the initial Gaussian PDF [23]. We use the unscented Kalman filter to propagate each Gaussian mode of the

GMM for the same system (2.20). Define

$$\begin{aligned}
\hat{\mathbf{x}}_0^{ja} &= \begin{bmatrix} (\hat{\mathbf{x}}_0^j)^T & 0 & 0 \end{bmatrix}^T, \\
P_0^{ja} &= \begin{bmatrix} \Sigma_0^j & 0 & 0 \\ 0 & K_{\mathbf{w}} & 0 \\ 0 & 0 & K_{\mathbf{v}} \end{bmatrix}_{L\times L}, \quad L = 2d+q,
\end{aligned} \tag{2.24}$$

for $j = 1, \ldots, M$ and $n = 1, \ldots \infty$, and sigma points

$$\chi_{n-1}^{ja} = \begin{bmatrix} \hat{\mathbf{x}}_{n-1}^{ja} & \hat{\mathbf{x}}_{n-1}^{ja} \pm \sqrt{(L+\lambda)P_{N-1}^{ja}} \end{bmatrix}, \tag{2.25}$$

for $j = 1, \ldots, M$ and $n = 1, \ldots \infty$, where $\lambda = \alpha^2(L+\kappa) - L$ is a scaling parameter. The superscript a is used to signify the state appended with process and measurement noise. α determines the spread of the sigma points around the mean and is usually set to a small positive value, e.g., 0.001 [59]. κ is a secondary parameter usually set to 0. β is used to incorporate prior knowledge about the distribution; $\beta = 2$ is optimal for a Gaussian distribution [59]. Define

$$\begin{aligned}
W_0^{(m)} &= \frac{\lambda}{L+\lambda}, \\
W_0^{(c)} &= \frac{\lambda}{L+\lambda} + 1 - \alpha^2 + \beta, \\
W_i^{(m)} &= W_i^{(c)} = \frac{1}{2(L+\lambda)}.
\end{aligned} \tag{2.26}$$

χ^{ja}_{n-1} has three parts: χ^{jx}_{n-1} is the first d rows; χ^{jw}_{n-1} is the next d rows; and χ^{jv}_{n-1} is the last q rows. w^j_0, $\hat{\mathbf{x}}^j_0$ and Σ^j_0 are defined by EM and BIC. The GMM-UKF is as follows, for $j = 1, \ldots, M$ and $n = 1, \ldots, \infty$:

$$\begin{aligned}
\chi^{jx}_{n|n-1} &= f(\chi^{jx}_{n-1}) + \chi^{jw}_{n-1} \text{ calculated columnwise,}\\
\bar{\mathbf{x}}^j_n &= \sum_{i-0}^{2L} W^{(m)}_i \chi^{jx}_{i,n|n-1},\\
&\quad \text{where } i \text{ represents the } i^{\text{th}} \text{ column,}\\
S^j_n &= \sum_{i-0}^{2L} W^{(c)}_i \left[\chi^{jx}_{i,n|n-1} - \bar{\mathbf{x}}^j_n\right]\left[\chi^{jx}_{i,n|n-1} - \bar{\mathbf{x}}^j_n\right]^T,\\
\mathcal{Y}^j_{n|n-1} &= C\chi^{jx}_{n|n-1} + \chi^{jv}_{n-1},\\
\bar{\mathbf{y}}^j_n &= \sum_{i-0}^{2L} W^{(m)}_i \mathcal{Y}^j_{i,n|n-1},\\
P^j_{\tilde{y}_n\tilde{y}_n} &= \sum_{i-0}^{2L} W^{(c)}_i \left[\mathcal{Y}^j_{i,n|n-1} - \bar{\mathbf{y}}^j_n\right]\left[\mathcal{Y}^j_{i,n|n-1} - \bar{\mathbf{y}}^j_n\right]^T,\\
P^j_{x_n y_n} &= \sum_{i-0}^{2L} W^{(c)}_i \left[\chi^{jx}_{i,n|n-1} - \bar{\mathbf{x}}^j_n\right]\left[\mathcal{Y}^j_{i,n|n-1} - \bar{\mathbf{y}}^j_n\right]^T,\\
K^j_n &= P^j_{x_n y_n}(P^j_{\tilde{y}_n\tilde{y}_n})^{-1},\\
\hat{\mathbf{x}}^j_n &= \bar{\mathbf{x}}^j_n + K^j_n(\mathbf{y}_n - \bar{\mathbf{y}}^j_n),\\
\Sigma^j_n &= S^j_n - K^j_n P^j_{\tilde{y}_n\tilde{y}_n}(K^j_n)^T,\\
w^j_n &= \frac{w^j_{n-1}\mathcal{N}(\mathbf{y}_n; \bar{\mathbf{y}}^j_n, P^j_{\tilde{y}_n\tilde{y}_n})}{\sum_{m=1}^M w^m_{n-1}\mathcal{N}(\mathbf{y}_n; \bar{\mathbf{y}}^m_n, P^m_{\tilde{y}_n\tilde{y}_n})} \qquad (2.27)
\end{aligned}$$

The estimate is again given by $\hat{\mathbf{x}}_n = \sum_{j=1}^{M} w^j_n \hat{\mathbf{x}}^j_n$.

Since UKF uses heuristics for choosing the sigma points, the following assumption is required for the error bound analysis.

Assumption 2.2. $\delta_j I \leq P^j_{x_n y_n}, P^j_{\tilde{y}_n \tilde{y}_n} < \gamma_j I$, where $\delta_j, \gamma_j \geq 0$, for all $n \geq 0$ and $j = 1, \cdots, M$

Remark 2.7. Assumption 2.2 essentially requires that the sample covariances of state and observation sigma points for each Gaussian mode remain bounded.

Theorem 2.4. Consider the system (2.20) with possibly non-Gaussian noise $\mathbf{W}_n$. If Assumption 2.2 is satisfied, C has full column rank, and $\mathbf{V}_n$ is bounded by b_V with probability P, then $\|\tilde{\mathbf{x}}_n\| = \|\mathbf{x}_n - \hat{\mathbf{x}}_n\|$ is bounded by $b_{\tilde{x}}(P)$ with probability P, where $b_{\tilde{x}}(P)$ is an increasing function of b_V and $\hat{\mathbf{x}}$ is obtained from the GMM-UKF algorithm (2.27).

Proof. Define

$$\begin{aligned}
\tilde{\mathbf{f}}^j_n &= \sum_{i=0}^{2L} W_i^{(m)} \left[f(\mathbf{x}_{n_1}) - f(\chi^{jx}_{i,n|n-1}) + \mathbf{W}_n - \chi^{jw}_{n-1} \right], \\
\tilde{\mathbf{V}}^j_n &= \sum_{i=0}^{2L} W_i^{(m)} \left[V_n - \chi^{jv}_{n-1} \right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
\tilde{\mathbf{x}}_n &= \sum_{j=1}^{M} w^j_n \tilde{\mathbf{x}}^j_n \\
&= \frac{1}{\beta_n} \sum_{j=1}^{M} w^j_{n-1} \frac{1}{\sqrt{\det(2\pi P^j_{\tilde{y}_n \tilde{y}_n})}} \\
&\quad \times e^{-\frac{1}{2}(C\tilde{\mathbf{f}}^j_n + \tilde{\mathbf{V}}^j_n)^T (P^j_{\tilde{y}_n \tilde{y}_n})^{-1} (C\tilde{\mathbf{f}}^j_n + \tilde{\mathbf{V}}^j_n)} \\
&\quad \times \left((I - K^j_n C)\tilde{\mathbf{f}}^j_n - K^j_n \tilde{\mathbf{V}}^j_n \right),
\end{aligned} \tag{2.28}$$

where $\beta_n = \sum_{m=1}^{M} w_{n-1}^m \mathcal{N}(\mathbf{y}_n; C\overline{\mathbf{x}}_n^m, CS_n^m C^T + K_V)$. Since $P_{\tilde{y}_n\tilde{y}_n}^j$ is bounded, C has full column rank, and $\mathbf{V}_n$ (hence $\tilde{\mathbf{V}}_n^j$ as well) is bounded with probability P, proceeding in the same way as the proof of Theorem 2.3, $\tilde{\mathbf{x}}_n$ will also be bounded for GMM-UKF by some bound $b_{\tilde{x}}(P)$. □

Remark 2.8. Both GMM-EKF and GMM-UKF are suggested as an alternative to particle filtering for nonlinear estimation. However, since both EKF and UKF are approximate methods for propagation of each Gaussian mode and accumulate some error over time, particle filters, when applied with suitable resampling method and well-chosen heuristics, may perform better. But particle filters are computationally intensive and there is no stochastic bound for estimation error. The GMM-EKF and GMM-UKF provide a computationally cheap way to perform nonlinear filtering with non-Gaussian noise with some guarantee on error boundedness, which is utilized in the analysis of the output feedback control.

2.3.3 Dynamic output feedback control

To control the nonlinear system (2.20), suppose we design a state-feedback controller $u(\mathbf{x})$, and drive it with the estimated state derived by the filter. The closed-loop system takes the form,

$$\begin{aligned} \mathbf{x}_{n+1} &= f(\mathbf{x}_n) + u(\hat{\mathbf{x}}_n) + \mathbf{W}_n \\ &= F(\mathbf{x}_n) + u(\hat{\mathbf{x}}_n) - u(\mathbf{x}_n) + \mathbf{W}_n \qquad (2.29) \\ \mathbf{y}_n &= C\mathbf{x}_n + \mathbf{V}_n,\ n \geq 1, \end{aligned}$$

where $\mathbf{W}_n$ is non-Gaussian additive noise and $F(\mathbf{x}_n) = f(\mathbf{x}_n) + u(\mathbf{x}_n)$.

To obtain *practical stability* [57], i.e., the ultimate-boundedness of the state under a suitable feedback law, we need some additional assumptions as follows.

Assumption 2.3. The nominal system $\mathbf{x}_{n+1} = F(\mathbf{x}_n)$ is uniformly asymptotically stable on an open ball $\mathcal{B}_r$ of radius r centered at 0 and $\exists$ a C^1 (i.e., differentiable) Lyapunov function [57] $V : \mathbb{Z}_+ \times \mathcal{B}_r \to \mathbb{R}$ that satisfies

$$\alpha_1(\|\mathbf{x}_n\|) \leq V(n, \mathbf{x}_n) \leq \alpha_2(\|\mathbf{x}_n\|), \tag{2.30}$$

$$\Delta V_N(n, \mathbf{x}_n) \leq -\alpha_3(\|\mathbf{x}_n\|), \tag{2.31}$$

where $\alpha_i, i = 1, 2, 3$ are class $\mathcal{K}$ functions and $\Delta V_N(n, \mathbf{x}_n) = V(n+1, \mathbf{x}_{n+1}) - V(n, \mathbf{x}_n)$ under the nominal system [60]

Remark 2.9. Under Assumption 2.3, the chosen feedback law stabilizes the original noise-free system with proper state feedback, which is valid, since choosing a stabilizing feedback law is essential in any output feedback.

Assumption 2.4. There exists, $p \in \mathbb{R}_+$ and $M > 0$ such that $\left\|\dfrac{\partial u}{\partial \mathbf{x}}\right\| \leq M \left\|\mathbf{x}\right\|^{p-1}$ and, with the application of the mean value inequality, we get $\|u(\hat{\mathbf{x}}_n) - u(\mathbf{x}_n)\| \leq M b_{\tilde{x}}(P)$ from Assumption 2.1

Assumption 2.5. $\|\mathbf{W}_n\| < b_W(P)$ with probability P.

Assumption 2.5 characterizes the process noise and $b_W(P)$ can be readily obtained from the PDF or cumulative distribution of the process noise (only the latter if $\mathbf{W}_n$ is not absolutely continuous). With these assumptions, Theorem 2.5 presents the state-boundedness result for a nonlinear non-Gaussian system with GMM-EKF or GMM-UKF output feedback.

Theorem 2.5. Consider the nonlinear feedback control system (2.29) with non-Gaussian process noise and GMM-EKF (resp. GMM-UKF) state estimate $\hat{\mathbf{x}}_n$. If Assumptions 2.1 and 2.3-2.5 (resp. 2.2 and 2.3-2.5) are satisfied, then $\|\mathbf{x}_n\|$ is bounded with an ultimate bound b_x with probability P, where b_x is a function of M, $b_{\tilde{x}}(P)$, and $b_W(P)$ as defined above.

Proof. We use $V(n, \mathbf{x}_n)$ from Assumption 2.2 for the perturbed system. Thus,

$$\Delta V_\pi(n, \mathbf{x}_n) = \Delta V_N(n, \mathbf{x}_n) + \delta V_\pi(n, \mathbf{x}_n),$$

where $\delta V_\pi(n, \mathbf{x}_n) = V(n+1, F(\mathbf{x}_n) + u(\hat{\mathbf{x}}_n) - u(\mathbf{x}_n) + \mathbf{W}_n) - V(n+1, F(\mathbf{x}_n))$, and subscript π stands for the perturbed system. Since $V \in C^1$, $\exists\, l > 0$ such that

$$\begin{aligned} |\delta V_\pi(n, \mathbf{x}_n)| &\leq l \, \|u(\hat{\mathbf{x}}_n) - u(\mathbf{x}_n) + \mathbf{W}_n\| \\ &\leq l(\|u(\hat{\mathbf{x}}_n) - u(\mathbf{x}_n)\| + \|\mathbf{W}_n\|). \end{aligned}$$

Now, from Assumptions 2.1 and 2.3-2.5 (resp. 2.2 and 2.3-2.5 for GMM-UKF),

$$\begin{aligned}
\Delta V_\pi(n, \mathbf{x}_n) &= \Delta V_N(n, \mathbf{x}_n) + \delta V_\pi(n, \mathbf{x}_n) \\
&\leq -\alpha_3(\|\mathbf{x}_n\|) + lM(b_{\tilde{x}}(P) + b_W(P)), \text{ wp } P \\
&= -(1-\theta)\alpha_3(\|\mathbf{x}_n\|) - \theta\alpha_3(\|\mathbf{x}_n\|) \\
&\quad + lM(b_{\tilde{x}}(P) + b_W(P)), \text{wp } P \text{ and } \theta \in (0,1) \\
&\leq -(1-\theta)\alpha_3(\|\mathbf{x}_n\|), \\
&\quad \forall \ \|\mathbf{x}_n\| \geq \alpha_3^{-1}\left(\frac{lM(b_{\tilde{x}}(P) + b_W(P))}{\theta}\right),
\end{aligned}$$

wp P. Hence $\|\mathbf{x}_n\|$ for the system (2.29) is uniformly ultimately bounded by probability P, with ultimate bound

$$b_x(P) = \alpha_1^{-1} \circ \alpha_2 \circ \alpha_3^{-1}\left(\frac{lM(b_{\tilde{x}}(P) + b_W(P))}{\theta}\right), \text{ wp } P.$$

□

2.4 Numerical simulations

The performance of GMM-KF and its nonlinear variants is illustrated here numerically with various discrete-time stochastic dynamical systems. the performance of the corresponding output feedback stabilization is also demonstrated.

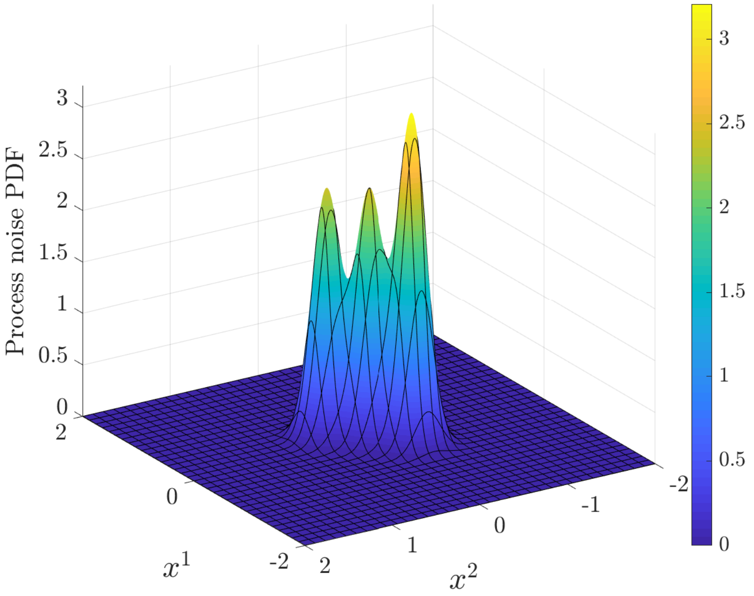


Figure 2.2: The PDF of the process-noise $\mathbf{W}_n$

2.4.1 GMM-KF Filtering and Output Feedback for Linear Systems

The effectiveness of the GMM-KF is shown for the system

$$\mathbf{x}_{n+1} = \begin{bmatrix} 1.0 & 0.9 \\ -0.5 & 1.2 \end{bmatrix} \mathbf{x}_n + \begin{bmatrix} 1.0 & 0 \\ 0 & 1.0 \end{bmatrix} \mathbf{u}_n + \mathbf{W}_n, \tag{2.32}$$

where the process noise $\mathbf{W}_n$ are zero mean and i.i.d. The noise is realized by independently sampling from a Gaussian mixture distribution with three modes centered at $\mu_1 = [-0.3, -0.3]^T$, $\mu_2 = [0, 0]^T$ and $\mu_3 = [0.3, 0.3]^T$. The covariance matrices of these modes are $\Sigma_i = 0.02I$, with $i = 1, 2, 3$ and I is the 2×2 identity matrix. The weights are given by $w = [0.4, 0.3, 0.3]$. The PDF of $\mathbf{W}_n$ is presented in Fig. 2.2. The system (2.32) is unstable and $\|\mathbf{x}_n\|$ increases exponentially with $\mathbf{u}_n = 0$ with n. The linear observation

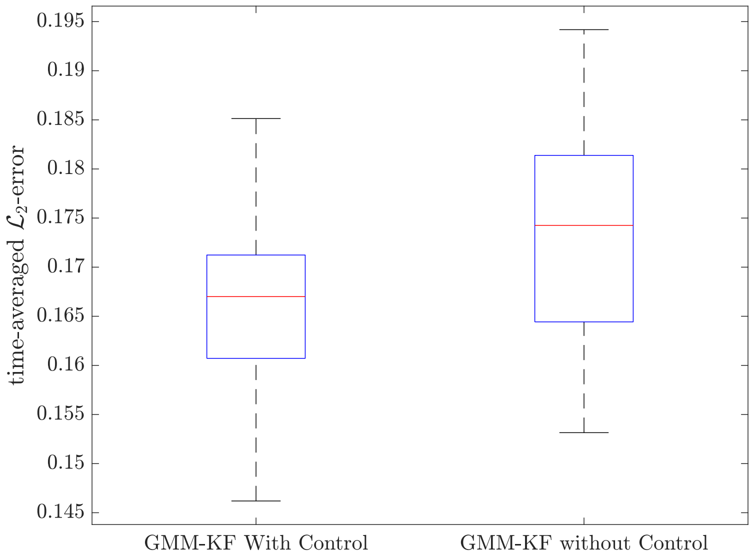


Figure 2.3: $\mathcal{L}_2$ error of GMM-KF estimate with and without feedback

model is

$$\mathbf{y}_n = C\mathbf{x}_n + \mathbf{V}_n,$$

where C is an identity matrix and $\mathbf{V}_n$ is a zero mean Gaussian measurement noise with covariance matrix $\text{diag}(0.1, 0.1)$. For feedback control, we use $\mathbf{u}_n = -\begin{bmatrix} 0.5 & 0 \\ 0 & 0.7 \end{bmatrix} \hat{\mathbf{x}}_n$, so that the eigenvalues of $A - BK$ lie within the unit circle. Time-averaged $\mathcal{L}_2$ error of the GMM-KF filter from 50 independent Monte-Carlo run is presented in Fig. 2.3. The observer-based control effectively stabilizes the system as shown in Fig. 2.4(a). Since $|\lambda_{max}(A - BK)| < 1$, $\mathbb{E}\,\|\mathbf{x}_{n+1}\|$ remains bounded as $n \to \infty$ if $\mathbb{E}\,\|\mathbf{x}_n - \hat{\mathbf{x}}_n\|$ is bounded, which is indeed true by Theorem 2 with probability P depending on the measurement noise.

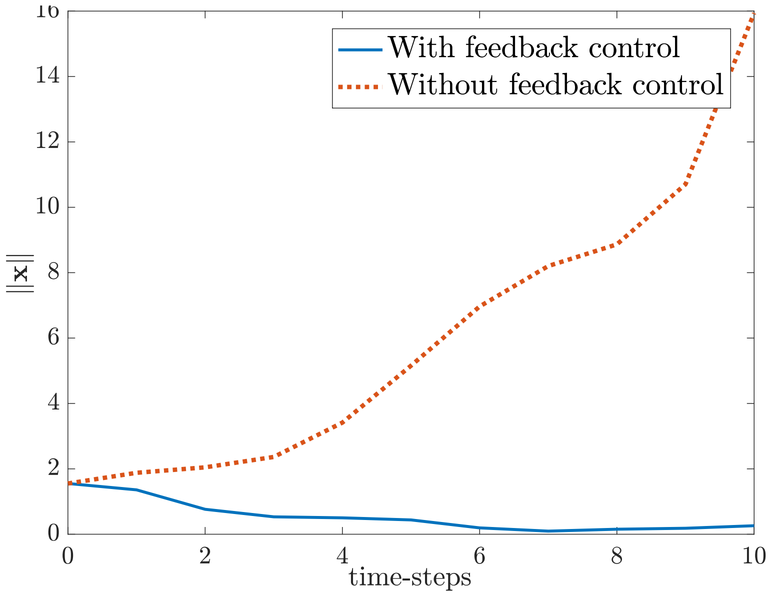


Figure 2.4: Stability of GMM-KF output feedback for linear system (2.32)

2.4.2 GMM-EKF/UKF Filtering and Output Feedback for Nonlinear Systems

The GMM-EKF and GMM-UKF have been tested on three different nonlinear systems. First, we use a 2D nonlinear chaotic map attributed to the discrete-time version of the Duffing Oscillator [61]. The map is

$$\begin{aligned} x^1_{n+1} &= x^2_n + w^1_n \\ x^2_{n+1} &= -a_2 x^1_n + a_1 x^2_n - (x^2_n)^3 + w^2_n \\ \mathbf{y}_n &= C\mathbf{x}_n + \mathbf{V}_n. \end{aligned} \tag{2.33}$$

The map (2.33) depends on the two constants a_1 and a_2, usually set to $a_1 = 2.75$ and $a_2 = 0.2$ to produce chaotic behaviour [61].

Then we use a stable nonlinear power map given by

$$\begin{aligned} x_{n+1}^1 &= x_n^2 + w_n^1 \\ x_{n+1}^2 &= 0.7x_n^1 x_n^2 + w_n^2 \\ \mathbf{y}_n &= C\mathbf{x}_n + \mathbf{V}_n, \end{aligned} \tag{2.34}$$

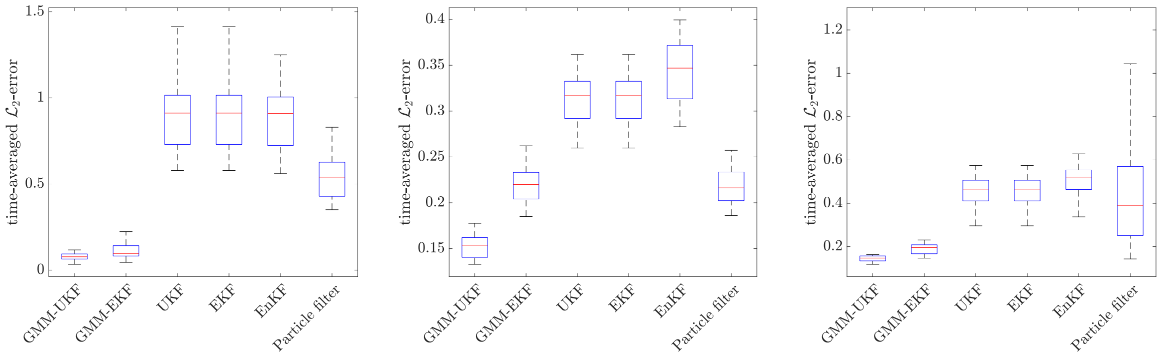
where the process measurement noise and the observation model stays the same.

Lastly, we use a sinusoidal map, i.e.,

$$\begin{aligned} x_{n+1}^1 &= x_n^1 \sin(x_n^1) + x_n^2 \cos(x_n^2) + w_n^1 \\ x_{n+1}^2 &= x_n^1 \sin(x_n^2) + x_n^2 \cos(x_n^1) + w_n^2 \\ \mathbf{y}_n &= C\mathbf{x}_n + \mathbf{V}_n. \end{aligned} \tag{2.35}$$

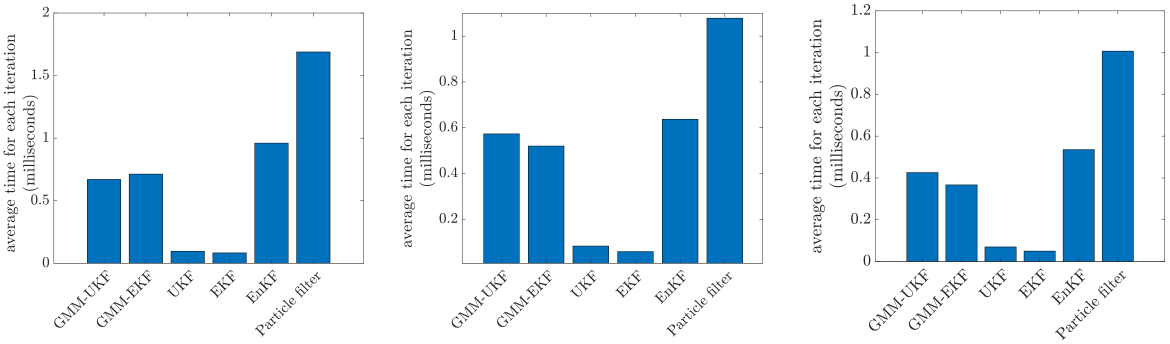
In (2.33)-(2.35),the process noise $\mathbf{W}_n = [w_n^1 \ w_n^2]^T$ are i.i.d. random vectors drawn from the same non-Gaussian distribution as in the system (2.32). The superscripts here denote the components of the vector. The observation model is the same one used for linear system (2.32) and C is an identity matrix.

GMM-EKF and GMM-UKF are compared against the EKF, UKF, EnKF, and a particle filter (PF) with systematic resampling [62] method and 1000 particles. For the comparison, we applied all six filters from the same initial conditions and initial uncertainties as used. In the GMM-UKF framework, the parameters values are $\alpha = 0.001$ and $\beta = 2$ as in [59]. The result is given in Fig. 2.5 in the form of time-averaged $\mathcal{L}_2$ error for 50 independent Monte-Carlo runs. Average run-time for each iteration of these five



((a)) ((b)) ((c))

Figure 2.5: Time-averaged $\mathcal{L}_2$ error for 50 independent Monte-Carlo runs: (a) System (2.33), (b) System (2.34), (c) System (2.35)



((a)) ((b)) ((c))

Figure 2.6: Average run-time per iteration for 50 independent Monte-Carlo runs: (a) System (2.33), (b) System (2.34), (c) System (2.35)

filters is demonstrated in Fig. 2.6. GMM-based filters performs significantly better than EKF, UKF, and EnKF in terms of $\mathcal{L}_2$ error, and similar or slightly better than the more computationally intensive particle filter. GMM-based filter also consumes less CPU-time as depicted in Fig. 2.6 than a particle filter. Fig. 2.7 and 2.8 show the actual estimated trajectories for systems (2.33) and (2.35).

The effectiveness of GMM-EKF in feedback control is illustrated in Fig. 2.9 with the

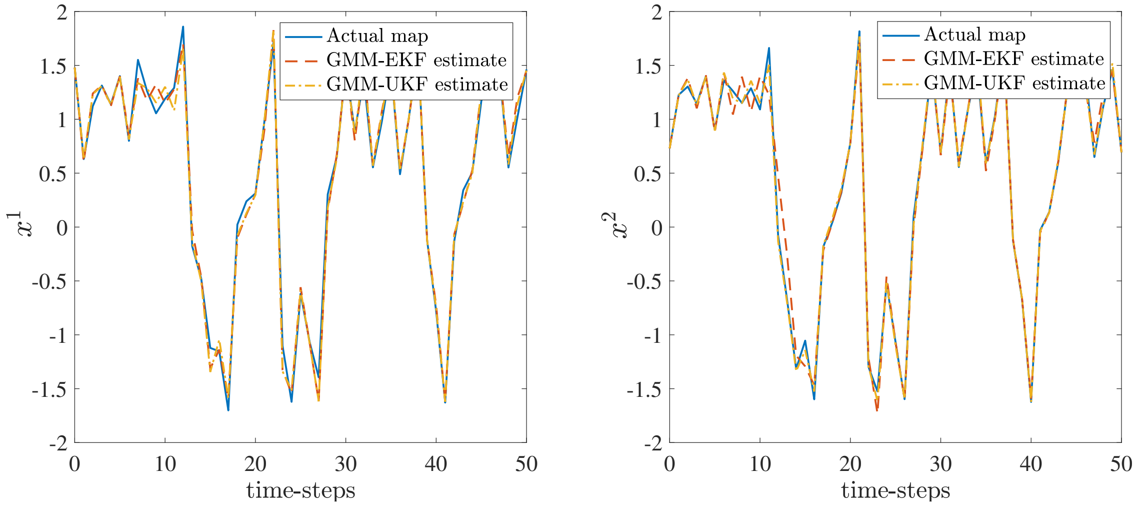


Figure 2.7: Actual and estimated trajectories for noisy Duffing system (2.33)

system

$$\begin{aligned} x_{n+1}^1 &= x_n^2 + w_n^1 \\ x_{n+1}^2 &= -1.1 x_n^1 x_n^2 + w_n^2 + u_n \\ \mathbf{y}_n &= C\mathbf{x}_n + \mathbf{V}_n, \end{aligned} \tag{2.36}$$

where $\mathbf{W}_n$ is a non-Gaussian i.i.d. process noise with the distribution as described in the system (2.32) along with the same observation model. The dynamic output feedback law $u_n = 0.3\hat{x}_n^1\hat{x}_n^2$ effectively stabilizes the system (2.36) as shown in Fig. 2.9. (The feedback using GMM-UKF has the same trend, though is not shown here.)

The $\mathcal{L}_2$ errors shown in Figures 2.3 and 2.5 do not blow up, as predicted by the probabilistic bounds derived in Theorems 2.1, 2.3, and 2.4. As the analytical bound $b_{\tilde{x}}(P)$ is qualitative and probability-dependent, it is not illustrative to quantitatively compare

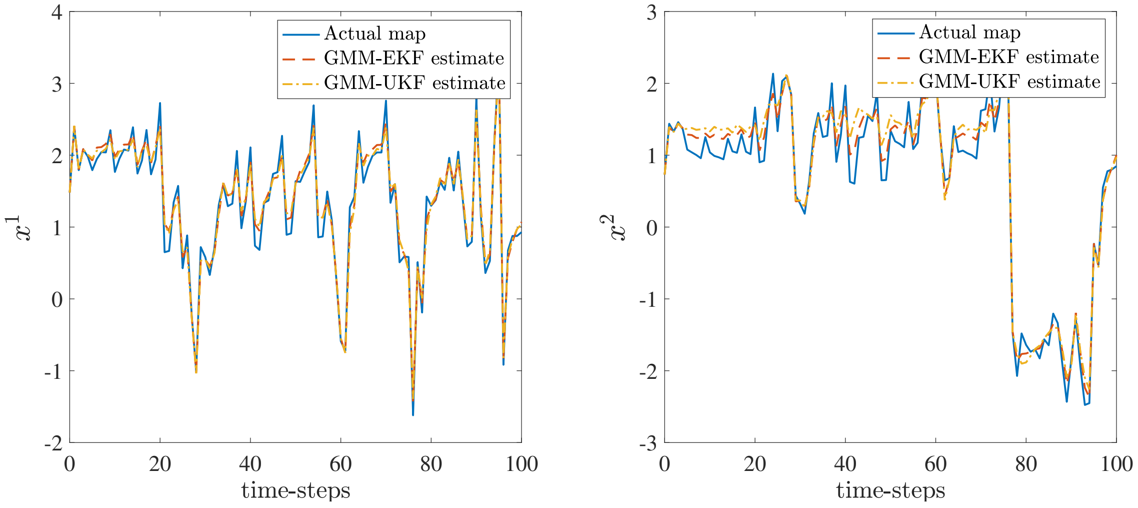


Figure 2.8: Actual and estimated trajectories for system (2.35)

this bound with the numerical $\mathcal{L}_2$ error.

