PROBABILISTIC METHODS FOR MODEL VALIDATION

A Dissertation

by

ABHISHEK HALDER

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Raktim Bhattacharya
Committee Members,	John L. Junkins
	Suman Chakravorty
	P. R. Kumar
Head of Department,	Rodney Bowersox

May 2014

Major Subject: Aerospace Engineering

ABSTRACT

This dissertation develops a probabilistic method for validation and verification (V&V) of uncertain nonlinear systems. Existing systems-control literature on model and controller V&V *either* deal with linear systems with norm-bounded uncertainties, *or* consider nonlinear systems in set-based and moment based framework. These existing methods deal with model *invalidation* or *falsification*, rather than assessing the quality of a model with respect to measured data. In this dissertation, an axiomatic framework for model validation is proposed in probabilistically relaxed sense, that instead of simply invalidating a model, seeks to quantify the "degree of validation".

To develop this framework, novel algorithms for uncertainty propagation have been proposed for both deterministic and stochastic nonlinear systems in continuous time. For the deterministic flow, we compute the time-varying joint probability density functions over the state space, by solving the Liouville equation via method-of-characteristics. For the stochastic flow, we propose an approximation algorithm that combines the method-of-characteristics solution of Liouville equation with the Karhunen-Loève expansion of process noise, thus enabling an indirect solution of Fokker-Planck equation, governing the evolution of joint probability density functions. The efficacy of these algorithms are demonstrated for risk assessment in Mars entry-descent-landing, and for nonlinear estimation. Next, the V&V problem is formulated in terms of Monge-Kantorovich optimal transport, naturally giving rise to a metric, called Wasserstein metric, on the space of probability densities. It is shown that the resulting computation leads to solving a linear program at each time of measurement availability, and computational complexity results for the same are derived. Probabilistic guarantees in average and worst case sense, are given for the valida-

tion oracle resulting from the proposed method. The framework is demonstrated for nonlinear robustness verification of F-16 flight controllers, subject to probabilistic uncertainties.

Frequency domain interpretations for the proposed framework are derived for linear systems, and its connections with existing nonlinear model validation methods are pointed out. In particular, we show that the asymptotic Wasserstein gap between two single-output linear time invariant systems excited by Gaussian white noise, is the difference between their average gains, up to a scaling by the strength of the input noise. A geometric interpretation of this result allows us to propose an intrinsic normalization of the Wasserstein gap, which in turn allows us to compare it with classical systems-theoretic metrics like ν-gap. Next, it is shown that the optimal transport map can be used to automatically refine the model. This model refinement formulation leads to solving a non-smooth convex optimization problem. Examples are given to demonstrate how proximal operator splitting based computation enables numerically solving the same. This method is applied for finite-time feedback control of probability density functions, and for data driven modeling of dynamical systems.

To Nilanjana

ACKNOWLEDGEMENTS

This thesis would not be possible without the generosity and help of many people, some who have directly contributed in shaping up this dissertation, and some whose way of thinking have indirectly influenced my growth as a researcher, and as a person. First and foremost, I thank my advisor Dr. Raktim Bhattacharya for agreeing to supervise my thesis at a difficult time, and providing me enough freedom to pursue my interests. I will cherish the countless hours of arguments we had over the years, on topics both technical and philosophical. When I look back, I realize, allowing this space and time to a graduate student is the rarest thing in today's frenzied life in academia. Thanks Raktim for all the candor and camaraderie.

I am indebted to my committee members for sharing many discussions. Dr. Chakravorty has helped me with numerous technical issues during this research, Dr. Junkins has shared his wisdom in pursuing career in academia, Dr. Kumar has provided insightful feedback on this work. I have learned a lot about the risk analysis in Mars entry-descent-landing (EDL) from Dr. J. (Bob) Balaram at NASA JPL, whom I had the privilege to work with during 2009-2010. Dr. Pramod Khargonekar at University of Florida was very kind to arrange a one-on-one meeting with me in Summer 2012, where he not only offered suggestions to improve certain technical results, but also pointed directions that helped focus my research. His infectious positive energy and encouragement were truly uplifting. On the same summer, I was hosted by Dr. Prashant Mehta at the Coordinated Science Lab at UIUC, as a visiting researcher for two months. I am grateful to Dr. Mehta for his kind support, and I look forward to continue this collaboration.

In Summer 2013, I interned at Eaton Corporation, Eden Prairie, Minnesota,

working on model based analysis and control problems in UPS, an area I had no background on. I thank my advisor Dr. Yigang Wang for patiently teaching me power electronics, and the entire Eaton GRT team at Eden Prairie for making me feel home. Special thanks must go to Dr. Wang and Dr. Vishal Mahulkar for allowing me to explain my dissertation research, and exploring collaborative opportunities.

Two courses at Texas A&M have influenced me the most in developing my taste in research. The *Approximation Theory* course by Dr. Ronald Devore, taught me how to visualize almost anything: be it Banach spaces, or Kolmogorov complexity. The *Stochastic Systems* course by Dr. Kumar was most relevant to my interests and offered exceptional clarity to many concepts. Outside classroom, I learned a lot from my interactions with fellow graduate students. Siming Zhao taught me La'Salle invariance principle and how to properly hold chopsticks, Mrinal Kumar taught me the Fokker-Planck equation, Parikshit Dutta taught me polynomial chaos and how to drive a car. I thank Baljeet, Kooktae, Eric, and Ali for many research discussions.

Credits must go to the staffs at the TAMU Aerospace Engineering department, specially Karen, Lisa, Rose, Jerry, Wanda and Mitchell, who ensured all my official paperwork and travel go smooth. This research was partially supported by funding from NASA JPL and NSF. I also thank Dr. Daniele Mortari for providing me the opportunity to be the TA for his Fall 2013 *Numerical Methods* course.

Last but not the least, I acknowledge my wife Nilanjana. Over all these years, whenever I felt dejected, she has always put an effort to cheer me up, and spent countless hours to make me stay positive. She is the reason I carry on. I dedicate this dissertation to her.

NOMENCLATURE

a.k.a.	Also Known As
approx.	Approximately
a.s.	Convergence in Almost Sure Sense
CDF	Cumulative Distribution Function
DMD	Dynamic Mode Decomposition
DQMOM	Direct Quadrature Method Of Moments
EDL	Entry-Descent-Landing
FPA	Flight Path Angle
GBM	Geometric Brownian Motion
gsLQR	Gain-scheduled Linear Quadratic Regulator
GWN	Gaussian White Noise
i.e.	*id est* (Latin), that is (English)
iff	if and only if
i.i.d.	(Statistically) independent and identically distributed
KL Expansion	Karhunen-Loève Expansion
KLMOC	"First KL, then MOC" Algorithm
KS Test	Kolmogorov-Smirnov Test
LP	Linear Program
LQR	Linear Quadratic Regulator
LTI	Linear Time Invariant
LTV	Linear Time Varying
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo

MIMO	Multiple Input Multiple Output
MISO	Multiple Input Multiple Output
MOC	Method Of Characteristics
MRI	Magnetic Resonance Imaging
m.s.	Convergence in Mean Square Sense
n/a	Not Applicable
NMR	Nuclear Magnetic Resonance
ODE	Ordinary Differential Equation
OU Process	Ornstein-Uhlenbeck Process
PC	Polynomial Chaos
PDE	Partial Differential Equation
PDF	Probability Density Function
PF Operator	Perron-Frobenius Operator
PMF	Probability Mass Function
POD	Proper Orthogonal Decomposition
PRVC	Probabilistically Robust Validation Certificate
PSD	Power Spectral Density
PWN	Poisson White Noise
PWVC	Probabilistically Worst-case Validation Certificate
QFPE	Quantile Fokker-Planck Equation
QMC	Quasi Monte Carlo
r.m.s.	Root Mean Square
SDE	Stochastic Differential Equation
SISO	Single Input Single Output
SOS	Sum Of Squares

SQP	Sequential Quadratic Program
TCI	Transportation Cost Inequality
v.s.	Versus
V&V	Validation and Verification
w.r.t.	With Respect To

Notation

$\mathbb{N}$	The set of natural numbers
$\mathbb{R}$	The set of real numbers
$\mathbb{R}^+$	The set of *positive* real numbers
$\mathbb{R}_0^+$	The set of *nonnegative* real numbers, $\mathbb{R}_0^+ \triangleq \mathbb{R}^+ \cup \{0\}$
$\mathbb{E}[\cdot]$	Expectation operator
$\mathrm{cov}(\cdot,\cdot)$	Covariance operator
$\mathbb{P}(\cdot)$	Probability of an event
$\sharp$	Push-forward of a probability measure
$^\top$	Matrix transpose
*	Complex conjugate
$^{\mathrm{H}}$	Matrix conjugate transpose
†	Moore-Penrose pseudo-inverse of a matrix
$\otimes$	Kronecker product
∇	Gradient operator w.r.t. spatial variables
$\mathrm{Hess}(\cdot)$	Hessian operator, $\mathrm{Hess}(\cdot) \triangleq \nabla\nabla^\top$
$\circ$	Composition operator
$\langle \cdot\,, \cdot \rangle_F$	Frobenius inner product
$\ker(.)$	Kernel of a linear operator

$\mathrm{Im}\,(.)$	Image of a linear operator
$\mathbf{0}$	Zero vector of appropriate dimensions
$\mathbf{1}$	Vector of ones, of appropriate dimensions
$\mathbb{1}$	Indicator function
I_n	Identity matrix of size $n \times n$
Id	Identity vector map of appropriate dimensions
$\mathrm{diag}\,(\cdot)$	Diagonal matrix of appropriate dimensions
$\mathrm{vec}\,(\cdot)$	Vectorization operator
$\mathrm{tr}\,(\cdot)$	Trace of a matrix
$\det(\cdot)$	Determinant of a matrix
$\mathrm{supp}\,(\cdot)$	Support of a function
$\Lambda_i(\cdot)$	i^{th} eigenvalue of a linear operator
$e_i(\cdot)$	i^{th} eigenfunction of a linear operator
$\lambda_i(\cdot)$	i^{th} eigenvalue of a matrix
$\lambda_{\max}(\cdot), \lambda_{\min}(\cdot)$	Maximum and minimum eigenvalue of a matrix
λ_p	Intensity parameter of Poisson counting process $N(t)$
$\mathrm{Re}\,(z)$	Real part of the complex number z
$\mathrm{Im}\,(z)$	Imaginary part of the complex number z
$\#$	Cardinality
$\wedge$	Minimum function, i.e. $a \wedge b =$ minimum of a and b
$\widehat{\ }$	Model variable
$\sim$	Distributional law, i.e. $X \sim \xi \Leftrightarrow X$ has joint PDF ξ
n_s	Dimension of the state space
n_o	Dimension of the output space
n_p	Dimension of the parameter space

n_w	Dimension of the process noise vector
n_v	Dimension of the measurement noise vector
t	Time, $t \in \mathbb{R}_0^+$
Δt	Time step
t_f	Final time
x	State vector, $x \in \mathbb{R}^{n_s}$
y	Output vector, $y \in \mathbb{R}^{n_o}$
p	Parameter vector, $p \in \mathbb{R}^{n_p}$
$\widetilde{x}$	Extended state vector, $\widetilde{x} = \begin{bmatrix} x & p \end{bmatrix}^\top \in \mathbb{R}^{n_s+n_p}$
x_0	Initial state vector, $x_0 \triangleq x\,(t=0)$
μ_x	State mean vector, $\mu_x \triangleq \mathbb{E}\left[x\right] \in \mathbb{R}^{n_s}$
μ_y	Output mean vector, $\mu_y \triangleq \mathbb{E}\left[y\right] \in \mathbb{R}^{n_o}$
Σ_x	State covariance matrix, $\Sigma_x \triangleq \mathbb{E}\left[\left(x-\mu_x\right)\left(x-\mu_x\right)^\top\right] \in \mathbb{R}^{n_s \times n_s}$
Σ_y	Output covariance matrix, $\Sigma_y \triangleq \mathbb{E}\left[\left(y-\mu_y\right)\left(y-\mu_y\right)^\top\right] \in \mathbb{R}^{n_y \times n_y}$
$\xi_0\left(\widetilde{x}_0\right)$	Joint PDF over initial states and parameters $\widetilde{x}_0 \triangleq \begin{bmatrix} x_0 & p \end{bmatrix}^\top$
$\xi\left(\widetilde{x}(t), t\right)$	Joint PDF over *true* extended state space $\widetilde{x}(t)$ at time t
$\widehat{\xi}\left(\widehat{\widetilde{x}}(t), t\right)$	Joint PDF over *model predicted* extended state space $\widehat{\widetilde{x}}(t)$ at time t
$\eta\left(y(t), t\right)$	Joint PDF over *true* output space $y(t)$ at time t
$\widehat{\eta}\left(\widehat{y}(t), t\right)$	Joint PDF over *model predicted* output space $\widehat{y}(t)$ at time t
$\mathcal{W}, \widehat{\mathcal{W}}$	True and model process noise vector, respectively
$\mathcal{V}, \widehat{\mathcal{V}}$	True and model measurement noise vector, respectively
h	Altitude
ζ	Latitude
λ	Longitude
V	Mars-relative velocity

γ	Flight path angle
χ	Velocity azimuth angle measured from North
σ_b	Bank angle
σ	Standard deviation
ρ	Martian atmospheric density
ρ_0	Reference-level density
B_c	Ballistic coefficient
$\frac{C_L}{C_D}$	Lift-to-drag ratio
R_0	Mean equatorial radius of Mars
GM	Gravitational constant for Mars
g	Acceleration due to gravity $\approx \frac{GM}{(R_0+h)^2}$
Ω_r	Rotational angular velocity for Mars
Ω_s	Sample space of a random vector
ω_s	Sample realization of a random vector, i.e. $\omega_s \in \Omega_s$
ω	Temporal frequency
v_c	Normalizing velocity constant $= \sqrt{\frac{\mu}{R_0}}$, where $\mu = gR_0^2$
$\delta(.)$	Dirac Delta
δ_{ij}	Kronecker Delta, i.e. $\delta_{ij} = 1$ if $i = j$, else zero
${}_rF_s$	Generalized hypergeometric function
$B(\alpha, \beta)$	Complete beta function, $B(\alpha,\beta) \triangleq \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\, dt$
$B(x;\alpha,\beta)$	Incomplete beta function, $B(x;\alpha,\beta) \triangleq \int_0^x t^{\alpha-1}(1-t)^{\beta-1}\, dt$
$I_x(\alpha,\beta)$	Regularized incomplete beta function, $I_x(\alpha,\beta) \triangleq \frac{B(x;\alpha,\beta)}{B(\alpha,\beta)}$
$I_t^{-1}(\alpha,\beta)$	Inverse of the beta CDF
$\Gamma(z)$	Gamma function, $\Gamma(z) \triangleq \int_0^\infty x^{z-1}e^{-x}\, dx$, $\mathrm{Re}(z) > 0$
$\Psi(z)$	Digamma function, $\Psi(z) \triangleq \frac{d}{dz}\log\Gamma(z)$

$\text{erf}^{-1}(\cdot)$	Inverse error function
wno(.)	Winding number
$\|\cdot\|_F$	Frobenius norm
$\mathcal{A}(\cdot)$	Arcsine PDF
$\mathcal{B}(\alpha, \beta)$	Beta PDF with parameters $\alpha, \beta > 0$, over $x \in [0, 1]$
$\mathcal{B}(a, b, \alpha, \beta)$	Shifted beta PDF with parameters $\alpha, \beta > 0$, over $x \in [a, b]$
$\mathcal{N}(\mu, \Sigma)$	Normal PDF with mean vector μ, covariance matrix Σ
$\mathcal{U}([a, b])$	Uniform PDF over $x \in [a, b]$

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

During the last two decades, engineering research and development has shifted emphasis from the conventional cycle of "designing, prototyping and testing" to "modeling and simulation", enabling faster and cheaper execution of a project from conceptualization to delivery. One major driving force behind model based design is that the modern engineering systems are becoming increasingly modular, and have complex interconnections in multiple spatial and temporal scales. Often, one has good scientific understanding of the individual components, but the collective behavior is not well understood from a design point of view. As a result, model based design and analysis are performed in mostly ad-hoc manner, and there is a need for foundational science that can guide towards scalable architecture for designing a large scale system. This dissertation concerns with a particular aspect of model based design and analysis: validation and verification (V&V). In this introductory Chapter, we first provide some background on the model validation problem, followed by a literature review. Then we describe the probabilistic model validation setup proposed in this dissertation, and finally list the contributions of the same.

1.1 Background on Model Validation

A model serves as a mathematical abstraction of the physical system, providing a framework for system analysis and controller synthesis. Since such mathematical representations are based on assumptions specific to the process being modeled, it is important to quantify the reliability to which the model is consistent with the physical observations. Model quality assessment is imperative for applications where the model needs to be used for prediction (e.g. weather forecasting, stock market) or safety-critical control design (e.g. aerospace, nuclear, systems biology) purposes.

Here it is important to realize that a model can only be validated against experimental observations, not against another model. Thus a *model validation problem* can be stated as: *given a candidate model and experimentally observed measurements of the physical system, how well does the model replicate the experimental measurements?* It has been argued in the literature [4, 5, 6, 7] that the term 'model validation' is a misnomer since it would take infinite number of experimental observations to do so. Hence the term 'model invalidation' or 'falsification' [8] is preferred. In this dissertation, instead of hard invalidation, we will consider the validation/invalidation problem in a probabilistically relaxed sense. Before proceeding further, it may be helpful to put the model validation and refinement framework considered in this dissertation, in a wider scientific context. We list here two examples from physics, to argue that validation and refinement are constructive iterative processes [9, 10] fundamental to scientific and technological advancement.

One of the early examples of model validation/invalidation came from Nicolaus Copernicus in 1543, who proposed the heliocentric model that opposed the century-old geocentric model proposed by Aristotle and Ptolemy. However, experimental *validation* for Copernicus's model had to wait until the invention of telescope in 17^{th} century that led to Galileo's observation of four moons of Jupiter, and phases of Venus in 1610, and Giovanni Zupi's observation of phases of Mercury in 1639. Based on the astronomical observation of Tycho Brahe, the Copernicus's model was *refined* by Johannes Kepler by proposing elliptical orbits for planets in 1609. Kepler's laws were later explained through Newton's law of gravitation in 1687. It is instructive to note that no single observation was enough to validate a model or to gain its acceptance.

Another example of the recursive model (in)validation and refinement comes from atomic physics. Following the discovery of electrons in 1897, J.J. Thomson proposed

the plum pudding model of atom, advocating that the negatively charged electrons are embedded in a positively charged "pudding" inside an atom. In 1909, this model was *invalidated* by the goldfoil experiment of Hans Geiger and Ernest Marsden. Thomson's plum pudding model was *refined* by Ernest Rutherford in 1911, upon the discovery of nucleus. Rutherford model was in turn *refined* by Neils Bohr in 1913 to explain the stability of the motions of subatomic particles, and eventually led to the development of quantum mechanics.

In 21^{st} century, technological progress has arguably outpaced the rate at which scientific theories are proposed or refined. Unlike Copernicus, we now have an abundance of observational data, thanks to the tight integration of control, communication and computation in modern lifestyle. However, a foundational science of model validation and refinement is lacking in systems-control literature, that can account this data deluge in a disciplined manner, without sacrificing the mathematical rigor. This dissertation is an attempt to address that.

1.2 Related Literature

Broadly speaking, there have been three distinct frameworks in which the model validation problem has been attempted till now. *First* is a discrete formulation in *temporal logic framework* [11] which has been extended to account probabilistic models [11, 12]. *Second* is the $\mathcal{H}_\infty$ *control framework* where time-domain [6, 13, 14], frequency domain [5, 15] and mixed domain [16] model validation methods have been studied extensively assuming structured norm-bounded uncertainty in linear dynamics setting. The *third* framework involves deductive inference based on barrier certificates [7] which was shown to encompass a large class of nonlinear models including differential-algebraic equations [17], dynamic uncertainties described by integral quadratic constraints [18], stochastic [19] and hybrid dynamics [20].

In statistical setting, model validation has been addressed from system identification perspective [21, 22] where the main theme is to validate an identified nominal model through correlation analysis of the residuals. A polynomial chaos framework has also been proposed [23] for model validation. Gevers *et. al.* [24] have connected the robust control framework with prediction error based identification for frequency-domain validation of linear systems. In another vein, using Bayesian conditioning, Lee and Poolla [25] showed that for *parametric* uncertainty models, the statistical validation problem may be reduced to the computation of relative weighted volumes of convex sets. However, for *nonparametric* models: "the situation is significantly more complicated" [25] and to the best of our knowledge, has not been addressed in the literature. Recently, in the spirit of weak stochastic realization problem [26], Ugrinovskii [27] investigated the conditions for which the output of a stochastic nonlinear system can be realized through perturbation of a nominal stochastic *linear* system.

In practice, one often encounters the situation where a model is either proposed from physics-based reasoning or a reduced order model is derived for computational convenience. In either case, the model can be linear or nonlinear, continuous or discrete-time, and in general, it's not possible to make any a-priori assumption about the noise. Given the experimental data and such a candidate model for the physical process, our task is to answer: "to what extent, the proposed model is valid?" In addition to quantify such degree of validation, one must also be able to demonstrate that the answer is *provably correct* in the face of uncertainty. This brings forth the notions of *probabilistically robust model validation* and *probabilistically worst-case model validation*. In this dissertation, we will propose to construct such a *robust validation certificate*, and *worst-case validation certificate*, guaranteeing the performance of probabilistic model validation algorithm.

1.3 Problem Setup

1.3.1 Intuitive Idea

The proposed framework is based on the evolution of densities in output space, instead of evolution of individual trajectories, as in the Lyapunov framework. Intuitively, characteristics of the input to output mapping is revealed by the growth or depletion of trajectory concentrations in the output space. Growth in concentration, or increased density, defines regions in where the trajectories accumulate. This corresponds to regions with slow time scale dynamics or time invariance. Similarly, depletion of concentration in a set implies fast-scale dynamics or unstable manifold. We refer the readers to [28] for an introduction to analysis of dynamical systems using trajectory densities. This idea of comparing dynamical systems based on density functions, have been presented before by Sun and Mehta [29] in the context of filtering, and by Georgiou [30] in the context of matching power spectral densities.

Given the experimental measurements of the physical system in the form of a time-varying distribution (such as histograms), we propose to compare the *shape* or *concentration profile* of this measured output density, with that predicted by the model. At every instant of time, if the model-predicted density matches with the experimental one "*reasonably well*" (to be made precise later in the paper), we conclude that the model is validated with high *confidence* (to be computed for guaranteeing quality of inference).

1.3.2 Why Compare Densities Instead of Trajectories

The rationale behind comparing the distributional shapes for model validation comes from the fact that the presence of uncertainties mask the difference between individual output realizations. Uncertainties in initial conditions, parameters and noise result different realizations of the trajectory or integral curve of the dynamical

system. Regions of high (low) concentration of trajectories correspond to regions of high (low) probability. Thus a model validation procedure should naturally aim to compare concentrations of the trajectories between the measurements and model-predictions, instead of comparing individual realizations of them, which would be meaningful only in the absence of uncertainties.

We would like to point out that in some applications, the measurement naturally arises in the form of a distribution. This includes (i) *process industry applications* like measurement made at the wet end of papermaking machines [31, 32] that involves the fibre length and filler size distribution sensed via vision sensors, (ii) *Nuclear Magnetic Resonance (NMR) spectroscopy and Imaging (MRI) applications* where the measurement variable is magnetization distribution [33], (iii) *neuroscience applications* where the measurement variable is the distribution of frequency across a collection of neurons [34], and (iv) *social systems* where the measurement variable could be an ensemble of crowd [35] sensed via cameras or motion detectors. Notice that for (i) and (iii), distributional measurement is a design choice; for (ii) it is motivated by technological limitations of sensing individual magnetization states where the number of states are of the order of Avogadro number 6×10^{23}; and for (iv) individual measurement may raise privacy concerns.

1.3.3 Why Compare Densities Instead of Moments or Sets

Density based model validation provides natural advantages over moment based or set containment methods for the following reasons. Moment based methods can be erroneous for nonlinear non-Gaussian systems, as two different trajectory densities may provide the same correlation information. This can be circumvented by including higher order moments, but such a strategy is computationally untractable since for a nonlinear dynamical system with non-Gaussian PDF evolution, the number of

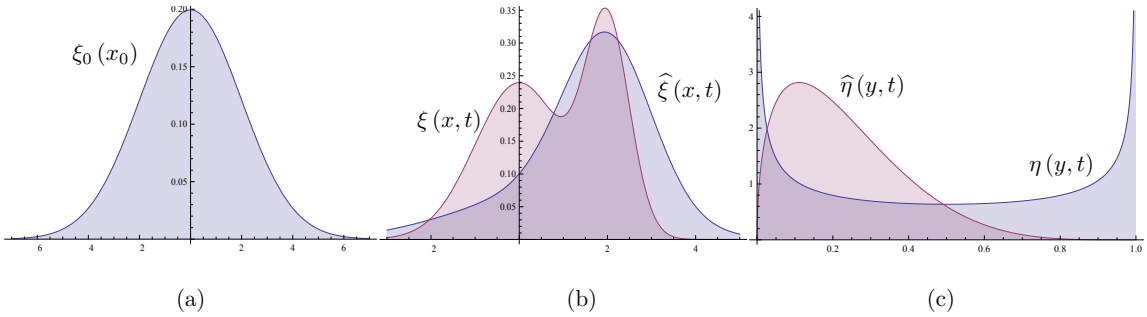


Figure 1.1: The proposed model validation framework compares experimentally observed output PDF $\eta(y,t)$ with the model-predicted output PDF $\widehat{\eta}(\widehat{y},t)$, the comparison being made with respect to some suitable metric at each instant of measurement availability. The state dynamics evolves the initial joint PDF $\xi_0(x_0)$ (Fig. 1.1 (a)) to instantaneous joint state PDFs $\xi(x,t)$ and $\widehat{\xi}(\widehat{x},t)$ (Fig. 1.1 (b)). The associated output PDFs $\eta(y,t)$ and $\widehat{\eta}(y,t)$ may share the same support ($[0,1]$ as shown in Fig. 1.1 (c)), but have different shapes. Hence, instead of matching output supports, we propose matching output PDFs at all times, for validating a model.

admissible moments at any fixed time, is not known *a priori*. This is typically referred as the "moment closure problem". On the other hand, set containment arguments can be erroneous since it is possible that at a given time, two systems have trajectory densities with identical supports but different concentrations (Fig. 1.1 (c)). In Fig. 1.1 and thereafter, we use the notation $\xi_0(.)$ to denote the joint PDF over initial conditions (x_0) and parameters. At time $t > 0$, $\xi(.,t)$ and $\widehat{\xi}(.,t)$ denote joint PDFs over instantaneous states and parameters, for the true and model dynamics, respectively. Similarly, $\eta(.,t)$ and $\widehat{\eta}(.,t)$, respectively denote joint PDFs over output spaces y and $\widehat{y}$ at time t, for the true and model dynamics. The symbol $\widetilde{x}$ is used to denote the extended state vector obtained by augmenting the state (x) and parameter (p) vectors.

A proposed model is validated, if the "distance" between its predicted density and the measured density, remains below a user-specified tolerance level, which need

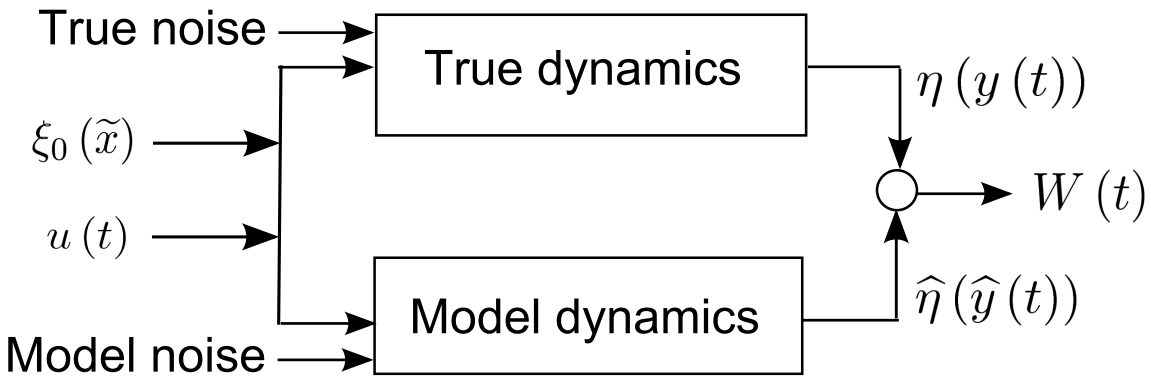


Figure 1.2: Block diagram for the proposed model validation formulation.

not be fixed over time. For example, take-off and landing are critical operational segments during the flight of a commercial aircraft, and it's unacceptable to have a controller that does not guarantee the robust performance for these critical time-segments with very high probability. This motivates the computation of probability of validation as part of the model validation oracle.

1.3.4 Methodology and Organization

In this subsection, we formalize the ideas presented above. Fig. 1.1 and 1.2 show the outline of the model validation framework proposed here. In this formulation, the systems under comparison are excited with a *known* input signal $u\left(t\right)$, and an initial PDF $\xi_0\left(\widetilde{x}_0\right)$, supported over the extended state space $\widetilde{x} := \{x,\ p\}^\top$, where the states $x \in \mathbb{R}^{n_s}$, and the parameters $p \in \mathbb{R}^{n_p}$. Given the PDF $\eta\left(y\left(t\right)\right)$ supported over the true output space $y \in \mathbb{R}^{n_o}$, and a candidate model, we compute and then compare the model predicted output PDF $\widehat{\eta}\left(\widehat{y}\left(t\right)\right)$, with $\eta\left(y\left(t\right)\right)$ at each instances of measurement availability $\{t_j\}_{j=1}^{\tau}$. Thus, one can think of two main steps of such

a model validation framework. These are:

1. evolving $\xi_0(\widetilde{x}_0)$ using the proposed model, to compute $\widehat{\eta}(\widehat{y}(t))$,

2. measuring an appropriate notion of distance, denoted as $W(t)$ in Fig. 1.2, between $\eta(y(t))$ and $\widehat{\eta}(\widehat{y}(t))$ at $\{t_j\}_{j=1}^{\tau}$.

The first step, namely the evolution of uncertainty in the form of joint PDF, subject to a nonlinear dynamics, will be treated in Chapter 2 and 3. The second step, namely the notion of distributional comparison and the construction of validation certificates, will be the topic of Chapter 4. Some case studies of the proposed probabilistic V&V framework will be presented in Chapter 5, followed by systems-theoretic results in Chapter 6. Chapter 7 will extend the proposed V&V framework to automatically refine the models based on data. Chapter 8 will conclude this dissertation.

1.4 Contributions of This Dissertation

With respect to the existing V&V literature, the contributions of this dissertation are as follows.

1. Instead of interval-valued structured uncertainty (as in $\mathcal{H}_\infty$ control framework) or moment based uncertainty (as in parametric statistics framework), this paper deals with model validation in the sense of nonparametric statistics. Uncertainties in the model are quantified in terms of the PDFs of the associated random variables. We argue that such a formulation offers several advantages. *Firstly*, we show that model uncertainties in the parameters, initial states and input disturbance, can be propagated accurately by spatio-temporally evolving the joint state and output PDFs. Since experimental data usually come in the form of histograms, it's a more natural quantification of uncertainty than specifying sets [7] to which the trajectories are contained at each instant of

time. However, if needed, such sets can be recovered from the supports of the instantaneous PDFs. *Secondly*, as we'll see in Chapter 4, instead of simply invalidating a model, our methodology allows to estimate the probability that a proposed model is valid or invalid. This can help to decide which specific aspects of the model need further refinement. Hard invalidation methods don't cater such constructive information. *Thirdly*, the framework can handle both discrete-time and continuous-time nonlinear models which need not be polynomial. Previous work like [7] dealt with semi-algebraic nonlinearities and relied on sum of squares (SOS) decomposition [36] for computational tractability. From an implementation point of view, the approach presented in this paper doesn't suffer from such conservatism.

2. Due to the uncertainties in initial conditions, parameters, and process noise, one needs to compare output ensembles instead of comparing individual output realizations. This requires a metric to quantify closeness between the experimental data and the model in the sense of distribution. We propose *Wasserstein distance* to compare the output PDFs and argue why commonly used information-theoretic notions like *Kullback-Leibler divergence* may not be appropriate for this purpose.

3. We show that the uncertainty propagation through continuous-time dynamics, for both deterministic and stochastic case, can be done via numerically efficient meshless algorithms, even when the model is high-dimensional and strongly nonlinear. Moreover, we outline how to compute the Wasserstein distance in such settings. Further, bringing together ideas from analysis of randomized algorithms, we give sample-complexity bounds for robust validation inference.

Table 1.1: List of peer-reviewed publications resulting from this dissertation research.

Chapters	Conference papers	Journal papers
2	[37]	[38]
3	[39, 40]	n/a
4	[41]	[42]
5	[43]	[44]
6	[45, 46]	n/a
7	[47]	n/a

The peer-reviewed publications resulting from this dissertation research, are summarized in Table 1.1. Next, we will focus on the first step of our model validation formulation, namely uncertainty propagation. The forthcoming Chapter 2 will deal with uncertainty propagation for *deterministic flow*.

2. UNCERTAINTY PROPAGATION FOR DETERMINISTIC FLOW

In this Chapter, we consider the propagation of uncertainties in initial conditions and parameters, subject to a *deterministic* nonlinear dynamics. As outlined in the previous Chapter, our motivation to solve the uncertainty propagation problem stems from model validation. However, in many science and engineering applications, uncertainty propagation itself is a problem of interest, for predictive risk assessment. Such applications include quantifying the landing footprint uncertainty in Mars landing [48, 49], weather prediction [50, 51], uncertainty quantification in fluid dynamics computations [52], and seismic loss prediction [53].

Traditionally, a Monte Carlo (MC) based dispersion analysis is carried out for this purpose where one simulates a large number of trajectories for randomly sampled initial conditions and parameter values. If most or all of such trajectories remain inside the safety margin, one can at best hope for the system safety and reliability without any quantitative guarantee whatsoever. Usually the engineers responsible for subsystem models identify the uncertainty bounds and decide about the sampling strategy based on their experience. Clearly, brute force MC simulation is not the best approach for safety critical uncertainty analysis. Moreover, for high dimensional and nonlinear dynamics like spacecraft entry-descent-landing (EDL), MC simulations are tremendously expensive as one strives to simulate individual trajectories one by one for uncertainties in hundreds of states and parameters and their combinations. In spite of all these drawbacks, MC simulations remain the state-of-the-art tool for dispersion analysis among the engineers for two reasons. *First*, the ease of implementation and *second*, lack of any rigorous alternative which is scalable and accurate.

Being aware of the computational scalability problem of MC simulations (see e.g. Chapter 1 in [2]), researchers have pursued different methods for uncertainty propagation in nonlinear dynamical systems. We will classify these methods in two broad categories: *parametric* (where one evolves the statistical moments) and *non-parametric propagation of uncertainty* (where one evolves the full PDF). There have been three major directions in parametric propagation of uncertainty, as listed below.

1. The simplest method in this category assumes a *linear time invariant* description of the dynamics of the form $\dot{x}(t) = Ax(t)$, where the state vector at time $t > 0$ is $x(t) \in \mathbb{R}^{n_s}$. One then sets for propagating the state mean vector $\mu_x(t)$, and covariance matrix $\Sigma_x(t)$, using the well known (see Appendix A) equations $\dot{\mu}_x(t) = A\mu_x(t)$ and $\dot{\Sigma}_x(t) = A\Sigma_x(t) + \Sigma_x(t)A^\top$. If the initial uncertainty is Gaussian, then evolving the mean and covariance is indeed enough (see Appendix A). This is because of two facts: (i) the dynamics $\dot{x}(t) = Ax(t)$ preserves Gaussianity, and (ii) mean and covariance are sufficient statistics for a Gaussian PDF. However, the major drawback of this approach is that most dynamical systems of practical interest (e.g Mars EDL) have nonlinear dynamics in the trajectory level, and even if the initial PDF is Gaussian, the nonlinearity manifests in the non-Gaussian joint PDF evolution. As a result, the linear Gaussian framework is too far to fit for practical uncertainty propagation. Nevertheless, this framework has been attempted in nonlinear problems like EDL (see Chapter 13, p. 425–475 in [54]).