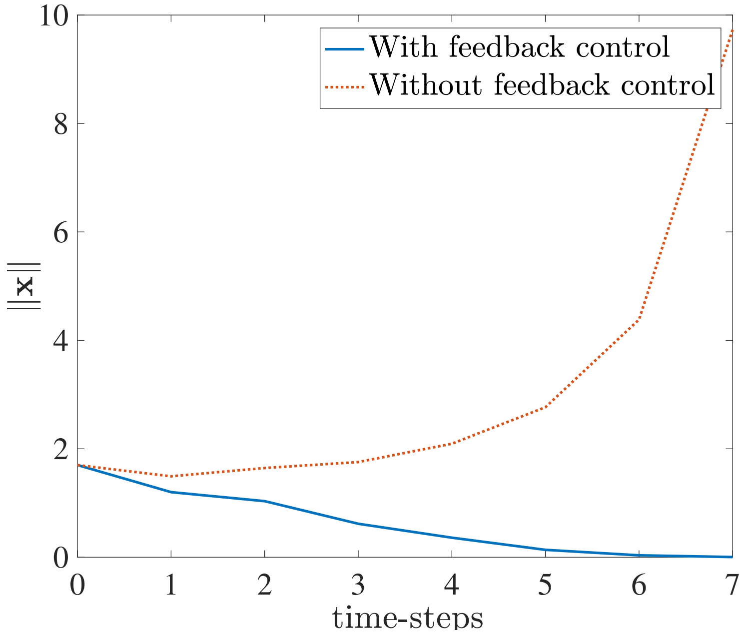


Figure 2.9: GMM-EKF observer-based feedback effectively stabilizes the nonlinear, non-Gaussian system (2.36)

CHAPTER 3

Density transport in dynamical systems: Perron Frobenius operator and its approximation

In the last chapter we have used the Gaussian mixture model in conjunction with Kalman filter variants to propagate and update the density of states. In this chapter, however, we transfered our attention to a different tool, namely the operator theoretic approach, for the same task. The two main candidates for this approach are the Perron-Frobenius and Koopman operators. While the Koopman operator is useful in studying observables, its dual, the Perron-Frobenius (PF) operator, acts on the space of densities. Hence the latter is important when dealing with uncertainties in the system, especially when the likelihood of the state is given in the form of a probability density function under a suitable absolutely continuous probability measure. PF operator, being able to transport density in a dynamical system, is important in Bayesian estimation problems. The approximated

PF operator can be used to transport density with lesser computational effort than solving a partial differential equation. Hence an accurate and efficient approximation technique of the PF operator has become necessary.

In this chapter, we combine Ulam's method and constrained EDMD [32] to remove the problem of long time steps in Ulam's method [28] and incorporate the accuracy of EDMD. The basis functions are chosen as the characteristic functions of the grids over the state-space. It is shown analytically that the PF operator projected onto this basis constructs a Markov chain, and its eigenfunctions can be approximated from the eigenvectors of the resultant stochastic matrix. The results are derived for both deterministic and time-homogeneous stochastic systems.

The proposed method is used for motion update for nonlinear estimation in a Bayesian framework. An application of the estimation using approximate PF operator for output feedback control of a pitching airfoil that maximizes lift using only the lift-coefficient measurements from the embedded pressure sensors is also presented.

3.1 Perron-Frobenius operator

Historically, transfer operators like the Perron-Frobenius (PF) operator are used in *ergodic theory* to study measure-theoretic characterization. The PF operator is described below in the context of both deterministic and stochastic systems.

3.1.1 Deterministic systems

Let $\mathbb{X} \subset \mathbb{R}^d$ be a compact manifold and $f : \mathbb{X} \to \mathbb{X}$ be a smooth time-invariant vector field. Consider the autonomous time-invariant ODE

$$\dot{x} = f(x). \tag{3.1}$$

Let $\Phi_f : \mathbb{R} \times \mathbb{X} \to \mathbb{X}$ be the flow map of the ODE (3.1), i.e., $\Phi_f(t, x_0)$ is a solution of the ODE (3.1) with the initial condition $x(0) = x_0$.

Definition 3.1. A semigroup of operator $\mathcal{P}^\tau : \tau \geq 0$ is said to be the *Perron Frobenius* (PF) operator if $\mathcal{P}^\tau : L^1(\mathbb{X}) \to L^1(\mathbb{X})$ is defined by [28]

$$\mathcal{P}^\tau \rho(\cdot) = \rho \circ \Phi_f(-\tau, \cdot)|\det(D_x \Phi_f(-\tau, \cdot))|, \tag{3.2}$$

where D_x denotes the Jacobian with respect to the space variable x.

If $\rho(\cdot)$ is a probability density function (PDF) with respect to an absolutely continuous probability measure ν, then $\mathcal{P}^\tau \rho$ is another PDF with respect to the absolutely continuous probability measure $\nu \circ \Phi_f(-\tau, \cdot)$. Specifically,

$$\int\limits_B \mathcal{P}^\tau \rho d\nu = \int\limits_{\Phi_f(-\tau, B)} \rho d\nu, \tag{3.3}$$

for any ν-measurable set B [63]. The PF operator translates a probability density function with time according to the flow of the dynamics. It can be shown that the generator of this semigroup $\mathcal{P}^\tau \rho$ is given by $-\nabla \cdot (\rho f)$ [28].

Now suppose we define a time-varying PDF $\tilde{\rho}(t,x) \triangleq \mathcal{P}^t \rho(x)$. Then it can be shown that $\tilde{\rho}(t,x)$ satisfies the PDE [28]

$$\begin{aligned} \frac{\partial \tilde{\rho}}{\partial t} &= -\nabla \cdot (\tilde{\rho} f) \qquad (3.4) \\ \tilde{\rho}(0,x) &= \rho(x). \qquad (3.5) \end{aligned}$$

3.1.2 Stochastic systems

Let $\mathbb{X}$ and f be defined as before. Let $g : \mathbb{X} \to \mathbb{X}$ be another smooth time-invariant vector field. Let $(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mathbb{P})$ be a probability triple with an absolutely continuous probability measure $\mathbb{P}$ where $\mathcal{B}(\mathbb{X})$ is the Borel sigma field of $\mathbb{X}$. Consider the following time-homogeneous Itô stochastic differential equation

$$dx_t = f(x_t)dt + g(x_t)dw_t,\ t > 0, \qquad (3.6)$$

where $x_t \in \mathbb{X}$ is a random process and w_t, $t > 0$, is a standard Wiener process. The stochastic dynamics (3.6) arise when there is process noise in the system defined by the ODE (3.1). The process noise enters the system as the diffusion term $g(x_t)$ in the Itô SDE. Here a flow map, being another random process, is difficult to define, but nonetheless the probability density may still be translated with a linear operator defined in terms of the transition density function.

Definition 3.2. The *transition density function* $p_\tau : \mathbb{X} \times \mathbb{X} \to \mathbb{R}_+$ is defined as [63]

$$\mathbb{P}[x_{t+\tau} \in \mathbb{B} | x_t = x] = \int\limits_{\mathbb{B}} p_\tau(x, m) dm, \, \forall \mathbb{B} \in \mathcal{B}(\mathbb{X}). \tag{3.7}$$

The transition density function is the infinite-dimensional counterpart of the transition matrix entries for a Markov chain. Now we can define the stochastic PF operator.

Definition 3.3. Let $\rho \in L^1(\mathbb{X})$ be a probability density function. The *Perron-Frobenius* semigroup of operator $\mathcal{P}^\tau$, $\tau > 0$, is defined as $\mathcal{P}^\tau : L^1(\mathbb{X}) \to L^1(\mathbb{X})$ [63], such that

$$\mathcal{P}^\tau \rho(\cdot) = \int\limits_{\mathbb{X}} p_\tau(m, \cdot) \rho(m) dm. \tag{3.8}$$

The PF operator $\mathcal{P}^\tau$ as defined here is also a linear operator and $\mathcal{P}^\tau \rho$ has an infinitesimal generator $-\nabla \cdot (\rho f) + \dfrac{1}{2} \nabla^2 (g^2 \rho)$ [63].

3.2 Numerical estimation of the Perron-Frobenius operator

To obtain an explicit numerical estimate of the infinite-dimensional PF operator, we need to project it onto a suitable finite-dimensional basis of functions. As the PF operators for a specific dynamics also form a semigroup, we need to fix a time step size to extract one representative of the semigroup. The latter amounts to the discretization of the dynamics, generally performed using Galerkin methods employed by converting the operator equations (3.2), (3.8) into a weakly approximated form, usually a Markov chain.

3.2.1 Weak approximation

The weak approximation of the operator equation (3.2) or (3.8) is done by projecting it on a suitable set of basis functions. Here, the basis functions are chosen to be a family of characteristic functions of a partition of the state space, scaled accordingly by their Lebesgue measure. Let $D = \{B_i : i = 1, \ldots, M\}$ be a partition of the state space $\mathbb{X}$, usually a set of fine grids. The basis function ψ_i is taken to be $\psi_i \triangleq \chi_{B_i}$, where

$$\chi_{B_i}(x) = \begin{cases} 1, & \text{if } x \in B_i \\ 0, & \text{otherwise.} \end{cases}$$

The basis functions $\{\psi_1, \ldots, \psi_M\}$ are orthogonal. Now define a projection $\pi_M : L^1(\mathbb{X}) \to \text{sp}\{\psi_i, \ldots \psi_M\}$ to project ρ onto the span of these basis functions by

$$\pi_M \rho = \sum_{i=1}^{M} \left(\frac{1}{m(B_i)} \int\limits_{B_i} \rho \, dm \right) \psi_i = \sum_{i=1}^{M} \frac{p^i}{m(B_i)} \psi_i, \tag{3.9}$$

where $m(\cdot)$ is the Lebesgue measure on $\mathbb{R}^d$ and $p^i \triangleq \int\limits_{B_i} \rho \, dm$ are the weights of basis function ψ_i. Since the choice of basis is done before approximation, the projected density ρ is usually expressed as a vector $\mathbf{p} = (p^1, \ldots, p^n)$. This projection in turns restricts the infinite-dimensional operator $\mathcal{P}^\tau$ to a stochastic matrix P_τ given by

$$
\begin{aligned}
P_{\tau,ij} &= \mathbb{P}[\Phi_f(t,x) \in B_j | x \in B_i] \\
&= \frac{\mathbb{P}[\Phi_f(t,x) \in B_j, x \in B_i]}{\mathbb{P}[x \in B_i]} \\
&= \frac{m(B_i \cap \Phi_f(-\tau, B_j))}{m(B_i)},
\end{aligned}
\tag{3.10}
$$

for deterministic system (3.2) where $\Phi_f(-\tau, B_j) \triangleq \{\Phi_f(-\tau, x) : x \in B_j\}$. In the last equality, the measure is changed to a standard Lebesgue measure, assuming we sample from a uniform distribution at $t = 0$ when computing $P_{\tau,ij}$.

For stochastic system (3.8), since the flow map is a random process, the P_τ matrix is given by

$$
\begin{aligned}
P_{\tau,ij} &= \mathbb{P}[x_{t+\tau} \in B_j | x_t \in B_i] \\
&= \mathbb{P}[x_\tau \in B_j | x_0 \in B_i] \text{ , time homogeneity of (3.6)} \\
&= \frac{\mathbb{P}[x_\tau \in B_j, x_0 \in B_i]}{\mathbb{P}[x_0 \in B_i]} \\
&= \frac{k\mathbb{P}[x_\tau \in B_j, x_0 \in B_i]}{m(B_i)},
\end{aligned}
\tag{3.11}
$$

where $k > 0$ is the normalization factor such that the probability $\mathbb{P}[x_0 \in B_i] = \frac{m(B_i)}{k}$ corresponding to a uniform initial distribution.

With this formulation, the approximation of the PF operator is equivalent to the approximation of P_τ. The weak approximation, in effect, turns the PDF ρ into a Probability Mass Function (PMF) on each grid B_i, and the operator (3.2) becomes a Markov state transition equation [28]

$$\mathbf{p}_\tau = \mathbf{p}_0 P_\tau, \tag{3.12}$$

where $\mathbf{p}_\tau$ and $\mathbf{p}_0$ are the projection of $\mathcal{P}^\tau\rho$ and ρ, respectively. Note that $\mathbf{p}_0$ and $\mathbf{p}_\tau$ are valid PMF since P_τ is a stochastic matrix. To see this, consider the projection,

$$\begin{aligned} \pi_M \mathcal{P}^\tau(\pi_M \rho) &= \pi_M \mathcal{P}^\tau \left(\sum_{i=1}^{M} \frac{p^i}{m(B_i)} \psi_i \right) \\ &= \sum_{i=1}^{M} \frac{p^i}{m(B_i)} \pi_M \left(\mathcal{P}^\tau \psi_i \right). \end{aligned} \tag{3.13}$$

Now, since $\{\psi_1, \ldots, \psi_M\}$ are orthogonal,

$$\pi_M \left(\mathcal{P}^\tau \psi_i \right) = \sum_{j=1}^{M} \frac{w_{ij}}{m(B_j)} \psi_j, \tag{3.14}$$

where the coefficients w_{ij} are given by

$$\begin{aligned} w_{ij} &= \int\limits_{\mathbb{X}} \mathcal{P}^\tau(\psi_i(x))\psi_j(x)dx \\ &= \int\limits_{\mathbb{X}} \mathcal{P}^\tau(\chi_{B_i}(x))\chi_{B_j}(x)dx \\ &= \int\limits_{B_j} \mathcal{P}^\tau(\chi_{B_i}(x))dx. \end{aligned} \tag{3.15}$$

For deterministic system (3.1), the coefficients are

$$w_{ij} = \int\limits_{B_j} \mathcal{P}^\tau(\chi_{B_i}(x))dx$$

$$
\begin{aligned}
&= \int\limits_{\Phi_f(-\tau,B_j)} \chi_{B_i}(x)dx, \text{ from (3.3)} \\
&= m(B_i \cap \Phi_f(-\tau, B_j)).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\pi_M \mathcal{P}^\tau(\pi_M \rho) &= \sum_{i=1}^{M} \frac{p^i}{m(B_i)} \sum_{j=1}^{M} \frac{m(B_i \cap \Phi_f(-\tau, B_j))}{m(B_j)} \psi_j \\
&= \sum_{j=1}^{M} \frac{1}{m(B_j)} \sum_{i=1}^{M} p^i P_{\tau,ij} \psi_j, \qquad (3.16)
\end{aligned}
$$

where $P_{\tau,ij} = \dfrac{m(B_i \cap \Phi_f(-\tau, B_j))}{m(B_i)}$ from (3.10).

For stochastic system (3.6)

$$
\begin{aligned}
w_{ij} &= \int\limits_{B_j} \mathcal{P}^\tau(\chi_{B_i}(x))dx \\
&= \int\limits_{B_j} \int\limits_{\mathbb{X}} p_\tau(y,x)\chi_{B_i}(y)dydx, \text{ from (3.8)} \\
&= \int\limits_{\mathbb{X}} \int\limits_{B_j} p_\tau(y,x)dx\chi_{B_i}(y)dy, \text{ by Fubini's theorem} \\
&= \int\limits_{\mathbb{X}} \mathbb{P}[x_{t+\tau} \in B_j | x_t = y]\chi_{B_i}(y)dy, \text{ from (3.7)} \\
&= \int\limits_{\mathbb{X}} \mathbb{P}[x_\tau \in B_j | x_0 = y]\chi_{B_i}(y)dy, \text{ time-homogeneity} \\
&= \int_{B_i} \mathbb{P}[x_\tau \in B_j | x_0 = y]kd\mathbb{P}(y) \\
&= k\mathbb{P}[x_\tau \in B_j, x_0 \in B_i].
\end{aligned}
$$

The uniform change of measure $dy = kd\mathbb{P}(y)$ with normalization factor $k > 0$ results from the fact that initial distribution (i.e., distribution of x_0) is uniform. Hence, for the stochastic system,

$$\begin{aligned}\pi_M \mathcal{P}^\tau(\pi_M \rho) &= \sum_{i=1}^{M} \frac{p^i}{m(B_i)} \sum_{j=1}^{M} \frac{k\mathbb{P}[x_\tau \in B_j, x_0 \in B_i]}{m(B_j)} \psi_j \\ &= \sum_{j=1}^{M} \frac{1}{m(B_j)} \sum_{i=1}^{M} p^i P_{\tau,ij} \psi_j, \qquad (3.17)\end{aligned}$$

which is same as (3.16) and differs only in the definition of $P_{\tau,ij} = \dfrac{k\mathbb{P}[x_\tau \in B_j, x_0 \in B_i]}{m(B_i)}$ from (3.11).

Now define $\sum_{i=1}^{M} p^i P_{\tau,ij} = p_\tau^j$, and $\mathbf{p}_\tau = (p_\tau^1, \ldots, p_\tau^M)$, which implies

$$\begin{aligned}\pi_M \mathcal{P}^\tau(\pi_M \rho) &= \sum_{j=1}^{M} \frac{p_\tau^j}{m(B_j)} \psi_j \qquad (3.18)\\ \mathbf{p}_\tau &= \mathbf{p} P_\tau. \qquad (3.19)\end{aligned}$$

This weak approximation shows that the eigenvalue-eigenvectors of the stochastic matrix P_τ can be used to approximate the eigenvalues and eigenfunctions of $\mathcal{P}^\tau$ by restricting it on the span of the basis functions.

Theorem 3.1. If $(\lambda, \mathbf{p})$ is an eigenvalue-(left) eigenvector pair of P_τ, then λ is also an eigenvalue of the restricted operator $\pi_M \mathcal{P}^\tau$ with eigenfunction $\varphi \triangleq \sum_{i=1}^{M} \frac{p^i}{m(B_i)} \psi_i$.

Proof. Since $(\lambda, \mathbf{p})$ is an eigenvalue-(left) eigenvector pair of P_τ,

$$\mathbf{p}P_\tau = \lambda \mathbf{p}. \tag{3.20}$$

Now from Eq. (3.18),

$$\pi_M \mathcal{P}^\tau \varphi = \sum_{j=1}^{M} \frac{p_\tau^j}{m(B_j)} \psi_j, \text{ since } \pi_M \varphi = \varphi \text{ here,}$$

and

$$\mathbf{p}_\tau = \mathbf{p}P_\tau.$$

But since $\mathbf{p}P_\tau$, i.e., $\mathbf{p}_\tau = \lambda \mathbf{p}$, we get

$$\pi_M \mathcal{P}^\tau \varphi = \sum_{j=1}^{M} \frac{\lambda p^j}{m(B_j)} \psi_j = \lambda \sum_{j=1}^{M} \frac{p^j}{m(B_j)} \psi_j = \lambda \varphi. \tag{3.21}$$

Therefore λ is also an eigenvalue of the restricted operator $\pi_M \mathcal{P}^\tau$ with eigenfunction φ. □

3.2.2 Contstrained Ulam dynamic mode decomposition

Ulam's method uses a Monte Carlo approach to numerically estimate the Markov state transition matrix P_τ. Within each B_i, a set of N test points $x_{i,1}, \ldots, x_{i,N}$ are defined and numerically integrated to obtain $\Phi_f(\tau, x_{i,k})$, $k = 1, \ldots, n$, i.e., their final positions along the trajectories of the ODE (3.1) or the SDE (3.6). The estimated P_τ is given by [28]

$$P_{\tau,ij} \approx \frac{\#\{k : x_{i,k} \in B_i, \Phi_f(\tau, x_{i,k}) \in B_j\}}{N}. \tag{3.22}$$

The choice of τ is important and depends on the resolution of the partition D in this method. If the resolution is coarse, i.e., too few grid-cells (M is small), and τ is also small, then many of the test points will not leave their original grid cell B_i, and the estimated P_τ will be close to the identity matrix.

Extended dynamic mode decomposition (EDMD) [33] is a method to extract the modes of a complex dynamical system by solving a least-squares problem. EDMD estimates the eigenvalues and eigenfunctions of the Koopman operator [64], the dual of the PF operator, which operates on the space of L^∞ observables. The Koopman semigroup of operator $\mathcal{K}^t : L^\infty(\mathbb{X}) \to L^\infty(\mathbb{X})$ is defined as

$$(\mathcal{K}^t \varphi)(\cdot) = \varphi \circ \Phi_f(t, \cdot). \tag{3.23}$$

If we fix the time step $t = \tau$, then the ODE (3.1) becomes an iterative map $x((k+1)\tau) = \Phi_f(\tau, x(k\tau))$, and we can drop τ and define $\Phi_f(\tau, x) \triangleq F(x)$. The discrete-time dynamics become

$$x_{k+1} = F(x_k). \tag{3.24}$$

The time-discretized version of Koopman operator is $\mathcal{K}^\tau \varphi(\cdot) = \varphi \circ F(\cdot)$. In EDMD, just like Ulam's method, the infinite-dimensional operator $\mathcal{K}^\tau$ is projected onto a finite-dimensional basis in $L^\infty(\mathbb{X})$ to represent it as a matrix U. Let $\{\psi_1, \ldots, \psi_M\}$ be the basis functions and, like in Ulam's method, we define $\pi_M : L^\infty(\mathbb{X}) \to \text{sp}\{\psi_i, \ldots \psi_M\}$ to project φ onto the span of these basis functions. Then

$$\varphi(x) = \sum_{i=1}^{M} a_i \psi_i(x) \tag{3.25}$$

$$\mathcal{K}^\tau \varphi(x) = \sum_{i=1}^{M} b_i \psi_i(x) + r, \tag{3.26}$$

with residue r. Now, since φ is an observable, we can observe $\{\varphi(x_0), \ldots \varphi(x_{n+1})\}$ for any $n > 0$, where x_i are from the discretized dynamics (3.24). So we can estimate the matrix K by the least-squares formulation

$$U = \Psi_{x_0}^\dagger \Psi_{x_1}, \tag{3.27}$$

where $\Psi_{x_0,ij} = \psi_i(x_j)$ and $\Psi_{X_1,ij} = \psi_i(x_{j+1})$, $i = 1, \ldots, M$, and $j = 0, \ldots, n$.

In the same light, the weak approximation of the PF operator can be thought of as projecting onto the basis function $\psi_i = \dfrac{1}{m(B_i)} \chi_{B_i}$. Since the basis functions are related to the density of states, and cannot be readily observed, we need the help of Monte Carlo simulation.

Let $\mathbf{p}_0, \ldots, \mathbf{p}_{n+1}$ be $n+2$ subsequent PMF resulting from the operation of $\mathcal{P}^\tau$ on the initial density ρ projected on the $sp\{\psi_1, \ldots, \psi_M\}$. We know from (3.12) that $\mathbf{p}_{k+1} = \mathbf{p}_k P_\tau$. Define

$$\Psi_0 = [\mathbf{p}_0^T, \ldots, \mathbf{p}_n^T]^T, \quad \Psi_1 = [\mathbf{p}_1^T, \ldots, \mathbf{p}_{n+1}^T]^T. \tag{3.28}$$

Therefore, from the Markov relation, $\Psi_1 = P_\tau \Psi_0$, we have

$$P_\tau \approx \Psi_1 \Psi_0^\dagger. \tag{3.29}$$

But since we cannot observe directly the values of $\mathbf{p}_0, \ldots, \mathbf{p}_{n+1}$, we need to estimate them from Monte Carlo sampling. For this, we start from a uniform distribution of N particles in each of the M grids at $t = 0$. Then we successively integrate them forward for time interval τ to get their position at $t = \tau$. The number of particles in B_i will change from N to some value d_1^i, where the subscript 1 represents the first time step. Similarly we go on integrating for n more successive time steps to get the number of particles in each grid d_j^i for each time step j. Let $\mathbf{d_j} = (d_j^1, \ldots, d_j^M)$. Then empirically $\hat{\mathbf{p}}_j \triangleq \dfrac{1}{MN}\mathbf{d}_j \approx \mathbf{p}_j$ for $j = 0, \ldots, n+1$. Now, define empirical data matrices

$$\hat{\Psi}_0 = [\hat{\mathbf{p}}_0^T, \ldots, \hat{\mathbf{p}}_n^T]^T, \quad \hat{\Psi}_1 = [\hat{\mathbf{p}}_1^T \ldots, \hat{\mathbf{p}}_{n+1}^T]^T. \tag{3.30}$$

Next find P_τ that minimizes the error between $\hat{\Psi}_1$ and $\hat{\Psi}_0 P_\tau$, using a constrained least-squares formulation:

$$\begin{aligned} \underset{P_\tau}{\text{minimize}} \quad & \left\| \hat{\Psi}_1 - \hat{\Psi}_0 P_\tau \right\|_F \\ \text{subject to} \quad & P_{\tau,ij} \geq 0,\ i,j \in \{1, \ldots, M\} \\ & \sum_{j=1}^{M} P_{\tau,ij} = 1,\ i \in \{1, \ldots, M\}. \end{aligned} \tag{3.31}$$

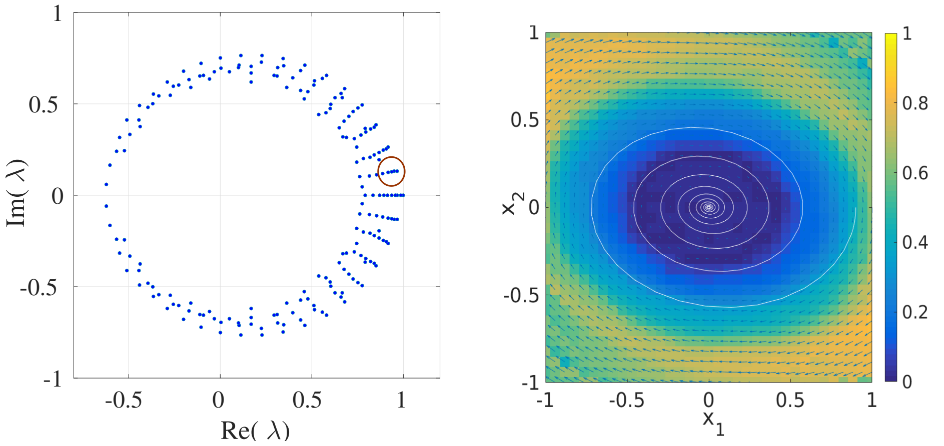
The problem (3.31) is a convex quadratic programming problem and yields a unique minimum that can be solved using gradient-descent or interior point methods. Unlike Ulam's method, this method is a multi-pass approach, which gives more accuracy with short time steps. It also has a distinct advantage over Ulam's method in stochastic systems, since there a very small increment in Wiener noise with a short time step. By solving for P_τ and exploiting Theorem 3.1, we can approximate the eigenvalues and eigenfunctions of the infinite-dimensional operator $\mathcal{P}^\tau$. Henceforth we refer to the problem (3.31) as Constrained Ulam DMD (CU-DMD).

3.3 Numerical simulations

CU-DMD is demonstrated on three different dynamical systems. First, consider a second-order linear system:

$$\dot{x} = Ax, \tag{3.32}$$

where $x \in \mathbb{R}^2$ and $A = \begin{bmatrix} 0 & 1 \\ -a & -b \end{bmatrix}$. The parameters $a = 0.5$ and $b = 0.1$ are chosen to produce damped oscillations. We have chosen time step $\tau = 0.1$ and a 40-by-40 grid, i.e., a total of 1600 grid-cells. For this system, if $\lambda_{1,2}$ are the eigenvalues of $\exp(\tau A)$ with left eigenvectors $\mathbf{w}_{1,2}$, then $\lambda' = k\lambda_1 + l\lambda_2$ for $k, l \in \mathbb{R}$ is a Koopman eigenvalue with eigenfunction $\phi(x) = (\mathbf{w}_1^* x)^k (\mathbf{w}_2^* x)^l$ [29], [43]. Moreover $\phi(x)$ is also a PF eigenfunction with eigenvalue $\lambda = -\lambda' - \mathrm{tr}(A)$ [65], if $|\lambda| \leq 1$. In this case $\lambda_2 = \overline{\lambda}_1$ and $\mathrm{tr}(A) = -0.1$. Choosing $(k, l) = (-0.8, 0)$ and $(k, l) = (-0.4, -0.3)$ so that $|\lambda| < 1$ produces two PF eigenvalues. Table 3.1 compares the exact eigenfunctions with those computed by CU-DMD using the error averaged over the state space. CU-DMD outperforms Ulam's method



((a)) ((b))

Figure 3.1: Second-order linear system (3.32): (a) dominant spectra and (b) eigenfunction corresponding to $\lambda = 0.89 + i0.06$ (red circle)

for time step $\tau = 0.1$. Moreover, the dominant 200 eigenvalues and the approximated eigenfunction corresponding to $\lambda = 0.89 - i0.06$ are shown in Fig. 3.1.

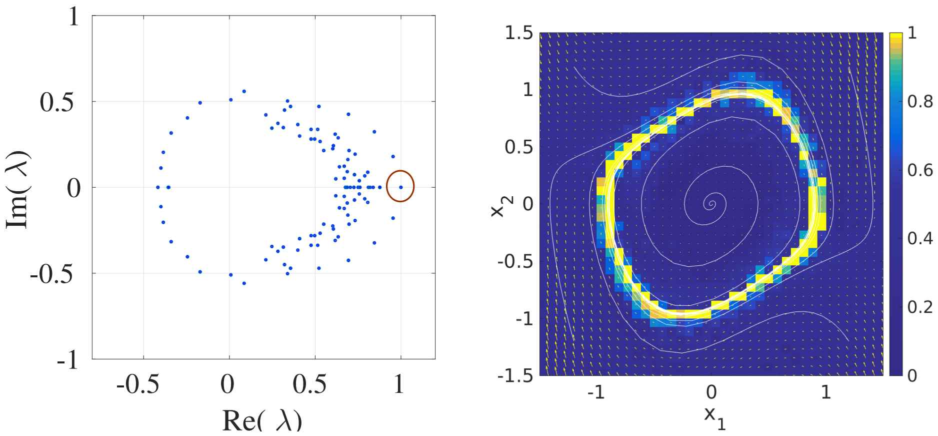
P-F Eigenvalue	Error from Ulam's method	Error from CU-DMD
$0.89 \pm i0.06$	0.45	0.11
$0.79 \pm i0.05$	0.51	0.14

Table 3.1: Error comparison for linear system (3.32)

Next, consider the (scaled) Van der Pol oscillator system:

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= m(c - x_1^2)x_2 - x_1. \end{aligned} \tag{3.33}$$

where $m = 2$ and $c = 0.2$. The time step is $\tau = 0.1$ and the grid is 40-by-40. The approximate spectra of the 100 dominant eigenvalues and the eigenfunction corresponding to $\lambda = 1$ are shown in Fig. 3.2. The eigenfunction traces out a limit cycle, as expected.



((a)) ((b))

Figure 3.2: Van der Pol system (3.33): (a) dominant spectra with unity eigenvalue circled and (b) the corresponding eigenfunction

Then CU-DMD is applied to a double gyre system [28]:

$$\begin{aligned} \dot{x}_1 &= -\pi a \sin\left(\frac{\pi x_1}{s}\right)\cos\left(\frac{\pi x_2}{s}\right) - \mu x_1 \\ \dot{x}_2 &= -\pi a \cos\left(\frac{\pi x_1}{s}\right)\sin\left(\frac{\pi x_2}{s}\right) - \mu x_2, \end{aligned} \tag{3.34}$$

where $a = 0.2$, $s = 1$, and $\mu = 0.1$. The time step $\tau = 0.1$ and a grid of 40-by-20 are chosen. Fig. (3.3) shows the approximated eigenfunction corresponding to $\lambda = 0.96$.

To demonstrate CU-DMD for a stochastic system, we chose a time-homogeneous Itô stochastic differential equation of the form

$$dx_t = f(x_t)dt + \sqrt{2\sigma}dw_t,\ t > 0, \tag{3.35}$$

where dw_t is the standard Wiener increment and $\sigma = 0.1$ is the variance parameter. The drift field $f(x)$ is the deterministic Van der Pol system (3.33). The dominant spectra and the eigenfunction corresponding to the unity eigenvalue for the stochastic Van der Pol

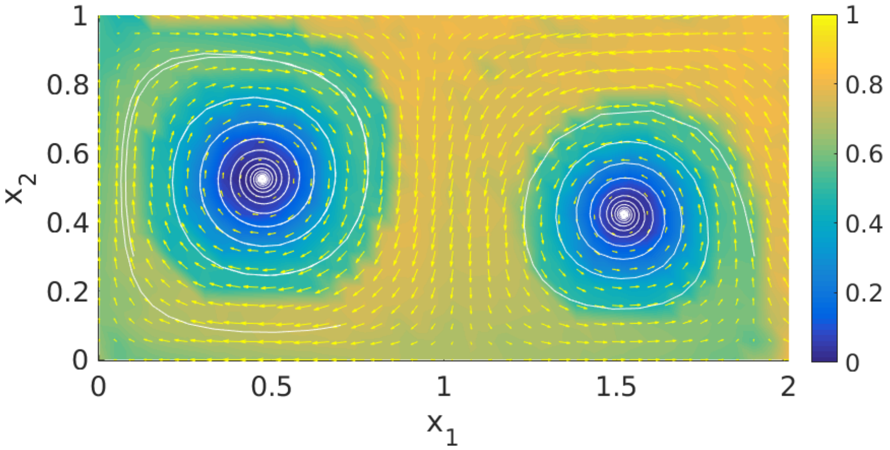


Figure 3.3: Approximated eigenfunction corresponding to $\lambda = 0.96$ for double-gyre system (3.34)

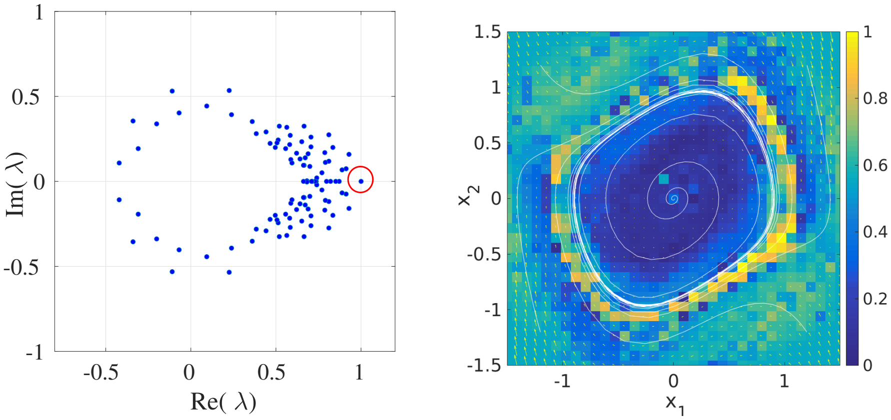
system are shown in the Fig. 3.4; the eigenfunction is not entirely concentrated on the limit-cycle due to the diffusion term.

Diffusive behavior is also demonstrated using a double gyre system [28] corrupted with a Wiener noise. Here the drift field

$$f(x) = \begin{bmatrix} -\pi a \sin\left(\frac{\pi x_1}{s}\right) \cos\left(\frac{\pi x_2}{s}\right) - \mu x_1 \\ -\pi a \cos\left(\frac{\pi x_1}{s}\right) \sin\left(\frac{\pi x_2}{s}\right) - \mu x_2 \end{bmatrix}, \tag{3.36}$$

where $a = 0.2$, $s = 1$, and $\mu = 0.1$. For the deterministic double gyre system, the eigenfunction corresponding to unity is singular (concentrated on the stable equilibria) and not shown here due to space constraint. For the stochastic case, the eigenfunction spreads out from the stable equilibria as illustrated in Fig. 3.5.

A similar result is obtained when the drift equation is set to the double-gyre system



((a)) ((b))

Figure 3.4: The stochastic Van der Pol system (3.33) with noise (3.35): (a) dominant spectra with unity eigenvalue circled and (b) the corresponding eigenfunction

(3.34). Fig. (3.5) shows the eigenfunction corresponding to unity is not singular (i.e., concentrated on the fixed points), but is spread around the equilibrium points. In Fig. (3.3), the eigenfunction corresponding to eigenvalue less than unity is shown since the eigenfunction with eigenvalue unity will be singular and concentrated on the grid-cell containing the equilibrium points. Clearly this is not the case for stochastic system, as demonstrated in Fig. (3.5).

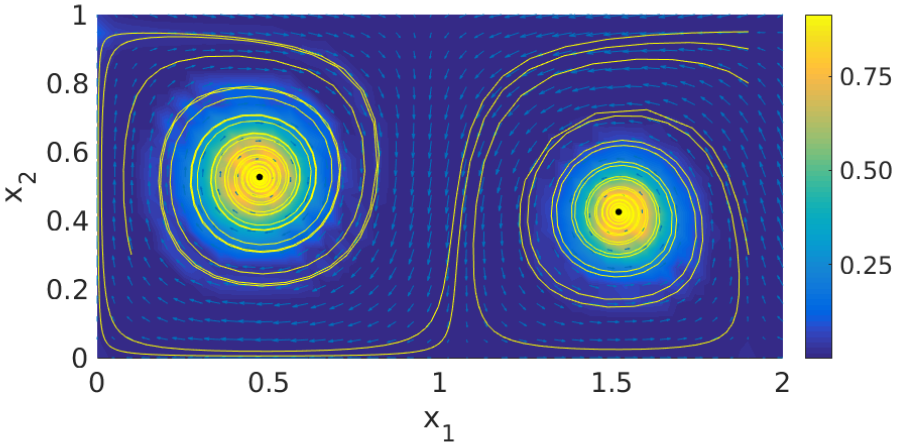


Figure 3.5: The stochastic double-gyre system (3.36): approximated eigenfunction corresponding to unity eigenvalue

3.4 Application: Output feedback control for lift maximization of a pitching airfoil

As an application of nonlinear estimation and feedback control by approximate Perron-Frobenius (PF) operator, this chapter presents the output feedback control of a pitching airfoil that maximizes lift using only the lift-coefficient measurements from the embedded pressure sensors. Unsteady aerodynamics is currently driving research at the interface of fluid dynamics and control theory for low Reynolds number aircraft such as micro air vehicles (MAVs). The regulation and control of unsteady behavior is crucial for maintaining the stability of an MAV, which necessitates accurate modeling of their flight surfaces. When an airfoil is repeatedly pitched, unsteady flow features develop delaying the onset of stall. Moreover, a rapidly pitching airfoil will typically stall at a higher angle of attack than a statically pitched airfoil. While unsteady or dynamic stall has catalyzed the development of model characterization in fluid dynamics, an optimal control approach has not yet emerged. Periodic pitching induces hysteresis loops in the lift coefficient, whose amplitudes are proportional to pitching frequency and airfoil geometry. Earlier works [1, 66, 67] have utilized linear control laws to maximize the steady-state lift, whereas here we seek to maximize the time-averaged lift in this chapter using a nonlinear control law.

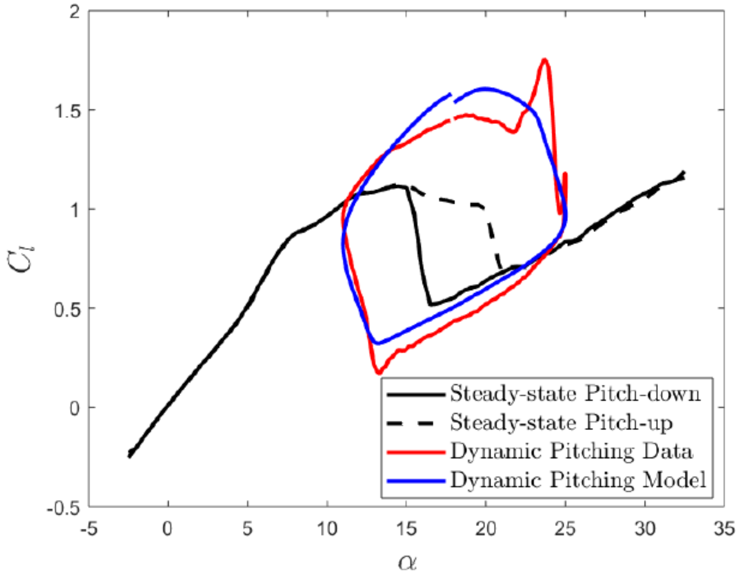


Figure 3.6: Dynamic stall characteristics of an airfoil, comparing model behavior with experimental results. An airfoil's pitch coefficient enters hysteresis loops near its stall angle as separation regions expand and collapse again [1].

The Goman-Khrabrov (GK) model is a low-order, nonlinear representation of the effect of unsteady flow features on lift based on two coupled first-order ordinary differential equations. Fig. 3.6 demonstrates the accuracy of the GK model in characterizing both static and dynamic pitching maneuvers for a wing, where dynamic pitching typically involves a periodic pitch rate. The lift hysteresis loops exhibited in Fig. 3.6 are typical of this type of pitching behavior [66, 67, 68]. Although current literature implements the GK model for purposes of active flow control and stabilization in unsteady flow, we examine the opposite approach: i.e., implementing a destabilizing behavior for the purpose of lift maximization. As [67] suggests, a time-varying control input, e.g., sinusoidal actuation, can reduce hysteresis effects and thereby decrease lift perturbations.

This section presents a state and output feedback control law that stabilizes a limit cycle in the closed-loop Goman-Khrabrov model for flow separation. The optimal unsteady behavior compares favorably with the best-case steady-state behavior. The dynaic output feedback is designed using a recursive Bayesian filter that assimilates noisy measurements of lift. The filter is implemented for the nonlinear dynamics using the CU-DMD-based approximation of PF operator[48]. The output feedback control performance is comparable to the state-feedback case when the measurement noise variance is sufficiently small.

3.4.1 Background on model-based dynamics and estimation

A data-driven approach to studying unsteady flow features is effective at reducing the dimensionality of modeling representations [1, 66, 67, 69]. For example, the Goman-Khrabrov model provides a mechanism for characterizing flow separation on a wing based on static lift data. Section 3.4.1.1 introduces an analytic equation for the lift and drag coefficients and a modified state-space form of the Goman-Khrabrov model that permits a control input. Section 3.4.1.2 introduces the preliminaries of the recursive Bayesian filter and the Perron-Frobenius density transport operator so that the modified GK model may be implemented using output feedback.

3.4.1.1 The Goman-Khrabrov model

Here we utilize an adaptation of the Goman-Khrabrov (GK) model [68] for unsteady flow around an airfoil undergoing dynamic stall. The GK model is a two-state dynamical system consisting of an internal representation of the flow stagnation point $x \in [0, 1]$ and the angle of attack $\alpha \in S^1$ whose time evolution is defined by the following system of equations [68]:

$$\begin{aligned} \tau_1 \dot{x} + x &= f_0(\alpha - \tau_2 \dot{\alpha}) \\ \dot{\alpha} &= u, \end{aligned} \tag{3.37}$$

where the time constants τ_1 and τ_2 are the flow settling times and u is the control input. The stagnation function describes the state of separation as deduced from an airfoil's alignment to pre- and post-stall lift curves. The time rate of change for a dynamically pitching airfoil maps predictably as a function of angle of attack, but at higher or lower angles depending on the pitch rate.

The GK model is applicable to a wide class of airfoils. The stagnation function f_0 may be determined experimentally in order to capture airfoil geometry and flow constants, including the Reynolds number. While there is no known mathematical framework to derive f_0 from first principles, the boundedness of f_0 in the interval $[0, 1]$ can be used to produce an analytic approximation that captures its form and behavior. We elect to use an arctangent function with an argument shift because of its simplicity when differentiated.

The stagnation function used here is

$$f_0(\alpha - \tau_2\dot{\alpha}) = \frac{1}{2} - \frac{1}{\pi}\arctan\left(K_s(\alpha - \tau_2\dot{\alpha} - \phi)\right), \tag{3.38}$$

where K_s and ϕ are additional tuning parameters set according to stall characteristics. Fig. 3.7 compares this stagnation function to experimental data.

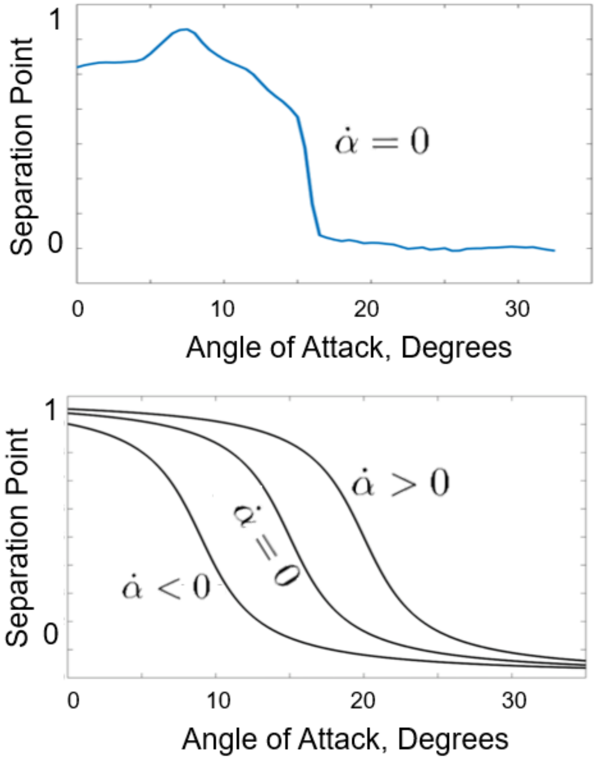


Figure 3.7: Comparison of stagnation function from (a) towing-tank data [1] and (b) analytic model using the arctangent

A generalized model of the coefficient of lift C_L at high angles of attack as a function of the GK-model states x and α is [66, 67]

$$C_L = g_1(\alpha)x + g_2(\alpha)(1 - x), \tag{3.39}$$

where the functions $g_1(\alpha)$ and $g_2(\alpha)$ represent attached flow and fully separated flow, respectively. Note that Eq. (3.39) is linear in x. Goman and Khrabrov also present an analytic approximation for the coefficient of lift, based on assumptions that the separated flow is modeled by linear cavitation theory and Kirchhoff's zone of constant pressure [68]. In the original GK-model, the coefficient of lift is nonlinear in x and α, i.e., [68]

$$C_L = \frac{\pi}{2}(1 + \sqrt{x})^2 \sin\alpha. \tag{3.40}$$

While the general representation of C_L (3.39) is more readily applicable to airfoil data, we elect to model lift behavior here using (3.40) despite the nonlinearity, since it is analytic and can be differentiated directly.

Although thin airfoil theory fails to provide a closed-form analytic model for drag on an airfoil [70], the following model captures the dramatic increase in drag introduced by high angles of attack. Let C_{D0} represent the parasitic and form drag. The coefficient of drag is [70]

$$C_D = C_{D0} + \frac{C_L^2}{\pi e\,\mathrm{AR}}. \tag{3.41}$$

While the exact induced drag characteristics of the wing used in testing are not known precisely, the behavior of the drag coefficient is enough to characterize the drag produced at large angles of attack [70]. Small zero-lift drag is neglected to reflect the comparatively small effect of skin friction on the airfoil. A conservative span efficiency factor of $e = 0.7$ is used here to represent a non-elliptical planform.

The internal variable x can be derived from existing lift curve data for pre-stall and post-stall curves, where $x = 1$ represents fully attached flow and $x = 0$ denotes that the flow is fully separated. For example, a value of $x = 1$ returns a lift curve (3.40) identical to an airfoil's standard pre-stall trend. x is not a direct measurement of flow stagnation, as no information related to free-stream velocity, static pressure, or vorticity exists in this system. Rather, it can be thought of a representation of the flow attachment based on prior data-driven testing of the steady-state airfoil. Intermediate values between 0 and 1 reflect transient conditions. From (3.40), the nonlinear dependence of C_L on both α and x suggests that period oscillations in these parameters may produce lift that is higher, on average, than at constant α and x. This realization motivates a time-averaged, lift-maximizing control design using periodic pitching.

Let $\mathbf{z} = (z_1, z_2) \triangleq (x, \alpha)$ and $y \triangleq C_L$. In state-space form, the modified GK model is

$$\begin{aligned} \dot{z}_1 &= -\tfrac{1}{\tau_1} z_1 + \tfrac{1}{\tau_1}\left[\tfrac{1}{2} - \tfrac{1}{\pi}\arctan\left(K_s(\alpha - \tau_2\dot{\alpha} - \phi)\right)\right] \\ \dot{z}_2 &= u \\ y &= \tfrac{\pi}{2}(1 + \sqrt{z_1})^2 \sin z_2. \end{aligned} \tag{3.42}$$

In the sequel, we refer to the dynamics in (3.42) as $\dot{\mathbf{z}} = \mathbf{F}(\mathbf{z}, u)$. With regard to the input u, it is possible to use a sinusoidal controller in the form $u = A\sin(2\pi f t)$ to vary the pitch rate directly [67]. However, we seek below to derive a state feedback control of the form $u = u(\mathbf{z})$. Then we seek to implement this control using measurements of the output y to estimate z_1, assuming z_2 is known.

3.4.1.2 Recursive Bayesian filtering

In order to implement an output feedback control for the system (3.42), we implement a recursive Bayesian filter to estimate $z_1 = x$ from measurements of $y = C_L$, assuming $z_2 = \alpha$ is known. This form of a filter is chosen because of the nonlinear dynamics and nonlinear output function in (3.42). A Bayesian framework allows the unknown state x to be inferred from the noisy measurements of the lift coefficient C_L together with the prior information accumulated in a probability density function [71]. In practice, measurements of the coefficient of lift of an airfoil can be obtained experimentally in a towing tank via a spring-gauge.

The first stage of a Bayesian filter, i.e., the prediction step, propagates the accumulated density according to the system dynamics. Because of the nonlinear dynamics, prediction is accomplished here by an offline approximation of the Perron-Frobenius (PF) transfer operator for this system [48]. We employ the discretized version of the PF operator that works on a discretized prior. The prior density $\rho(z_1)$ is discretized using a grid on the state-space arranged in a vector resembling a probability mass function (PMF) $\mathbf{p}$. The discretized PF operator in form of a matrix P^τ operates over the PMF $\mathbf{p}$ to yield the updated prior $\mathbf{p}_\tau = \mathbf{p}P^\tau$, according to the dynamics. The time step of the motion update is small to keep the discretization error sufficiently small. The discretized approximation of the PF operator is computed using the CU-DMD algorithm.

The second stage of the Bayesian filter, i.e., the measurement step, involves the calculation of the posterior estimate using the conditional probability density known as the likelihood function. Let $\tilde{y}$ represent the noisy measurement of y. The Bayesian filter recursively calculates the posterior estimate $p(z_1|\tilde{y})$ according to Bayes' theorem:

$$\rho(z_1|\tilde{y}) = \frac{\rho(\tilde{y}|z_1)\ \rho(z_1)}{\rho(\tilde{y})}. \tag{3.43}$$

For simplicity, assume that the lift-measurement noise is zero-mean Gaussian and the initial (prior) distribution is uniform. Also, the instantaneous angle-of-attack z_2 is known precisely, i.e., measured with zero variance.

3.4.2 Feedback control design for the Goman-Khrabrov Model

The Goman-Khrabrov model provides a reduced-order framework to design a state-feedback control that maximizes the time-averaged lift. This section describes a nonlinear state-feedback control law that drives the airfoil dynamics to a limit cycle whose parameters can be optimized using average lift as a metric.

In contrast to an open-loop sinusoidal input, a state-feedback control is more robust in the presence of disturbances and model errors. Moreover, the closed-loop system is autonomous, which permits rigorous analysis in the phase plane.

To begin, consider the linear state-feedback controller, $u = -k_1 z_1 - k_2 z_2$, which drives the system to a static equilibrium point. However, using the GK lift coefficient (3.40), no steady pitch angle achieves the same lift performance as periodic pitching (see Fig. 3.8), which also avoids stall. Hence, we seek to stabilize a limit cycle by *destabilizing* the static equilibrium point. To accomplish this objective, consider the nonlinear state-feedback law

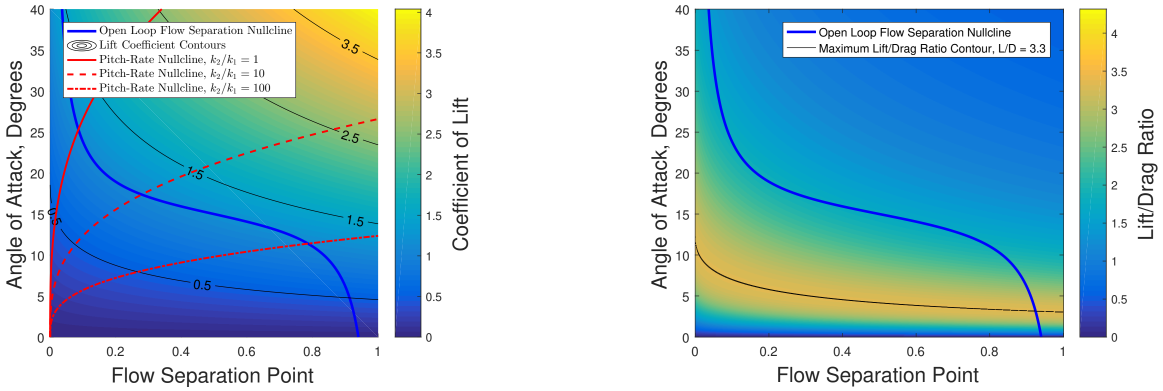


Figure 3.8: (Left) Lift coefficient in the phase plane. The open-loop flow separation nullcline (blue) represents the set of possible steady-state equilibrium points, which do not access the high lift regions (yellow). (Right) Color map depicting the lift-to-drag ratio as a function of the lift coefficient

$$u = k_1 z_1 - k_2 z_2^3, \tag{3.44}$$

where k_1 and k_2 are positive control gains. Note that setting $k_1 < 0$ emulates stabilizing the linear control, using a cubic term for z_2 instead of a linear term.

The closed-loop system (3.42) with the control (3.44) stabilizes a limit cycle for certain values of the gains k_1 and k_2. Fig. 3.8(left) depicts the nullclines for angle-of-attack, i.e., $\dot{z}_2 = 0$, for several values of the ratio k_2/k_1 superimposed on the heat map of the lift coefficient over the state space. Also shown is the open-loop nullcline for the flow separation point, i.e., $\dot{z}_1|_{u=0} = 0$. Fig. 3.8(right) shows the same nullcline for the lift-to-drag ratio.

In order to establish the existence of a stable limit cycle in the closed-loop system, observe that the equilibrium points of the GK model satisfy $\dot{z}_1 = 0$ and $\dot{z}_2 = 0$, which implies

$$\frac{k_2}{k_1} z_2^3 = f_0(z_2). \tag{3.45}$$

Let z_2^* represent a solution to the equilibrium condition (3.45). The Jacobian of the closed-loop system evaluated at z_2^* is

$$A \triangleq \left.\frac{\partial \mathbf{F}}{\partial \mathbf{z}}\right|_{z_2 = z_2^*} = \begin{bmatrix} \tau_1^{-1}\left(\dfrac{\tau_2^{-1}\pi^{-1}K_s k_1}{K_s(z_2^* - \phi)^2 + 1} - 1\right) & \dfrac{\tau_1^{-1}\tau_2^{-1}\pi^{-1}K_s(-3k_2(z_2^*)^2 - \tau_2)}{K_s(z_2^* - \phi)^2 + 1} \\ k_1 & -3k_2(z_2^*)^2 \end{bmatrix}. \tag{3.46}$$

The eigenvalues of A are

$$\lambda_{1,2} = \frac{1}{2}\left(\operatorname{tr} A \pm \sqrt{\operatorname{tr}^2 A - 4 \det A}\right). \tag{3.47}$$

If k_1 is allowed to vary while k_2 is held fixed, then the real and imaginary parts of the eigenvalue may be considered functions of the bifurcation parameter k_1, i.e.,

$$\lambda_{1,2} = \sigma(k_1) \pm \omega(k_1). \tag{3.48}$$

Fig. 3.9 illustrates how the eigenvalues of the closed-loop system cross the imaginary axis with nonzero velocity as the bifurcation parameter k_1 is increased. This behavior is consistent with a supercritical Hopf bifurcation [72], which gives rise to a stable limit cycle that corresponds to the desired oscillating trajectory. Fig. 3.10(left) shows the shape of

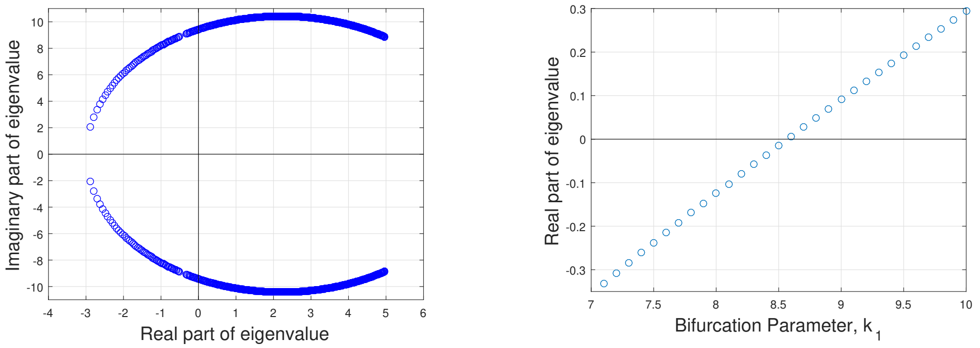


Figure 3.9: Poles of the linearized, closed-loop system as a function of the bifurcation parameter k_1. A Hopf bifurcation occurs when the real part of the eigenvalue crosses the imaginary axis away from zero.

the limit cycle in phase space. Note that it contains an unstable equilibrium point, and is thus attracting for interior points as well as the local neighborhood outside of it.

Figure 3.10(right) depicts the average-lift over all control gains k_1 and k_2 that stabilize a permissible limit cycle, i.e., a limit cycle that remains in the region of model validity. (Without loss of generality, assume an upper limit of 50° for the angle of attack; the analysis can be performed for any such limit.) To determine the optimal gains, time series simulation data of (z_1, z_2) were computed for a 100×100 grid of gain values.

The optimal gain for the model parameters is $(k_1, k_2) = (11.75, 22.8)$, which produces a time-averaged lift coefficient of $\bar{C}_L = 2.0$. The steady-state maximum possible lift coefficient is $\bar{C}_L = 1.43$ (see Fig. 3.8). Therefore, the nonlinear (unsteady) control performs approximately 40% better than linear (steady) control, using lift as a metric. However, the linear control performs better when using as metric the average ratio of lift to drag.

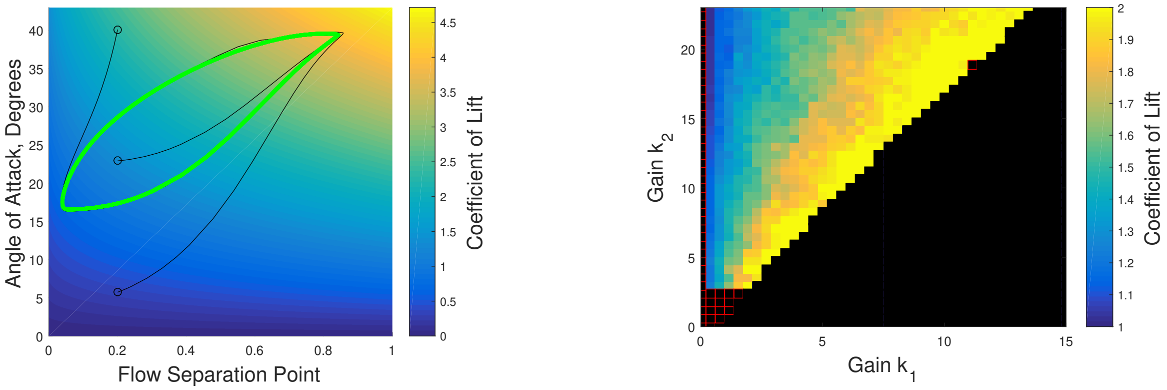


Figure 3.10: (Left) The stable limit cycle (green) in the closed-loop system attracts solutions that start inside or outside the limit cycle (black); (Right) Average lift performance of the closed-loop system for gain values that stabilize a limit cycle. Black squares indicate gains that fail to result in a permissible limit cycle. Red borders indicate gains that fail to produce a limit cycle at all.

3.4.3 Output feedback control

The nonlinear control designed above is a state-feedback control that requires the precise values of the state variables to be available. However, state feedback control of this system is difficult or impossible to implement, because the flow-separation point z_1 is not measured directly. Here we design a filter to recursively estimate z_1 from noisy measurements of the lift coefficient $y = C_L$. Then we implement an output-feedback controller by replacing the state variable in the nonlinear controller with the estimated value $\hat{z}_1$.

Consider the state space equations (3.42). Let η be zero-mean uncorrelated Gaussian measurement noise with variance σ_M^2. The state and output equations have the general

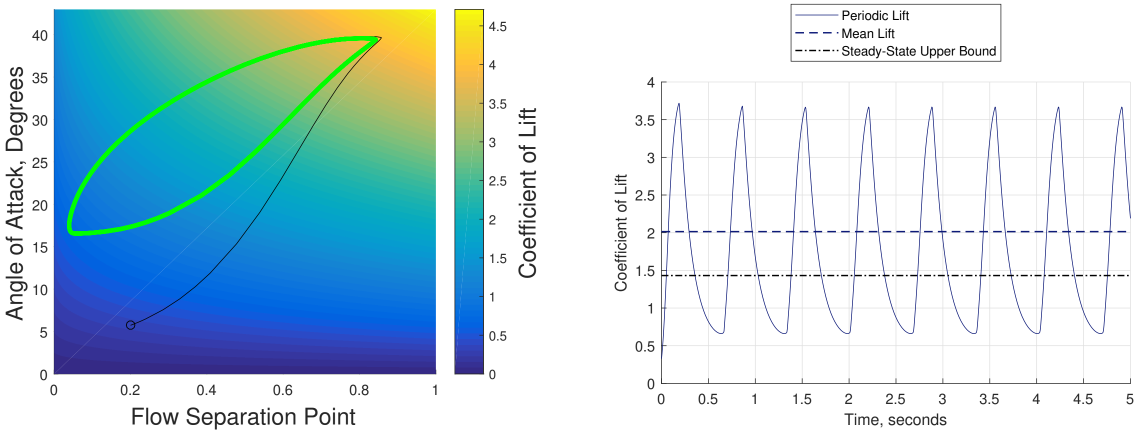


Figure 3.11: The lift-optimizing limit cycle (left) yields an average lift coefficient 40% higher than the steady-state upper bound (right).

form

$$\begin{aligned} \dot{\mathbf{z}} &= \mathbf{F}(\mathbf{z}, u) \\ \tilde{y} &= y + \eta, \end{aligned} \tag{3.49}$$

where $\tilde{y}$ denotes the measured signal plus noise. These equations form the basis of a two-stage Bayesian filter with output feedback: the first equation describes the time evolution of the system states and the second allows them to be estimated from measurements. Assumed that the angle of attack $\alpha = z_2$ is measured precisely. With these assumptions, the likelihood function of the measurement $\tilde{y}$ given the flow separation point z_1 is

$$\rho(\tilde{y}|z_1) = \frac{1}{\sqrt{2\pi\sigma_M^2}} \exp\left(-\frac{(\tilde{y} - y(\mathbf{z}))^2}{2\sigma_M^2}\right). \tag{3.50}$$

To apply the Bayesian framework, we use motion and measurement updates periodically over time. In the motion update, the prior $\rho(z_1)$ is propagated using the dynamics (3.42) and the propagated prior is updated using Bayes' rule (3.43) when the measurement $\tilde{y}$

arrives. The posterior is normalized to integrate to unity over the state space.