2. In *polynomial chaos* (PC) method, one derives a set of deterministic ODEs using either Galerkin projection [55] or the stochastic collocation [56] and then solves that set of ODEs. Although this method can handle nonlinear dynamics with non-Gaussian uncertainties, one ends up solving a higher dimensional

state space problem, which becomes intractable for a realistic problems with large number of states and parameters. Further, the method is difficult to apply for high nonlinearities [57] and computational performance degrades due to the finite-dimensional approximation of the probability space, if long-term statistics is desired.

3. Another method in this category is called the *direct quadrature method of moments* (DQMOM) [58] where the PDF is approximated as a sum of Dirac delta functions with evolving parameters. This method suffers from the Hausdorff moment problem (see Chapter 1, p. 1–22 in [59]). There are some variations of this method where the PDF is expressed as a weighted sum of few constituent PDFs, referred as 'partial PDF's and propagates them instead of the Dirac delta functions [60].

On the other hand, non-parametric propagation of uncertainties can be done in two ways: *approximate method* and *direct method*.

1. *Approximate method* is one where one tries to estimate (in nonparametric sense) the underlying PDF. The method aims to approximate the solution of the PDF transport equation. This method is widely exercised in statistics community [61] under the name of *kernel density estimation*, although most applications there concern with static data. In a dynamical system, optimal values of the parameters must be determined at every instant of time. Many special cases of this, may be constructed depending on the type of kernel function and the criteria for optimization. A least-square error minimization set up is described in [62] and is shown to have good computational performance. Further generalizations are possible by considering general basis functions which are not density functions themselves. However, this too can suffer from high computational

cost arising due to the explicit enforcement of normality constraint and moment closure constraint at each step of the optimization procedure. Moreover, for high dimensional state spaces, recursively performing constrained optimization becomes extremely challenging.

2. In *direct method*, one works with the PDF transport equation and instead of approximating its solution, strives to solve that equation directly. In the absence of process noise (which is the case for deterministic flow), this transport equation reduces to the *stochastic Liouville equation* [50], which is a quasi-linear PDE, first order in both space and time. This equation describes the time evolution of the joint PDF over the state space, which itself is changing due to the known dynamics. In this Chapter, we argue that the stochastic Liouville equation can be easily solved in such direct way using the method of characteristics (MOC). Since all the statistics can be derived from the PDF, from an information point of view, it is definitely superior than parametric propagation methods. Further, since we will be looking to solve the Liouville equation directly, the solutions will automatically satisfy the criteria to be PDF. Hence the conditions like moment closure or normality constraints are not required to be enforced explicitly.

Our objective in this Chapter is to demonstrate that an MOC computation to solve the Liouville equation is numerically attractive and it does provide the necessary rigor for a statistically consistent uncertainty quantification for nonlinear non-Gaussian problems. To see why it is the case, one must realize that in solving the Liouville equation, one propagates the joint PDF prescribed at the initial time subject to the deterministic dynamics. In MC method, one randomly picks a single initial condition and computes the trajectory and then repeats the process. In

the Liouville equation method, instead of *individual realizations* (initial conditions and/or parameters), one propagates the *ensemble of realizations*. Just like the continuity equation in fluid mechanics transports the fluid mass in configuration space, the Liouville equation transports the probability mass in phase space.

This Chapter is organized as follows. The Liouville equation will be described in the next Section along with the MOC formulation. The framework will be discussed through illustrative examples and numerical algorithm. Then, the following Section will present a case study to demonstrate the applicability of the proposed method in a realistic problem setting. We have chosen the case study as the risk analysis for Mars hypersonic entry. The last section will be devoted to compare the numerical performance of the proposed method vis-a-vis with MC simulations.

2.1 The Liouville Equation

Consider the continuous-time nonlinear model with state dynamics given by the ODE $\dot{\widehat{x}} = \widehat{f}(\widehat{x}, \widehat{p})$, where $\widehat{x}(t) \in \widehat{\mathcal{X}} \subseteq \mathbb{R}^{\widehat{n}_s}$ is the state vector, $\widehat{p} \in \widehat{\mathcal{P}} \subseteq \mathbb{R}^{\widehat{n}_p}$ is the parameter vector, the dynamics $\widehat{f}(., \widehat{p}) : \widehat{\mathcal{X}} \mapsto \mathbb{R}^{\widehat{n}_s}\ \forall\, \widehat{p} \in \widehat{\mathcal{P}}$, and is at least locally Lipschitz . It can be put in an extended state space form

$$\dot{\widehat{\widetilde{x}}} = \widehat{\widetilde{f}}\left(\widehat{\widetilde{x}}\right),\ \widehat{\widetilde{x}} \in \widehat{\mathcal{X}} \times \widehat{\mathcal{P}} \subseteq \mathbb{R}^{\widehat{n}_s + \widehat{n}_p},\ \widehat{\widetilde{f}} = \left\{ \begin{matrix} \widehat{f}_{\widehat{n}_s \times 1} \\ \mathbf{0}_{\widehat{n}_p \times 1} \end{matrix} \right\}. \tag{2.1}$$

The output equation can be written as

$$\widehat{y} = \widehat{h}\left(\widehat{\widetilde{x}}\right), \quad \widehat{h} : \widehat{\mathcal{X}} \times \widehat{\mathcal{P}} \mapsto \widehat{\mathcal{Y}}, \tag{2.2}$$

where $\widehat{y}(t) \in \widehat{\mathcal{Y}} \subseteq \mathbb{R}^{n_o}$ is the output vector. If uncertainties in the initial conditions ($x_0 := x(0)$) and parameters ($\widehat{p}$) are specified by the initial joint PDF $\xi_0(\widetilde{x})$, then

the evolution of uncertainties subject to the dynamics (2.1), can be described by evolving the joint PDF $\widehat{\xi}\left(\widehat{\widetilde{x}},t\right)$ over the extended state space. Such spatio-temporal evolution of $\widehat{\xi}\left(\widehat{\widetilde{x}},t\right)$ is governed by the *Liouville equation* given by (see Section 7.6 in [28])

$$\frac{\partial\widehat{\xi}}{\partial t} \;=\; \mathscr{L}_{\mathrm{LE}}\widehat{\xi} = D_1\widehat{\xi} = -\nabla.\left(\widehat{\xi}\widehat{f}\right) = -\sum_{i=1}^{\widehat{n}_s}\frac{\partial}{\partial\widehat{x}_i}\left(\widehat{\xi}\widehat{f}_i\right), \tag{2.3}$$

which is a quasi-linear PDE, first order in both space and time. Notice that, the spatial Liouville operator $\mathscr{L}_{\mathrm{LE}}$ is a drift operator D_1 that describes the *advection* of the PDF in extended state space. The Liouville equation (2.3) essentially states that the spatio-temporal evolution of the joint PDF occurs in a way that preserves the total probability mass. Putting differently, if we pick up a control volume in the extended state space, the net flux of probability mass must be zero since no realizations are created or destroyed (no source or sink).

Once (2.3) is solved, the output PDF $\widehat{\eta}\left(\widehat{y},t\right)$ can be computed from the state PDF as

$$\widehat{\eta}\left(\widehat{y},t\right) = \sum_{j=1}^{\nu}\frac{\widehat{\xi}\left(\widehat{\widetilde{x}}_j^{\star}\right)}{\left|\det\left(\mathcal{J}\left(\widehat{\widetilde{x}}_j^{\star}\right)\right)\right|}, \tag{2.4}$$

where $\widehat{\widetilde{x}}_j^{\star}$ is the j^{th} root of the inverse transformation of (2.2) with $j = 1, 2, \ldots, \nu$, and $\mathcal{J}$ is the Jacobian of this inverse transformation.

2.1.1 Method-of-Characteristics Formulation

In this section, we briefly describe the MOC and show how that helps in reducing a linear or quasi-linear PDE to an ODE along the characteristics. Application of MOC to nonlinear PDEs can be found in [63] (Chapter 2, p. 36–41).

Consider a PDE of the form

$$\sum_{i=1}^{n} a_i \left(z_1, z_2, \ldots, z_n, \vartheta\right) \frac{\partial \vartheta}{\partial z_i} = \Gamma \left(z_1, z_2, \ldots, z_n, \vartheta\right) \tag{2.5}$$

with ϑ being the dependent variable and $z_1, z_2, \ldots, z_n$ are the n independent variables. The characteristic curves corresponding to (2.5) are given by the Lagrange-Charpit equations [64]

$$\frac{dz_1}{a_1 \left(z_1, z_2, \ldots, z_n, \vartheta\right)} = \ldots = \frac{dz_n}{a_n \left(z_1, z_2, \ldots, z_n, \vartheta\right)} = \frac{d\vartheta}{\Gamma \left(z_1, z_2, \ldots, z_n, \vartheta\right)}. \tag{2.6}$$

Geometrically, this means that the $(n+1)$ dimensional vector field

$$\mathscr{F} := \left(a_1 \left(z_1, z_2, \ldots, z_n, \vartheta\right), \ldots, a_n \left(z_1, z_2, \ldots, z_n, \vartheta\right), \Gamma \left(z_1, z_2, \ldots, z_n, \vartheta\right)\right)$$

is tangent to the surface $\vartheta = \vartheta \left(z_1, z_2, \ldots, z_n\right) \; \forall \left\{z_1, z_2, \ldots, z_n\right\}^{\top} \in \mathbb{R}^n$. In other words, the solution of the PDE (2.5) is an $(n+1)$ dimensional surface $\vartheta \left(z_1, z_2, \ldots, z_n\right)$ that can be constructed as the union of the integral curves (or characteristic curves given by (2.6)) of the vector field $\mathscr{F}$.

To derive the characteristic curves for the Liouville equation, we put (2.3) in a form similar to (2.5), using product rule of differentiation, to obtain

$$\left(\sum_{i=1}^{\widehat{n}_s} \widehat{f}_i \frac{\partial \widehat{\xi}}{\partial \widehat{x}_i}\right) + \frac{\partial \widehat{\xi}}{\partial t} = -\widehat{\xi} \sum_{i=1}^{\widehat{n}_s} \frac{\partial \widehat{f}_i}{\partial \widehat{x}_i}. \tag{2.7}$$

From (2.6), it readily follows that the characteristic curves for (2.7) are given by

$$\frac{d\widehat{x}_1}{\widehat{f}_1} = \frac{d\widehat{x}_2}{\widehat{f}_2} = \ldots = \frac{d\widehat{x}_{\widehat{n}_s}}{\widehat{f}_{\widehat{n}_s}} = \frac{dt}{1} = \frac{d\widehat{\xi}}{-\widehat{\xi} \sum_{i=1}^{\widehat{n}_s} \frac{\partial \widehat{f}_i}{\partial \widehat{x}_i}}. \tag{2.8}$$

The equation above shows that the characteristic curves for the Liouville equation are nothing but the trajectories of the dynamics given by (2.1). Notice that the parameters, by definition, do not change with time, i.e. $\dfrac{d\widehat{p}_i}{dt} = 0$, $\forall i = 1, \ldots, \widehat{n}_p$.

It can be noted from (2.8) that using MOC, along the trajectory, one can reduce the Liouville PDE (2.3) to an ODE of the form

$$\frac{d\widehat{\xi}}{dt} = -\widehat{\xi}\, \nabla \cdot \widehat{f}, \tag{2.9}$$

where $\nabla \cdot \widehat{f}$ is the trace of the Jacobian of the underlying dynamics and hence, evolves with time. If the initial state and parametric uncertainties are specified in terms of a joint PDF $\xi_0\left(\widetilde{x}_0\right) \triangleq \widehat{\xi}\left(\widehat{\widetilde{x}}\left(0\right), 0\right)$, then we can write the solution of (2.9) as

$$\widehat{\xi}\left(\widehat{x}, t\right) = \xi_0\left(\widetilde{x}_0\right) \, \exp\left(- \int_0^t \nabla \cdot \widehat{f}\left(\widehat{\widetilde{x}}\left(\tau\right)\right) \, d\tau \right). \tag{2.10}$$

The exponential in (2.10) is formally known as the *ordered exponential* [65] and is analogous to the Dyson operator of the *quantum Liouville equation* [66] in statistical quantum mechanics. Since the ordered exponential is a ratio of the instantaneous and initial PDFs, one may interpret it as a likelihood ratio [67].

Notice that $\nabla \cdot \widehat{f}$ being the divergence of the vector field $\widehat{f}$, is a measure of the rate of change of the phase space (Lebesgue) volume. For example, if $\widehat{f}$ is linear time invariant, then $\nabla \cdot \widehat{f}$ must be a constant. Depending on the sign of this constant, the phase space volume can expand (expanding flow) or contract (contractive flow) exponentially fast or may remain constant (volume-preserving flow). The Liouville Theorem [68] tells us that the case of divergence-free vector field ensures that the system is Hamiltonian. Notice that, a nonlinear vector field can be Hamiltonian too.

In general, it is hardly possible to analytically evaluate the integral in (2.10) and

thus mandates numerical solution. Once the solution for the joint PDF $\widehat{\xi}\left(\widehat{\widetilde{x}}\left(t\right),t\right)$ is obtained, one can find the marginal PDFs by integrating out the other states over their respective domains, namely

$$\widehat{\xi_i} \triangleq \widehat{\xi}\left(\widehat{\widetilde{x}}_i,t\right) = \int_{\mathcal{D}_1} ... \int_{\mathcal{D}_{i-1}} \int_{\mathcal{D}_{i+1}} ... \int_{\mathcal{D}_{\widehat{n}_s+\widehat{n}_p}} \widehat{\xi}\left(\widehat{\widetilde{x}},t\right) d\widehat{\widetilde{x}}_1 ... d\widehat{\widetilde{x}}_{i-1} d\widehat{\widetilde{x}}_{i+1} ... d\widehat{\widetilde{x}}_{\widehat{n}_s+\widehat{n}_p} \quad (2.11)$$

where $\mathcal{D}_i$ is the domain of the i^{th} extended state variable at time t. Here it's important to realize that since the domain in the state space is deforming with time, one must know the instantaneous domain to carry out the integration in (2.11). This will be explained in more details later in this Chapter, when we perform the numerical simulations for our case study.

2.1.2 Examples

We now provide some examples to clarify the ideas presented above. Specifically, we want to illustrate how MOC enables the solution of Liouville PDE (2.3) by solving the ODE initial value problem (2.9) along the trajectories. For ease of understanding, we drop the "hat" notation for the time-being.

2.1.2.1 1D Example

Let's consider the 1D dynamics $\dot{x}(t) = f\left(x(t)\right) = -x^2$, with initial condition $x\left(0\right) = x_0$. By direct integration, we can solve this initial value problem as

$$x\left(x_0,t\right) = \frac{x_0}{1+tx_0}. \quad (2.12)$$

Consequently, we have

$$\nabla \cdot f \;\; = \;\; -2x\left(t\right) = -\frac{2x_0}{1+tx_0}$$

$$\Rightarrow \exp\left(-\int_0^t \nabla \cdot f\left(x\left(\tau\right)\right) d\tau\right) = \exp\left(2x_0 \int_0^t \frac{d\tau}{1+\tau x_0}\right) = \left(1+tx_0\right)^2 \quad (2.13)$$

which, from (2.10), leads to

$$\xi\left(x,t\right) = \xi_0\left(x_0\right)\ \left(1+tx_0\right)^2. \quad (2.14)$$

Now we can find $x_0 = x_0\left(x,t\right)$ using (2.12) as

$$x_0 = \frac{x}{1-tx}, \quad (2.15)$$

and substitute this to (2.14) to yield

$$\xi\left(x,t\right) = \xi_0\left(\frac{x}{1-tx}\right)\ \left(1+\frac{tx}{1-tx}\right)^2 = \frac{\xi_0\left(\frac{x}{1-tx}\right)}{\left(1-tx\right)^2}. \quad (2.16)$$

Fig. 2.1 shows the spatio-temporal evolution of the PDF $\xi\left(x,t\right)$ according to (2.16), when the initial PDF ξ_0 is chosen to be a standard normal distribution $\mathcal{N}\left(0,1\right)$. The plot illustrates the rise of the PDF peak with time, accompanied with a shrinkage of its support. As $t \to \infty$, the PDF tends to become a Dirac delta. This is not surprising since the origin being the unique equilibrium of this dynamics, in the asymptotic limit, all probability mass gets in that "sink". Thus the support of the stationary distribution has Lebesgue measure zero.

Another interesting observation from Fig. 2.1 is the existence of multiple peaks, which become more prominent as t increases. This is due to the fact that the vector field $\dot{x} = -x^2$ is locally Lipschitz and the unique solution (2.12) has finite escape time at $t = -\frac{1}{x_0}$. Since $t \in \mathbb{R}_0^+$, the finite escape time comes into play for $x_0 < 0$. On the other hand, $x(t) = 0$ is always a solution. Thus, the MOC solution of the Liouville equation captures the non-uniqueness of the flow.