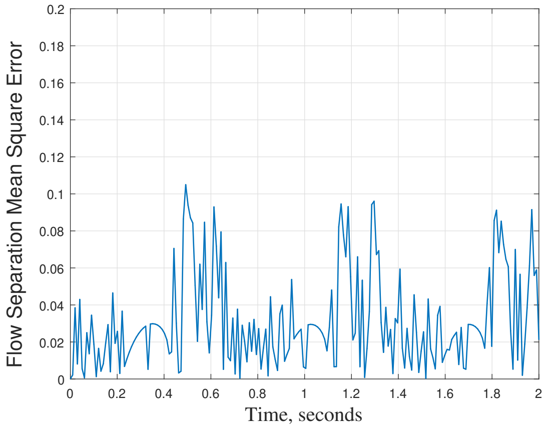


Figure 3.12: Mean-squared error of flow separation point using output feedback calculated over 100 Monte-Carlo simulations. Each simulation corresponds to 3 orbits of the limit cycle.

Because the equations of motion are nonlinear, we use a discretized version of the Perron-Frobenius (PF) operator [48, 26] with time step $\tau = 0.01$ and a size 30 grid on $z_1 = x$ for the motion update step. Constrained Ulam Dynamic Mode Decomposition (CU-DMD) [48] calculates the PF operator for 30 different control signals corresponding to the 30 grid points on the z_1 axis. For each $\tau = 0.01$, the discretized prior vector p is updated using the discretized PF matrix P^τ, and the mode of the prior $\hat{z}_1 = \underset{z_1}{\text{argmax}} \rho(z_1)$ is used as the estimate for the control design. After the Bayesian measurement update (3.43), the mode of the posterior $\hat{z}_1 = \underset{z_1}{\text{argmax}} \rho(z_1|\tilde{y})$ is used as the estimate as well. For each time step τ, the corresponding output-feedback control $u(\hat{z}_1, z_2)$ is used instead of the state-feedback control. Fig. 3.12 shows the mean-square error between the output- and the state-feedback controls for 100 Monte Carlo trials.

An important requirement of the output-feedback control is the preservation of the limit-cyclic behavior, which allows lift augmentation to continue indefinitely. Fig. 3.13 shows that the output-feedback control with discretized PF operator preserves the limit-cyclic structure of the state-feedback control, thereby preserving the high average lift. Adherence to the optimal limit cycle in each iteration is a function of the measurement noise covariance, as depicted in Fig. 3.13. Fig. 3.14 compares the lift coefficient time series obtained from the trajectories depicted in Fig. 3.13.

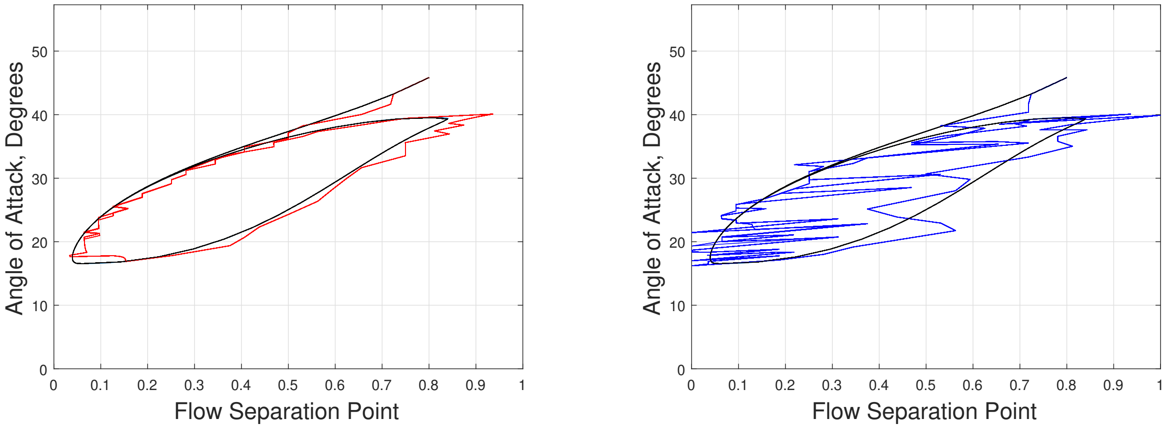


Figure 3.13: Limit cyclic behavior is preserved with output feedback, although adherence to the state feedback trajectory is a function of measurement noise σ_L: (left) $\sigma_M = 0.01$; (right) $\sigma_M = 0.1$

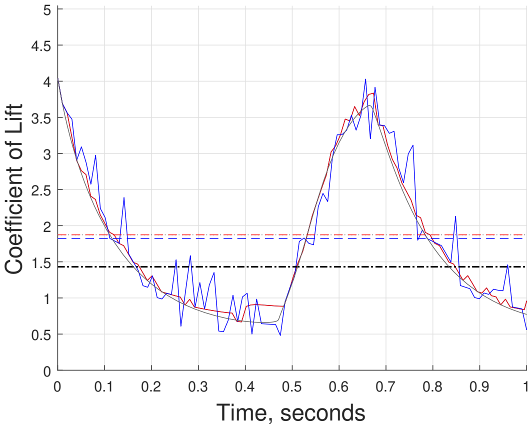


Figure 3.14: Coefficient of lift versus time for the trajectories in Fig. 3.13 (gray: state feedback, red: low noise, blue: high noise) and the optimal steady-state coefficient of lift (black). Time averages for each trajectory are shown with dashed lines of the same color.

Part II

Estimation and control of unmodeled or partially modeled systems

CHAPTER 4

Data-driven estimation using an echo-state neural network equipped with an ensemble Kalman filter

This chapter develops a data-driven sparse estimation method by combining the strength of a neural network with an existing nonlinear filtering algorithm. An echo-state network (ESN) is chosen as the recurrent neural engine for modeling unknown dynamics, because it can be trained quickly with limited computational resources. The ESN adopts an input-output neural network with a randomly generated recurrent reservoir. Linear regression determines the output weights. It has been shown that an ESN with fading memory can universally model nonlinear dynamics [39, 40]. During the training phase, a full measurement of the states is typically utilized as the training data. Once the ESN is trained to reasonably model the dynamics, it is used to generate the motion update of the data-driven estimation. Since the ESN models a nonlinear dynamical system, a nonlinear data-assimilation method is required for the measurement update. While the EKF and UKF perform well in model-based scenarios, the computation of the linearized dynamics is challenging for an ESN. The ensemble Kalman filter [73] is thus chosen for the measurement update for its strength in representing the posterior distribution of states by its sample means and covariances. The incorporation of an ensemble Kalman filter in the feedback loop of an ESN improves measurement assimilation in comparison to its ESN-only counterpart [2], because the former accounts for the measurement noise with the help of a traditional Bayesian framework and assimilates a series of measurements over the testing phase, whereas the latter uses the current measurement only [2].

4.1 Echo-state networks: a universal predictor

Echo-state networks (ESNs) are a special kind of fixed recurrent neural network (RNN) in which a large, random, and fixed RNN is driven by the input signal. The nonlinear response signals thus induced in the neurons are then linearly combined to match a desired output signal. The random, fixed network is called a reservoir and the technique is also known as reservoir computing (RC) [39].

An ESN is composed of three principal components: a linear input layer $\mathbf{u}$ with m input nodes, a recurrent nonlinear reservoir network $\mathbf{r}$ with n neurons, and a linear output layer $\mathbf{y}$ with p output nodes. The reservoir network evolves with the following dynamics [39]

$$\mathbf{r}(t+\Delta t) = (1-\alpha)\mathbf{r}(t) + \alpha\psi(W\mathbf{r}(t) + W_{in}\mathbf{u}(t)), \tag{4.1}$$

where W is the $n \times n$ reservoir weight matrix, W_{in} is the $n \times m$ input weight matrix, $\mathbf{u}$ is the m-dimensional input signal, and $\mathbf{y}$ is the p-dimensional output signal. The time step δt is chosen according to the sampling interval of the training data. The parameter $\alpha \in (0, 1]$ is called the leakage rate, which forces the reservoir to evolve more slowly as $\alpha \to 0$. The activation function ψ is usually a sigmoid function, e.g., $\tanh(\cdot)$ or a logistic function. The output is taken as a linear combination of the reservoir states [39], i.e.,

$$\mathbf{y}(t) = W_{out}\mathbf{r}(t), \tag{4.2}$$

where W_{out} is the $p \times n$ output weight matrix. The weights W_{in} and W are initially randomly drawn and then held fixed. The weight W_{out} is adjusted during the training process. The reservoir weight matrix W is usually kept sparse for computational efficiency.

During the training phase, the ESN is driven by an input sequence $\{\mathbf{u}(t_1), \ldots, \mathbf{u}(t_N)\}$ that yields a sequence of reservoir states $\{\mathbf{r}(t_1), \ldots, \mathbf{r}(t_N)\}$. The reservoir states are stored in a matrix $\mathbf{R} = [\mathbf{r}(t_1), \ldots, \mathbf{r}(t_N)]$. The correct outputs $\{\mathbf{y}(t_1), \ldots, \mathbf{y}(t_N)\}$, which are part of the training data, are also arranged in a matrix $\mathbf{Y} = [\mathbf{y}(t_1), \ldots, \mathbf{y}(t_N)]$. The training is carried out by a linear regression with Tikhonov regularization as follows [37]:

$$W_{out} = (\mathbf{R}\mathbf{R}^T + \beta\mathbf{I})^{-1}\mathbf{R}\mathbf{Y}, \tag{4.3}$$

where $\beta > 0$ is a regularization parameter that ensures non-singularity.

Remark 4.1. For an ESN to be an universal approximator, i.e., to realize every nonlin-

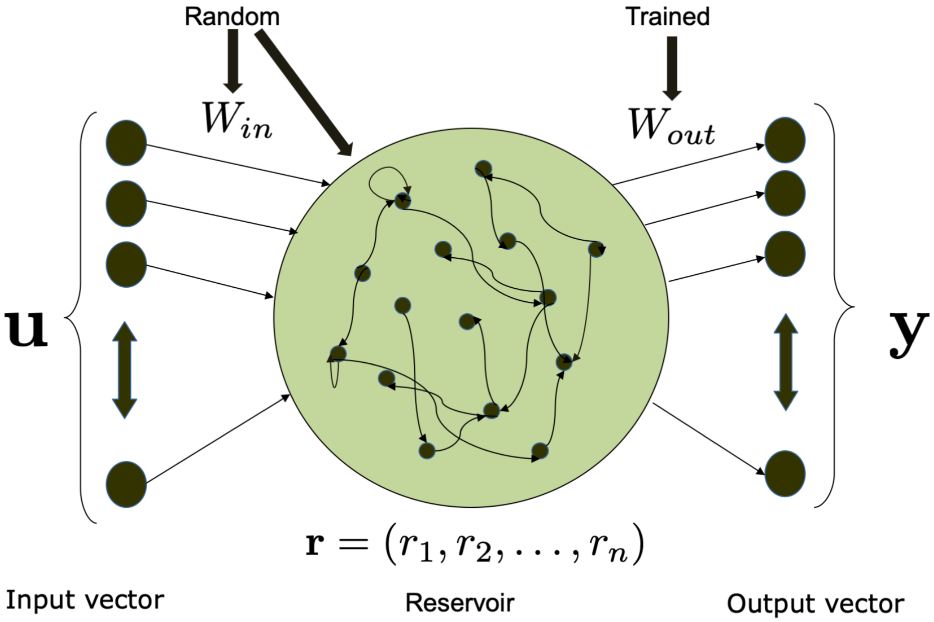


Figure 4.1: The basic schema of an ESN with the input, reservoir, and output layers

ear operator with bounded memory arbitrarily accurately, it must satisfy the echo state property (ESP) [37]. Intuitively, the ESP states that the reservoir will asymptotically wash out any information from initial conditions. For the $\tanh(\cdot)$ activation function, it is empirically observed that the ESP is granted for any input if the spectral radius of W is smaller than unity [37]. To ensure this condition is met, the randomly generated W used here is normalized by its spectral radius.

Remark 4.2. An ESN is advantageous over other kinds of RNN when a cheap, fast, and adaptive training is required since its training does not require back propagation through time.

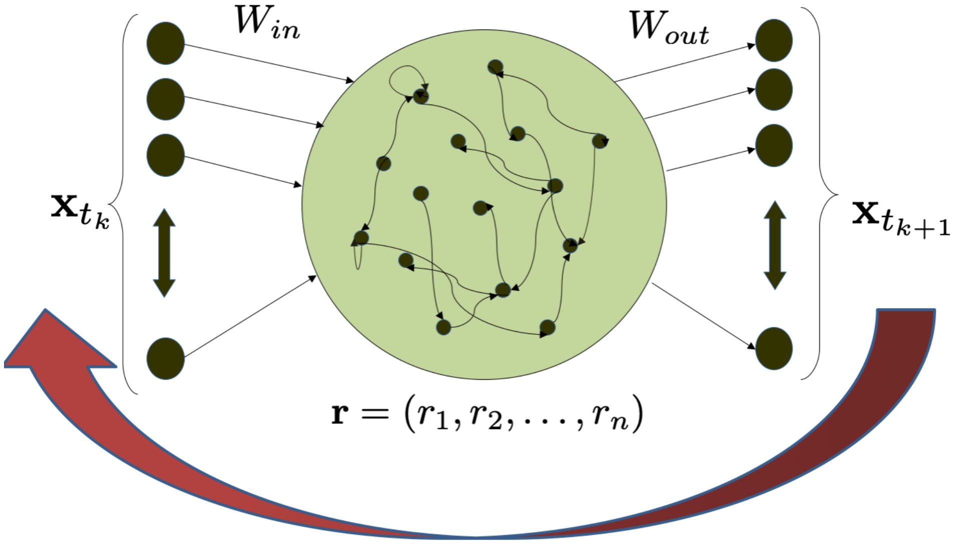


Figure 4.2: A free-running ESN trained for time-series prediction

4.2 Kalman filtering with Echo-state networks: a sparse estimation technique

An ESN can be trained to predict a time series $\{\mathbf{x}_{t_i} \in \mathbb{R}^d : i \in \mathbb{N}\}$ generated by a dynamical system by setting $\mathbf{u}(t)$ and $\mathbf{y}(t)$ as the current and next state value (i.e., $\mathbf{x}_{t_k}$ and $\mathbf{x}_{t_{k+1}}$) respectively. The network is trained for a certain training length N of the time series data $\{\mathbf{x}_{t_i}, i = 1, \ldots, N\}$, and then can run freely by feeding the output $\mathbf{y}_{t_k}$ back to the input $\mathbf{u}_{t_{k+1}}$ of the reservoir. In this case, both $\mathbf{u}$ and $\mathbf{y}$ have the same dimension d as that of the time series data. This setup is shown in Fig. 4.2, where a trained ESN is used to predict the next states of a dynamical time series starting from an initial condition.

Although a sufficiently large free-running ESN trained with enough data can predict a dynamical system reasonably well [2], [38], it has some shortcomings. The initial input to the ESN during the free-running phase must match the exact time series data or the initial condition on the trajectory that is predicted. But for a data-driven estimation problem, the state might not be fully observable. Moreover, the free-running ESN does not take into account any change in the availability of possibly sparse observations that are available during the testing phase. A solution to this problem is presented in [2] where an ESN is trained with all state measurements available, and then predicts the states with only a limited subset of them measured. However, this method, called a *reservoir observer*, uses the ESN's internal connections in the testing phase to assimilate the measurement of the current time step only, and does not take measurement noise into consideration.

We propose an alternative method for measurement assimilation by adding an ensemble Kalman filter (EnKF) [74] in the feedback loop of the ESN (Fig. 4.3). The EnKF block takes sparse observations and uses the state forecast from the reservoir output to generate a state estimate for feedback to the reservoir input. The ensemble Kalman filter is realized as follows. For time-step $k = 0$, an ensemble $\mathbf{X}_{t_0} = [\mathbf{x}_{t_0}^{(1)}, \ldots, \mathbf{x}_{t_0}^{(M)}]$ is chosen from a Gaussian distribution with an ensemble covariance R_x. Then for $k = 0, 1, \ldots,$ the following steps are computed:

$$\begin{aligned}
\mathbf{X}_{t_k} &= [\mathbf{x}_{t_k}^{(1)}, \ldots, \mathbf{x}_{t_k}^{(M)}] \qquad (4.4)\\
\overline{\mathbf{x}}_{t_{k+1}}^{(i)} &= W_{out}\psi(W\mathbf{r}_{t_k} + W_{in}\mathbf{x}_{t_k}^{(i)}), \ \text{for } i = 1, \ldots, M\\
\mathbf{X}_{t_{k+1}}^{f} &= [\overline{\mathbf{x}}_{t_{k+1}}^{(1)}, \ldots, \overline{\mathbf{x}}_{t_{k+1}}^{(M)}]
\end{aligned}$$

These steps carry out the motion update for the ensemble using the ESN. The superscript (i) denotes the i^{th} ensemble member. The forecast ensemble is collected in the $\mathbf{X}^f_{t_{k+1}}$ matrix. Next, the observations are assimilated through an ensemble Kalman filter as follows:

$$
\begin{aligned}
\mathcal{Y}_{t_{k+1}} &= h(\mathbf{X}^f_{t_{k+1}}) \\
P_{xy}(t_{k+1}) &= (\mathbf{X}^f_{t_{k+1}} - \overline{\mathbf{X}}^f_{t_{k+1}})(\mathcal{Y}_{t_{k+1}} - \overline{\mathcal{Y}}_{t_{k+1}})^T \\
P_{yy}(t_{k+1}) &= (\mathcal{Y}_{t_{k+1}} - \overline{\mathcal{Y}}_{t_{k+1}})(\mathcal{Y}_{t_{k+1}} - \overline{\mathcal{Y}}_{t_{k+1}})^T \\
K_{t_{k+1}} &= P_{xy}(t_{k+1})P_{yy}(t_{k+1})^{-1} \\
\mathbf{X}^{\hat{}}_{t_{k+1}} &= \mathbf{X}^f_{t_{k+1}} + K_{t_{k+1}}(\mathbf{Y}_{t_{k+1}} - \mathcal{Y}_{t_{k+1}}) \\
\hat{\mathbf{x}}_{t_{k+1}} &= \frac{1}{M}\sum_{i=1}^{M} \mathbf{x}^{(i)}_{t_{k+1}},
\end{aligned}
\tag{4.5}
$$

where $\mathbf{Y}_{t_{k+1}} = [\mathbf{y}_{t_{k+1}}, \ldots, \mathbf{y}_{t_{k+1}}]$ is a matrix constructed by stacking M copies of the true observation. $P_{xy}(t_{k+1})$ denotes the sample cross-covariance between the states and the observation, whereas $P_{yy}(t_{k+1})$ denotes the sample observation covariance for the ensemble. The sample mean is taken as the state estimate $\hat{\mathbf{x}}_{t_{k+1}}$.

Remark 4.3. An ESN with an ensemble Kalman filter for the measurement update is particularly useful for estimation problems where the state is fully observable during the training phase but only partial and noisy measurements can be obtained during the testing phase. Some of these applications include prediction of atmospheric quantities [75] and flow estimation over an airfoil [69]. Knowledge of initial condition is not required.

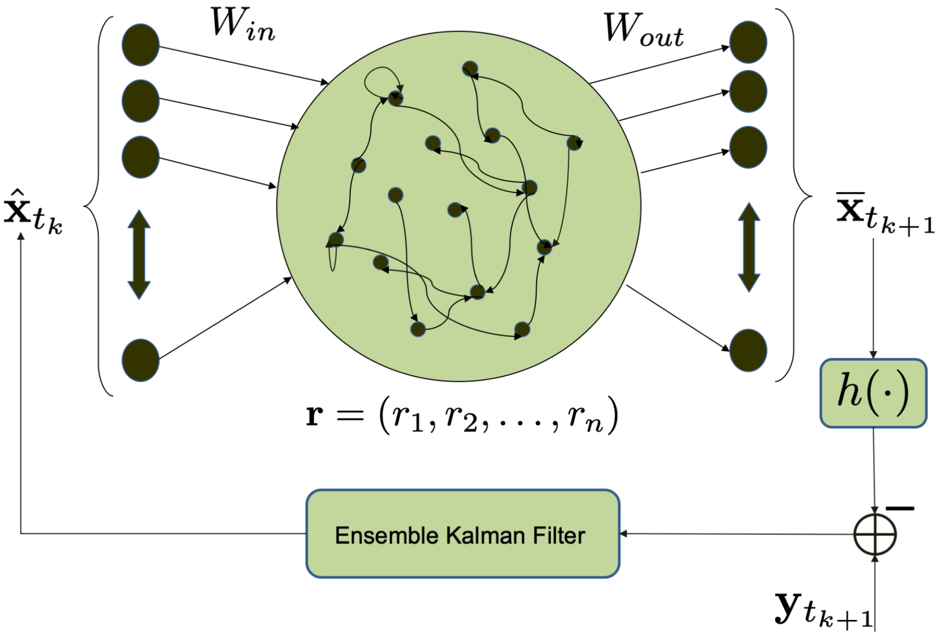


Figure 4.3: An ESN with a Kalman filter in the loop

Remark 4.4. The use of an ensemble Kalman filter enables a Bayesian framework for assimilation of noisy measurements. It also improves the estimation accuracy by integrating measurements over time rather than merely using the current measurement [2].

4.3 Examples

This section illustrates the ESN-based sparse estimation on three data-assimilation problems. The first two are time series generated by chaotic dynamical systems. The last one is a real-time series of traffic flow data obtained from the Numina sensor nodes [76] installed on the University of Maryland campus.

4.3.1 Lorenz system

We tested the ESN with the ensemble Kalman filter to estimate a time series generated by the Lorenz system:

$$\begin{aligned} \dot{x}_1 &= \sigma(x_2 - x_1) \\ \dot{x}_2 &= x_1(\rho - x_3) - x_2 \\ \dot{x}_3 &= x_1 x_2 - \beta x_3, \end{aligned} \tag{4.6}$$

where $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$ produces chaotic behavior. The time-series of $\mathbf{x}(t_k) = [x_1(t_k), x_2(t_k), x_3(t_k)]$ is available during the training phase with $\Delta t = 0.1$. An ESN with 1000 reservoir nodes is trained with data spanning 1000 time-steps. The reservoir weight matrix is constructed as the adjacency matrix of an Erdös-Rényi graph $G(n, p)$ where $n = 1000$ is the number of the reservoir nodes and $p = 0.01$ denotes the probability that an edge is present independent of the other edges. Being the adjacency matrix of an undirected graph, W is symmetric. In the testing phase, only x_2 is observed with an i.i.d. additive zero mean Gaussian noise of covariance 0.01. This measurement is fed into the ensemble Kalman filter with an ensemble size of 100. The reservoir nodes are initialized with random initial conditions and an initial guess of the time series is chosen. The estimated signal is compared with the true signal in Fig 4.4.

Remark 4.5. The introduction of the ensemble Kalman filter in the ESN feedback loop enables it to accurately estimate the time-series signal even if the initial error is large,

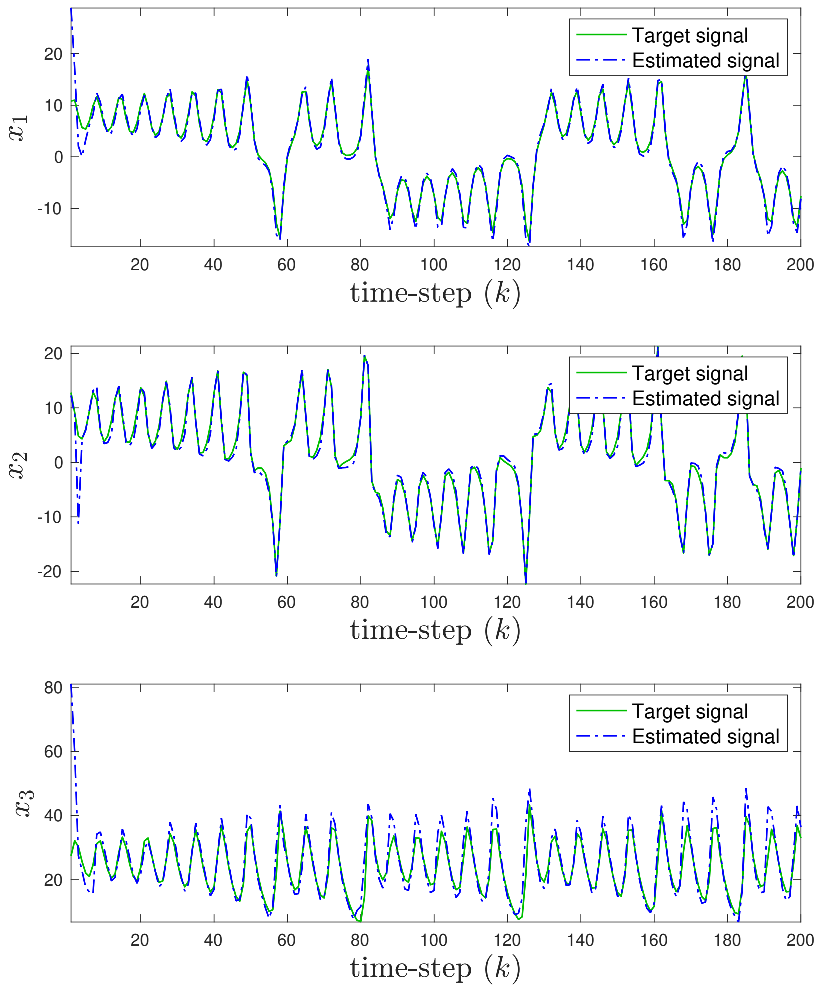


Figure 4.4: Estimation of the time-series using an ESN with an EnKF for the Lorenz system

for example if the testing ohase does not start immediately after the training phase. A free-running ESN can only predict the time series with a sufficiently accurate initial condition.

The comparison of the $\mathcal{L}_2$ error between a free-running ESN predictor and the ESN-EnKF driven by sparse measurements is depicted in Fig. 4.5. It is evident that an ESN with sparse measurement performs significantly better than its free-running counterpart.

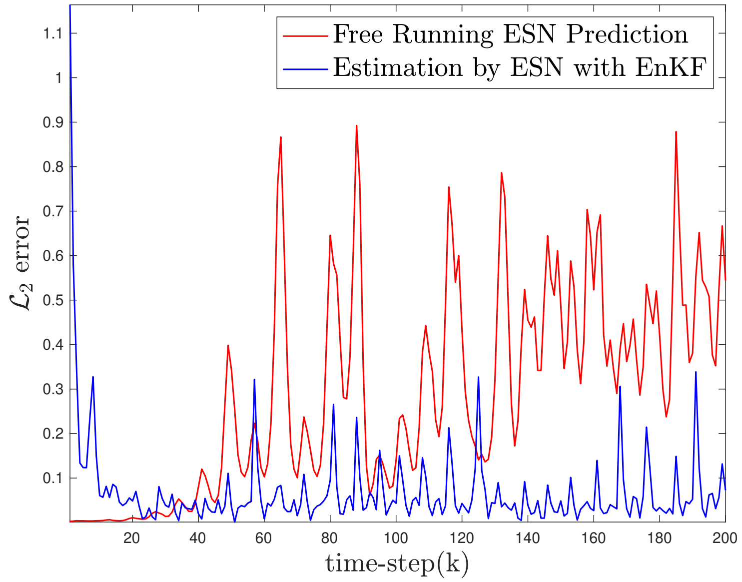


Figure 4.5: Comparison of $\mathcal{L}_2$ error between a free-running ESN and an ESN with an ensemble Kalman filter. Note the EnKF does not require the initial condition to be known.

4.3.2 Lorenz-96 model

Next, the ESN-EnKF estimation algorithm is tested with the Lorenz-96 model [77], a spatially correlated high-dimensional chaotic system developed by E. N. Lorenz in 1996 to describe the variation of an atmospheric quantity of interest, such as temperature and vorticity, at discrete locations on a periodic lattice representing a latitude circle on the earth. The state variables are coupled spatially and their equations of motion include contributions relevant to fluid systems including a quadratic nonlinearity, dissipation, and a constant external forcing. This model has been widely used as a model for atmospheric prediction and the study of spatiotemporal chaos. Mathematically, Lorenz-96 is a linear

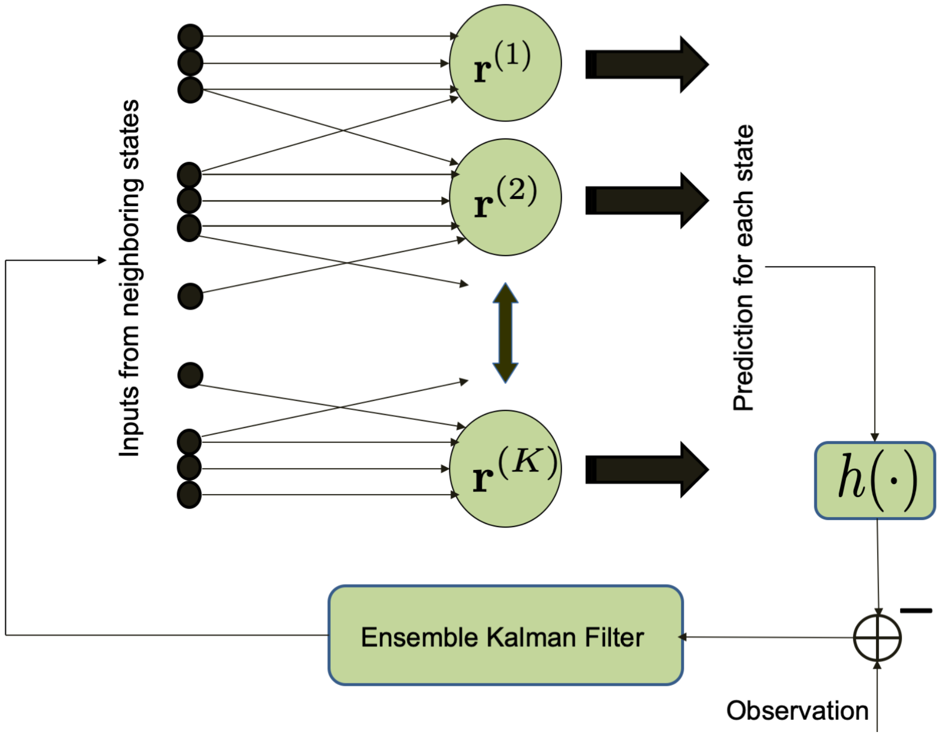


Figure 4.6: The parallel ESN scheme with ensemble Kalman filter

lattice of K variables, where the the dynamics of the i^{th} variable is

$$\dot{x}_i = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \tag{4.7}$$

assuming $x_{-1} = x_{K-1}$, $x_0 = x_K$, and $x_{K+1} = x_1$. The parameter F is a forcing constant with $F = 8$ being a common value causing chaotic behavior.

Let $K = 40$ be the total number of lattice points; which can be thought of as the sensors on a latitude circle that measure the atmospheric quantity of interest. For a large number of state variables the size of the ESN reservoir required to predict the system using a single reservoir must also be large. This makes the single reservoir prediction intractable for the large values of K. In order to mitigate the problem, the local nature of

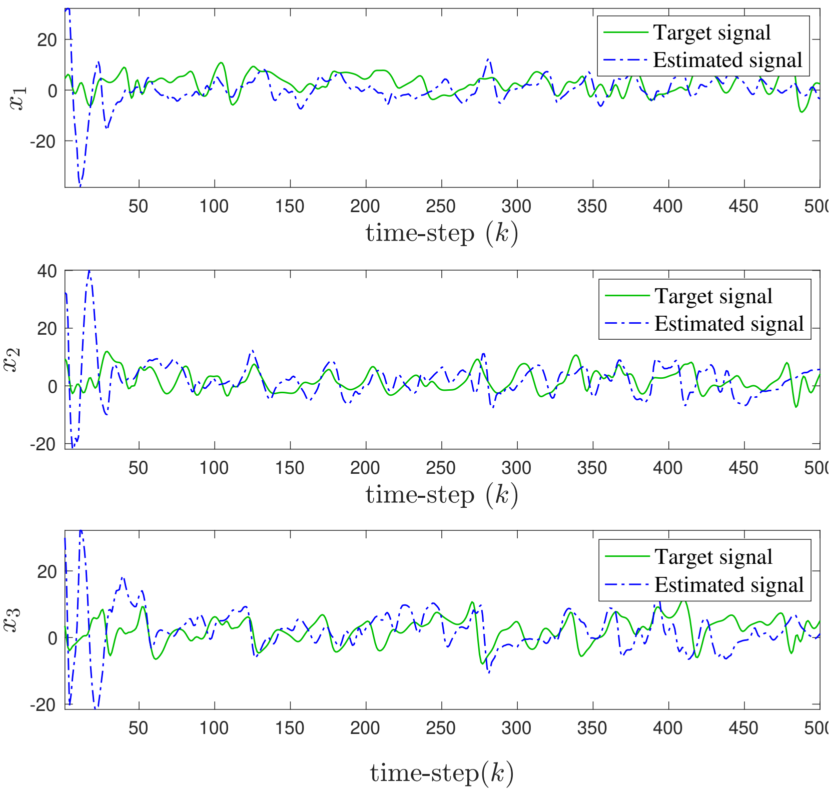


Figure 4.7: Lorenz-96 time series estimation for first three states, which are not measured during the testing phase

the interactions among the state variables x_i in (4.7) is utilized. From (4.7), x_i depends only on its neighbors x_{i-2} to x_{i+1} as in [38]. A parallel set of ESNs is used, each of which predicts the state-variable x_i. ESN i takes input from the states x_{i-2} to x_{i+1} and produces the prediction for x_i. This scheme is depicted in Fig. 4.6.

The ESNs are trained for $N = 2000$ time-steps. Each of these ESNs can be trained in parallel, thereby reducing the computation time. We have compared the performance of the proposed estimation algorithm with a free-running parallel ESN scheme. Here, a random 50% of the lattice points are assumed measurable during the Kalman filter update. The measurements are corrupted by additive i.i.d. zero mean Gaussian noise

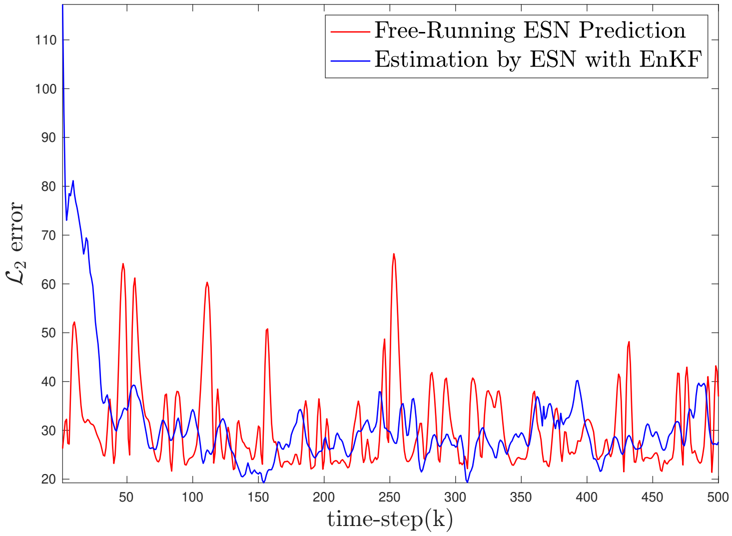


Figure 4.8: Lorenz-96 $\mathcal{L}_2$ error comparison between free-running parallel ESNs with and without an Ensemble Kalman filter

with covariance 0.01 and assimilated by an ensemble Kalman filter with an ensemble size of 100. A comparison between some of the estimated time-series signal and the true data is shown in Fig. 4.7. The $\mathcal{L}_2$ error comparison between the free-running parallel ESN structure and the ESN-EnKF with sparse measurements is presented in Fig. 4.8. The ESN-EnKF algorithm is further compared with the reservoir observer, a model-free prediction scheme for unmeasured variables [2]. In the reservoir observer, a subset of the state variables is measured and fed into the parallel ESNs at each step, but without the measurement update step used in the ensemble Kalman filter. The average correlation between the true and estimated time series with 20 independent Monte-Carlo trials for both the proposed algorithm and the reservoir observer is demonstrated in Fig. 4.9. The proposed algorithm significantly outperforms the reservoir observer, especially when only a moderate number of states are observable.

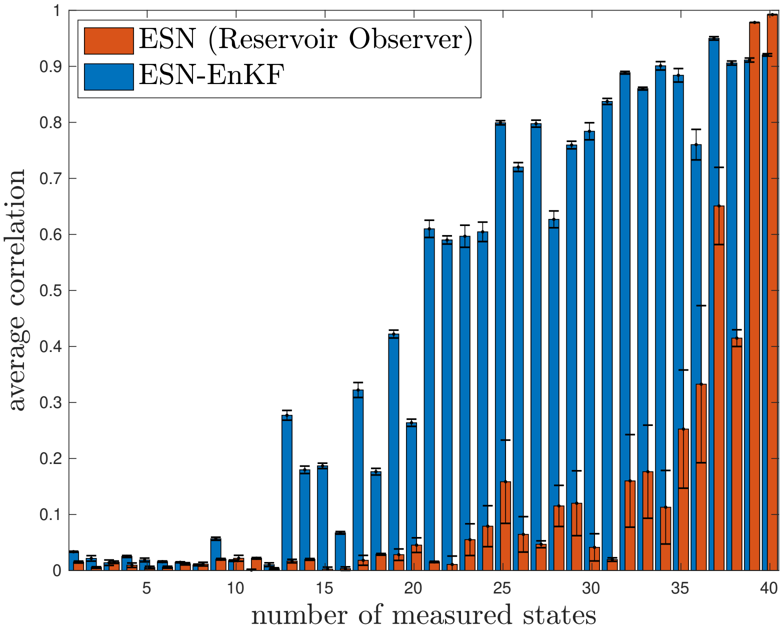


Figure 4.9: Correlation between estimated data and the actual data for the Lorenz-96 model with error-bars: estimation by parallel ESNs with an ensemble Kalman filter has a higher average correlation then the prediction with parallel ESNs by a reservoir observer [2].

Remark 4.6. Lorenz-96 is an example where an ESN-based approach is greatly improved by assimilating sparse measurements through an ensemble Kalman filter. This insight has applications in atmospheric and oceanic data assimilation where the sensor measurements are distributed sparsely in time.

4.3.3 Prediction of traffic congestion in a road network

The proposed algorithm is now applied to a data set of traffic counts obtained from Numina sensors [76] at five different road-intersections on the University of Maryland campus. Fig. 4.11 presents the road network along with the sensor locations. Each

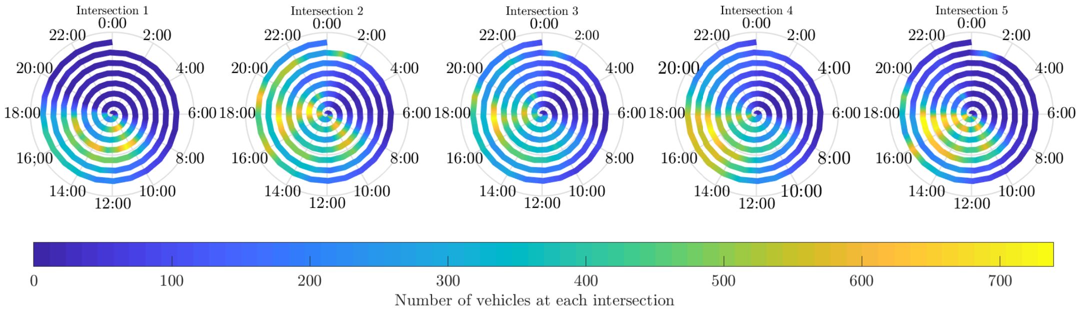


Figure 4.10: Traffic congestion pattern of five intersections over a single week, Each revolution denotes a day of the week with times marked as angles; the number of vehicles is denoted by the colormap. The daily pattern of peak congestion between mornings and afternoons is evident.

sensor counts the number of vehicles, pedestrians, and bicycles at those intersections and records them in a central server. An ESN of 1000 reservoirs are trained from hourly traffic congestion data (total number of vehicles) in all five of these intersections for two months. The training and testing timeline is presented in Fig. 4.12. Since the number of vehicles is non-negative, the activation function ψ is modified to be a rectified tanh function with the negative part set to zero. The network is then tested for a week with only one sensor active. The ensemble Kalman filter is modified to have positive ensemble members only. The estimates are rounded to the nearest positive integer. Fig. 4.13 shows the traffic congestion estimator's performance. The network is able to predict the daily variation of the traffic and the congestion peaks quite well. The algorithm is also compared against the reservoir observer [2] with different numbers of available sensors. The average correlation between the estimated and true traffic congestion time series is computed for 20 independent Monte-Carlo trials and presented in Fig. 4.14 for both the proposed predictor and the reservoir observer that uses an ESN only without the Kalman

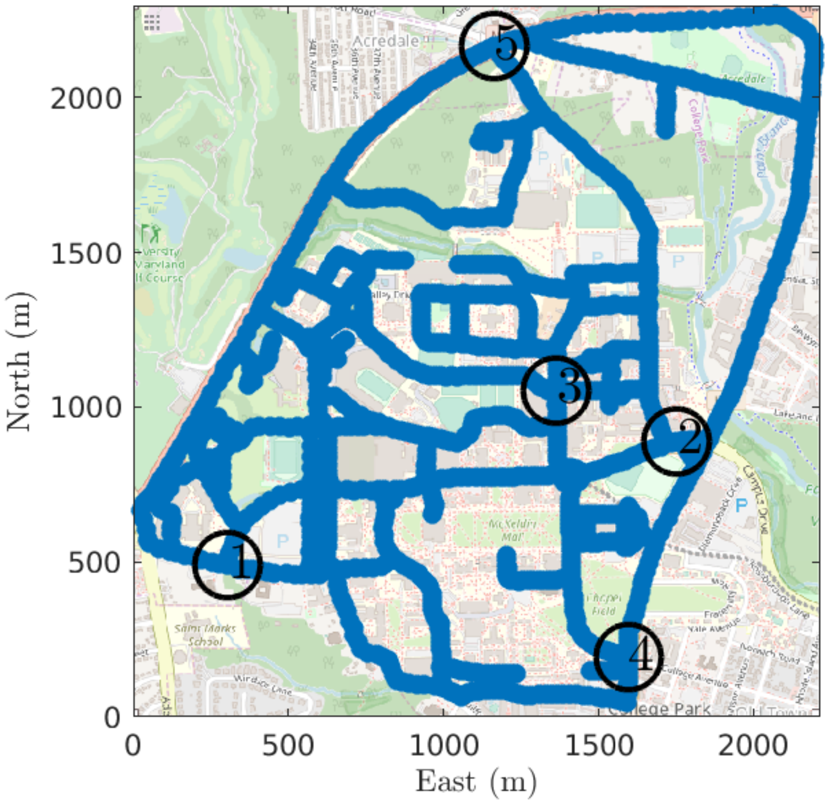


Figure 4.11: University of Maryland road network with Numina sensors

filter measurement update. The Kalman-filter-driven ESN has significantly higher average correlation for partially observable cases.

Remark 4.7. The measurement noise for the traffic sensors are not Gaussian since they can only report positive integer values, which may account for the relatively large prediction error while the congestion is low.

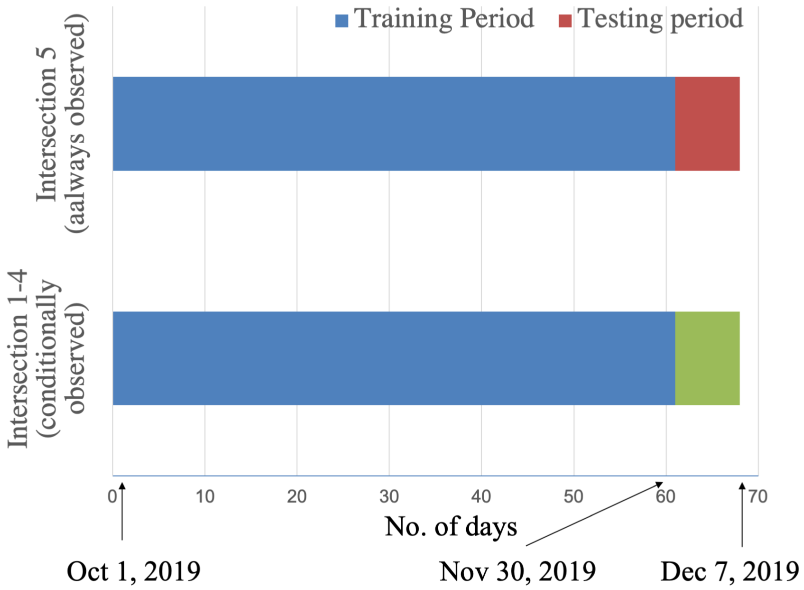


Figure 4.12: Schematic diagram of traffic data training and testing. Red intersections are always observed and green intersections are conditionally observed.

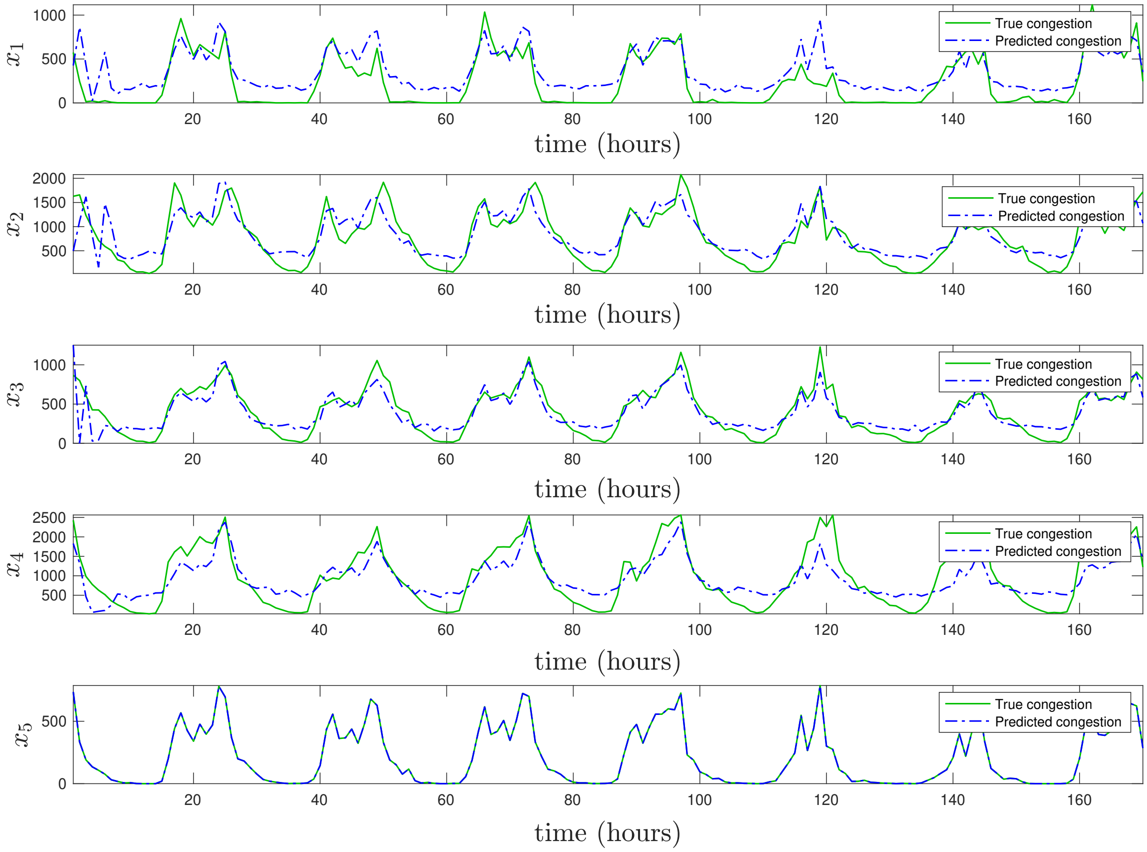


Figure 4.13: Estimation of traffic congestion in all five nodes with only the fifth one observed (x_i denotes the estimated number of vehicles at the i^{th} intersection.)

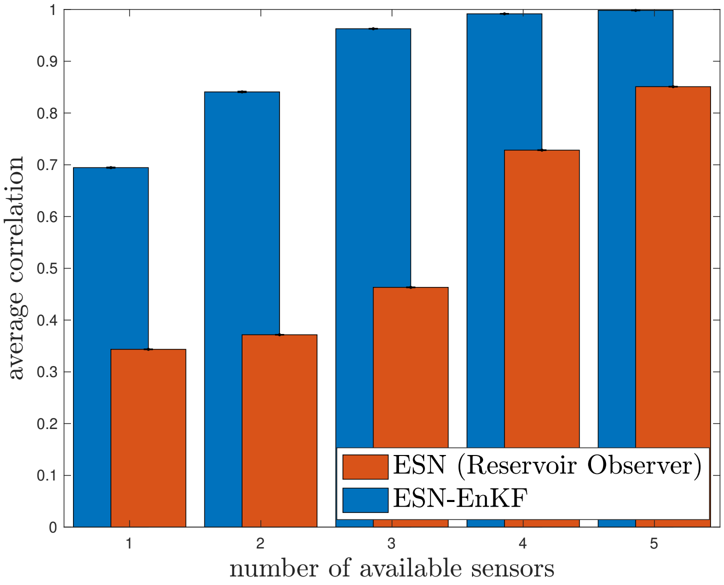


Figure 4.14: Correlation between estimated and actual data for traffic congestion: estimation by ESN with an ensemble Kalman filter has a higher average correlation than the prediction with an ESN by a reservoir observer [2].

Chapter 5

Bilinearization, controllability, and optimal control of a control affine system: a Koopman spectral approach

We present a data-driven modeling and control-design of a control-affine system in this chapter to complement the data-driven estimation algorithm presented in chapter 4. A Koopman theoretic framework is utilized for modeling and control of a control-affine system here.

Operator-theoretic methods essentially work by embedding finite-dimensional dynamics in an infinite-dimensional function space in which functions evolve under a linear operator. While in chapter 3 we discussed the density transport by approximating the Perron-Frobenius (PF) chapter, this chapter makes an effort to utilize its dual, tha Koopman operator to control a nonlinear system. The spectral property of the Koopman operator is well analyzed, see, e.g., [41]. The Koopman operator offers effective methods to characterize a nonlinear system in terms of stability [43] and linearization [44].

The application of the Koopman method to actuated systems has proven to be difficult, because the Koopman operator changes its spectral properties with the actuation signals. Proctor et al. introduces a method [45] to incorporate the control input in Koopman framework by redefining the Koopman operator with two arguments. Williams et al. [42] bridges the gap of the analysis and simulation by providing a method to determine the spectral property of the Koopman operator of the underlying unforced system from the data of the actuated dynamics. However, neither of these methods permit the linearization of full-scale actuated systems with Koopman spectra.

Surana et al. [46] proposed a framework for designing an observer for a discrete-time unactuated nonlinear system. The same framework is extended into continuous time with control-affine dynamics in [47]. Reference [47] introduces the Koopman canonical transform (KCT) using Koopman eigenfunctions, which transforms the (nonlinear) dynamics into an observer form. This chapter utilizes the KCT to transform a control-affine nonlinear system into a bilinear one and derives the sufficient conditions for such a reduction. The bilinearization thus obtained is global and does not rely on the neighbourhood of the operating point or trajectory. An approximate bilinearization technique is also derived when the sufficient conditions do not hold. The bilinear system is then used for controllability analysis and designing a stabilizing control based on the Myhill semigroup theory and Lie-algebraic methods described in [78], [79]. The bilinear representation is then used to derive an optimal control for the original system using Pontryagin's Principle [80]. This method does not require knowledge of the drift vector field since the KCT may be derived with time-series data from the open-loop system.

5.1 Koopman overview and motivation

Consider a dynamical system on a compact manifold $\mathbb{X} \subseteq \mathbb{R}^d$,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \tag{5.1}$$

where $\mathbf{x} \in \mathbb{X}$ and $\mathbf{f} : \mathbb{X} \to \mathbb{X}$. Let $\Phi(t, \mathbf{x})$ be the flow map of the system (5.1). Let $\mathcal{F}$ be the space of all complex-valued observables $\varphi : \mathbb{X} \to \mathbb{C}$. The continuous time Koopman operator is defined as $\mathcal{K}^t : \mathcal{F} \to \mathcal{F}$ such that

$$(\mathcal{K}^t \varphi)(\cdot) = \varphi \circ \Phi(t, \cdot). \tag{5.2}$$

Unlike the original system, the Koopman operator is linear over its arguments, i.e., observable functions, and therefore can be characterized by its eigenvalues and eigenfunctions. A function $\phi : \mathbb{X} \to \mathbb{C}$ is an eigenfunction of $\mathcal{K}^t$ if

$$(\mathcal{K}^t \phi)(\cdot) = e^{\lambda t} \phi(\cdot), \tag{5.3}$$

with eigenvalue $\lambda \in \mathbb{C}$. The infinitesimal generator of $\mathcal{K}^t$, i.e., $\lim_{t \to 0} \dfrac{\mathcal{K}^t - I}{t}$, is $\mathbf{f} \cdot \nabla = L_{\mathbf{f}}$ [43], where $L_{\mathbf{f}}$ is the Lie derivative with respect to $\mathbf{f}$. The infinitesimal generator satisfies the eigenvalue equation

$$L_{\mathbf{f}}\phi = \lambda\phi. \tag{5.4}$$

Hence, the time-varying observable $\psi(t, \mathbf{x}) \triangleq \mathcal{K}^t\varphi(\mathbf{x})$ is the solution of the Partial Differential Equation (PDE) [43]

$$\begin{aligned} \frac{\partial\psi}{\partial t} &= L_{\mathbf{f}}\psi, \\ \psi(0, \mathbf{x}) &= \varphi(\mathbf{x}). \end{aligned} \tag{5.5}$$

In spite of its linearity, the Koopman operator is infinite dimensional, and has an infinite number of eigenfunctions. In fact, if ϕ_1 and ϕ_2 are eigenfunctions of $\mathcal{K}^t$ with eigenvalues λ_1 and λ_2, respectively, then $\phi_1^k\phi_2^l$ is also an eigenfunction with eigenvalue $k\lambda_1 + l\lambda_2$ for any $k, l \in \mathbb{N}$. Moreover, the Koopman operator, being infinite dimensional, may contain continuous and residual spectra with a generalized eigendistribution [31]. However, the discussions in this paper are restricted to the point spectra of the Koopman operator.

Let $\mathbf{g}(\cdot) \in \mathcal{F}^p$, $p \in \mathbb{N}$, be a vector-valued observable. The observable $\mathbf{g}$ can be expressed in terms of Koopman eigenfunctions $\phi_i(\cdot)$ as follows:

$$\mathbf{g}(\cdot) = \sum_{i=1}^{\infty} \phi_i(\cdot)\mathbf{v}_i^{\mathbf{g}}, \tag{5.6}$$

where $\mathbf{v}_i^{\mathbf{g}} \in \mathbb{R}^p, i = 1, 2, \ldots,$ are called the *Koopman modes* of the observable $\mathbf{g}(\cdot)$. Koopman modes form the projection of the observable on the span of Koopman eigenfunctions [81]. The Koopman eigenvalues and eigenfunctions are properties of the dynamics only, whereas the Koopman modes depend on the observable.

5.1.1 Motivation for bilinearization: The function space

Consider the control affine system

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}) + \sum_{i=1}^{m} \mathbf{g}_i(\mathbf{x}) u_i \\ \mathbf{y} &= \mathbf{h}(\mathbf{x}), \end{aligned} \tag{5.7}$$

where $\mathbf{x} \in \mathbb{X} \subseteq \mathbb{R}^d$, $u_i \in \mathbb{R}$, for $i = 1, \ldots, m$ and $\mathbf{y} \in \mathbb{R}^p$. Here the control input enters only through the control vector fields $\mathbf{g}_i$. The objective is to transform the system into an appropriate basis in higher or possibly infinite dimension, so that the drift and control vector fields become linear. Let $\psi(t, \mathbf{x})$ be defined as in (5.5). Note that $\psi(t, \mathbf{x})$ gives the evolution of the observable quantity with time along the trajectory. Applying (5.5) to the system (5.7), the evolution PDE is given by

$$\begin{aligned} \frac{\partial \psi}{\partial t} &= L_{\mathbf{f}} \psi + \sum_{i=1}^{m} u_i L_{\mathbf{g}_i} \psi, \\ \psi(0, \mathbf{x}) &= \varphi(\mathbf{x}), \end{aligned} \tag{5.8}$$

where $L_{\mathbf{g}_i} \triangleq \mathbf{g}_i \cdot \nabla$, $i = 1, \ldots, m$ are the corresponding Lie derivatives and hence are linear operators on the space of ψ. The PDE system (5.8) appears qualitatively similar to the standard bilinear system, i.e., $\dot{\mathbf{x}} = A\mathbf{x} + \sum_{i=1}^{m} B_i \mathbf{x} u_i$. The system (5.8) differs only in the fact that $L_{\mathbf{f}}$ and $L_{\mathbf{g}_i}$ are infinite dimensional operators operating over function space. The idea here is to choose a suitable collection of scalar observables to transform the original system into a new state space to get a bilinear form (possibly in infinite dimensions) and then to project these Lie derivative operators on a finite-dimensional subspace to get a finite-dimensional bilinear system that approximates the original system (5.7).

5.1.2 Koopman canonical transform

In order to choose a suitable collection of observables that reliably describes the original system, we seek basis functions on which we can project a wide variety of vector-valued observables. A natural choice is to use the eigenfunctions of the Koopman operator, since these functions, when operated on by the Koopman infinitesimal generator, get multiplied by a scalar only. For this transformation, we use the Koopman canonical transform (KCT) defined in [47]. The KCT relies on the point spectra of the Koopman operator related to the drift vector field, and it suffices because the continuous spectrum is empty for most systems near an attractor [43]. For the system (5.7) we investigate the Koopman eigenvalues and eigenfunctions of the unactuated dynamics, i.e.,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \tag{5.9}$$

and the flow associated with it. Let $\lambda_i, \phi_i(\cdot)$ for $i = 1, 2, \ldots$ be the eigenvalue-eigenfunction pairs of the Koopman operator associated with the system (5.9). Surana [47] introduced the KCT to transform the dynamics (5.7) using the eigenfunctions ϕ_i in possibly a higher-dimensional space. To enable us to use KCT, [47] mentions the following assumption.

Assumption 5.1. $\exists\ \phi_i, i = 1, 2, \ldots, n$, such that

$$\mathbf{x} = \sum_{i=1}^{n} \phi_i(\mathbf{x})\mathbf{v}_i^{\mathbf{x}}, \;\; \mathbf{h}(\mathbf{x}) = \sum_{i=1}^{n} \phi_i(\mathbf{x})\mathbf{v}_i^{\mathbf{h}},$$

where $\mathbf{v}_i^{\mathbf{x}} \in \mathbb{C}^d$ and $\mathbf{v}_i^{\mathbf{h}} \in \mathbb{C}^p$.

This assumption implies that the state vector and the output function can be described in terms of a finite number of Koopman eigenfunctions. With sufficiently large n, Assumption 5.1 is likely to be satisfied. Otherwise $\mathbf{x}$ and $\mathbf{h}(\mathbf{x})$ can be well approximated by n eigenfunctions as in the case of a Fourier series.

The KCT is defined by the transformation $T(\mathbf{x})$ [47]:

$$\begin{aligned} T(\mathbf{x}) &= [\tilde{\phi}_1(\mathbf{x}), \ldots, \tilde{\phi}_n(\mathbf{x})]^T \\ \tilde{\phi}_i(\mathbf{x}) &= \phi_i(\mathbf{x}), \text{ if } \phi_i : \mathbb{X} \to \mathbb{R} \\ (\tilde{\phi}_i(\mathbf{x}), \tilde{\phi}_{i+1}(\mathbf{x}))^T &= (2Re(\phi_i(\mathbf{x})), -2Im(\phi_i(\mathbf{x})))^T, \\ & \quad \text{if } \phi_i : \mathbb{X} \to \mathbb{C} \end{aligned} \tag{5.10}$$

Following the transformation $\mathbf{z} = T(\mathbf{x})$, the system (5.7) in the new coordinates becomes [47]

$$\begin{aligned} \mathbf{x} &= C^{\mathbf{x}}\mathbf{z}, \\ \dot{\mathbf{z}} &= D\mathbf{z} + \sum_{i=1}^{m} L_{\mathbf{g}_i} T(\mathbf{x}) u_i|_{\mathbf{x}=C^{\mathbf{x}}\mathbf{z}}, \\ \mathbf{y} &= C^{\mathbf{h}}\mathbf{z}, \end{aligned} \tag{5.11}$$

where $C^{\mathbf{x}} = [\tilde{\mathbf{v}}_1^{\mathbf{x}}|\ldots|\tilde{\mathbf{v}}_n^{\mathbf{x}}]$ and $C^{\mathbf{h}} = [\tilde{\mathbf{v}}_1^{\mathbf{h}}|\ldots|\tilde{\mathbf{v}}_n^{\mathbf{h}}]$ with $\tilde{\mathbf{v}}_i^{\mathbf{x}} = \mathbf{v}_i^{\mathbf{x}}$, if ϕ_i is real-valued, and

$[\tilde{\mathbf{v}}_i^{\mathbf{x}}, \tilde{\mathbf{v}}_{i+1}^{\mathbf{x}}] = [\operatorname{Re}\mathbf{v}_i^{\mathbf{x}}, \operatorname{Im}\mathbf{v}_i^{\mathbf{x}}]$, if ϕ_i is complex-valued. $\tilde{\mathbf{v}}_i^{\mathbf{h}}$ are defined similarly. $D \in \mathbb{R}^{n\times n}$ is a block diagonal matrix with diagonal entry $D_{i,i} = \lambda_i$ if ϕ_i is a real-valued eigenfunction, or $\begin{bmatrix} D_{i,i} & D_{i,i+1} \\ D_{i+1,i} & D_{i+1,i+1} \end{bmatrix} = |\lambda_i| \begin{bmatrix} \cos(\angle\lambda_i) & \sin(\angle\lambda_i) \\ -\sin(\angle\lambda_i) & \cos(\angle\lambda_i) \end{bmatrix}$ if ϕ_i is complex.