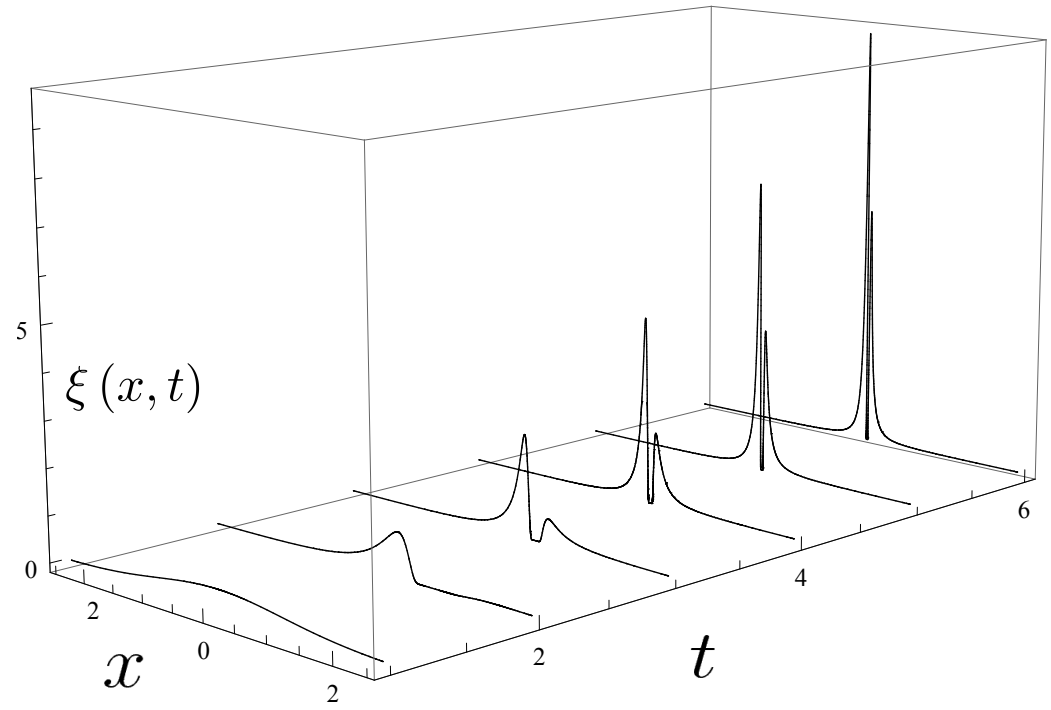


Figure 2.1: Snapshots for the PDF evolution according to (2.16), at initial and five consecutive times, starting with a standard normal PDF $\mathcal{N}(0,1)$.

2.1.2.2 2D Example

Next, consider a planar vector field

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, x_2) = -x_1 - \frac{2x_2}{\log(x_1^2 + x_2^2)}, \\ \dot{x}_2 &= f_2(x_1, x_2) = -x_2 + \frac{2x_1}{\log(x_1^2 + x_2^2)}, \end{aligned} \tag{2.17}$$

with given initial conditions $x_1(0) = x_{10}$ and $x_2(0) = x_{20}$. Looking at the form of the dynamics, we convert (2.17) from cartesian to polar coordinates using the standard transformation $r\dot{r} = x_1\dot{x}_1 + x_2\dot{x}_2$ and $\dot{\theta} = \dfrac{x_1\dot{x}_2 - x_2\dot{x}_1}{x_1^2 + x_2^2}$, to obtain

$$\dot{r} = f_r(r, \theta) = -r,$$

$$\dot{\theta} \;=\; f_\theta(r,\theta) = \frac{1}{\log r}, \tag{2.18}$$

purely as a matter of working convenience. The initial conditions for (2.18) are $r_0 \triangleq r(0) = \sqrt{x_{10}^2 + x_{20}^2}$ and $\theta_0 \triangleq \theta(0) = \arctan\dfrac{x_{20}}{x_{10}}$. From the polar equations, it immediately follows that as $t \to \infty$, $r(t) \to 0$ and $|\theta(t)| \to \infty$ implying that the origin is a globally asymptotically stable spiral for this nonlinear system (Fig. 2.2). Notice however that a linear stability analysis predicts the origin to be a stable star. In fact, one can easily solve (2.18) to get the trajectory in closed form

$$r(r_0,t) = r_0 e^{-t}, \qquad \theta(\theta_0,t) = \theta_0 + \log\left(\frac{\log r_0}{\log r_0 - t}\right), \tag{2.19}$$

which corroborates the asymptotic behavior mentioned above. Further, one can compute

$$\begin{aligned}\nabla\cdot f = \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} \;&=\; \frac{1}{r}\left(\frac{\partial}{\partial r}(r f_r) + \frac{\partial f_\theta}{\partial \theta}\right) = \frac{1}{r}\left(\frac{\partial}{\partial r}\left(-r^2\right) + \frac{\partial}{\partial\theta}\left(\frac{1}{\log r}\right)\right) = -2 \\ &\Rightarrow\; \exp\left(-\int_0^t \nabla\cdot f(r(\tau),\theta(\tau))\,d\tau\right) = e^{2t}. \end{aligned} \tag{2.20}$$

Notice that the coordinate transformation rule for the divergence operator $\dfrac{\partial f_1}{\partial x_1} + \dfrac{\partial f_2}{\partial x_2} = \dfrac{1}{r}\left(\dfrac{\partial}{\partial r}(r f_r) + \dfrac{\partial f_\theta}{\partial\theta}\right)$, simplifies the computation. Alternatively, one could get the same result in Cartesian coordinates, by deriving

$$\begin{aligned}\frac{\partial f_1}{\partial x_1} \;&=\; -1 - 2x_2\frac{\partial}{\partial x_1}\left(\frac{1}{\log\left(x_1^2+x_2^2\right)}\right) = -1 + \frac{4x_1x_2}{\left(x_1^2+x_2^2\right)\left(\log\left(x_1^2+x_2^2\right)\right)^2}, \\ \frac{\partial f_2}{\partial x_2} \;&=\; -1 + 2x_1\frac{\partial}{\partial x_2}\left(\frac{1}{\log\left(x_1^2+x_2^2\right)}\right) = -1 - \frac{4x_1x_2}{\left(x_1^2+x_2^2\right)\left(\log\left(x_1^2+x_2^2\right)\right)^2}. \end{aligned}$$

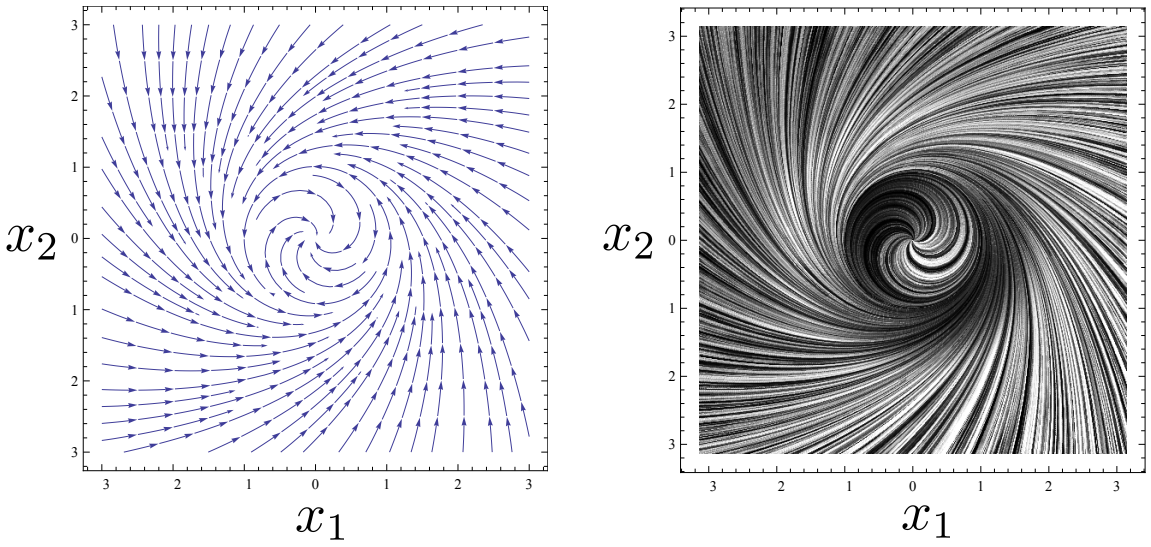


Figure 2.2: Vector field (left) and an ensemble of trajectories in the phase space (right) for the nonlinear system given by (2.17).

From (2.19), we also get

$$r_0\left(r,t\right) = re^t, \qquad \theta_0\left(\theta,t\right) = \theta - \log\left(\frac{\log r + t}{\log r}\right). \tag{2.21}$$

Thus, (2.20) and (2.21) results

$$\xi\left(r,\theta,t\right) = \xi_0\left(r_0,\theta_0\right)\ e^{2t} = \xi_0\left(re^t,\ \theta - \log\left(\frac{\log r + t}{\log r}\right)\right)\ e^{2t}. \tag{2.22}$$

If the initial conditions are sampled from a uniform PDF, then the transient PDFs resemble the phase portrait of Fig. 2.2, converging toward a Dirac delta at the origin. To examine the case for non-uniformly sampled initial conditions, an initial PDF is taken which has a high probability around $\theta = 0$ and is symmetric about the same. The polar plots of Fig. 2.3 shows the PDF contours at $t = 0, 0.2, 0.5, 1.0, 1.4$ and 2.0 respectively, for the dynamics given by (2.17). It can be observed that the support

of the transient PDFs shrink progressively and spirally converge toward the origin, which is consistent with the trajectory level analysis. The red (blue) color denotes high (low) probability.

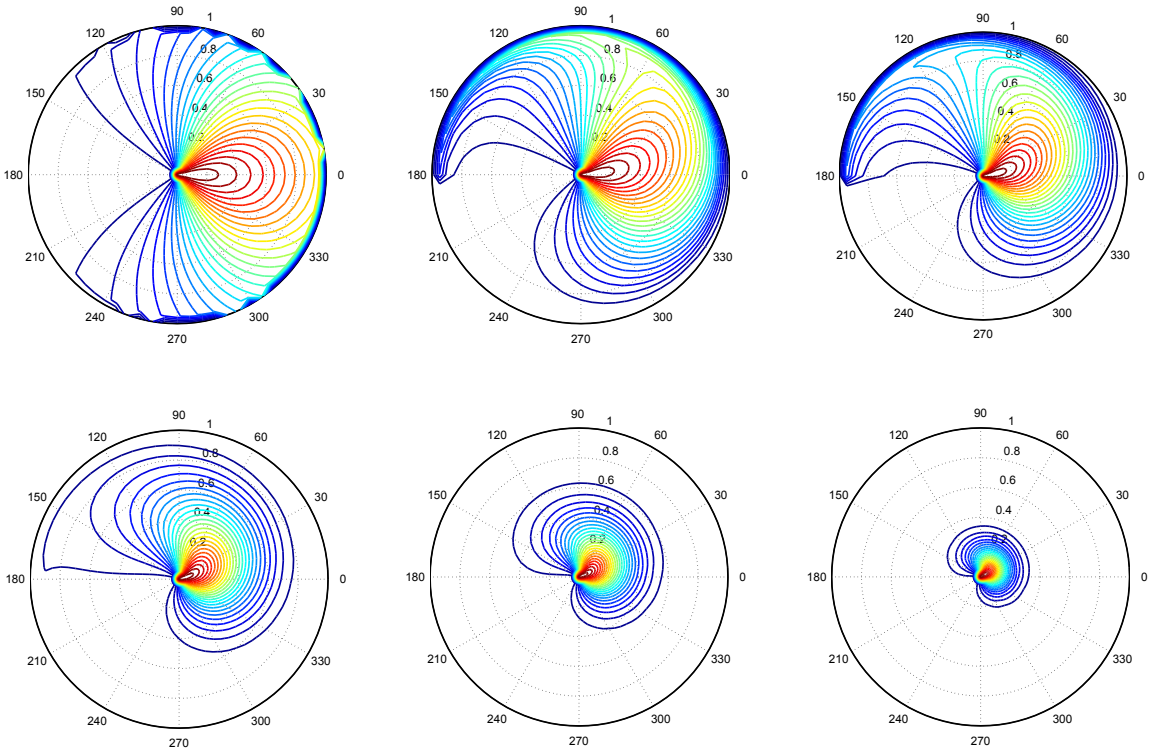


Figure 2.3: Contours of the PDF $\xi(r, \theta, t)$ computed from (2.22) at $t = 0$ (top left), 0.2 (top middle), 0.5 (top right), 1.0 (bottom left), 1.4 (bottom middle) and 2.0 (bottom right), respectively.

Remark 1. *The two simple examples given above illustrate how MOC solves the Liouville equation. In MOC, the initial value problem (2.9) is solved along the characteristics, which in case of Liouville equation, are the integral curves or trajectories of the flow. Thus the integral in (2.10) is a path integral computed along each trajectory (see Fig. 2.4). To further clarify this, consider a divergence-free vector field.*

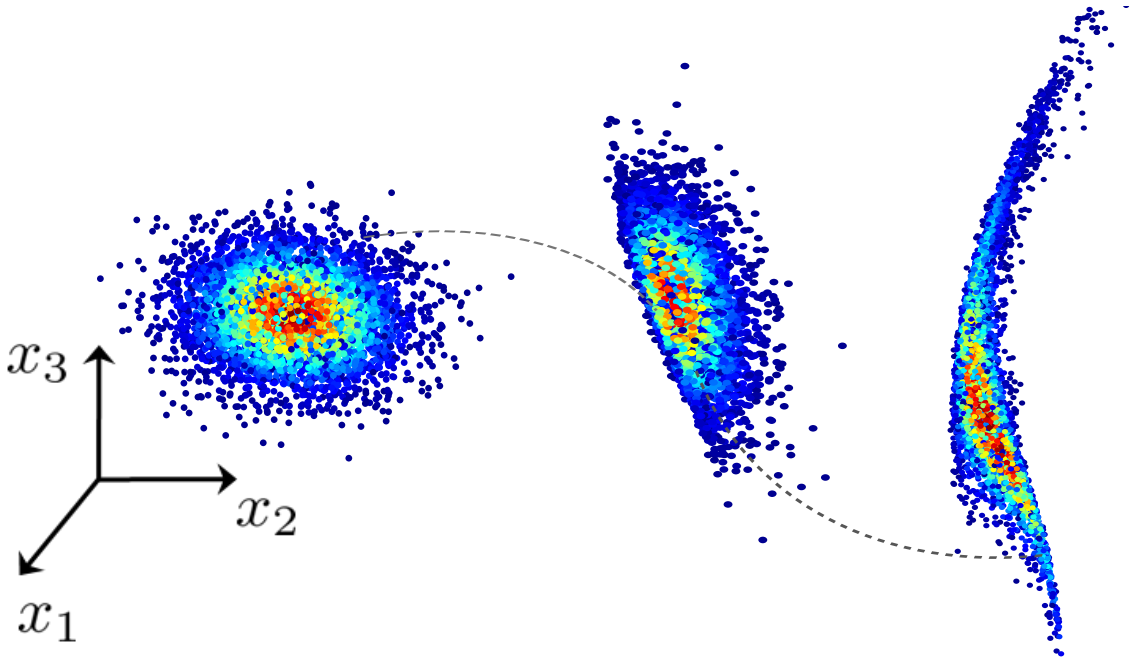


Figure 2.4: In MOC based solution of the Liouville equation, along each sample trajectory, the probability weights are updated during dynamics propagation. In MC method, one tries to reconstruct a histogram to *approximate* such weight distribution, as a post-processing step. So the main advantage of using Liouville equation compared to MC is the ability to update *exact* probability weights "on the fly" and hence, the samples are 'colored' so to speak, with the color-value being proportional to the value of the instantaneous joint PDF. Red (blue) denotes high (low) value.

Then (2.9) tells us that the joint PDF remains constant as long as we are "riding" a particular trajectory. The value of this constant is different along a different trajectory. Thus, a volume-preserving flow, in general, does admit a spatio-temporally evolving PDF. For the same reason, in the above examples, initial condition was computed as a function of the current state and time to substitute for $x(0)$ *in* $\xi(x(0), 0)$ *(see (2.15), (2.16) and (2.22)).*

Remark 2. *For the Liouville equation, since the trajectories are same as the characteristic curves and trajectories can't intersect (due to uniqueness), the solutions of*

the MOC are no where discontinuous.

2.1.3 Algorithms

The numerical algorithm for implementing the MOC method to solve Liouville equation, comprises of three main modules as described below.

1. **Initial sampling**: The initial uncertainties are specified by an initial joint PDF ξ_0. Once the initial joint PDF is known, one needs to generate a pre-specified number of samples such that they best represent that prescribed joint PDF ξ_0. For the case of uniform initial distribution, one may do a grid-based discretization *or* for high dimensions, opt for a pseudo-random number generator using low discrepancy sequences like *Halton sequence* [2] to avoid the 'curse of dimensionality'. Some preliminary comparative simulation results along these lines were reported in [37]. In this dissertation, samples from uniform initial PDF were generated using multi-dimensional Halton sequence. For non-uniform initial PDFs, one needs to use probability integral transform (e.g. Box-Muller transform in case of normal distribution) methods [69]. However, one must resort to the Markov Chain Monte Carlo [70, 71] (MCMC) techniques to achieve better computational performance for sampling any general initial PDF in high dimensions [72].

2. **Uncertainty propagation**: The samples from the initial PDF ξ_0 are propagated according to the given deterministic dynamics, and the MOC ODE (2.9) is solved at each time step, for each of these samples, as illustrated in Fig. 2.4. Notice that solving (2.9) along one trajectory, is independent of the other, and hence the formulation is a natural fit for parallel implementation. Algorithm 1 details this method for uncertainty propagation.

Algorithm 1 Uncertainty Propagation via MOC Solution of the Liouville PDE

Require: The initial joint PDF $\xi_0(x_0, \widehat{p})$, dynamics (2.1), number of samples N, final time t_f, time step Δt.

Generate N scattered samples $\{x_{0i}, \widehat{p}_i\}_{i=1}^N$ from the initial PDF $\xi_0(x_0, \widehat{p})$ ▷ Using MCMC

Evaluate the samples $\{x_{0i}, \widehat{p}_i\}_{i=1}^N$ at $\xi_0(x_0, \widehat{p})$, to get the point cloud $\{x_{0i}, \widehat{p}_i, \xi_{0i}\}_{i=1}^N$

for $t = 0 : \Delta t : t_f$ **do** ▷ Index for time

  for $i = 1 : 1 : N$ **do** ▷ Index for samples

    Numerically integrate the dynamics (2.1) ▷ Propagate states to obtain $\{\widehat{x}_i(t)\}_{i=1}^N$

    Compute $\nabla \cdot \widehat{f}$ ▷ either analytically or numerically, evaluated at $\widehat{x}_i(t)$

    Numerically integrate the characteristic ODE (2.9) ▷ Propagate joint PDF values to

    ▷ obtain $\{\widehat{\xi}_i(t) \triangleq \widehat{\xi}(\widehat{x}_i(t), \widehat{p}_i, t)\}_{i=1}^N$

  end for

end for ▷ Get time-varying probability-weighted scattered data $\{\widehat{x}_i(t), \widehat{p}_i, \widehat{\xi}_i(t)\}_{i=1}^N$ for each t

3. **Marginal PDF computation**: As the samples from the initial PDF ξ_0 are propagated according to dynamics $\widehat{f}$, the joint PDF $\widehat{\xi}$ at any given time $t > 0$, is represented by the instantaneous distribution of those evolved samples. Because of the nonlinear dynamics, such a distribution, in general, is a scattered data set residing over the extended state space. To compute the marginal PDFs from these joint PDFs, one needs to isolate a snapshot of interest and integrate out the dimensions other than whose marginal is sought. In Section 2.1.1, we briefly touched upon the fact that because of dynamics, the domain or the support of the joint PDF deforms with time, and the integration for marginal computation needs to be carried out over few dimensions of this instantaneous domain. This brings forth the problem of integration over high dimensional scattered data.