Remark 5.1. Although Assumption 5.1 is stated as a requirement in [47] to express the state $\mathbf{x}$ and the observation $\mathbf{h}(\mathbf{x})$ in a closed form by the Koopman eigenfunctions, it will be shown later that we can drop this assumption and, for most systems, can express any continuous function of the state $\mathbf{x}$ arbitrarily accurately by the linear combination of Koopman eigenfunctions.

5.2 Koopman-induced bilinearization of a control affine system

To establish the bilinearizability of the system (5.11), we need to analyze the control vector fields of the original system. In the transformed system, the control enters through the transformed vector field $L_{\mathbf{g}_i}T(\mathbf{x})|_{\mathbf{x}=C^{\mathbf{x}}\mathbf{z}}$. Note that $L_{\mathbf{g}_i}$ is the infinitesimal Koopman operator with respect to control vector field $\mathbf{g}_i$.

5.2.1 Bilinearizability of the Koopman Canonical Transform

Definition 5.1. An *invariant subspace* of a linear mapping $T : V \to V$ from a vector space V to itself is a subspace $W \subseteq V$ such that $T(W) \subseteq W$.

Theorem 5.1. The system (5.11) (hence the system (5.7) as well) is bilinearizable in a countable (possibly infinite) basis if the eigenspace of $L_{\mathbf{f}}$, i.e., the Koopman operator corresponding to the drift vector field is an invariant subspace of $L_{\mathbf{g}_i}, i = 1, \ldots, m$, i.e., the Koopman operators related to the control vector fields.

Proof. If the hypothesis is true, then we can choose eigenfunctions of $L_{\mathbf{f}}$, $\{\phi_j : j = 1, 2, \ldots\}$, such that $L_{\mathbf{g}_i}\phi_k \in \text{span}\{\phi_j : j = 1, 2, \ldots\}$, $\forall\, i = 1, \ldots, m$, $k = 1, 2, \ldots$. This choice is guaranteed, because $\text{span}\{\phi_j : j = 1, 2, \ldots\}$, i.e., the eigenspace of $L_{\mathbf{f}}$, is invariant under $L_{\mathbf{g}_i}$, $i = 1, \ldots, m$. So, $\forall\, k = 1, 2, \ldots$, we have $L_{\mathbf{g}_i}\phi_k = \sum_{j=1}^{\infty} v_j^{\mathbf{g}_i}\phi_j$, where $v_j^{\mathbf{g}_i} \in \mathbb{R}$. Now taking $T(\mathbf{x})$ as in Eq. (5.10) but without imposing the finite n condition, we get

$$L_{\mathbf{g}_i}T(\mathbf{x}) = \sum_{j=1}^{\infty} \mathbf{v}_j^{\mathbf{g}_i}\phi_j(\mathbf{x}) = \sum_{j=1}^{\infty} \tilde{\mathbf{v}}_j^{\mathbf{g}_i}\tilde{\phi}_j(\mathbf{x}),$$

where $\tilde{\mathbf{v}}_j^{\mathbf{g}_i} \in \mathbb{R}^d$ and $\tilde{\phi}_j$ are defined as in Eq. (5.10). Now define $B_i^{\infty} = [\tilde{\mathbf{v}}_1^{\mathbf{g}_i}|\tilde{\mathbf{v}}_2^{\mathbf{g}_i}|\ldots]$. Then, with $\mathbf{z} = T(\mathbf{x})$, the system (5.11) can be expressed as

$$\dot{\mathbf{z}} = D\mathbf{z} + \sum_{i=1}^{m} B_i^{\infty}\mathbf{z}u_i, \tag{5.12}$$

which completes the proof. □

Although Theorem 5.1 gives the condition for the bilinearizability of the control-affine system using KCT with a countable number of eigenfunctions, it still does not solve the problem with infinitely many eigenfunctions. However, for an approximate result we can truncate the number of eigenfunctions to only the dominant ones. This linear approximation, unlike the Jacobian approach, is global, i.e., it is valid over the manifold $\mathbb{X}$ on which the dynamics (5.7) is defined.

Corollary 5.1.1. The systems (5.7) and (5.11) are bilinearizable if the drift vector field $\mathbf{f} \equiv 0$, i.e., if the system (5.7) a pure control-affine system.

Proof. The proof follows from the fact that every function $\phi(\cdot) \in \mathcal{F}$ is an eigenfunction of $L_{\mathbf{f}}$ with $\mathbf{f} \equiv 0$ corresponding to the zero eigenvalue. Hence the whole space $\mathcal{F}$ is the eigenspace of $L_{\mathbf{f}}$, which is of course invariant under $L_{\mathbf{g}_i}$, $\forall\, i = 1, \dots, m$. Therefore, from Theorem 5.1, the system is bilinearizable. □

Theorem 5.1 and Corollary 1 essentially embed the finite-dimensional nonlinear dynamics (5.7) in a higher, possibly infinite-dimensional linear system (5.12). There are other embedding techniques that deal with Hermite polynomials, e.g., Carleman embedding [82], but that technique works only on analytic nonlinearities. The method with Koopman eigenfunctions works on a wide varieties of systems, and can be characterized in terms of the range and eigenspace of the corresponding Koopman operator.

For a finite-dimensional bilinearization of the system (5.7), we need a stronger assumption than invariance of the eigenspace of $L_{\mathbf{f}}$. The invariant subspace must be spanned by a finite number of Koopman eigenfunctions, which is the statement of Theorem 5.2.

Theorem 5.2. Suppose $\exists\, \{\phi_j : j = 1, \dots, n\}$, $n \in \mathbb{N}$, $n < \infty$ such that ϕ_j, $j = 1, \dots, n$ are the Koopman eigenfunctions of the unactuated system (5.9) and $\operatorname{span}\{\phi_1, \dots, \phi_n\}$ forms an invariant subspace of $L_{\mathbf{g}_i}$, $i = 1, \dots, m$. Then the system (5.7) and, in turn system (5.11), are bilinearizable with an n dimensional state space.

Proof. The hypothesis dictates that $L_{\mathbf{g}_i}\phi_k \in \operatorname{span}\{\phi_j : j = 1, \dots, n\}\, \forall\, i = 1, \dots, m$, $k =$

$1, \ldots, n$. Therefore, we conclude

$$L_{\mathbf{g}_i}\phi_k = \sum_{j=1}^{n} v_j^{\mathbf{g}_i}\phi_j,\ \ k = 1, \ldots, n,$$

where $v_j^{\mathbf{g}_i} \in \mathbb{R}$. Now consider $T(\mathbf{x})$ as defined in Eq. (5.10). Its Lie derivatives with respect to the control vector fields are

$$L_{\mathbf{g}_i}T(\mathbf{x}) = \sum_{j=1}^{n} \mathbf{v}_j^{\mathbf{g}_i}\phi_j(\mathbf{x}) = \sum_{j=1}^{n} \tilde{\mathbf{v}}_j^{\mathbf{f}_i}\tilde{\phi}_j(\mathbf{x}),$$

where $\tilde{\mathbf{v}}_j^{\mathbf{g}_i} \in \mathbb{R}^d$ and $\tilde{\phi}_j$ are defined as in Eq. (5.10). Now, as in Theorem 5.1, define $B_i \triangleq [\tilde{\mathbf{v}}_1^{\mathbf{g}_i}|\tilde{\mathbf{v}}_2^{\mathbf{g}_i}|\ldots|\tilde{\mathbf{v}}_n^{\mathbf{g}_i}]$. The difference from the B_i^∞ in the proof of Theorem 5.1 is that B_i is not only countable but also a finite-dimensional matrix. Now with transformed coordinate $\mathbf{z} = T(\mathbf{x})$, the transformed system looks like

$$\dot{\mathbf{z}} = D\mathbf{z} + \sum_{i=1}^{m} B_i\mathbf{z}u_i, \tag{5.13}$$

with $\mathbf{z} \in \mathbb{R}^n$, $n < \infty$. $\square$

Though the hypothesis of Theorem 5.2 is difficult to satisfy, we can always include more eigenfunctions ϕ_j in the span so that $\|L_{\mathbf{g}_i}\phi_j - \sum_{j=1}^{n} v_j^{\mathbf{g}_i}\phi_j\|$ becomes sufficiently small to satisfy the invariance condition of Theorem 5.1. Note that usually $n \gg d$, i.e., this method of bilinearization lifts the original dynamics (5.7) to a much higher dimensional state space. The resulting bilinear system is easier to work with in terms of controllability analysis and designing a stabilizing control than the original nonlinear system. The bilinear system defined by (5.13) will be refered as the Koopman bilinear form (KBF) in the sequel.

5.2.2 Approximate bilinearization of the Koopman canonical transform

Under the condition of Theorem 5.2, we have

$$L_{\mathbf{g}_i} T(\mathbf{x}) = B_i T(x),\ i = 1 \ldots, m, \tag{5.14}$$

where $B_i \in \mathbb{R}^{n \times n}$. If this condition is not satisfied, we need to approximate B_i with respect to a suitable metric. For example, either a least-squares technique based on time series data might be used or an analytical tool can be developed, if the control vector fields are known. In this section, the latter approach yields a closed-form approximation of B_i that depends on the control vector fields.

Remark 5.2. Although we assume that the control vector fields are known, the technique can be carried out without knowledge of the drift vector field. The time series data for the unactuated system may be used to approximate the Koopman eigenfunctions ϕ_i that are necessary for constructing the KCT.

To determine B_i for an approximate bilinear system, we use the $\mathcal{L}^2$ functional norm. Since the control vector fields and the Koopman eigenfunctions might not belong to $\mathcal{L}^2_m(\mathbb{X})$, we truncate the state-space $\mathbb{X}$ into a set $\mathbf{X}$ of finite measure so that $m(\mathbf{X}) < \infty$, where $m : \mathbb{X} \longrightarrow \mathbb{R}^+$ is the Lebesgue measure.

Assumption 5.2. Let $\phi_i \in \mathcal{L}_m^\infty(\mathbf{X})$ for $i = 1, \ldots, n$ and $g_i \in \mathcal{L}_m^\infty(\mathbf{X})$ for $i = 1, \ldots, m$.

Under this assumption, $\phi_i \in \mathcal{L}_m^2(\mathbf{X})$ and $g_i \in \mathcal{L}_m^2(\mathbf{X})$, since we have a finite measure space.

Remark 5.3. The $\mathcal{L}^2$ norm is chosen specifically because the space of $\mathcal{L}^2(\mathbf{X})$ functions is a Hilbert space that yields a inner-product of the form

$$\langle \psi_1, \psi_2 \rangle_{\mathcal{L}_m^2(\mathbf{X})} \triangleq \int\limits_{\mathbf{X}} \psi_1 \psi_2 dm,$$

where $\psi_1, \psi_2 \in \mathcal{L}_m^2(\mathbf{X})$.

We pose the approximation of B_i as the following quadratic optimization problem:

$$\begin{gathered} \underset{B_i}{\text{minimize}} \, \| L_{\mathbf{g}_i} T(\mathbf{x}) - B_i T(x) \|^2_{\mathcal{L}_m^2(\mathbf{X})} \\ \text{for } i = 1, \ldots, m. \end{gathered} \tag{5.15}$$

The optimization problem (5.15) is an unconstrained convex quadratic program on B_i. A solution to this problem is obtained analytically in the next theorem.

Theorem 5.3. The optimization problem (5.15) yields a minimum B_i^* characterized as

$$QB_i^{*T} = R_i, \; Q \in \mathbb{R}^{n \times n}, \; R_i \in \mathbb{R}^{n \times n}, \tag{5.16}$$

where

$$Q_{kl} = \langle \tilde{\phi}_k, \tilde{\phi}_l \rangle_{\mathcal{L}_m^2(\mathbf{X})}$$

and

$$(R_i)_{kl} = \left\langle \frac{\partial \tilde{\phi}_l}{\partial \mathbf{x}} \mathbf{g}_i, \tilde{\phi}_k \right\rangle_{\mathcal{L}^2_m(\mathbf{X})}.$$

Proof. Let the j^{th} row of B_i be $\mathbf{b}^j$, i.e., $B_i = [\mathbf{b}_1^T, \ldots, \mathbf{b}_n^T]^T$. Then the cost function becomes

$$\begin{aligned}
J(B_i) &\triangleq \|L_{\mathbf{g}_i} T(\mathbf{x}) - B_i T(x)\|^2_{\mathcal{L}^2_m(\mathbf{X})} \\
&= \int\limits_{\mathbf{X}} \left(\frac{\partial T(\mathbf{x})}{\partial \mathbf{x}} \mathbf{g}_i - B_i T(x) \right)^T \left(\frac{\partial T(\mathbf{x})}{\partial \mathbf{x}} \mathbf{g}_i - B_i T(x) \right) dm \\
&= \sum_{j=1}^{n} \int\limits_{\mathbf{X}} \left(\frac{\partial \tilde{\phi}_j(\mathbf{x})}{\partial \mathbf{x}} \mathbf{g}_i - \mathbf{b}_j T(x) \right)^2 dm \\
&= \sum_{j=1}^{n} J_j(\mathbf{b}_j), \qquad (5.17)
\end{aligned}$$

where $J_j(\mathbf{b}_j) \triangleq \int\limits_{\mathbf{X}} \left(\frac{\partial \tilde{\phi}_j(\mathbf{x})}{\partial \mathbf{x}} \mathbf{g}_i - \mathbf{b}_j T(x) \right)^2 dm$, $j = 1, \ldots, n$. So it suffices to find the minima for each $J_j(\mathbf{b}_j)$. Now for each j, J_j is a convex quadratic function of $\mathbf{b}_j$ that is bounded below by zero. Hence $\mathbf{b}_j^*$ is a minima for $J_j(\mathbf{b}_j)$ if and only if

$$\left. \frac{\partial J_j}{\partial (\mathbf{b}_j)_k} \right|_{\mathbf{b}_j = \mathbf{b}_j^*} = 0, \qquad (5.18)$$

for $k = 1, \ldots, n$, where $(\mathbf{b}_j)_k$ denotes the k^{th} component of the vector $\mathbf{b}_j$.

Solving (5.18) for every $k \in 1, \ldots, n$, we get a system of equations

$$Q \mathbf{b}_j^{*T} = \mathbf{r}_j, \qquad (5.19)$$

where $Q_{kl} = \int\limits_{\mathbf{X}} \tilde{\phi}_k \tilde{\phi}_l dm = \langle \tilde{\phi}_k, \tilde{\phi}_l \rangle_{\mathcal{L}^2_m(\mathbf{X})}$ and $(\mathbf{r}_j)_k = \int\limits_{\mathbf{X}} \frac{\partial \tilde{\phi}_j}{\partial \mathbf{x}} \mathbf{g}_i \tilde{\phi}_k dm = \left\langle \frac{\partial \tilde{\phi}_j}{\partial \mathbf{x}} \mathbf{g}_i, \tilde{\phi}_k \right\rangle_{\mathcal{L}^2_m(\mathbf{X})}$, $k, l \in 1, \ldots, n$. Note that $\mathbf{r}_j$ is a column vector, whereas $\mathbf{b}_j$ is a row vector. Now combining the condition (5.19) for $j = 1, \ldots, n$, we get

$$QB_i^{*T} = R_i, \tag{5.20}$$

where $R_i = [\mathbf{r}_1 | \ldots | \mathbf{r}_n]$. From now on, we will denote the optimal B_i without the asterisk for the sake of convenience. □

Remark 5.4. With the optimal B_i from Theorem 5.3, the bilinear representation of the system (5.7) is

$$\begin{aligned} \mathbf{x} &= C^{\mathbf{x}} \mathbf{z} \\ \dot{\mathbf{z}} &= D\mathbf{z} + \sum_{i=1}^{m} B_i \mathbf{z} u_i \\ \mathbf{y} &= C^{\mathbf{h}} \mathbf{z}. \end{aligned} \tag{5.21}$$

The Koopman bilinear form (5.21) formed from the Theorem 5.3 is available even if the conditions of Theorem 5.2 is not met.

Remark 5.5. Although B_i in (5.21) gives the optimal bilinear representation of the system (5.7) in the truncation $\mathbf{X}$ of the state space $\mathbb{X}$ with respect to the $\mathcal{L}^2$ norm, the truncation impacts the degree of optimality and should be chosen depending on the system.

5.2.3 Convergence of the Koopman bilinear form to the original system

While describing the Koopman bilinear form (5.21), the implicit assumption so far is that there are sufficiently many Koopman eigenfunctions to represent the state, observation equations, and the control vector fields. But the crucial question is whether an arbitrary continuous function of the state can be approximated by the linear combination of the Koopman eigenfunctions. It turns out that it is possible for most cases by using the approach of a non-recurrent surface [83].

Definition 5.2. A set $\Gamma \in \mathbb{X}$ is called non-recurrent with respect to the dynamical flow map $\Phi(t, \mathbf{x})$ and time $T > 0$, if $\mathbf{x} \in \Gamma \implies \Phi(t, \mathbf{x}) \notin \Gamma$, $\forall\, t \in (0, T]$.

Reference [83] shows that for any function $g \in \mathcal{C}(\Gamma)$ and $\lambda \in \mathbb{C}$, an eigenfunction for Koopman operator can be defined as

$$\phi_{\lambda,g}(\mathbf{x}) = e^{-\lambda\tau(\mathbf{x})} g(\Phi(\tau(\mathbf{x}), \mathbf{x})),\ \forall\, \mathbf{x} \in X_T, \tag{5.22}$$

where $\tau(\mathbf{x}) = \inf\limits_{t\in\mathbb{R}}\{t |\ (\mathbf{t}, \mathbf{x}) \in \Gamma\}$ and $X_T = \bigcup\limits_{t\in[0,T]} \Phi(t, \Gamma) = \bigcup\limits_{t\in[0,T]} \{\Phi(t, \mathbf{x}_0) | \mathbf{x}_0 \in \Gamma\}$.

Remark 5.6. The time $\tau(\mathbf{x})$ defines the first instant that the trajectory $\Phi(t, \mathbf{x})$ hits the non-recurrent set Γ backwards starting from $\mathbf{x}$.

Remark 5.7. The functions $g \in \mathcal{C}(\Gamma)$ are called boundary functions [83] and can be viewed as dictionary functions required to generate the set of eigenfunctions.

The next question is whether the set of eigenfunctions $\phi_{\lambda,g}$ are rich enough to represent the state $\mathbf{x}$ and observation function $\mathbf{h}(\mathbf{x})$ by linear combination. The next theorem clarifies this point.

Theorem 5.4. Let $\Lambda \subset \mathbb{C}$ be an arbitrary nontrivial set of the complex numbers and define $L(\Lambda) \triangleq \{\sum_{k=1}^{N} \alpha_k \lambda_k | \lambda_k \in \Lambda, \alpha_k \in \mathbb{N}_0, N \in \mathbb{N}\}$. Let $\Gamma \subset \mathbb{X}$ be a non-recurrent set closed in $\mathbb{R}^d$ and $G \subset \mathcal{C}(\Gamma)$ denote a dense unital subalgebra (i.e., G is closed under multiplication and contains a multiplicative identity) of continuous functions on Γ. Assume that the system state $\mathbf{x}$ does not blow up in $t \in (0, T]$. Then

$$\Phi_{\Lambda,G} \triangleq \{\phi_{\lambda,g} | \lambda \in L(\Lambda), g \in G\}$$

is dense in $\mathcal{C}(X_T)$. Hence for any $\epsilon > 0$ and $\xi \in \mathcal{C}(X_T)$, $\exists\ \phi_1, \dots, \phi_n \in \Phi_{\Lambda,G}$ such that

$$\sup_{\mathbf{x} \in X_T} \left| \xi(\mathbf{x}) - \sum_{i=1}^{n} v_i \phi_i(\mathbf{x}) \right| < \epsilon.$$

Proof. Since there is no finite escape time in $(0, T]$, X_T must be bounded. Now consider a sequence $\mathbf{x}_i \in X_T$ with limit $\mathbf{x}^* \in \mathbb{R}^d$. Therefore $\exists$ a sequence $(t_i, \mathbf{x}_i^0) \in [0, T] \times \Gamma$ such that $\mathbf{x}_i = \Phi(t_i, \mathbf{x}_i^0)$. Since $[0, T] \times \Gamma$ compact, $\exists$ a convergent subsequence that converges to $(t^*, \mathbf{x}^{0^*})$. By the continuity of the flow map, we have $\mathbf{x}^* = \Phi(t, \mathbf{x}^{0^*}) \in X_T$. Hence X_T is closed, which implies it is compact. From (5.22)

$$\phi_{\lambda,g}(\mathbf{x}) = e^{-\lambda \tau(\mathbf{x})} g(\Phi(\tau(\mathbf{x}), \mathbf{x})),\ \forall\, \mathbf{x} \in X_T.$$

Since G is a dense unital subalgebra of continuous functions over a compact set Γ, by the

Stone-Weierstrass theorem [84] it separates points, i.e., $\forall\ \mathbf{x}_0, \mathbf{y}_0 \in \Gamma$, $\mathbf{x}_0 \neq \mathbf{y}_0$, $\exists\ g \in G$ such that $g(\mathbf{x}_0) \neq g(\mathbf{y}_0)$. Furthermore, since trajectories cannot cross, $\forall\ \mathbf{x}, \mathbf{y} \in X_T$, $\mathbf{x} \neq \mathbf{y}$, $\Phi(\tau(\mathbf{x}), \mathbf{x}) \neq \Phi(\tau(\mathbf{y}), \mathbf{y})$, which in turn implies that $g(\Phi(\tau(\mathbf{x}), \mathbf{x})) \neq g(\Phi(\tau(\mathbf{y}), \mathbf{y}))$. This fact, together with $\lambda = 0 \in L(\Lambda)$, readily shows that $\Phi_{\Lambda,G}$ also separates points and contains a constant function. It is also evident from the construction that $\Phi_{\Lambda,G}$ is a unital subalgebra of $\mathcal{C}(X_T)$ with X_T compact. Hence $\Phi_{\Lambda,G}$ is dense in $\mathcal{C}(X_T)$. □

Remark 5.8. The unital subalgebra can be chosen in various ways. The space of polynomials with the identity function is one such choice.

Remark 5.9. Whether a closed non-recurrent set or surface can be found is an interesting question. Reference [83] shows that a $(d-1)$-dimensional closed surface Γ exists if we can find a finite set of points $\hat{\Gamma} = \{\mathbf{x}^1, \ldots, \mathbf{x}^M\}$ with no two points lying on the same trajectory and the flow of (5.1) can be rectified, i.e., there exists a diffeomorphism $h : Y' \subset \mathbb{R}^d \longrightarrow D \subset \mathbb{X}$ through which (5.1) is conjugate to $\dot{\mathbf{y}} = (0, \ldots, 0, 1)^T$. Then $\Gamma \supset \hat{\Gamma}$.

Remark 5.10. The dynamics (5.1) are rectifiable locally in the neighbourhood of a non-singular (i.e., the vector field does not become zero) point [85]. However, this property breaks down when we have a singularity, i.e., a fixed point nearby.

Remark 5.11. In the region of attraction of a asymptotically stable equilibrium point, the flow is not rectifiable. But the existence of a non-recurrent surface can readily be demonstrated with the level curves of the Lyapunov function by the converse Lyapunov theorem [86]. The same is true for an unstable equilibrium with no stable manifold.

5.3 Data-Driven Construction of the Koopman Bilinear Form

Until now, analytical representations of the Koopman eigenfunctions ϕ_i for the drift vector-field have been assumed. But in practical cases, the Koopman eigenfunctions and eigenmodes may need to be calculated from data. For this purpose, consider extended dynamic mode decomposition [32] is used here. For the autonomous system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$, time-series data is generated from the experiment or simulation and arranged in snapshot pairs with time interval Δt as follows:

$$\begin{aligned} \overline{X} &= [\mathbf{x}_1, \ldots, \mathbf{x}_N] \\ \overline{Y} &= [\mathbf{y}_1, \ldots, \mathbf{y}_N], \end{aligned} \tag{5.23}$$

with $\mathbf{y}_i = \Phi_f(\Delta t, \mathbf{x}_i)$, where $\Phi_f : [0, \infty) \times \mathbb{X} \longrightarrow \mathbb{X}$ is the flow-map for $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$.

Now a dictionary of observable functions $\mathcal{D} = [h_1, \ldots, h_K]$ is chosen such that $h_i \in \mathcal{F}$. The span of these dictionary functions is denoted as $\mathcal{F}_{\mathcal{D}} \subseteq \mathcal{F}$. Define a vector-valued function $\mathbf{H} : \mathbb{X} \longrightarrow C^n$, where

$$\mathbf{H}(\mathbf{x}) = [h_1(\mathbf{x}), \ldots, h_n(\mathbf{x})]. \tag{5.24}$$

The choice of dictionary is an open question, but a discussion of some pragmatic choices can be found in [32]. Our objective is to project a Koopman eigenfunction ϕ onto the dictionary $\mathcal{D}$. The projection is derived by a least-square optimization problem as follows

[32]:

$$\underset{\mathbf{K}}{\text{minimize}} \, \|\mathbf{G}\mathbf{K} - \mathbf{A}\|_F^2 \tag{5.25}$$

where

$$\mathbf{G} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{H}(\mathbf{x}_i)^* \mathbf{H}(\mathbf{x}_i),$$

$$\mathbf{A} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{H}(\mathbf{x}_i)^* \mathbf{H}(\mathbf{y}_i)$$

with $\mathbf{K}$, $\mathbf{G}$, $\mathbf{A} \in \mathbb{C}^{K \times K}$. The solution to this problem [32],

$$\mathbf{K} \triangleq \mathbf{G}^{\dagger} A \tag{5.26}$$

gives the finite-dimensional approximation of the Koopman operator $\mathcal{K}$, where $\mathbf{G}^{\dagger}$ is the pseudoinverse of $\mathbf{G}$. As a consequence, if ξ_j is the j^{th} eigenvector of $\mathbf{K}$ with the eigenvalue μ_j, then the EDMD approximation of an eigenfunction of $\mathcal{K}$ is [32]

$$\phi_j = \mathbf{H}\xi_j,$$

corresponding to the Koopman eigenvalue $\lambda_j = \dfrac{\ln \mu_j}{\Delta t}$.

Using EDMD, we compute n such eigenvalues and eigenfunctions for the drift vector field from simulations or experimental data. These eigenfunctions are used to perform the KCT and derive the Koopman bilinear form. The $C^{\mathbf{x}}$ matrix is approximated as

$$C^{\mathbf{x}} \approx \overline{X} T_{\mathbf{x}}^{\dagger},$$

where $T_{\mathbf{x}} \triangleq [T(\mathbf{x}_1), \ldots, T(\mathbf{x}_N)]$.

5.4 Reachability analysis of the Koopman bilinear form

This section is devoted to reachability analysis of the KBF and requires the concatenation-semigroup structure of the control signals [78]. First, the concepts of reachability and reachable sets are briefly described for bilinear systems.

Definition 5.3. Given a bilinear system

$$\dot{\mathbf{z}} = D\mathbf{z} + \sum_{i=1}^{m} B_i \mathbf{z} u_i,\ \mathbf{z}(0) = \mathbf{z}_0,$$

a point $\mathbf{z}_d$ is said to be *reachable in time* T if $\exists$ an input $\mathbf{u} : t \in [0, T] \mapsto \mathbf{u}(t) \in \mathbb{R}^m$ such that $\mathbf{z}(T) = \mathbf{z}_d$. If $\mathbf{z}_d$ is reachable in time T for all $T > 0$, then $\mathbf{z}_d$ is said to be *reachable*. Moreover, the set of all reachable point of the system, i.e., $\{\mathbf{z}_d : \mathbf{z}_d \text{ is reachable}\}$ is called the *reachable set*.

Next the semigroup is defined and explained in terms of piecewise continuous control signal.

Definition 5.4. A *semigroup* is an algebraic structure consisting of a set S and a binary operation "$\circ$" defined as $\circ : S \times S \to S$, i.e., $x \circ y \in S$, $\forall\, x, y \in S$ such that it satisfies the associative property:

$$\forall\, x, y, z \in S,\ (x \circ y) \circ z = x \circ (y \circ z).$$

Let $\mathbf{u}^i$ be an m-dimensional piecewise continuous control signal. We can form a semigroup from the set $\{\mathbf{u}^i(\cdot)|\mathbf{u}^i : \mathbb{R}^+ \to \mathbb{R}^m, \mathbf{u}^i \text{ piecewise continuous}\}$ with the concatenation operation. The concatenation operation looks like

$$\mathbf{u}^1 \circ \mathbf{u}^2 = \begin{cases} \mathbf{u}^1(t), & t \in [0, t_1) \\ \mathbf{u}^2(t - t_1), & t \in [t_1, t_2) \end{cases} \tag{5.27}$$

Denote this semigroup as U^m. Each $\mathbf{u} \in U^m$, when applied to the dynamics (5.13), generates a one-to-one continuous map from $\mathbb{R}^n$ into $\mathbb{R}^n$ in terms of a flow map. Let T^n be the semigroup of all such maps with the composition operation. The system (5.13) defines a homomorphism H from U^m into T^n [78]. The image of U^m under H is called the Myhill semigroup [78] of the system. The maps of the Myhill semigroup are, in fact, the flow maps of the system with a particular piecewise-continuous control signal $\mathbf{u} \in U^m$ and, therefore, provide all the information about the dynamics.

In general, for an arbitrary nonlinear system, these maps are difficult to obtain analytically and yield little practical value. But for the bilinear system (5.13), the Myhill semigroup maps are the matrices $Z \in \mathbb{R}^{n \times n}$, satisfying the matrix differential equation

$$\begin{aligned} \dot{Z}(t) &= DZ(t) + \sum_{i=1}^{m} B_i Z(t) u_i, \\ Z(0) &= I, \end{aligned} \tag{5.28}$$

with $\mathbf{z}(t) = Z(t)\mathbf{z}(0)$ for any $\mathbf{z}(0) \in \mathbb{R}^n$. Therefore, given any initial state $\mathbf{z}_0$, the states reachable from $\mathbf{z}_0$ are given by all the points in $\mathbb{R}^n$ that can be generated by $Z(t)\mathbf{z}_0$ for some $t \geq 0$, where $Z(t)$ satisfies (5.28). Consequently, the controllability of the system

(5.13) can be characterized by the controllability of the matrix differential equation (5.28). The controllability of a bilinear matrix system has been studied widely [78], [79], [87], [88], exploiting the characteristic of matrices as operators and the corresponding Lie-algebraic structures.

The Lie bracket of $\mathbb{R}^{n\times n}$ matrices is defined as

$$[\cdot,\cdot] : \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n},$$

$$[X, Y] \mapsto XY - YX.$$

Any space of matrices over a field F (usually $\mathbb{R}$ or $\mathbb{C}$) closed under the Lie bracket operation forms a Lie algebra. The dimension l of a Lie algebra is its dimension as a vector space over F. The matrix exponentials of all elements of a matrix Lie algebra along with usual matrix multiplication forms a matrix Lie group associated with the algebra. For example, all $\mathbb{R}^{n\times n}$ matrices form a Lie algebra with dimension $l = n^2$. The corresponding group is known as the *General linear group* and denoted as $\mathrm{GL}(n, \mathbb{R})$, which corresponds to the multiplicative group of invertible $n \times n$ real matrices. Let $\{X_i : i = 1, \ldots, n\}_A$ denote as the smallest Lie algebra containing $\{X_i : i = 1, \ldots, n\}$ and $\{\exp\{X_i\} : i = 1, \ldots, n\}_G$ be the smallest Lie group containing $\{\exp\{X_i\} : i = 1, \ldots, n\}$. Also $\forall\, A, B \in \mathbb{R}^{n\times n}$ and $k = 0, 1, \ldots,$ define $\mathrm{ad}_A^{k+1} B \triangleq [A, \mathrm{ad}_A^k B]$ with $\mathrm{ad}_A^0 B \triangleq B$. The reachability results are stated below using the notation of Lie groups and algebras.

The necessary and sufficient conditions for reachability in the KBF (5.13) are described next. First, it is derived for the drift-free matrix bilinear system, then it is extended to the matrix systems with drift. Finally, the relationship between the reachability in matrix bilinear system (5.28) and KBF (5.13) is explained.

5.4.1 Conditions for reachability

To describe the reachability of a drift-free matrix bilinear system, we need a factorization lemma by Wei and Norman [89], which gives the solution of a bilinear system locally in terms of matrix exponentials.

Lemma 5.1. Consider the drift-free matrix differential equation in $\mathbb{R}^{n\times n}$,

$$\dot{Z}(t) = \sum_{i=1}^{m} B_i Z(t) u_i(t),\ \ Z(0) = I, \tag{5.29}$$

where $Z \in \mathbb{R}^{n\times n}$. Let l be the dimension of the Lie algebra $\{B_i : i = 1, \ldots, m\}_A$. Then there exists a neighbourhood of $t = 0$ in which the solution of (5.29) may be expressed in the form

$$Z(t) = \prod_{i=1}^{l} \exp(h_i(t)B_i),$$

where $\{B_i : i = 1, \ldots, l\}$ is the extension of $\{B_i : i, \ldots, m\}$ to a basis of $\{B_i : i = 1, \ldots, m\}_A$.

The next two theorems are presented in [79]. For the sake of completeness, both of the proofs are included.

Theorem 5.5. [79] Consider the drift-free matrix differential equation (5.29) from Lemma 5.1, which corresponds to the system (5.13) with $\mathbf{f} \equiv 0$ and B_i as defined in Theorem

5.2. $Z_1 \in \mathbb{R}^{n\times n}$ is in the reachable set of (5.29) if and only if $Z_1 \in \{\exp\{\{B_i : i = 1,\dots,m\}_A\}\}_G$, i.e., it lies within the smallest group generated by the matrix exponential of the elements of the smallest algebra generated by the control matrices.

Proof. Sufficiency: Let $Z_1 \in \{\exp\{\{B_i : i = 1,\dots,m\}_A\}\}_G$. Then from [79], Z_1 can be written as a finite product

$$Z_1 = \prod_{k=1}^{m} \exp(B_{i_k} a_k),$$

where $B_{i_k} \in \{B_i : i = 1,\dots,m\}$ and $a_k \in \mathbb{R}\,\forall\, k = 1,\dots,m$. For a $T > 0$, partition $[0,T]$ into m equal intervals $[t_{k-1}, t_k]$, $k = 1,\dots,m$. Define $\tau \triangleq t_k - t_{k-1}$.

Now for $t \in [t_{k-1}, t_k)$, choose $u_{i_{m-k+1}} = \dfrac{a_{m-k+1}}{\tau}$ and $u_i \equiv 0\,\forall\, i \neq i_{m-k+1}$. Hence, for each interval $[t_{k-1}, t_k)$ the system becomes

$$\dot{Z}(t) = \frac{a_{m-k+1}}{\tau} B_{i_{m-k+1}} Z(t),$$

with state transition matrix $\exp\left(t\dfrac{a_{m-k+1}}{\tau} B_{i_{m-k+1}}\right)$. Therefore

$$\begin{aligned} Z(T) &= \exp\left(a_1 B_{i_1}\right) \cdot \ldots \cdot \exp\left(a_m B_{i_m}\right) Z(0) \\ &= \prod_{k=1}^{m} \exp\left(a_k B_{i_k}\right), \end{aligned}$$

since $Z(0) = I$. Hence Z_1 is reachable using piecewise continuous control $\mathbf{u}(t)$.

Necessity: Let Z_1 be reachable in time T, i.e., $Z_1 = Z(T)$. From Lemma 5.1, we have a partition $0 = t_0 < t_1 < \ldots < t_p = T$ such that for $t \in [t_k - 1, t_k)$, $k = 1, \ldots, p$,

$$Z(t) = \prod_{i=1}^{l} \exp(h_i(t)B_i),$$

where $\{B_i : i = 1, \ldots, l\}$ is the extension of $\{B_i : i, \ldots, m\}$ to a basis of $\{B_i : i = 1, \ldots, m\}_A$. Therefore,

$$\begin{aligned} Z_1 &= Z(T) \\ &= \prod_{k=1}^{p} \prod_{i=1}^{l} \exp(h_i(t_k - t_{k-1})B_i), \end{aligned}$$

i.e., a product of the matrix exponentials on the Lie algebra $\{B_i : i = 1, \ldots, m\}_A$. Hence $Z_1 \in \{\exp\{\{B_i : i = 1, \ldots, m\}_A\}\}_G$. □

The result of the Theorem 5.5 is somewhat incomplete in that the drift term is absent. The following theorem describes one way in which this constraint can be relaxed.

Theorem 5.6. [79] Consider the matrix differential equation in $\mathbb{R}^{n \times n}$,

$$\dot{Z}(t) = DZ(t) + \sum_{i=1}^{m} B_i Z(t) u_i(t),\ Z(0) = I, \tag{5.30}$$

where D and B_i are defined as in the proof of Theorem 5.2. Assume $[\mathrm{ad}_D^j k B_i, B_j] = 0$ for $i, j = 1, \ldots, m$ and $k = 0, 1, \ldots, n^2 - 1$. Let $\mathcal{L} = \mathrm{span}\{\mathrm{ad}_D^k B_i : i = 1, \ldots, m,\ k = 0, 1, \ldots, n^2 - 1\}$. Then Z_1 is reachable at time t_1 through continuous controls if and only if $\exists\, L \in \mathcal{L}$ such that

$$Z_1 = \exp(t_1 D)\exp(L).$$

Proof. Necessity: Let Z_1 is reachable at time t_1, i.e., $Z(t_1) = Z_1$. Using the Baker-Hausdorff formula [79],

$$\begin{aligned}
& [\exp(tD)B_i\exp(-tD)\,,\,B_j] \\
= \; & \left[B_i + [D,B_i]t + \frac{1}{2}[D,[D,B_i]]t^2 + \dots\,,\,B_j\right] \\
= \; & [B_i,B_j] + [\mathrm{ad}_D B_i, B_j] + \frac{1}{2}[\mathrm{ad}_D^2 B_i, B_j] \\
& + \frac{1}{6}[\mathrm{ad}_D^3 B_i, B_j] + \dots \\
= \; & \sum_{k=0}^{\infty}\frac{1}{k!}[\mathrm{ad}_D^k B_i, B_j]. \qquad (5.31)
\end{aligned}$$

Now we know $[\mathrm{ad}_D^k B_i, B_j] = 0$ for $k = 0, \dots, n^2-1$. However, $\mathrm{ad}_D(\cdot)$ is a linear operator from an n^2-dimensional space to itself. Hence, by the Cayley-Hamilton theorem [79], $\mathrm{ad}_D^k(\cdot)$ for $k \geq n^2-1$ are linear combinations of the first n^2-1 powers. Hence, from (5.31), $[\exp(tD)B_i\exp(-tD)\,,\,B_j]$ vanishes identically.

Also we get

$$\begin{aligned}
0 \; &= \; \exp(tD)B_i\exp(-tD)B_j - B_j\exp(tD)B_i\exp(-tD) \\
&= \; \exp(\sigma D)(\exp(tD)B_i\exp(-tD))B_j\exp(-\sigma D) \\
& \quad - \exp(\sigma D)B_j(\exp(tD)B_i\exp(-tD))\exp(-\sigma D),
\end{aligned}$$

for some arbitrary $\sigma \in \mathbb{R}$. Let $\beta \triangleq t+\sigma$ and $\gamma = \sigma$. Therefore, for all β and $\gamma \in \mathbb{R}$,

$$[\exp(\beta D)B_i \exp(-\beta D), \exp(\gamma D)B_i \exp(-\gamma D)] = 0.$$

Now in order to solve the matrix differential equation(5.30), let $Y(t) \triangleq \exp(-tD)Z(t)$. The (5.30) becomes

$$\dot{Y}(t) = \left(\sum_{i=1}^{m} u_i(t) \exp(-tD)B_i \exp(tD) \right) Y(0), \tag{5.32}$$

which is of the form $\dot{Y}(t) = B(t)Y(t)$ with $B(t) \triangleq \sum\limits_{i=1}^{m} u_i(t) \exp(tD)B_i \exp(-tD)$ and $[B(t), B(\sigma)] = 0$ for all t and σ. The solution of this system can be written [90] as $\exp \int_0^t B(\sigma)d\sigma$, i.e.,

$$\begin{aligned} Y(t) &= \exp\left(\sum_{i=1}^{m} \int\limits_0^t u_i(\sigma) \exp(-\sigma D)B_i \exp(\sigma D)d\sigma \right) Y(0) \\ &= \exp(L(t))Y(0), \end{aligned} \tag{5.33}$$

where $L(t) \triangleq \left(\sum\limits_{i=1}^{m} \int\limits_0^t u_i(\sigma) \exp(-\sigma D)B_i \exp(\sigma D)d\sigma \right)$. In (5.33), each term in the summation can be written as

$$\begin{aligned} &\int\limits_0^t u_i(\sigma) \exp(-\sigma D)B_i \exp(\sigma D)d\sigma \\ = &\int\limits_0^t (B_i - [D, B_i]\sigma + \frac{1}{2}[D,[D,B_i]]\sigma^2 - \ldots)u_i(\sigma)d\sigma \end{aligned}$$

$$= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \mathrm{ad}_D^k B_i \left(\int_0^t \sigma^k u_i(\sigma) d\sigma \right),$$

i.e., as a linear combination of $\mathrm{ad}_D^k B_i$ for $k = 0, 1, \ldots$. But, again by the Cayley-Hamilton theorem, $\mathrm{ad}_D^k B_i$ for $k > n^2 - 1$ is the linear combination of the previous $n - 1$ powers. Therefore, the term within the exponential in (5.33), $L(t) \in \mathcal{L} = \mathrm{span}\{\mathrm{ad}_D^k B_i : i = 1, \ldots, m,\ k = 0, 1, \ldots, n^2 - 1\}$ for all $t > 0$. Choose $t = t_1$ and $L \triangleq L(t_1)$. Now, $Z(t_1) = \exp(tD)Y(t_1) = \exp(tD)\exp(L)Y(0) = \exp(tD)\exp(L)$ since $Z(0) = Y(0) = I$ with $L \in \mathcal{L}$.

Sufficiency: Notice that

$$\frac{d^k}{dt^k} \exp(-tD) B_i \exp(tD) \bigg|_{t=0} = (-1)^k \mathrm{ad}_D^k B_i. \tag{5.34}$$

[91] shows that the image space of the operator taking continuous functions u_i to $\mathbb{R}^{n\times n}$ according to the rule $x = L(u_i) = \int_0^{t_1} u_i(\sigma)\exp(-\sigma D)B_i \exp(\sigma D) d\sigma$ is spanned by the first n^2 derivatives of $\exp(-tD)B_i \exp(tD)$ (including the zeroth derivative) evaluated at zero. Using this fact with (5.33) and (5.34), we see that for each $L = \mathcal{L} = \mathrm{span}\{\mathrm{ad}_D^k B_i : i = 1, \ldots, m,\ k = 0, 1, \ldots, n^2 - 1\}$ and $t_1 > 0$, there exists a continuous $\mathbf{u} : [0, t_1] \to \mathbb{R}^m$ such that $Y(t_1) = \exp(L)Y(0) = \exp(L)$ since $Y(0) = Z(0) = I$. Hence $Z_1 \triangleq Z(t_1) = \exp(tD)Y(t_1) = \exp(tD)\exp(L)$ is reachable at time t_1. $\square$

Theorems 5.5 and 5.6 describes the reachable set for the bilinear matrix differential equations (5.29) and (5.28). The reachability of KBF (5.13) can be expressed with the help of Theorems 5.5 and 5.6, and is derived in Theorem 5.7.

Theorem 5.7. 1) Given a transformed state $\mathbf{z}_1$, if $\exists\, Z_1 \in \mathbb{R}^{n\times n}$ such that $\mathbf{z}_1 = Z_1\mathbf{z}_0$, then $\mathbf{z}_1$ is reachable from $\mathbf{z}_0$ in KBF (5.13) if Z_1 is reachable from $Z(0) = I$ in the matrix differential equation (5.28).

2) Conversely if $\mathbf{z}_1$ is reachable from $\mathbf{z}_0$ in KBF (5.13), then $\exists\, Z_1$ in the reachable set of the matrix differential equation (5.28) from $Z(0) = I$.

Proof. 1) We have $Z_1 \in \mathbb{R}^{n\times n}$ such that $\mathbf{z}_1 = Z_1\mathbf{z}_0$, and Z_1 is reachable at time T in the matrix differential equation (5.28) from $Z(0) = I$. Hence we have $Z(T) = Z_1$ in (5.28). Now define

$$\mathbf{z}(t) = Z(t)\mathbf{z}_0,\ t \geq 0,$$

where $Z(t)$ is the trajectory of the matrix differential equation (5.28). By differentiation we obtain

$$\begin{aligned}
\dot{\mathbf{z}}(t) &= \dot{Z}(t)\mathbf{z}_0 \\
&= DZ(t)\mathbf{z}_0 + \sum_{i=1}^{m} B_i Z(t)\mathbf{z}_0 u_i(t) \\
&= D\mathbf{z}(t) + \sum_{i=1}^{m} B_i \mathbf{z}(t) u_i(t),
\end{aligned}$$

which is of the same form as of KBF (5.13) with $\mathbf{z}(0) = \mathbf{z}_0$ and $\mathbf{z}(T) = Z(T)\mathbf{z}_0 = Z_1\mathbf{z}_0 = \mathbf{z}_1$. Hence $\mathbf{z}_1$ is reachable from $\mathbf{z}_0$ in KBF (5.13).

2) Now we have $\mathbf{z}_1$ reachable from $\mathbf{z}_0$ in KBF (5.13). By contradiction, assume $\nexists\ Z_1 \in \mathbb{R}^{n\times n}$ reachable from $Z(0) = I$ in the matrix differential equation (5.28) such that $\mathbf{z}_1 = Z_1\mathbf{z}_0$. Then at any $t > 0$, $Z(t)\mathbf{z}_0 \neq \mathbf{z}_1$. Like the previous part, define

$$\mathbf{z}(t) = Z(t)\mathbf{z}_0,\ t \geq 0,$$

where $Z(t)$ is the trajectory of the matrix differential equation (5.28). By differentiation we obtain

$$\begin{aligned} \dot{\mathbf{z}}(t) &= \dot{Z}(t)\mathbf{z}_0 \\ &= DZ(t)\mathbf{z}_0 + \sum_{i=1}^{m} B_i Z(t)\mathbf{z}_0 u_i(t) \\ &= D\mathbf{z}(t) + \sum_{i=1}^{m} B_i \mathbf{z}(t) u_i(t), \end{aligned} \tag{5.35}$$

which is of the same form as of KBF (5.13) with $\mathbf{z}(0) = \mathbf{z}_0$. But we know $\mathbf{z}_1$ is reachable in the system (5.35) from $\mathbf{z}_0$, i.e., $\exists\ T$ such that $\mathbf{z}(T) = \mathbf{z}_1$, and again from (5.35) $\mathbf{z}(T) = Z(T)\mathbf{z}_0$. Hence we have $Z_1 \triangleq Z(T)$, such that $\mathbf{z}_1 = Z_1\mathbf{z}_0$, which establishes the contradiction. $\square$

The controllability and reachable sets of the system (5.7) may be characterized by the transformed bilinearized system with D, $B_i, i = 1, \ldots, m$, from the Koopman Bilinear Form in Section 5.3 and 5.4. However, because the transformed KBF usually has more dimensions, it may not achieve complete controllability even when the original system does.

5.5 Optimal control of the Koopman bilinear form

Next, the Koopman bilinear form (KBF) representation of the original system (5.7) is used for optimal control. With the bilinear structure of the KBF (5.21), the complexity of designing an optimal feedback is significantly reduced. We follow the necessary condition given by Pontryagin's principle [80] to design a fixed endpoint feedback control for the KBF.

The fixed endpoint optimal control problem for the system (5.7) is given by

$$\begin{aligned} \underset{\mathbf{u}(t)}{\text{minimize}} \quad & \frac{1}{2}\int_{t_0}^{t_f} \left(\mathbf{x}^T Q \mathbf{x} + \mathbf{u}^T \mathbf{u}\right) dt \\ \text{subject to} \quad & \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \sum_{i=1}^{m} \mathbf{g}_i(\mathbf{x}) u_i, \\ & \mathbf{x}(t_0) = \mathbf{x}_0,\ \mathbf{x}(t_f) = \mathbf{x}_f, \\ & \mathbf{u} \in \mathcal{U}, \end{aligned} \tag{5.36}$$

$\mathcal{U}$ being the set of the admissible control and $Q = Q^T \succeq 0$. Now with the Koopman bilinear form (5.21), the problem (5.36) turns into

$$\begin{aligned} \underset{\mathbf{u}(t)}{\text{minimize}} \quad & \frac{1}{2}\int_{t_0}^{t_f} \left(\mathbf{z}^T C^{\mathbf{x}^T} Q C^{\mathbf{x}} \mathbf{z} + \mathbf{u}^T \mathbf{u}\right) dt \\ \text{subject to} \quad & \dot{\mathbf{z}} = D\mathbf{z} + \sum_{i=1}^{m} B_i \mathbf{z} u_i, \\ & \mathbf{z}(t_0) = T(\mathbf{x}_0),\ C^{\mathbf{x}} \mathbf{z}(t_f) = \mathbf{x}_f, \\ & \mathbf{u} \in \mathcal{U}. \end{aligned} \tag{5.37}$$

To apply Pontryagin's principle to problem (5.37), we construct the pre-Hamiltonian as

$$H(t,\mathbf{z},\mathbf{p},\mathbf{v}) = \mathbf{p}^T\left(D\mathbf{z} + \sum_{i=1}^{m} B_i\mathbf{z}v_i\right) - \mathcal{L}(\mathbf{z},\mathbf{v}), \tag{5.38}$$

with the Lagrangian $\mathcal{L}(\mathbf{z},\mathbf{v}) \triangleq \frac{1}{2}\mathbf{z}^T C^{\mathbf{x}^T} Q C^{\mathbf{x}}\mathbf{z} + \frac{1}{2}\mathbf{v}^T\mathbf{v}$ and the co-state equation

$$\dot{\mathbf{p}} = -\frac{\partial H^T}{\partial \mathbf{z}} = -\left(D + \sum_{i=1}^{m} B_i u_i\right)^T \mathbf{p} + C^{\mathbf{x}^T} Q C^{\mathbf{x}}\mathbf{z}. \tag{5.39}$$

If $\mathbf{u}^*$ solves the problem (5.37), then we have

$$\mathbf{u}^*(t) = \underset{\mathbf{v}}{\operatorname{argmax}}\, H(t,\mathbf{z},\mathbf{p},\mathbf{v}),$$

from Pontryagin's principle, and we define the Hamiltonian

$$\mathcal{H}(t,\mathbf{z},\mathbf{p}) \triangleq \max_{\mathbf{v}} H(t,\mathbf{z},\mathbf{p},\mathbf{v}) = H(t,\mathbf{z},\mathbf{p},\mathbf{u}^*(t)).$$

Hence, if an optimal control $\mathbf{u}^*$ exists, then

$$\left.\frac{\partial H}{\partial v_i}\right|_{\mathbf{v}=\mathbf{u}^*(t)} = 0, \text{ for } i = 1,\ldots,m$$

i.e.,

$$u_i^*(t) = \mathbf{p}^T(t) B_i \mathbf{z}(t). \tag{5.40}$$

The co-state equation (5.39) must also satisfy the transversality condition, i.e.,

$$\mathbf{p}(t_f) \perp \ker C^{\mathbf{x}}, \tag{5.41}$$

for the endpoint constraint.

To calculate the optimal control signal in (5.40), we need to calculate both the optimal state $\mathbf{z}(t)$ and co-state $\mathbf{p}(t)$ trajectories that satisfy the initial and final conditions with the state-feedback control input $u_i^*(t) = \mathbf{p}^T(t)B_i\mathbf{z}(t)$ for $i = 1, \ldots, m$. However, this poses a two-point boundary value problem with no prior knowledge of $\mathbf{p}(0)$. We solve this problem using the single shooting method. In this method, start from an initial guess of $\mathbf{p}(0)$ and solve (*shoot*) for $\mathbf{z}$ and $\mathbf{p}$. The error in the terminal condition $C^{\mathbf{x}}\mathbf{z}(t_f) - \mathbf{x}_f$ is formulated as a function of the guessed $\mathbf{p}(0)$. The transversality condition (5.41) is also checked for each such trajectory. This error function can then be solved with a nonlinear equation solver, e.g., Newton's method or the Levenberg-Marqardt algorithm to find the correct $\mathbf{p}_0$. If the solution fails to converge, then we conclude that no optimal solution is possible.

Remark 5.12. The cost function in problem (5.36) can be readily generalized into $\frac{1}{2}\int\limits_{t_0}^{t_f}\left(\mathbf{x}^TQ\mathbf{x} + \mathbf{u}^TR\mathbf{u}\right)dt$ with $R = R^T \succeq 0$. The co-state equation (5.39) remains the same, and the optimal control $\mathbf{u}^*$ becomes

$$\mathbf{u}^* = R^{-1}\mathbf{B}(\mathbf{z}, \mathbf{p}), \tag{5.42}$$

where $\mathbf{B}(\mathbf{x},\mathbf{p}) \triangleq [\mathbf{p}^T B_1\mathbf{z},\ldots,\mathbf{p}^T B_m\mathbf{z}]^T$.

5.6 Numerical simulation

To demonstrate the effectiveness of the bilinearization technique, we first choose the system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}_1(\mathbf{x})u_1 + \mathbf{g}_2(\mathbf{x})u_2, \tag{5.43}$$

where the drift $\mathbf{f}$ is

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} \lambda x_1 \\ \\ \mu x_2 + (2\lambda-\mu)cx_1^2 \end{pmatrix}.$$

This choice of $\mathbf{f}$ is inspired from [29] so that the eigenfunctions may be obtained by inspection. For the demonstration, we vary $\mathbf{g}_1$ and $\mathbf{g}_2$. For our purpose, we have used the following four Koopman eigenvalue-eigenfunction pairs for $L_{\mathbf{f}}$:

- $\phi_1(\mathbf{x}) = x_1$ with eigenvalue λ,
- $\phi_2(\mathbf{x}) = x_2 - cx_1^2$ with eigenvalue μ,
- $\phi_3(\mathbf{x}) = x_1^2$ with eigenvalue 2λ, and
- $\phi_4(\mathbf{x}) = 1$ with eigenvalue 0.

The Koopman canonical transformation is

$$\mathbf{z} = T(\mathbf{x}) = \begin{pmatrix} \phi_1(\mathbf{x}) \\ \phi_2(\mathbf{x}) \\ \phi_3(\mathbf{x}) \\ \phi_4(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 - cx_1^2 \\ x_1^2 \\ 1 \end{pmatrix}$$

and the matrix $D = \mathrm{diag}(\lambda, \mu, 2\lambda, 0)$.

5.6.1 Completely bilinearizable system

Now let us choose $\mathbf{g}_1$ and $\mathbf{g}_2$ such that the system becomes completely bilinearizable in four dimension according to Theorem 5.2. Let

$$\mathbf{g}_1(\mathbf{x}) = \begin{pmatrix} 1 \\ x_1^2 \end{pmatrix} \text{ and } \mathbf{g}_2(\mathbf{x}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Then

$$L_{\mathbf{g}_1} T(\mathbf{x}) = \begin{pmatrix} 1 \\ -2cx_1 + x_1^2 \\ 2x_1 \\ 0 \end{pmatrix} = B_1\mathbf{z},$$

where

$$B_1 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ -2c & 0 & 1 & 0 \\ 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Similarly $L_{\mathbf{g}_2} T(\mathbf{x}) = B_2 \mathbf{z}$ with

$$B_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

For this simulation, let $\lambda = 0.3$, $\mu = 0.2$ and $c = -0.5$. Apply $u_1 = \cos(2\pi t)$, a sinusoidal excitation, and $u_2 = -x_2 = -(z_2 + cz_1^2)$, a state feedback. Fig. 5.1 shows the original system response and the response from the bilinearized system after transforming back to the original coordinates. The responses are identical.

The controllability and reachable sets of the system (5.7) may be characterized by the transformed bilinearized system with D, B_1 and B_2. However, because the transformed KBF has more dimensions, it may not achieve complete controllability even when the original system does. It can be shown that the bilinearized system $\dot{\mathbf{z}}(t) = D\mathbf{z}(t) + B_1\mathbf{z}(t)u_1(t) + B_2\mathbf{z}(t)u_2(t)$ satisfies the hypothesis of Theorem 5.6. So we resort to finding $Z(t) \in \mathbb{R}^{n\times n}$ where $Z(t)$ satisfies the matrix differential equation (5.28) with

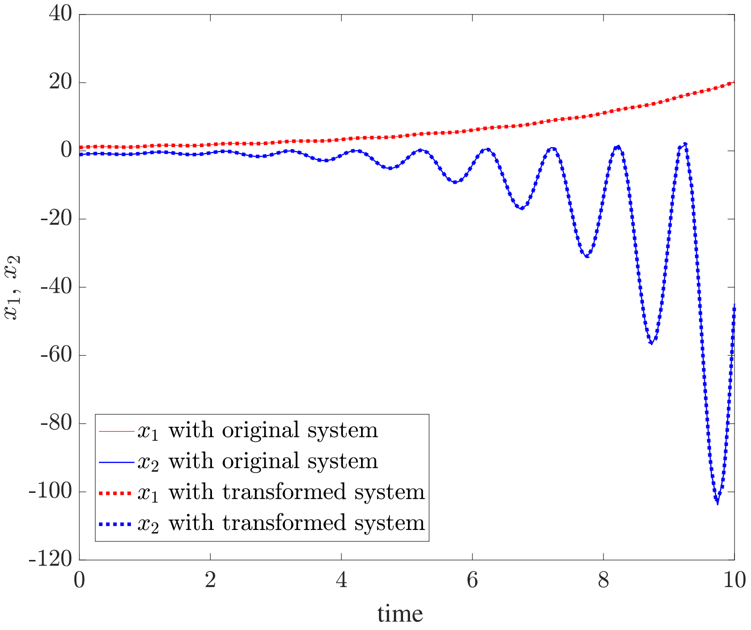


Figure 5.1: Simulation of the original and bilinearized system for (5.43) with exact bilinearization

$m = 2$ and $\mathbf{z}(t) = Z(t)\mathbf{z}(0)$. According to Theorem 5.6, $Z(t)$ must take the form

$$Z(t) = \exp(tD)\exp(L),$$

where $L \in \mathcal{L} = \operatorname{span}\{\operatorname{ad}_D^k B_i : i = 1, \ldots, m,\ k = 0, 1, \ldots, n^2 - 1\}$. By explicitly calculating $\exp(tD)\exp(L_1)$, where $L_1 = c_1 B_1 + c_2 B_2 \in \operatorname{span}\{\operatorname{ad}_D^k B_i : i = 1, \ldots, m,\ k = 0, 1, \ldots, n^2 -$

1}, we see that the resultant matrix $Z(t)$ is

$$\begin{bmatrix} e^{\lambda t} & 0 & 0 & c_1 e^{\lambda t} \\ -c_1 e^{\mu t}\left(2c - \dfrac{c_1}{2}\right) & e^{\mu t} & c_1 e^{\mu t} & e^{\mu t}\left(c_2 - c c_1^2 + \dfrac{c_1^3}{3}\right) \\ 2c_1 e^{2\lambda t} & 0 & e^{2\lambda t} & c_1^2 e^{2\lambda t} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

From any $\mathbf{z}_0$, we can achieve $\mathbf{z}(t) = Z(t)\mathbf{z}(0)$ and, therefore, any $z_1(t)$ and $z_2(t)$ can achieved by varying the scalars c_1 and c_2 as $z_4 \equiv 1$. So we have global controllability for the original system (5.43). However, the bilinearized system is not globally controllable, because we have no control authority over $z_4(t) \equiv 1$. To stabilize the system we choose $u_1(t) = -(\lambda + 0.5)\, z_1(t) = -(\lambda + 0.5)\, x_1(t)$ and $u_2(t) = -(2cz_1(t) - z_3(t))\, u_1(t) - (\mu + 0.5)\, z_2(t) = (2cx_1(t) - x_1^2(t))\, u_1(t) - (\mu + 0.5)(x_2(t) - cx_1^2(t))$. This effectively reduces the

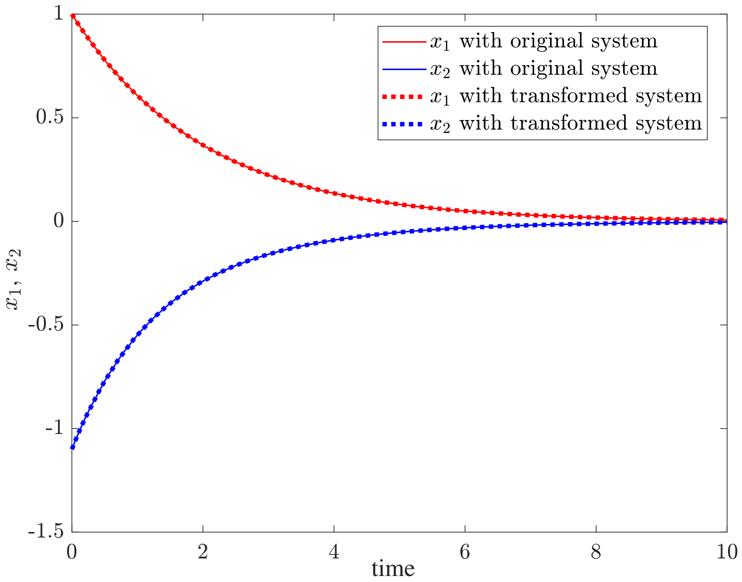
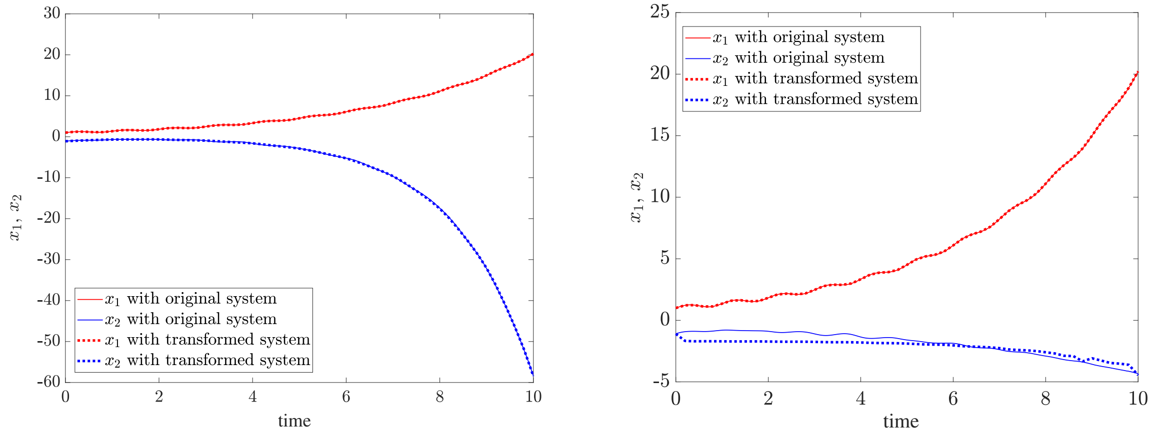


Figure 5.2: Feedback control of the system (5.43) using the bilinearization



((a)) ((b))

Figure 5.3: Approximate bilinearization of the system (5.43)

transformed system into $\dot{\mathbf{z}} = A\mathbf{z}$, where $A = \text{diag}(-0.5, -0.5, -0.5, 1)$. This input in turn feedback linearizes the original system (5.43) giving a strong connection between KBF and feedback linearizability of the system. The system response under this feedback is shown in Fig. 5.2.