   One way to tackle this problem is to interpolate these scattered data, which

itself is numerically challenging. Since the joint PDF values were computed directly by solving the Liouville equation, it's an interpolation problem as opposed to function approximation. Alternatively, one may attempt the numerical integration without interpolation. For this, one can sprinkle a new set of Halton points (preferably more than the number of samples) inside the bounding box of this static/time-frozen data and then use these newly sprinkled points as the quadrature points to carry out quasi-Monte Carlo (QMC) integration (see Chapter 2 in [2]). The computational cost associated with this approach comes from the evaluation of the joint PDF values at this new set of points, which can be determined by first back-integrating the dynamics and then forward integrating the MOC ODE for these quadrature points.

Notice that for computing marginals from MC simulations, one takes a frequentist approach and counts samples in the bins lying on the requisite slices. Since we are interested to compare the Liouville derived marginals with those obtained from MC, a similar (and computationally less heavy than described above) method can be employed to approximate the marginals from MOC computation. In the MOC computation, since there is a probability weight associated with each sample, one can do a binning similar to MC histograms. Only this time, instead of counting the number of samples in each bin, the bin weight can be assigned as the average of the joint pdf values of the samples in that bin (see Fig. 6 in [38]). In both cases, the individual bin weights need to be normalized with respect to the bin size and total probability weight. In this dissertation, all marginals computed from the joint PDFs propagated by the MOC ODE, are obtained in this fashion.

2.2 Case Study: Risk Analysis for Mars Hypersonic Entry

Almost all space mission uncertainty analysis have been done with MC simulations including Mars Pathfinder [48], METEOR [73] recovery module, Stardust [74] comet sample return capsule, Mars Microprobe [75], Mars Surveyor Program 2001 Orbiter and Lander [76] and the Mars Science Laboratory [49] mission. In fact, many important decisions in mission design have been historically driven by MC based dispersion analysis. Mars landing site selection [77], design of the Pathfinder aeroshell thermal protection system [78], and parachute deployment algorithm [48] are examples for the same. This heavy bias towards MC simulations among EDL analysis practitioners is partly due to its ease of implementation, and partly due to the scarcity of alternative analysis methods. Two primary EDL simulation frameworks which are seeing extensive use at present, are NASA Langley Research Center's (LaRC) Program to Optimize Simulated Trajectories II (POST2) [79] and NASA Jet Propulsion Laboratory's (JPL) Dynamics Simulator for Entry, Descent and Surface Landing (DESENDS) [80]. Both of these two presently rely on MC based dispersion analysis for EDL simulations. In this Section, we work out an uncertainty propagation case study for Mars hypersonic entry, using the MOC computation to solve the Liouville PDE, described in Section 2.1, and compare our results with MC simulations.

2.2.1 Vinh's Equation for Hypersonic Entry

We model the kinematics for hypersonic entry into the Mars atmosphere through Vinh's equations [81]. We will work with two different models of the same, a three state model where the kinematics is assumed to be purely longitudinal, and a more general six state model with lateral-longitudinal coupling. Both these models describe the trajectory of the center-of-mass of a spacecraft entering into the Mars

atmosphere.

2.2.1.1 Three State Model

Assuming the entire trajectory is contained in the longitudinal plane, one can write the following non-dimensionalized three state (h, V, γ) model for non-rotating spherical Mars with zero bank angle flight.

$$\dot{h} = V \sin\gamma, \tag{2.23a}$$

$$\dot{V} = -\frac{\rho R_0}{2B_c} V^2 - \frac{gR_0}{v_c^2} \sin\gamma, \tag{2.23b}$$

$$\dot{\gamma} = \frac{\rho R_0}{2B_c} \frac{C_L}{C_D} V + \frac{gR_0}{v_c^2} \cos\gamma \left(\frac{V}{1+h} - \frac{1}{V} \right). \tag{2.23c}$$

Here the model for Martian atmospheric density variation [82] is taken as

$$\rho = \rho_0 \exp\left(\frac{h_2 - hR_0}{h_1} \right), \tag{2.24}$$

where $h_2 = 20$ km and $h_1 = 9.8$ km. The mean equatorial radius of Mars will be taken as $R_0 = 3397$ km.

2.2.1.2 Six State Model

Here we present the more general form of Vinh's equations, which is a non-dimensionalized six state $(h, \zeta, \lambda, V, \gamma, \chi)$ model. This model takes the self-rotation rate (Ω_r) of the planet and the bank angle (σ_b) into account.

$$\dot{h} = V \sin\gamma, \tag{2.25a}$$

$$\dot{\zeta} = \frac{V \cos\gamma \sin\chi}{(1+h)}, \tag{2.25b}$$

$$\dot{\lambda} = \frac{V \cos\gamma \cos\chi}{(1+h)\cos\zeta}, \tag{2.25c}$$

$$
\begin{aligned}
\dot{V} &= -\frac{\rho R_0}{2B_c}V^2 - \frac{gR_0}{v_c^2}\sin\gamma \\
&+ \frac{R_0^2\Omega_r^2}{v_c^2}(1+h)\cos\zeta\left(\sin\gamma\cos\zeta - \cos\gamma\sin\zeta\sin\chi\right), \qquad (2.25d)\\
\dot{\gamma} &= \frac{\rho R_0}{2B_c}\frac{C_L}{C_D}V\cos\sigma_b + \frac{gR_0}{v_c^2}\cos\gamma\left(\frac{V}{1+h} - \frac{1}{V}\right), \qquad (2.25e)\\
\dot{\chi} &= \frac{\rho R_0}{2B_c}\frac{C_L}{C_D}\frac{V\sin\sigma_b}{\cos\gamma} - \frac{V\cos\gamma}{(1+h)}\tan\zeta\cos\chi + \frac{2R_0\Omega_r}{v_c}\left(\tan\gamma\cos\zeta\sin\chi - \sin\zeta\right) \\
&- \frac{R_0^2\Omega_r^2}{v_c^2}\frac{(1+h)}{V\cos\gamma}\sin\zeta\cos\zeta\cos\chi. \qquad (2.25f)
\end{aligned}
$$

Ω_r was calculated from the rotational time period of Mars, which is 24 hours 39 minutes and 35.24 seconds. The density variation is taken identical to the three-state model.

2.2.2 Numerical Simulations

Before solving the Liouville equation for the three state and six state Vinh's equation, we will examine certain restricted cases of the same. Since the three and six-state models, in general, require numerical solution for the PDF, considering specific cases will give us some physical understanding of the problem.

2.2.2.1 Horizontal Flight

For horizontal flight, $\gamma \approx 0$ and $h =$ constant. Therefore, only the second equation remains to be considered in (2.23), which becomes

$$
\begin{aligned}
\dot{V} &= -\frac{\rho R_0}{2B_c}V^2 \\
\Rightarrow \int_{V_0}^{V}\frac{dV}{V^2} &= -\frac{\rho R_0}{2B_c}\int_0^t dt \qquad \text{[since } h \text{ is constant, so is } \rho\text{]} \\
\Rightarrow V &= \frac{V_0}{1+\frac{\rho R_0}{2B_c}V_0 t} \qquad (2.26)
\end{aligned}
$$

which implies that V decreases monotonically with time. In this case, $\widehat{\xi}\,(V,t) = \xi_0\,(V_0) \exp\left(-\int_0^t -\frac{\rho R_0}{B_c} V\,(\tau)\; d\tau\right)$. Therefore,

$$\begin{aligned}\widehat{\xi}\,(V,t) = \xi_0\,(V_0)\left[1+\frac{\rho R_0}{2B_c}V_0 t\right]^2 &= \xi_0\,(V_0)\left(\frac{V_0}{V}\right)^2 \\ &= \xi_0\left(\frac{V}{1-\frac{\rho R_0}{2B_c}Vt}\right)\frac{1}{\left(1-\frac{\rho R_0}{2B_c}Vt\right)^2}. \quad (2.27)\end{aligned}$$

Thus, given an initial PDF ξ_0 describing the initial condition uncertainty, (2.27) provides an algebraic expression for determining the transient PDF $\widehat{\xi}\,(V,t)$ at any current time t and velocity V.

2.2.2.2 Vertical Flight

This special case concerns with the vertical descent ($\gamma = -\frac{\pi}{2}$) in a non-lifting trajectory. Consequently, we eliminate (2.23c) as all terms in it are identically zero. Hence, we are left with (2.23a), (2.23b) and (2.24). Substituting ρ in (2.23b) as a function of h, we get two first order coupled nonlinear ODEs in h and V, shown below.

$$\dot{h} = -V, \quad (2.28a)$$

$$\dot{V} = -K_1 V^2 e^{-\beta h} + K_2, \quad (2.28b)$$

with $K_1 = \frac{\rho_0 R_0}{2B_c} e^{h_2/h_1}$, $\beta = \frac{R_0}{h_1}$ and $K_2 = \frac{gR_0}{v_c^2}$. With $\nabla \cdot \widehat{f} = -2K_1 V e^{-\beta h}$, the Liouville equation needs to be solved numerically along with the above dynamics.

With nominal initial altitude $h_0 = 80$ Km and nominal initial velocity $V_0 = 3.5$ Km/sec, and assuming 5% uniform dispersion in both h_0 and V_0, the Liouville equation was solved numerically. The simulation was repeated for 15% uniform dispersion in both h_0 and V_0. In both the cases, 1000 samples were taken to represent

the trajectory ensemble. Fig. 2.5 and 2.6 show the color-coded scatter plots at $t = 19.13$ seconds in the $h\,V$ plane, with the color-value being proportional to the value of the bivariate joint PDF at that instant. As before, red denotes high and blue denotes low value of the joint PDF. Notice that, a larger dispersion in the initial conditions results in more spread in the point cloud at the same instant of time.

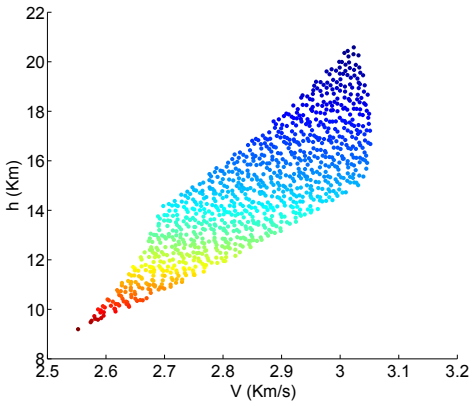


Figure 2.5: Scatterplot at $t = 19.13$ sec with 5% uniform initial condition uncertainties in h_0 and V_0.

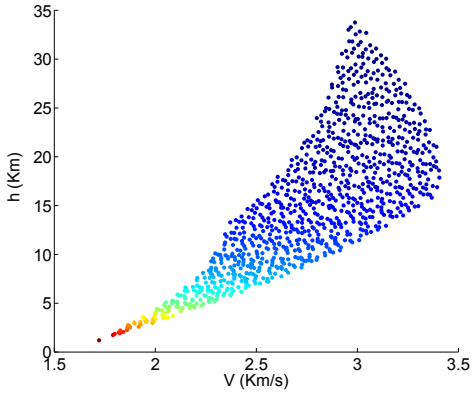


Figure 2.6: Scatterplot at $t = 19.13$ sec with 15% uniform initial condition uncertainties in h_0 and V_0.

2.2.2.3 Numerical Results for the Generic Vinh's Equation

The nominal initial conditions are taken to be $h_0 = 80$ Km, $\zeta_0 = 24.01°$N, $\lambda_0 = 341.03°$E, $V_0 = 3.5$ Km/s, $\gamma_0 = -2°$, and $\chi_0 = 0.0573°$. The nominal values of the parameters are taken as $B_c = 72.8$ Kg/m^2, $\rho_0 = 0.0019$ Kg/m^3, and $\dfrac{C_L}{C_D} = 0.3$. Since the models described in Section 2.2.1 are non-dimensionalized, numerical integration was performed in non-dimensional time $t = 0$ to $t = 0.7$ with the non-dimensional step-size $\Delta t = 0.01$. One can easily convert it back to the physical time by multiplying the non-dimensional time with a factor $\dfrac{R_0}{v_c}$. In this paper, results are

presented for two kinds of initial uncertainties, viz. 5% uniform dispersion in each variable, and Gaussian dispersion about the nominals with 10% variance along each dimension.

For the generic three state model (2.23), it is possible to visualize the joint PDF using the three dimensional color-coded scatter plots similar to Fig. 2.5 and 2.6. Such plots are shown in Fig. 2.7 for 1000 samples at $t = 0.05, 0.20, 0.30$ and 0.50, with both uniform (top row) and Gaussian (bottom row) initial condition uncertainties. It can be observed that at $t = 0.05$, the joint PDFs are slightly perturbed from the

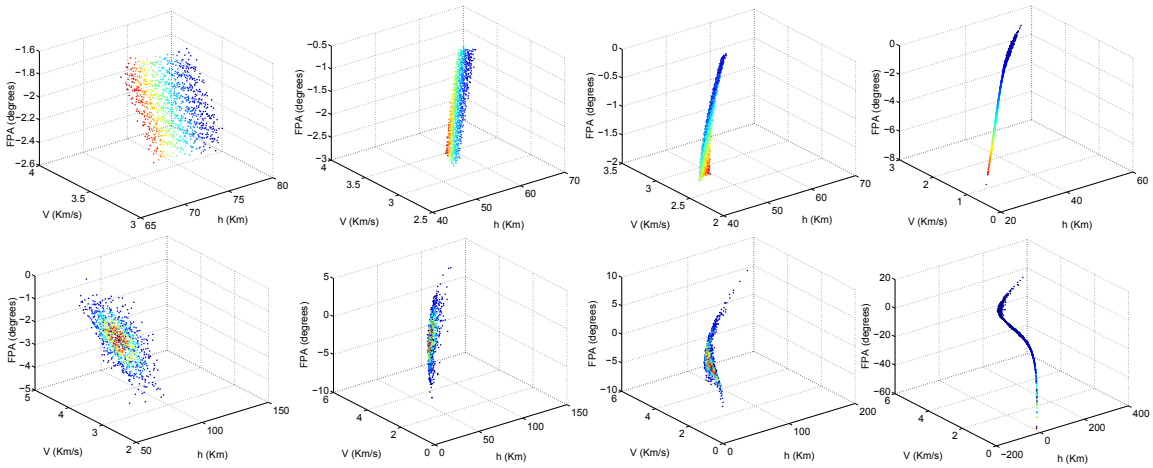


Figure 2.7: Scatter plots of the joint PDF $\widehat{\xi}(h, V, \gamma, t)$ at $t = 0.05, 0.20, 0.30$ and 0.50, respectively. Columns show different times, rows signify different initial PDFs (uniform for *top* and Gaussian for *bottom* row).

respective initials. As time progresses, the probability mass accumulates near zero altitude and zero velocity and the flight path angle assumes a steep value. This is in agreement with the physical intuition as the vehicle, with high probability, slows down through the lower part of the atmosphere.

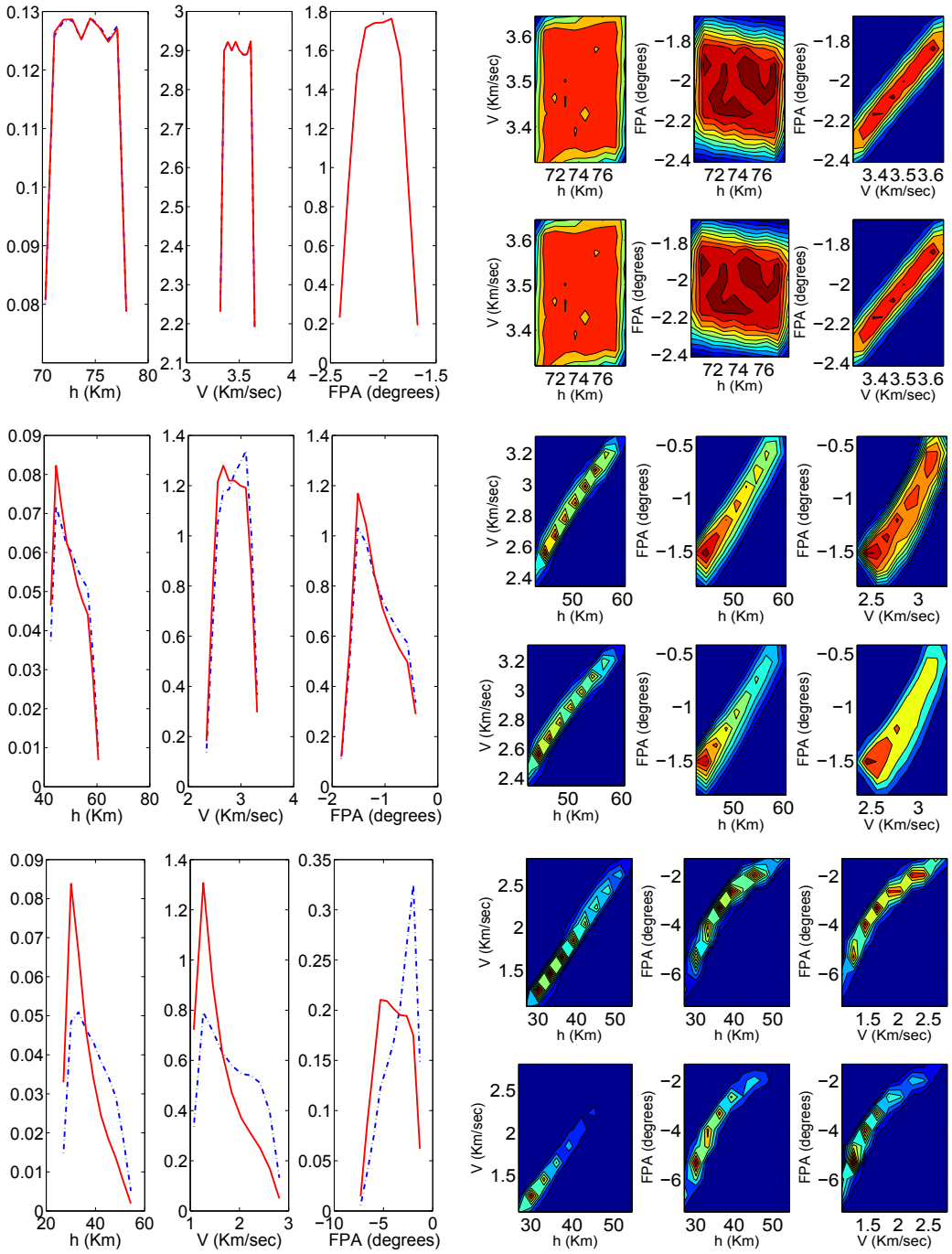


Figure 2.8: The univariate and bivariate marginals for the case of uniform initial condition uncertainty at $t = 0.05, 0.30$ and 0.50, respectively. The simulation is for three state Vinh's equations (2.23) with 5000 samples. For univariate marginals, MOC results are in *solid red* and MC results are in *dashed blue*. For bivariate marginals, MOC results are in the *bottom row* and MC results are in the *top row*.