5.6.2 Approximately bilinearized system

Choice of $\mathbf{g}_1(\mathbf{x}) = [1 \ \ \cos x_1]^T$ and $\mathbf{g}_2 = [0 \ 1]^T$[53] shows that this choice of $\mathbf{g}_i$ and $T(\mathbf{x})$ does not yield an exact bilinearization. However, we have used the optimization scheme (5.15) to obtain an approximate bilinearization with $\mathbf{X} = [0, 30] \times [-10, 0]$ and simulated the original and bilinear system for 10 seconds starting from $\mathbf{x}(0) = [1, -1.1]^T$. We applied $u_1 = \cos(2\pi t)$, a sinusoidal excitation, and $u_2 = -x_2 = -(z_2 + cz_1^2) = -(z_2 + cz_3)$, a state feedback. The result is presented in Fig. 5.3(a). The same simulation is presented after changing $\mathbf{g}_2(\mathbf{x}) = [0 \ x_2^2]$ and $\mathbf{X} = [0, 30] \times [-10, 0]$ in Fig. 5.3(b).

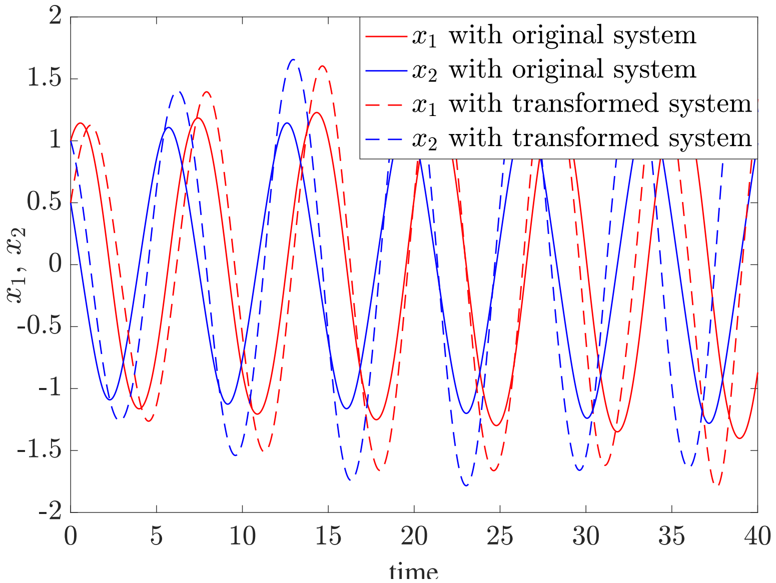


Figure 5.4: Comparison of original system and KBF response for pendulum (5.44)

5.6.3 Numerical construction of KBF and optimal control

To demonstrate the computation of the KBF numerically, consider the controlled pendulum with a destabilizing term:

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= 0.01x_2 - \sin x_1 + u,\end{aligned} \tag{5.44}$$

which has an unstable equilibrium at the origin. A dictionary of monomials up to 5^{th} degrees (i.e., 21 monomials) are used. The truncated state-space is chosen as $\mathbf{X} = [-1,1]\times[-1,1]$. To calculate the EDMD, time-series data for 10 seconds with 0.01s time step is used. The original system response and the corresponding transformed system response with the KBF are presented in Fig. 5.4. The optimal control signal u for the system (5.44)

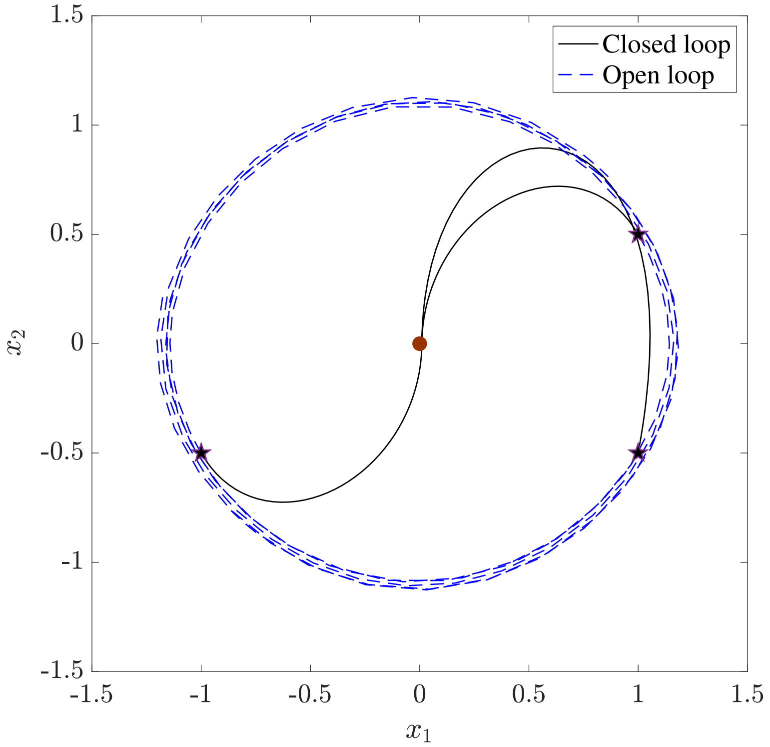


Figure 5.5: Closed-loop optimal control of the unstable pendulum (5.44) using KBF from different initial conditions marked with stars ($\star$)

is calculated with Pontryagin's principle and the shooting method to drive the system from different initial conditions to $\mathbf{x}_f = [0.01, -0.01]$ in $t = 2$ seconds. $Q = R = I_{2\times 2}$ is used in the cost function. The corresponding closed-loop trajectories as well as the open-loop counterparts in phase space are shown in Fig 5.5.

Last, a controlled Van Der Pol system is used to derive the corresponding KBF and optimal control design:

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= (1 - x_1^2)x_2 - x_1 + u \end{aligned} \tag{5.45}$$

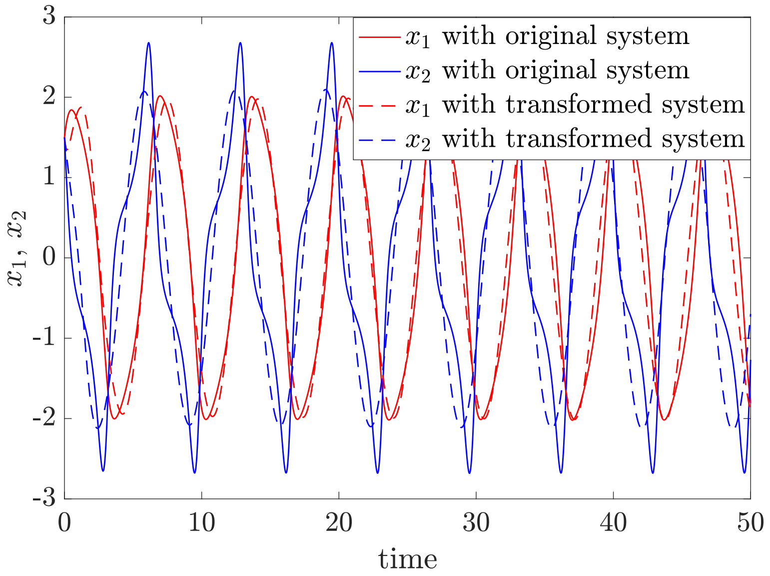


Figure 5.6: Comparison of original system and KBF response for Van der Pol oscillator (5.45)

Again a dictionary of monomials up to 5th degrees (i.e., 21 monomials) are used. The truncated state-space is chosen as $\mathbf{X} = [-3, 3] \times [-3, 3]$. The comparison of the original system and the transformed KBF without control are presented in Fig. 5.6. Then the shooting method is used to compute the optimal control with $Q = R = I_{2\times 2}$, with the objective of driving the system from different initial conditions to $\mathbf{x}_f = [0.01, -0.01]$ in $t = 1$ second. The corresponding closed-loop and open-loop trajectories are presented in Fig 5.7.

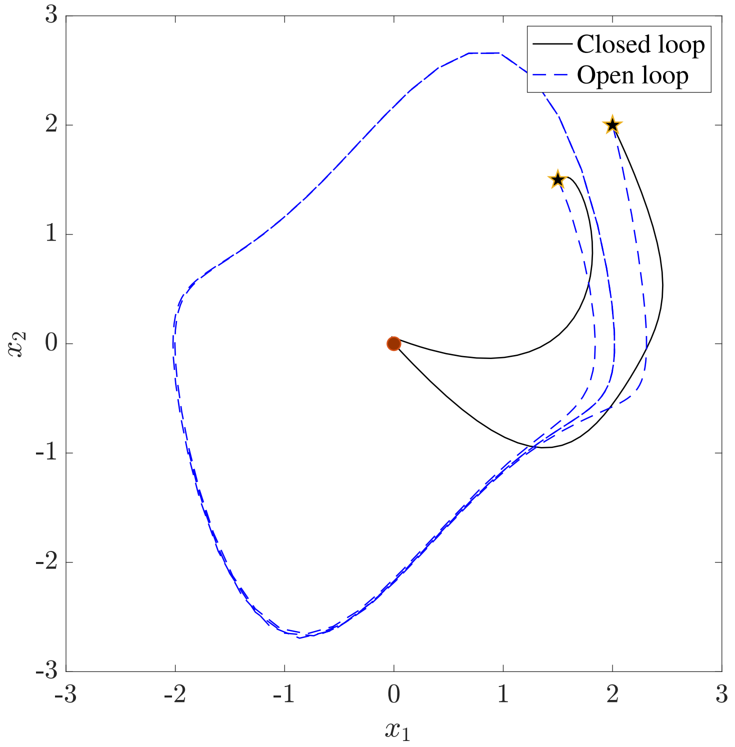


Figure 5.7: Closed-loop optimal control of Van der Pol oscillator (5.45) using KBF from different initial conditions marked with stars ($\star$)

Chapter 6

Conclusion

This dissertation develops several techniques for nonlinear estimation and output feedback in both model-based and partially modeled or model-free scenarios. The proposed methods utilize tools from Gaussian mixture model representation, transfer operator theory, recurrent neural networks, and optimal control theories.

6.1 Summary of contributions

This dissertation addresses the problem of nonlinear estimation and output feedback in model-based, model-free, and partially modeled scenarios. This dissertation divides the problems in two broad classes: model-based and model free estimation and output feedback. In model-based problems, a reasonably accurate dynamical model of the system is present which has to be driven by an output feedback to achieve stability. In model free scenario, the system is represented by a stream of data available from a huge number of sensors without a reliable model. A model-identification and subsequent controller design needs to be carried out there. For model-based approach, Gaussian mixture model with

Kalman filter is utilized for non-Gaussian estimation and output feedback. A Perron-Frobenius operator based motion update in a Bayesian estimation technique is also proposed for a system with a dynamical model. On the other hand, an echo-state network approach equipped with a Kalman filter is used for data-driven estimation of a complex systems. Finally, a Koopman spectral framework is used for the data-driven control of a control-affine nonlinear system.

6.1.1 Dynamic output feedback using the Gaussian mixture model Kalman filter

Chapter 2 addresses the problem of non-Gaussian estimation for linear and nonlinear feedback systems and presents the Gaussian mixture Kalman filter (GMM-KF) as an effective tool. Past literature [12, 13, 16, 17, 92, 93] has identified the advantage of adopting GMM in the filtering scheme in order to retain the potential non-Gaussian features of the posterior. However, they mostly rely on a heuristic choice of parameters and resampling at each step, thus making it analytically as well as computationally burdensome. To deal with the analytical tractability, the re-sampling is dropped and the concept of bounded with probability is introduced. The error-boundedness for GMM-KF, GMM-EKF, and GMM-UKF is proved for linear and nonlinear systems, including dynamic output feedback, under certain assumptions. The extension of GMM-KF for nonlinear systems is facilitated by the use of the extended Kalman filter (EKF) and the unscented Kalman filter (UKF).

6.1.2 Density transport in dynamical systems: Perron Frobenius operator approximation

Chapter 3 provides a new approximation method for computing eigenvalues and eigenfunctions of the Perron-Frobenius (PF) operator by combining the accuracy of extended Dynamic Mode decomposition (EDMD) and the Galerkin projection used in Ulam's method. The CU-DMD algorithm successfully approximates the PF operator and its eigenfunctions for smaller time steps than Ulam's Method. CU-DMD utilizes time-series data from Monte-Carlo simulations and constrained quadratic programming to generate a Markov state-transition matrix to approximate the PF operator. The analytical justifications for the Galerkin projection and the eigenfunction approximation from the basis functions are provided. The algorithm is demonstrated on several nonlinear systems with and without diffusive Wiener noise, and captures the modes of the system reasonably well. Furthermore, the approximate PF operator is used for the estimation and output-feedback control of a pitching airfoil in an unsteady flow. Since the coefficient of lift undergoes hysteresis at or near stall, the proposed output-feedback control generates by design a stable limit cycle. The resulting periodic pitching trajectory yields a 40% increase in time-averaged lift, compared to the optimal steady pitch angle. However, a steady pitch angle maximizes the lift-to-drag ratio. In ongoing work, we seek to compare observer-based feedback strategies for online estimation of the model parameters. The chosen strategy will be used to implement an adaptive gain selection using a table of optimal gains for the model parameters.

6.1.3 Data-driven estimation using an echos-state neural network equipped with an ensemble Kalman filter

Chapter 4 describes a data-driven sparse estimation technique for complex dynamical systems and uses it to estimate the states of nonlinear systems from time-series data. The method utilizes the echo-state network (ESN) for model identification from the time-series training data and an ensemble Kalman filter for data assimilation in the testing phase. The estimation is carried out in a data-driven way without a dynamic model. The method is applied to a real data set of traffic pattern on a road network at the University of Maryland campus to predict the traffic congestion at different intersections. The method is also extended to the Lorenz-96 model for atmospheric data assimilation with a parallel-reservoir ESN. In ongoing and future work, a data-driven controller design using the ESN will be investigated.

6.1.4 Bilinearization, controllability, and optimal control of a control affine system: a Koopman spectral approach

Chapter 5 proposes a bilinearization technique for a control-affine nonlinear system using Koopman spectral approach. The bilinear representation relies on the Koopman canonical transform (KCT) to project the system dynamics onto the Koopman eigenspace. The approximate bilinearization is achieved by solving a $\mathcal{L}^2$ norm-based quadratic optimization problem. It is proved that with some assumptions, any system can state can be approximated arbitrarily accurately by Koopman canonical transform and hence the bilinearization will converge. The resulting Koopman Bilinear Form is used for reachability analysis using Myhill semigroup structure. Finally, an optimal controller design is proposed for control-affine systems using the Koopman bilinear form and Pontryagin's principle. The complete procedure of the bilinearization and optimal control can be carried out with time-series data without explicit knowledge of the autonomous system dynamics, i.e., the drift vector field. The bilinearization method and optimal control design are demonstrated on several nonlinear systems including an unstable pendulum and the Van der Pol oscillator.

6.2 Directions for future research

6.2.1 Model-based estimation and output feedback of nonlinear systems

The GMM-based Kalman filtering requires the measurement equation to be linear with the Gaussian measurement noise profile. To employ the same technique for a general nonlinear measurement equation is still an open question. A general Bayesian filtering framework can be explouted for this purpose. The convergence of the PF operator approximation in the Hilbert space can be utilized for a convergence/error-bound analysis of the estimation and output feedback. On the application side,an online estimation of the model parameters can be carried out to for an adaptive output feedback control.

6.2.2 Data-driven estimation and output feedback for unmodeled or partially modeled systems

Recurrent neural networks (RNN), specifically the echo-state networks (ESN) show a great potential for data-driven optimal control of an unmodeled system and it is largely unexplored. An optimal choice dictionary for Koopman spectral approximation using the extended dynamic mode decomposition (EDMD) is still an open question. An RNN, being an universal approximator, can be employed for the dictionary learning and data-driven modeling of the control-affine nonlinear systems. The convergence rate of the Koopman bilinear form to the original nonlinear system is needed to assess its potential as a model-identification method.

Bibliography

[1] G. Sedky, A. Jones, and F. Lagor. Lift modeling and regulation for a finite wing during transverse gust encounters. *AIAA Scitech 2019 Forum*, pages 809–814, 2019.

[2] Z. Lu, J. Pathak, B. Hunt, M. Girvan, R. Brockett, and E. Ott. Reservoir observers: Model-free inference of unmeasured variables in chaotic systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(4):1054–1500, 2017.

[3] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[4] M. Zakai. On the optimal filtering of diffusion processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 11(3):230–243, Sep 1969.

[5] V. E. BeneŠ. Exact finite-dimensional filters for certain diffusions with nonlinear drift. *Stochastics*, 5(1-2):65–92, 1981.

[6] S. S. T. Yau and S.-T. Yau. Real time algorithm for nonlinear filtering problem. In *Proceedings of the 40th IEEE Conference on Decision and Control (Cat. No.01CH37228)*, volume 3, pages 2137–2142 vol.3, Dec 2001.

[7] D. Brigo, B. Hanzon, and F. L. Gland. A differential geometric approach to nonlinear filtering: the projection filter. In *Proceedings of 1995 34th IEEE Conference on Decision and Control*, volume 4, pages 4006–4011 vol.4, Dec 1995.

[8] R. Beard, J. Kenney, J. Gunther, J. Lawton, and W. Stirling. Nonlinear Projection Filter Based on Galerkin Approximation. *Journal of Guidance, Control, and Dynamics*, 22(2):258–266, mar 1999.

[9] B. Azimi-Sadjadi and P. S. Krishnaprasad. Approximate nonlinear filtering and its applications for GPS. In *Proceedings of the 39th IEEE Conference on Decision and Control (Cat. No.00CH37187)*, volume 2, pages 1579–1584 vol.2, Dec 2000.

[10] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002.

[11] J. V. Candy. Bootstrap particle filtering. *IEEE Signal Processing Magazine*, 24(4):73–85, July 2007.

[12] D. Alspach and H. Sorenson. Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Transactions on Automatic Control*, 17(4):439–448, Aug 1972.

[13] T. Sondergaard and P. F. J. Lermusiaux. Data assimilation with Gaussian mixture models using the dynamically orthogonal field equations. Part I: Theory and scheme. *Monthly Weather Review*, Volume 141:1737–1760, 2013.

[14] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[15] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 03 1978.

[16] L. Dovera and E. Della Rossa. Multimodal ensemble Kalman filtering using Gaussian mixture models. *Computational Geosciences*, 15(2):307–323, Mar 2011.

[17] M. Frei and H. R. Künsch. Mixture ensemble Kalman filters. *Computational Statistics & Data Analysis*, 58:127 – 138, 2013. The Third Special Issue on Statistical Signal Extraction and Filtering.

[18] T. Bengtsson, C. Snyder, and D. Nychka. Toward a nonlinear ensemble filter for high-dimensional systems. *Journal of Geophysical Research: Atmospheres*, 108(D24), 2003. 8775.

[19] K. Smith. Cluster ensemble Kalman filter. *Tellus A*, 59(5), 2007.

[20] P. Tagade, H. Seybold, and S. Ravela. Mixture ensembles for data assimilation in dynamic data-driven environmental systems. *Procedia Computer Science*, 29:1266 – 1276, 2014.

[21] M. I. Ribeiro. Kalman and Extended Kalman Filters: concept, derivation and properties. *Technical Report, Institute for Systems and Robotics, Instituto Superior Tecnico*, 2004.

[22] S. J. Julier and J. K. Uhlmann. New extension of the Kalman filter to nonlinear systems. In *AeroSense'97*, pages 182–193. International Society for Optics and Photonics, 1997.

[23] S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, 2004.

[24] G. Froyland, R. M. Stuart, and E. van Sebille. How well-connected is the surface of the global ocean? *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 24(3):033126, 2014.

[25] M. Dellnitz, G. Froyland, and O. Junge. The algorithms Behind GAIO — set oriented numerical methods for dynamical systems. In *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems*, pages 145–174, Berlin, Heidelberg, 2001. Springer.

[26] S. Klus, P. Koltai, and C. Schütte. On the numerical approximation of the Perron-Frobenius and Koopman operator. *Journal of Computational Dynamics*, 3(1):51–79, 2016.

[27] J. Ding, Q. Du, and T. Li. High order approximation of the Frobenius-Perron operator. *Applied Mathematics and Computation*, 53(2):151–171, 1993.

[28] G. Froyland and K. Padberg. Almost-invariant sets and invariant manifolds — connecting probabilistic and geometric descriptions of coherent structures in flows. *Physica D: Nonlinear Phenomena*, 238(16):1507 – 1523, 2009.

[29] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson. Spectral analysis of nonlinear flows. *Journal of Fluid Mechanics*, 641:115–127, 2009.

[30] P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5â 28, 2010.

[31] R. Mohr and I. Mezić. Construction of eigenfunctions for scalar-type operators via Laplace averages with connections to the Koopman operator. *ArXiv e-prints*, March 2014.

[32] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley. A data–driven approximation of the Koopman operator: extending Dynamic Mode Decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, Dec 2015.

[33] Q. Li, F. Dietrich, E. M. Bollt, and I. G. Kevrekidis. Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the Koopman operator. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(10):103111, 2017.

[34] B. Huang and U. Vaidya. Data-driven approximation of transfer operators: Naturally structured dynamic mode decomposition. In *2018 Annual American Control Conference (ACC)*, pages 5659–5664, 2018.

[35] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97, Nov 2012.

[36] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach,

K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

[37] H. Jaeger and H. Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80, 2004.

[38] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Phys. Rev. Lett.*, 120:024102, Jan 2018.

[39] W. Maass and H. Markram. On the computational power of circuits of spiking neurons. *Journal of Computer and System Sciences*, 69(4):593 – 616, 2004.

[40] L. Grigoryeva and J.-P. Ortega. Echo state networks are universal. *Neural Networks*, 108:495 – 508, 2018.

[41] I. Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1):309–325, 2005.

[42] M. O. Williams, M. S. Hemati, S. T. M. Dawson, I. G. Kevrekidis, and C. W. Rowley. Extending data-driven Koopman analysis to actuated systems. *IFAC-PapersOnLine*, 49(18):704–709, 2016.

[43] A. Mauroy and I. Mezić. Global stability analysis using the eigenfunctions of the Koopman operator. *IEEE Transactions on Automatic Control*, 61(11):3356–3369, Nov 2016.

[44] Y. Lan and I. Mezić. Linearization in the large of nonlinear systems and Koopman operator spectrum. *Physica D: Nonlinear Phenomena*, 242(1):42–53, 2013.

[45] J. Proctor, S. Brunton, and J. Kutz. Generalizing Koopman theory to allow for inputs and control. *SIAM Journal on Applied Dynamical Systems*, 17(1):909–930, 2018.

[46] A. Surana and A. Banaszuk. Linear observer synthesis for nonlinear systems using Koopman operator framework. *IFAC-PapersOnLine*, 49(18):716–723, 2016.

[47] A. Surana. Koopman operator based observer synthesis for control-affine nonlinear systems. In *55th IEEE Conference on Decision and Control*, pages 6492–6499, Dec 2016.

[48] D. Goswami, E. Thackray, and D. A. Paley. Constrained Ulam Dynamic Mode Decomposition: approximation of the Perron-Frobenius operator for deterministic and stochastic systems. *IEEE Control Systems Letters*, 2(4):809–814, Oct 2018.

[49] J. Lidard, D. Goswami, D. Snyder, G. Sedky, A. R. Jones, and D. A. Paley. *Output Feedback Control for Lift Maximization of a Pitching Airfoil.*

[50] D. Goswami and D. A. Paley. Non-Gaussian estimation and dynamic output feedback using the Gaussian mixture model Kalman filter. *submitted*, 2020.

[51] D. Goswami and D. A. Paley. Global bilinearization, reachability, and optimal control of control-affine nonlinear systems: a Koopman spectral approach. *submitted*, 2020.

[52] D. Goswami and D. A. Paley. Non-Gaussian estimation and observer-based feedback using the Gaussian Mixture Kalman and Extended Kalman Filters. In *2017 American Control Conference*, pages 4550–4555, May 2017.

[53] D. Goswami and D. A. Paley. Global bilinearization and controllability of control-affine nonlinear systems: A Koopman spectral approach. In *2017 IEEE 56th Annual Conference on Decision and Control*, pages 6107–6112, Dec 2017.

[54] C. Bishop. *Pattern recognition and machine learning.* Springer, New York, 2006.

[55] G. Casella and R. L. Berger. *Statistical Inference.* Duxbury, 2001.

[56] J. Walrand. Kalman filter: Convergence. *UC Berkeley, EE 226A Lecture Note.*

[57] H. K. Khalil. *Nonlinear Systems.* Prentice Hall, Upper Saddle River, (N.J.), 1996.

[58] W. J. Rugh. *Linear System Theory.* Prentice Hall, Upper Saddle River, (N.J.), 1996.

[59] E. A. Wan and R. V. D. Merwe. The Unscented Kalman Filter for nonlinear estimation. *Adaptive Systems for Signal Processing, Communications, and Control Symposium*, pages 153–158, 2000.

[60] C. Cruz-Hernández, J. Alvarez-Gallegos, and R. Castro-Linjares. Stability of discrete nonlinear systems under novanishing pertuabations: application to a nonlinear model-matching problem. *IMA Journal of Mathematical Control and Information*, 16(1):23–41, 1999.

[61] T. Kapitaniak. *Chaotic Oscillators: Theory and Applications.* World Scientific, 1992.

[62] R. Douc, O. Cappé, and E. Moulines. Comparison of resampling schemes for particle filtering. *CoRR*, abs/cs/0507025, 2005.

[63] S. Klus, F. Nüske, P. Koltai, H. Wu, I. Kevrekidis, C. Schütte, and F. Noé. Data-Driven Model Reduction and Transfer Operator Approximation. *Journal of NonLinear Science*, January 2018.

[64] B. O. Koopman. Hamiltonian systems and transformation in Hilbert space. *Proceedings of National Academy of Sciences*, 17:315–318, 1931.

[65] A. Lasota and M. Mackey. *Chaos, fractals, and noise: stochastic aspects of dynamics.* Applied mathematical sciences. Springer-Verlag, 1994.

[66] D. Williams, W. Kerstens, S. Buntain, V. Quach, J. Prieffer, R. King, G. Tadmor, and T. Colonius. Closed-loop control of a wing in an unsteady flow. *48th AIAA Aerospace Sciences Meeting*, 2010.

[67] D. Greenblatt, H. Muller-Vahl, S. Srimanta, D. Williams, and F. Reissner. Feedforward control for goman-khrabrov (g-k) model on pitching airfoils with flow control. *8th AIAA Flow Control Conference*, 2016.

[68] M. Goman and A. Khrabrov. State-space representation of aerodynamic characteristics of an aircraft at high angles of attack. *AIAA Journal*, 31(5):1109–1115, 1994.

[69] D. F. Gomez, F. D. Lagor, P. B. Kirk, A. H. Lind, A. R. Jones, and D. A. Paley. Data-driven estimation of the unsteady flowfield near an actuated airfoil. *Journal of Guidance, Control, and Dynamics*, 42(10):2279–2287, 2019.

[70] J. Anderson. *Fundamentals of Aerodynamics*. McGraw-Hill, New York, NY, 5 edition, 2007.

[71] Z. S. Chen. Bayesian filtering: From kalman filters to particle filters, and beyond. *Statistics: A Journal of Theoretical and Applied Statistics*, 182(1):1–69, 2003.

[72] J. E. Marsden and M. McCracken. *The Hopf Bifurcation and its Applications*. Springer-Verlag, New York, NY, 19 edition, 1976.

[73] T. Bengtsson, C. Snyder, and D. Nychka. Toward a nonlinear ensemble filter for high-dimensional systems. *Journal of Geophysical Research: Atmospheres*, 108(D24):n/a–n/a, 2003. 8775.

[74] P. L. Houtekamer and H. L. Mitchell. Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126(3):796–811, 1998.

[75] E. Ott, B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A*, 56(5):415–428, 2004.

[76] Numina mobility solutions. `https://numina.co`, 2019.

[77] E. Lorenz. Predictability: a problem partly solved. In *Seminar on Predictability, 4-8 September 1995*, volume 1, pages 1–18, Shinfield Park, Reading, 1995. ECMWF.

[78] R. W. Brockett. *Lie Algebras and Lie Groups in Control Theory*, pages 43–82. Springer Netherlands, Dordrecht, 1973.

[79] R. W. Brockett. System theory on group manifolds and coset spaces. *SIAM Journal on Control*, 10(2):265–284, 1972.

[80] V. G. Boltyanskiĭ, R. V. Gamkrelidze, and L. S. Pontryagin. On the theory of optimal processes. *Dokl. Akad. Nauk SSSR (N.S.)*, 110:7–10, 1956.

[81] M. Budisić, R. Mohr, and I. Mezić. Applied Koopmanism. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(4), 2012.

[82] K. Kowalski and W. H. Steeb. *Nonlinear Dynamical Systems and Carleman Linearization*, pages 73–102. World Scientific, 2011.

[83] M. Korda and I. Mezić. Optimal construction of Koopman eigenfunctions for prediction and control. *ArXiv e-prints*, 2018.

[84] M. H. Stone. The generalized Weierstrass approximation theorem. *Mathematics Magazine*, 21(4):167–184, 1948.

[85] R. Cooke and V. Arnold. *Ordinary Differential Equations.* Springer Textbook. Springer Berlin Heidelberg, 1992.

[86] H. Khalil. *Nonlinear Systems.* Pearson Education International Incorporated, 2015.

[87] R. F. Gantmacher. *Theory of Matrices.* Chelsea, New York, 1959.

[88] V. Jurdjevic and H. J. Sussmann. Control systems on Lie groups. *Journal of Differential Equations*, 12(2):313–329, 1972.

[89] J. Wei and E. Norman. On global representations of the solutions of linear differential equations as a product of exponentials. *Proceedings of the American Mathematical Society*, 15(2):327–334, 1964.

[90] J. Martin. Some results on matrices which commute with their derivatives. *SIAM Journal on Applied Mathematics*, 15(5):1171–1183, 1967.

[91] R. W. Brockett. *Finite Dimensional Linear Systems*, pages 67–122. John Wiley, New York, 1970.

[92] V. Vittaldev, R. P. Russell, and R. Linares. Spacecraft uncertainty propagation using Gaussian Mixture Models and Polynomial Chaos Expansions. *Journal of Guidance, Control, and Dynamics*, 39(12):2615–2626, 2016.

[93] K. J. DeMars and M. K. Jah. Probabilistic initial orbit determination using Gaussian Mixture Models. *Journal of Guidance, Control, and Dynamics*, 36(5):1324–1335, 2013.

Index