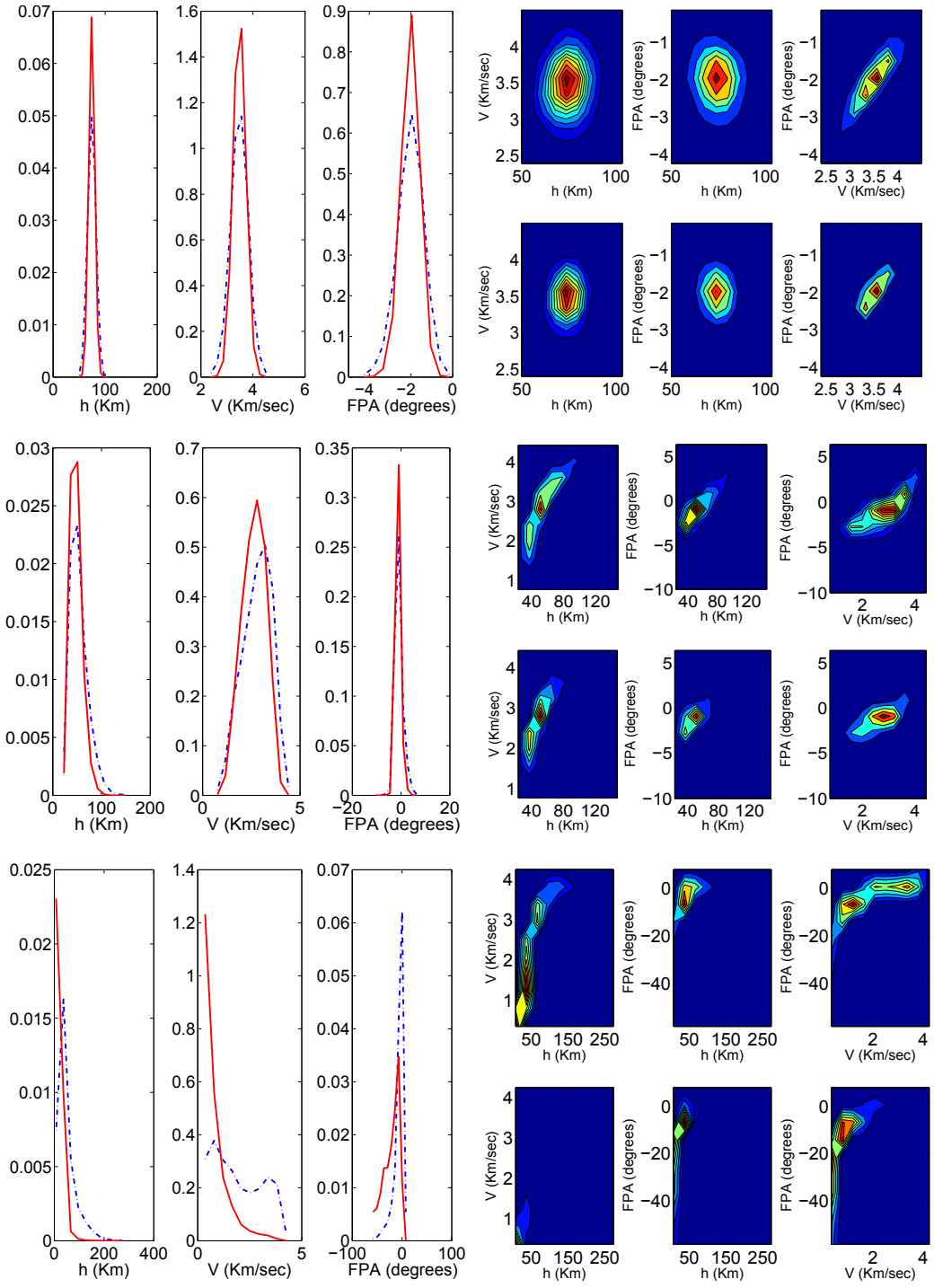


Figure 2.9: The univariate and bivariate marginals for the case of Gaussian initial condition uncertainty at $t = 0.05, 0.30$ and 0.50, respectively. The simulation is for three state Vinh's equations (2.23) with 5000 samples. Conventions for the MC and MOC plots are same as in the previous figure.

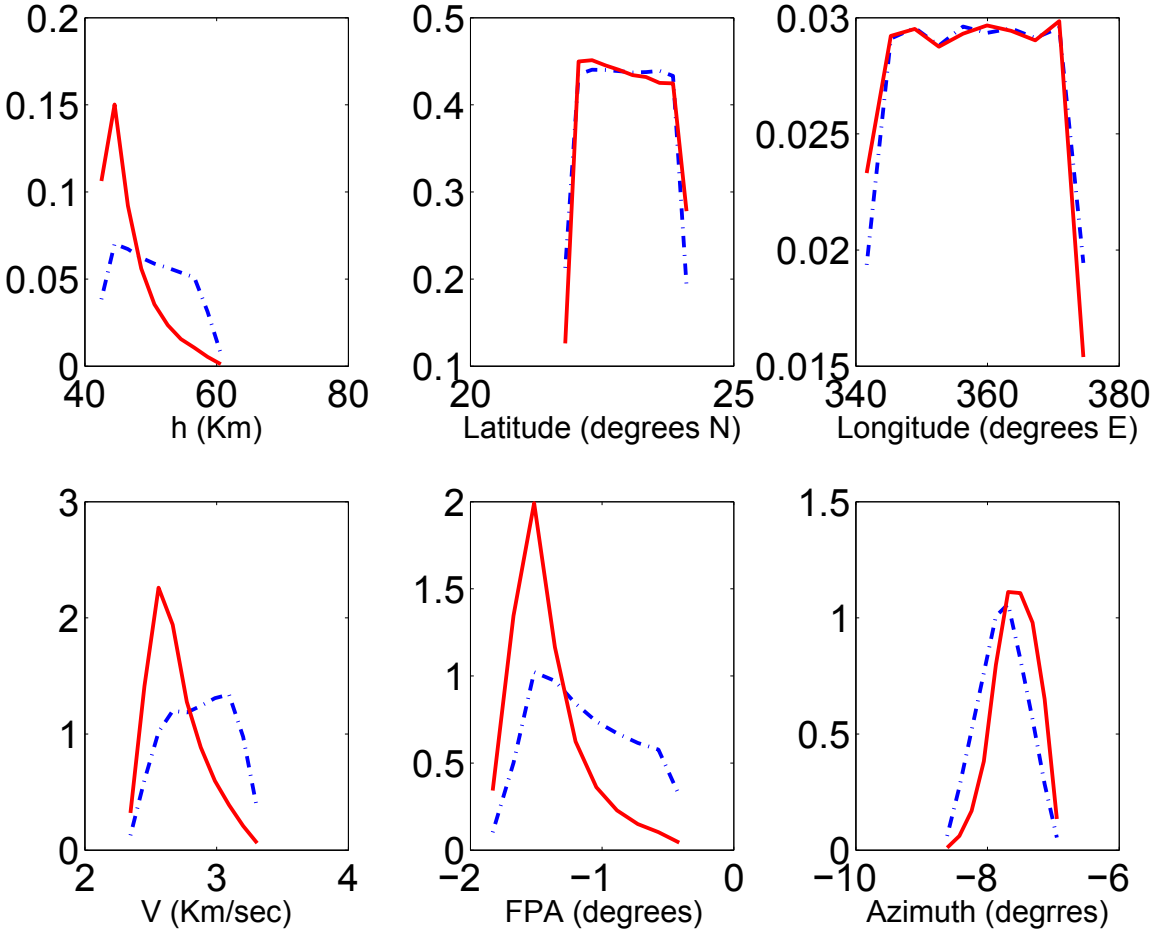


Figure 2.10: The univariate marginals for the case of uniform initial condition uncertainty at $t = 0.30$. The simulation is for six state Vinh's equations (3.66f) with 10,000 samples. MOC results are in *solid red* and MC results are in *dashed blue*.

Starting from the uniform initial joint PDF, the evolution of the univariate and bivariate marginals for the three state Vinh's equations, are shown in Fig. 2.8. The same for the Gaussian initial PDF are plotted in Fig. 2.9. The univariate MC (dashed blue) and MOC (solid red) PDFs are in good match. The bivariate marginals show the general trend that MOC-derived marginals (bottom row) capture the concentration of the probability mass well (by virtue of the probability weights obtained by solving the Liouville equation) while the MC bivariate marginals (top row) tend to smear it out (because of the histogram approximation). This can be seen, for example, in $V - \gamma$ bivariate plots. Similar trends can be observed for the six state model (3.66f). For brevity, in Fig. 2.10, we only show the snapshot of univariate PDFs at $t = 0.30$, for the six state model with uniform initial PDF.

The simulation results shown above bears testimony to the fact that with same number of simulations, MOC based approach can better resolve the instantaneous PDF compared to the MC method. This is not surprising since the former assigns explicit probability weights computed by solving the Liouville equation while the latter tries to approximate a PDF by constructing histograms.

2.2.2.4 Further Statistical Analysis

Next, we demonstrate two statistical analysis pertaining to EDL interest in the MOC framework. More analysis along this line can be found in Section VI of reference [38].

1. **Tracking Uncertainty:** It's of interest to compute the probability that the flight path angle (FPA) will be within a specified interval, i.e. $\gamma_{\min} \leqslant \gamma \leqslant \gamma_{\max}$. This problem is important in the context of tracking the spacecraft by a space-based antenna. Univariate FPA marginals (like those shown in Fig. 2.8, 2.9 and 2.10) can be computed at different times to calculate the tracking probabilities.

Such information can be crucial from mission design perspective.

2. **Landing Footprint Uncertainty:** Computing the landing footprint uncertainty has been one of the key aspects of EDL analysis. Important decisions like landing risk evaluation and trajectory correction maneuver design depend on it. A list of factors contributing toward landing footprint uncertainty, can be found in [77]. Almost all EDL analysis has been based on evolving a bivariate Gaussian in latitude and longitude and thereby characterizing a 3σ landing ellipse representing the landing footprint uncertainty. Historically, the landing ellipses have spanned hundreds of Km (see Fig. 2.11). However, depending on the initial uncertainty and system dynamics, the latitude-longitude bivariate marginal can be far from Gaussian, resulting the 3σ estimates unrealistic. Computing this marginal using MC method is not only computationally expensive but can be inaccurate, for reasons discussed before. Fig. 2.12 compares the latitude-longitude (ζ λ) bivariate marginal at the final time, computed for the six state model using MC (left) and MOC based Liouville equation (right) method. Notice that, the MOC based Liouville equation method (right in Fig. 2.12) predicts the landing footprint to be at 377 degrees E and 21.3 degrees N (approx.) with maximum probability and a very small dispersion around it. It ascertains that the landing probability everywhere else is zero. In contrast, MC method (left in Fig. 2.12) can at best predict a high probability around 357–381 degrees E and 20.5–21.4 degrees N and is unable to do any further refinement of the landing footprint uncertainty. Not surprisingly, such huge MC dispersion in latitude-longitude results 3σ landing ellipse spanning hundreds of Km. It's evident that MOC based Liouville equation method outperforms MC.

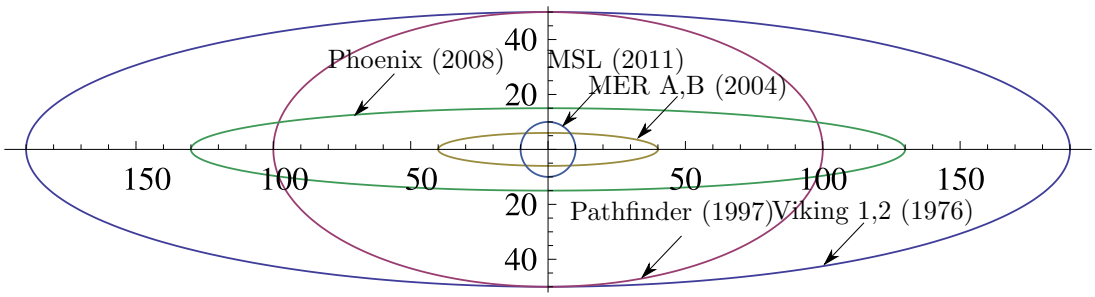


Figure 2.11: Schematic comparison of landing footprints of Mars missions. To make a comparison between their sizes, all ellipses are drawn with the same center and same orientation. The scale on each axis is in Km (data taken from [1]).

Table 2.1: Comparison of joint PDF computation over $\mathbb{R}^{\widehat{n}_s+\widehat{n}_p}$: MC vs. MOC

Attributes	MC simulation	MOC for Liouville equation
Concurrency	Offline post-processing	Online
Accuracy	Histogram approximation	Exact arithmetic
Spatial discretization	Grid based	Meshless
ODEs per sample	$\widehat{n}_s$	$\widehat{n}_s + 1$

2.3 Computational Performance Assessment Against Monte Carlo

In this section, we compare the computational performance between MC and the proposed MOC method. The main differences of the MOC computation from the MC simulation are summarized in Table 2.1. Next, we focus on comparing the numerical performance of MC and MOC, with respect to the three steps, viz. sampling the initial PDF, uncertainty propagation, and computing marginals.

2.3.1 Sampling Initial PDF

The initial distribution is specified as a PDF supported over the initial conditions and parameters. Then the question arises, how to generate a user-specified number

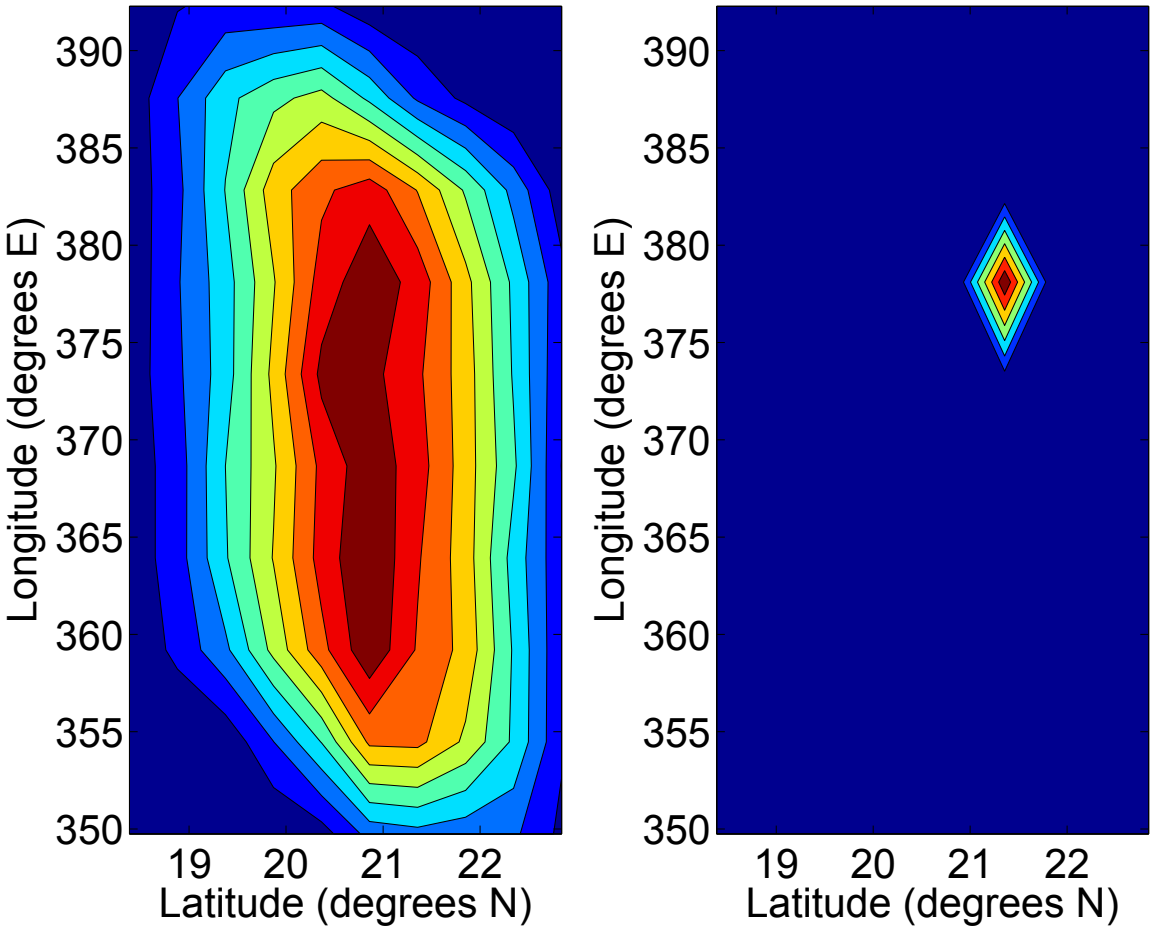


Figure 2.12: Comparison of the latitude-longitude (ζ λ) bivariate marginal PDF at the final time, from MC (left) and MOC based Liouville equation (right) method.

of samples from such prescribed initial joint PDF, and what can be said about the accuracy and computation time for the same?

2.3.1.1 Accuracy

If the initial density is jointly uniform, then multidimensional *Halton sequence* [83] provides a deterministic method to generate a prescribed number of samples. Since the construction is based on a deterministic algorithm (as opposed to randomized algorithm), it's often called a quasi Monte Carlo (QMC) method of sample generation. The Halton sequence is a multidimensional extension of *van der Corput sequence with respect to base* b, denoted by $\mathcal{V}_b$, whose n^{th} term is given by

$$\mathcal{V}_b(n) := \sum_{r=0}^{\infty} \frac{a_r(n)}{b^{r+1}}, \qquad n \in \mathbb{N}, \tag{2.29}$$

where $a_r(n)$ is the r^{th} digit of the b-adic expansion of

$$n - 1 = \sum_{r=0}^{\infty} a_r(n)\, b^r. \tag{2.30}$$

The n^{th} term of the Halton sequence in unit hypercube $[0,1]^{n_s}$ is defined as the n_s-tuple

$$\mathcal{H}_n := \left(\mathcal{V}_{b_1}(n), \ldots, \mathcal{V}_{b_{\widehat{n}_s}}(n)\right), \tag{2.31}$$

where the b_js are pairwise coprime, for $j = 1, \ldots, \widehat{n}_s$. Accuracy of the generated Halton points is quantified through the notion of *discrepancy*, denoted by $D_\nu(\mathcal{S}_{\widehat{n}_s}) \in (0,1)$, that intuitively measures the amount of irregularity in the generated $\widehat{n}_s$-dimensional sample set of cardinality ν i.e. $\mathcal{S}_{\widehat{n}_s} \triangleq \{\mathcal{H}_1, \ldots, \mathcal{H}_\nu\}$. Since the samples of uniform distribution are desired to be scattered evenly throughout the

domain, lower discrepancy is better.

Definition 1. *(**Discrepancy and star discrepancy**) Let $\mathscr{F}$ denote the nonempty family of subsets of $[0,1]^{\widehat{n}_s}$. The discrepancy $D_\nu\left(\mathcal{S}_{\widehat{n}_s}\right)$ of a sample set $\mathcal{S}_{\widehat{n}_s}$ of cardinality ν, is defined as*

$$D_\nu\left(\mathcal{S}_{\widehat{n}_s}\right) \triangleq \sup_{\mathcal{F}\in\mathscr{F}} \left| \frac{1}{\nu}\sum_{i=1}^{\nu} \chi\left(\mathcal{H}_i \in \mathcal{F}\right) - \mathrm{Vol}\left(\mathcal{F}\right) \right|. \tag{2.32}$$

If instead of $\mathscr{F}$, we use $\mathscr{G}$, the nonempty family of subsets of semi-open unit hypercube $[0,1)^{n_s}$, then the above defines star discrepancy $D_\nu^\star\left(\mathcal{S}_{\widehat{n}_s}\right)$ as

$$D_\nu^\star\left(\mathcal{S}_{\widehat{n}_s}\right) \triangleq \sup_{\mathcal{G}\in\mathscr{G}} \left| \frac{1}{\nu}\sum_{i=1}^{\nu} \chi\left(\mathcal{H}_i \in \mathcal{G}\right) - \mathrm{Vol}\left(\mathcal{G}\right) \right|. \tag{2.33}$$

The following theorem ensures that the multidimensional Halton sequence, as defined in (2.31), is of low discrepancy.

Theorem 1. *[2]*

$$\forall \nu \geqslant 1,\ D_\nu^\star\left(\mathcal{S}_{\widehat{n}_s}\right) < \frac{\widehat{n}_s}{\nu} + \frac{1}{\nu}\prod_{j=1}^{\widehat{n}_s}\left(\frac{b_j-1}{2\log b_j}\log\nu + \frac{b_j+1}{2}\right). \tag{2.34}$$

There are other low discrepancy sequences like Sobol, Faure and Niederreiter sequences [2], which suitably permute Halton sequences to lower the discrepancy for large $\widehat{n}_s$. If the initial joint density is other than uniform, then the availability of a low discrepancy uniform random number generator still comes handy for inverse transform sampling or other specialized methods [69].

Unfortunately, sampling non-uniform density through transform techniques suffers from computational inefficiency in high dimensions. Hence, to sample arbitrary

initial PDF in high dimensions, one must resort to an MCMC sampler. The core idea behind MCMC is to create a Markov chain whose stationary distribution is the one from where we want to sample. Two popular ways of doing this are *Metropolis-Hastings algorithm* [71, 84] and *Gibbs sampler* (also known as *Glauber dynamics* and *heat-bath algorithm*) [85, 86, 87]. Since MCMC guarantees that the constructed Markov chain *asymptotically* converges to the desired density, to adjudge the performance of MCMC, one can ask: how long must the chain be run to get sufficiently close to the target density? This rate of convergence, often called 'burn-in period' in MCMC literature, quantifies the performance of a sample generation algorithm. We describe it next.

2.3.1.2 Computational Time

Although numerically implementing Metropolis-Hastings and Gibbs sampler are straightforward, deriving rigorous bounds for burn-in period is a daunting task. In particular, sharp computable bounds are not available in general setting. Hence two approaches of tackling this problem have evolved in the statistics community. One is the practitioners' approach, where instead of computing an upper bound for burn-in period, a *convergence diagnostics* is performed to test whether the realized samples "seem to be stable" after "large enough" number of iterations. Some widely used convergence diagnostics can be found in [88].

The second approach strives to find a *rigorous upper bound for the burn-in period*. Here the distance between the n^{th} iterate of the Markov chain and the stationary target density is estimated in total variation norm. The idea is to upper bound this distance as a function of n, to answer how many steps are necessary to be ϵ close to the target density. This thread of research is an ongoing pursuit. In our context of continuous state space, most results (see section 3, [89], and [90]) impose some

restrictions on the Markov chain (e.g. uniform ergodicity, geometric ergodicity etc.).

Due to lack of practical and computable MCMC bounds in continuous state space, we simply denote the time required to sample as $T_{\text{gen}}\left(\nu, \widehat{n}_s\right)$, a function of number of samples ν, and dimension of the state space n_s. To make the comparison meaningful, we propagate the uncertainty with same set of MCMC samples, for both MC and PF method. This precludes T_{gen} from exponential dependence on $\widehat{n}_s$, i.e. grid-based curse of dimensionality. However, if the initial density is jointly uniform, then the QMC samplers of the previous subsection are in force, and runtime complexity results are known for them (see Table 2 in [91]).

2.3.2 Propagation of Uncertainty

The main idea here is that constructing joint PDF through MC histogram is an *approximate method* (piecewise constant approximation) while computing the same via MOC solution of Liouville equation, is an *exact method*. Except the truncation error of the integrator, for any generic nonlinear dynamics, the joint PDF weights updated through MOC computation of the Liouville PDE, is exact at the sample sites. The finite sample computation of MOC does not incur any loss of generality. To elicit this, consider a case where the user queries the instantaneous joint PDF value at a location of the extended state space where none of the finite samples have landed. To determine this value exactly, by back integrating the dynamics till $t = 0$, we can determine the initial condition this sample would have come from. If the corresponding initial condition is found to lie outside the support of the prescribed initial joint density, the joint PDF value at the query site is zero. Otherwise, the instantaneous joint PDF value can be computed exactly by forward integrating the Liouville equation along the characteristic curve for that single initial condition. Comparison of computational time between MC and MOC computation is detailed

in Section 4.2 of reference [91].

2.3.3 Computing Marginals

The comparison of marginal computation algorithms for MC and MOC are covered in detail in Section 5 of reference [91], where accuracy and runtime are compared in section 5.1 and 5.2 respectively. It can be noted that unlike propagation of joint PDF, marginal PDF computation is an approximation algorithm for both MC and MOC formalism. For the purpose of brevity, we skip the details here and refer the interested readers to [91].

2.4 Chapter Summary

In this Chapter, an algorithm is proposed for propagating probabilistic uncertainties subject to a deterministic flow. This situation arises when the deterministic model is known, but there are initial conditions and parametric uncertainties. The proposed algorithm computes the time-varying joint PDF supported over the state or output space of the deterministic model. This is accomplished by solving the method of characteristics ODE corresponding to the Liouville PDE, obviating the need for function approximation. This exact computation differs from histogram approximation in Monte Carlo, and dispenses the need for grid. The algorithm is demonstrated through various simple examples, and through a risk analysis case study for Mars EDL. Numerical performance comparison for the proposed method was done vis-a-vis with Monte Carlo.

3. UNCERTAINTY PROPAGATION FOR STOCHASTIC FLOW

In the previous Chapter, we dealt with propagation of uncertainties in initial conditions and parameters, subject to a *deterministic* flow. In this Chapter, we deal with propagation of uncertainties in initial conditions and parameters, subject to a *stochastic* flow. We first describe the Fokker-Planck equation, also known as forward Kolmogorov equation, that transports the probability mass in the state space, and outline the computational challenges in solving the same. Next, we combine the MOC computation of Liouville equation described in the previous Chapter with Karhunen-Loève (KL) expansion, to develop an algorithm that provably approximates the solution of the Fokker-Planck PDE. The efficacy of the proposed algorithm is demonstrated through its application to nonlinear estimation.

The motivation behind investigating a mixed parametric-nonparametric approach stems from the fact that the presence of process noise necessitates solving *second order* Fokker-Planck PDE, and consequently, MOC based exact arithmetic computation can not be achieved. This argument usually leads researchers to develop approximation algorithms to solve the second order Fokker-Planck PDE, which is an exact description of the problem. Here, instead, we derive an approximate ordinary differential equation (ODE) representation of the problem using KL expansion, and then solve this approximate problem in exact arithmetic (using MOC). Hence our approach differs from those which strive to numerically solve the Fokker-Planck PDE, in the sense that we propose to approximate the problem while the latter strives to approximate the solution. We prove that our "first KL, then MOC" algorithm is asymptotically consistent in distribution.

3.1 The Fokker-Planck a.k.a. Forward Kolmogorov Equation

On a probability space $(\Omega_s, \mathscr{F}, \mathbb{P})$ with filtration $\{\mathscr{F}_t\}_{t\geqslant 0}$, consider the Itô stochastic differential equations (SDEs)

$$\begin{aligned} d\widehat{\widetilde{x}}(t) &= \widehat{\widetilde{f}}\left(\widehat{\widetilde{x}}(t), t\right) dt + \widehat{g}_{\widehat{\widetilde{x}}}\, d\widehat{\mathcal{W}}(\omega_s, t), \qquad (3.1) \\ d\widehat{y}(t) &= \widehat{h}\left(\widehat{\widetilde{x}}(t), t\right) dt + \widehat{g}_{\widehat{y}}\, d\widehat{\mathcal{V}}(\omega_s, t), \qquad (3.2) \end{aligned}$$

where at time instance $t > 0$, the extended state vector $\widehat{\widetilde{x}}(t) \in \mathbb{R}^{\widehat{n}_s+\widehat{n}_p}$, and the measurement vector $\widehat{y}(t) \in \mathbb{R}^{n_o}$. Further, $\widehat{\mathcal{W}}(\omega_s, t) : \Omega_s \times \mathbb{R}^+ \mapsto \mathbb{R}^{\widehat{n}_w}$, $\widehat{\mathcal{V}}(\omega_s, t) : \Omega_s \times \mathbb{R}^+ \mapsto \mathbb{R}^{\widehat{n}_v}$ are mutually independent Wiener processes (a.k.a. Brownian motions) denoting process and measurement noise, of dimensions $\widehat{n}_w$ and $\widehat{n}_v$, respectively. For the drift vector fields $\widehat{\widetilde{f}}$ and $\widehat{h}(.)$, we have

$$\widehat{\widetilde{f}} : \mathbb{R}^{\widehat{n}_s+\widehat{n}_p} \times \mathbb{R}^+ \mapsto \mathbb{R}^{\widehat{n}_s+\widehat{n}_p}, \quad \widehat{h} : \mathbb{R}^{\widehat{n}_s+\widehat{n}_p} \times \mathbb{R}^+ \mapsto \mathbb{R}^{n_o}. \qquad (3.3)$$

Similarly, for the diffusion fields $\widehat{g}_{\widehat{\widetilde{x}}}$, and $\widehat{g}_{\widehat{y}}$, we have

$$\widehat{g}_{\widehat{\widetilde{x}}} : \mathbb{R}^{\widehat{n}_s+\widehat{n}_p} \times \mathbb{R}^+ \mapsto \mathbb{R}^{(\widehat{n}_s+\widehat{n}_p)\times\widehat{n}_w}, \quad \widehat{g}_{\widehat{y}} : \mathbb{R}^{\widehat{n}_s+\widehat{n}_p} \times \mathbb{R}^+ \mapsto \mathbb{R}^{\widehat{n}_o\times\widehat{n}_v}. \qquad (3.4)$$

The process noise is assumed to satisfy $\mathbb{E}\left[d\widehat{\mathcal{W}}_i\right] = 0$, $\mathbb{E}\left[d\widehat{\mathcal{W}}_i d\widehat{\mathcal{W}}_j\right] = Q_{ij}$; $i, j = 1, \ldots, \widehat{n}_w$. Similarly the measurement noise is assumed to satisfy $\mathbb{E}\left[d\widehat{\mathcal{V}}_i\right] = 0$, $\mathbb{E}\left[d\widehat{\mathcal{V}}_i d\widehat{\mathcal{V}}_j\right] = R_{ij}$; $i, j = 1, \ldots, \widehat{n}_v$. Then the Fokker-Planck equation that describes the spatio-temporal transport of state PDF $\widehat{\xi}\left(\widehat{\widetilde{x}}(t), t\right)$, is given by

$$\begin{aligned} \frac{\partial \widehat{\xi}}{\partial t} &= \mathscr{L}_{\text{FPE}}\widehat{\xi} \\ &= (D_1 + D_2)\,\widehat{\xi} \end{aligned}$$

$$
\begin{aligned}
&= -\nabla.\left(\widehat{\widetilde{\xi}}\widehat{\widetilde{f}}\right) + \left\langle \nabla\nabla^{\top}, \left(\widehat{g}_{\widehat{\widetilde{x}}} Q \widehat{g}_{\widehat{\widetilde{x}}}^{\top}\right)\widehat{\xi}\right\rangle_{F} \\
&= -\sum_{i=1}^{\widehat{n}_s} \frac{\partial}{\partial \widehat{x}_i}\left(\rho \widehat{f}_i\right) + \sum_{i=1}^{\widehat{n}_s}\sum_{j=1}^{\widehat{n}_s} \frac{\partial^2}{\partial \widehat{x}_i \partial \widehat{x}_j}\left(\left(\widehat{g}_{\widehat{\widetilde{x}}} Q \widehat{g}_{\widehat{\widetilde{x}}}^{\top}\right)_{ij}\widehat{\xi}\right). \qquad (3.5)
\end{aligned}
$$

It can be noticed that the spatial Fokker-Planck operator $\mathscr{L}_{\text{FPE}}$ can be written as a sum of a *drift operator* (D_1) and a *diffusion operator* (D_2). The drift term $D_1\,\widehat{\xi} = -\nabla.\left(\widehat{\widetilde{\xi}}\widehat{\widetilde{f}}\right) = -\left(\nabla.\widehat{\widetilde{f}} + \widehat{\widetilde{f}}\cdot\nabla\right)\widehat{\xi} = -\sum_{i=1}^{\widehat{n}_s}\frac{\partial}{\partial \widehat{x}_i}\left(\widehat{\xi}\widehat{f}_i\right)$, governs the advection of the PDF in the state space. The diffusion term $D_2\,\widehat{\xi} = \left\langle \nabla\nabla^{\top}, \left(\widehat{g}_{\widehat{\widetilde{x}}} Q \widehat{g}_{\widehat{\widetilde{x}}}^{\top}\right)\widehat{\xi}\right\rangle_{F} = \sum_{i=1}^{\widehat{n}_s}\sum_{j=1}^{\widehat{n}_s}\frac{\partial^2}{\partial \widehat{x}_i \partial \widehat{x}_j}\left(\left(\widehat{g}_{\widehat{\widetilde{x}}} Q \widehat{g}_{\widehat{\widetilde{x}}}^{\top}\right)_{ij}\widehat{\xi}\right)$, can be thought as a *Frobenius inner product*, denoted by the symbol $\left\langle .\,,.\right\rangle_F$, between the *Hessian operator* $\left(\nabla\nabla^{\top}\right)$ and the matrix $\left(\widehat{g}_{\widehat{\widetilde{x}}} Q \widehat{g}_{\widehat{\widetilde{x}}}^{\top}\right)\widehat{\xi}$. This term accounts for the smearing of the joint state PDF, due to process noise.

The Fokker-Planck equation (3.5) needs to be solved for $\widehat{\xi}\left(\widehat{\widetilde{x}}(t), t\right)$ with specified initial PDF ξ_0 supported over the space of initial states and parameters. Since (3.5) is a homogeneous PDE, it does not guarantee the solution $\widehat{\xi}$ to be non-negative, or to obey the normality constraint. Hence, the numerical methods to solve this PDE must explicitly enforce these constraints.

Notice that if the diffusion vector field $\widehat{g}_{\widehat{\widetilde{x}}} = 0$, then (3.5) reduces to the Liouville equation (2.3). On the other hand, if the drift vector field $\widehat{\widetilde{f}} = 0$, then (3.5) reduces to the heat equation.

3.1.1 Examples

We now work out few examples to demonstrate the solution of Fokker-Planck equation (3.5). Just like Section 2.1.2, for the sake of notational simplicity, we drop the "hat" symbol for the time being.

3.1.1.1 Linear Drift with Constant Diffusion: Ornstein-Uhlenbeck (OU) Process

Consider the multivariate Ornstein-Uhlenbeck (OU) process $x(t) \in \mathbb{R}^{n_s}$, that satisfies the Itô SDE with linear drift and constant diffusion, given by

$$dx(t) = Ax(t)\ dt + G\, d\mathcal{W}(\omega_s, t), \tag{3.6}$$

where $\mathbb{E}[d\mathcal{W}_i] = 0$, $\mathbb{E}[d\mathcal{W}_i d\mathcal{W}_j] = Q_{ij}$; $i, j = 1, \ldots, n_w$. Further, $A \in \mathbb{R}^{n_s \times n_s}$, and $G \in \mathbb{R}^{n_s \times n_w}$. Due to process noise, the dynamics (3.6) gives rise to a state PDF $\xi(x(t), t)$, governed by the Fokker-Planck equation (3.5), which in this case becomes

$$\frac{\partial \xi}{\partial t} = -\mathrm{tr}(A)\,\xi \ + \ Ax \cdot \nabla \xi \ + \ \sum_{i=1}^{n_s}\sum_{j=1}^{n_s} \left(GQG^\top\right)_{ij} \frac{\partial^2 \xi}{\partial x_i \partial x_j}. \tag{3.7}$$

The stationary state PDF $\xi_\infty(x)$ exists iff A is Hurwitz and (A, G) is a controllable pair, and is given by $\mathcal{N}(0, \Sigma_{x\infty})$, where $\Sigma_{x\infty}$ solves the Lyapunov equation $A\Sigma_{x\infty} + \Sigma_{x\infty}A^\top + GQG^\top = 0$.

The transient PDF depends on the initial condition x_0. It can be shown that OU process preserves Gaussianity. Also, no matter what kind of uncertainty x_0 has, the transient mean vector $\mu_x(t)$ and covariance matrix $\Sigma_x(t)$ for OU process are given by (directly taking expectation and covariance of (3.6), see Section 4.4.6 in [92], for example)

$$\dot{\mu}_x(t) \ = \ A\mu_x(t), \quad \mu_x(0) = \mu_0, \tag{3.8}$$

$$\dot{\Sigma}_x(t) \ = \ A\Sigma_x(t) + \Sigma_x(t)A^\top + GQG^\top, \quad \Sigma_x(0) = \Sigma_0, \tag{3.9}$$

which can be explicitly solved as

$$\mu_x(t) = e^{At}\mu_0, \tag{3.10}$$

$$\Sigma_x(t) = e^{At}\Sigma_0 e^{A^\top t} + \int_0^t e^{A(t-\tau)}GQG^\top e^{A^\top(t-\tau)}\,d\tau. \tag{3.11}$$

Thus, $x_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$ results $x(t) \sim \xi(x(t), t) = \mathcal{N}(\mu_x(t), \Sigma_x(t))$, where $\mu_x(t)$ and $\Sigma_x(t)$ are given by (3.10) and (3.11), respectively. Furthermore, a deterministic initial condition x_0 can be seen as a special case of Gaussian, i.e. $x_0 \sim \mathcal{N}(x_0, 0)$. Thus, starting from a deterministic initial condition, the state PDF for OU process at any time t, is $\mathcal{N}\left(e^{At}x_0, \int_0^t e^{A(t-\tau)}GQG^\top e^{A^\top(t-\tau)}\,d\tau\right)$. These solutions for $\xi(x(t), t)$ can be readily verified from (3.7).

3.1.1.2 Stochastic Lure System

Consider the Itô SDE

$$dx(t) = Ax(t)\,dt + G\,d\mathcal{W}(\omega_s, t) + bf\left(c^\top x\right)\,dt, \tag{3.12}$$

where $x(t), b, c \in \mathbb{R}^{n_s}$; $\mathcal{W} \in \mathbb{R}^{n_w}$. The function $f : \mathbb{R} \mapsto \mathbb{R}$ is assumed to be piecewise continuous. The matrix $A \in \mathbb{R}^{n_s \times n_s}$ is Hurwitz stable, and (A, G) is a controllable pair. As in the previous example, $\Sigma_{x\infty}$ denotes the solution of the Lyapunov equation $A\Sigma_{x\infty} + \Sigma_{x\infty}A^\top + GQG^\top = 0$. We say two vectors u and v are proportional, if there exists $\lambda > 0$ such that $u = \lambda v$. If the vectors b and $A\Sigma_{x\infty}c$ are proportional, then the stationary PDF of (3.12) is given by

$$\xi_\infty(x) = N \exp\left(-\frac{1}{2}x^\top \Sigma_{x\infty}^{-1} x - \lambda F\left(c^\top x\right)\right), \tag{3.13}$$

where $F(z) \triangleq \int_0^z f(\theta)\, d\theta$, and N is the normalization constant. Proof for this result and its application to RLC circuit with noisy resistor, can be found in [93]. Again, one can verify that (3.13) satisfies the stationary Fokker-Planck equation corresponding to the Itô SDE (3.12).

3.1.2 Computational Challenges in Solving the Fokker-Planck PDE

The Fokker-Planck equation (3.5) is a parabolic PDE, second order in space and first order in time. Being a second order PDE, it does not enjoy the MOC solution that we employed for the first order Liouville equation in the previous chapter. Thus, compared to Liouville equation (2.3), it is this second order diffusion term that makes the numerical solution of (3.5) difficult. Function approximation techniques for solving (3.5) usually suffer from the "curse of dimensionality" [94]. For this reason, grid-based finite element methods have poor scalability with the increase in dimension. On the other hand, the method of eigenfunction expansion can be used to solve (3.5) for some special cases [95], but for general nonlinear dynamics, suffers from poor rate of convergence [96]. Next, we present a new algorithm that approximates the solution of Fokker-Planck equation (3.5).

3.2 A New Approximation Algorithm

We propose to develop an uncertainty propagation methodology based on Karhunen-Loève (KL) expansion [97], and MOC solution of Liouville equation presented in the previous Chapter. KL expansion, represents any random process as an infinite sum of homogeneous products of functions of deterministic and stochastic variables. In the context of dynamical systems, it has primarily been used in model reduction and data analysis [98]. However, its application to uncertainty propagation has been limited.

3.2.1 Karhunen-Loève Expansion for Process Noise

KL expansion was derived independently by many researchers [99, 97, 100, 101, 102] to represent a stochastic process $Y(\omega_s, t)$ as a random linear combination of a set of orthonormal deterministic L_2 functions $\{e_i(t)\}_{i=1}^{\infty}$, i.e.

$$Y(\omega_s, t) = \sum_{i=1}^{\infty} Z_i(\omega_s)\, e_i(t), \tag{3.14}$$

where $Z_i(\omega_s)$ are random variables. The idea is similar to the Fourier series expansion, where a deterministic linear combination of orthonormal L_2 functions is used. Further, if we write $Z_i(\omega_s) = \sqrt{\Lambda_i}\zeta_i(\omega_s)$, where $\Lambda_i \in \mathbb{R}^+$ and $\{\zeta_i(\omega_s)\}_{i=1}^{\infty}$ is a sequence of random variables to be determined, then $\{\Lambda_i\}_{i=1}^{\infty}$ and $\{e_i(t)\}_{i=1}^{\infty}$ can be interpreted as the eigenvalues and eigenfunctions of the *covariance function* [103]

$$C(t_1, t_2) \triangleq \operatorname{cov}\left(Y(\omega_s, t_1) - \mathbb{E}[Y(\omega_s, t_1)], Y(\omega_s, t_2) - \mathbb{E}[Y(\omega_s, t_2)]\right), \tag{3.15}$$

that admits a spectral decomposition [104] of the form $C(t_1, t_2) = \sum_{i=1}^{\infty} \Lambda_i e_i(t_1)\, e_i(t_2)$. Since the covariance function is bounded, symmetric and positive-definite, the eigenvalue problem can be cast as a homogeneous Fredholm integral equation of second kind, given by

$$\int_{\mathcal{D}_t} C(t_1, t_2)\, e_i(t_1)\; dt_1 = \Lambda_i e_i(t_2). \tag{3.16}$$

Given the covariance function of a stochastic process, the eigenvalue-eigenfunction set can be found by solving (3.16), and the resulting expansion (3.14) converges to $Y(\omega_s, t)$ in mean-square sense [105]. The following four results will serve as examples to illustrate our formulation. First two results below are well known in the literature

[104]; we will prove the last two results.

Theorem 2. *(**KL expansion of Wiener process**) For Wiener process $\mathcal{W}(\omega_s, t)$ with variance σ^2, the eigenvalues and eigenfunctions $\{\Lambda_i, e_i(t)\}_{i=1}^{\infty}$, of its covariance function $C(t_1, t_2) = (t_1 \wedge t_2)$, $t_1, t_2 \in [0, T] \times [0, T]$, are given by*

$$\Lambda_i = \frac{4T}{\pi^2 (2i-1)^2}, \qquad e_i(t) = \sqrt{2} \sin\left(\left(i - \frac{1}{2}\right)\frac{\pi t}{T}\right), \tag{3.17}$$

and hence the KL expansion of $\mathcal{W}(\omega_s, t)$, is of the form

$$\mathcal{W}(\omega_s, t) \overset{\text{m.s.}}{=} \sqrt{2} \sum_{i=1}^{\infty} \zeta_i(\omega_s) \frac{\sin\left(\left(i - \frac{1}{2}\right)\frac{\pi t}{T}\right)}{\left(i - \frac{1}{2}\right)\frac{\pi}{T}}, \tag{3.18}$$

where $\zeta_i(\omega_s)$ are i.i.d. samples drawn from $\mathcal{N}(0, \sigma^2)$.

Corollary 3. *(**KL expansion of Gaussian white noise**) Since $d\mathcal{W}(\omega_s, t) = \mathrm{GWN}(\omega_s, t)\, dt$, the KL expansion for Gaussian white noise $\mathrm{GWN}(\omega_s, t)$ can be obtained by taking the derivative of (3.18) with respect to t, i.e.*

$$\mathrm{GWN}(\omega_s, t) \overset{\text{m.s.}}{=} \sqrt{2} \sum_{i=1}^{\infty} \zeta_i(\omega_s) \cos\left(\left(i - \frac{1}{2}\right)\frac{\pi t}{T}\right), \tag{3.19}$$

where the i.i.d. random variables $\zeta_i(\omega_s) \sim \mathcal{N}(0, \sigma^2)$.

Definition 2. *The compound Poisson process $\mathcal{P}(\omega_s, t)$ is defined as [106]*

$$\mathcal{P}(\omega_s, t) = \begin{cases} 0, & \text{if } N(t) = 0, \\ \sum_{j=1}^{N(t)} \mathcal{P}_j(\omega_s), & \text{if } N(t) > 0, \end{cases} \tag{3.20}$$

where $N(t)$ is a homogeneous Poisson counting process with intensity parameter

$\lambda_p > 0$, *and* $\mathcal{P}_j(\omega_s)$ *are i.i.d random variables drawn from* $\mathcal{N}(\mu, \sigma^2)$. *The choice of Gaussian distribution for* $\mathcal{P}_j(\omega_s)$ *is a working convenience, and could be generalized. If* $\mathcal{P}_j(\omega_s)$ *is chosen to be non-Gaussian, then* $\mathcal{P}(\omega_s, t)$ *is still called compound Poisson process as long as the chosen distribution for* $\mathcal{P}_j$ *is independent to that of the counting process* $\{N(t)\}_{t\geqslant 0}$.

Theorem 4. *(**KL expansion of compound Poisson process**) For the compound Poisson process* $\mathcal{P}(\omega_s, t)$, *if* $\{\Lambda_i, e_i(t)\}_{i=1}^{\infty}$ *are the eigenvalue-eigenfunction pairs of its covariance function* $C(t_1, t_2) = \lambda_p \sigma^2 (t_1 \wedge t_2) + (\lambda_p \mu)^2 t_1 t_2$, *where* $t_1, t_2 \in [0, T] \times [0, T]$, *then the KL expansion of* $\mathcal{P}(\omega_s, t)$, *is of the form*

$$\mathcal{P}(\omega_s, t) \overset{\text{m.s.}}{=} \sum_{i=1}^{\infty} \sqrt{\Lambda_i} \zeta_i(\omega_s) e_i(t), \tag{3.21}$$

where $\zeta_i(\omega_s)$ *are i.i.d. samples drawn from* $\mathcal{N}(0, 1)$. *Further,* Λ_i *solves*

$$\tan\left(\sigma T \sqrt{\frac{\lambda_p}{\Lambda_i}}\right) = \left[1 + \frac{1}{\lambda_p T}\left(\frac{\sigma}{\mu}\right)^2\right]\left(\sigma T \sqrt{\frac{\lambda_p}{\Lambda_i}}\right),\ \lambda_p, \sigma, \mu, T > 0;\ i \in \mathbb{N}, \tag{3.22}$$

and the eigenfunctions $e_i(t) = \frac{2}{\sqrt{\left[2T - \beta_i \sin\frac{2T}{\beta_i}\right]}} \sin\left(\frac{t}{\beta_i}\right)$, *where* $\beta_i \triangleq \sqrt{\frac{\Lambda_i}{\lambda_p \sigma^2}}$.

Proof. We first solve the Fredholm integral equation of second kind (3.16), associated with the covariance function

$$C(s, t) = \lambda_p \sigma^2 (s \wedge t) + (\lambda_p \mu)^2 st, \text{where } s, t \in [0, T] \times [0, T]. \tag{3.23}$$

This leads to

$$\Lambda_i e_i(s) = (\lambda_p \mu)^2 s \int_0^T t e_i(t)\, dt + \lambda_p \sigma^2 \int_0^s t e_i(t)\, dt + \lambda_p \sigma^2 s \int_s^T e_i(t)\, dt \tag{3.24}$$

$$
\begin{aligned}
\Rightarrow \Lambda_i e_i'(s) &= (\lambda_p \mu)^2 \int_0^T t e_i(t)\, dt + \lambda_p \sigma^2 \left[s e_i(s) + \int_0^s \frac{\partial (t e_i(t))}{\partial s}\, dt \right] \\
&\quad + \lambda_p \sigma^2 \int_s^T e_i(t)\, dt + \lambda_p \sigma^2 s \left[-e_i(s) + \int_s^T \frac{\partial (e_i(t))}{\partial s}\, dt \right], \qquad (3.25) \\
\Rightarrow \Lambda_i e_i'(s) &= (\lambda_p \mu)^2 \int_0^T t e_i(t)\, dt + \lambda_p \sigma^2 \int_s^T e_i(t)\, dt, \qquad (3.26) \\
\Rightarrow \Lambda_i e_i''(s) &= -\lambda_p \sigma^2 e_i(s), \qquad \text{let } \beta_i^2 \triangleq \frac{\Lambda_i}{\lambda_p \sigma^2}, \qquad (3.27) \\
\Rightarrow e_i(t) &= A \sin\left(\frac{t}{\beta_i}\right) + B \cos\left(\frac{t}{\beta_i}\right). \qquad (3.28)
\end{aligned}
$$

Now, $\Lambda_i e_i(0) = 0 \Rightarrow B = 0$; and $\int_0^T e_i^2(t)\, dt = 1 \Rightarrow A \int_0^T \sin^2\left(\frac{t}{\beta_i}\right)\, dt = 1 \Rightarrow A = \frac{2}{\sqrt{\left[2T - \beta_i \sin\frac{2T}{\beta_i}\right]}}$. Consequently, we get

$$
e_i(t) = \frac{2}{\sqrt{\left[2T - \beta_i \sin\frac{2T}{\beta_i}\right]}} \sin\left(\frac{t}{\beta_i}\right). \qquad (3.29)
$$

Now, we substitute $s = T$ in (3.26) and use (3.29), to obtain

$$
\frac{\Lambda_i}{\beta_i} \cos\frac{T}{\beta_i} = (\lambda_p \mu)^2 \beta_i^2 \sin\frac{T}{\beta_i} - (\lambda_p \mu)^2 T \beta_i \cos\frac{T}{\beta_i}, \qquad (3.30)
$$

which is a transcendental equation in Λ_i. Next, we substitute $s = T$ in (3.24) and use (3.29), to obtain

$$
\Lambda_i \sin\frac{T}{\beta_i} = \left[(\lambda_p \mu)^2 T + \lambda_p \sigma^2\right] \left[\beta_i^2 \sin\frac{T}{\beta_i} - T \beta_i \cos\frac{T}{\beta_i}\right], \qquad (3.31)
$$

which is another transcendental equation in Λ_i. Since $\beta_i = \sqrt{\frac{\Lambda_i}{\lambda_p \sigma^2}}$, solving for eigenvalues Λ_i boils down to solve for β_i as a function of the parameters $\lambda_p, \sigma, \mu, T >$

0, and $i \in \mathbb{N}$. From (3.30) and (3.31), we have

$$\begin{aligned}
\frac{1}{(\lambda_p \mu)^2} \frac{\Lambda_i}{\beta_i} \cos \frac{T}{\beta_i} &= \frac{1}{(\lambda_p \mu)^2 T + \lambda_p \sigma^2} \Lambda_i \sin \frac{T}{\beta_i}, && (3.32) \\
\Rightarrow \beta_i \tan \frac{T}{\beta_i} &= T + \frac{\sigma^2}{\lambda_p \mu^2}, \qquad \text{since } \Lambda_i \neq 0,\ \forall i \in \mathbb{N}, && (3.33) \\
\Rightarrow \tan x &= m\, x, \qquad \text{where } x \triangleq \frac{T}{\beta_i},\ \text{and } m \triangleq 1 + \frac{1}{\lambda_p T} \left(\frac{\sigma}{\mu} \right)^2. && (3.34)
\end{aligned}$$

Thus, solving for $x > 0$ is same as finding *positive abscissa* for intersections of $\tan x$ and a straight line passing through the origin with slope $> 45^\circ$ (since $m > 1$, from (3.34). Such intersections happen in *either first or fourth quadrant*, depending on the value of m. Hence, x will be a function of $(2i-1)\,\dfrac{\pi}{2}$, up to translation. Consequently, the KL expansion of compound Poisson process $\mathcal{P}(\omega_s, t)$ is given by

$$\mathcal{P}(\omega_s, t) = \sum_{i=1}^{\infty} \sqrt{\Lambda_i} \zeta_i(\omega_s)\, e_i(t), \tag{3.35}$$

where Λ_i solves

$$\tan\left(\sigma T \sqrt{\frac{\lambda_p}{\Lambda_i}} \right) = \left[1 + \frac{1}{\lambda_p T} \left(\frac{\sigma}{\mu} \right)^2 \right] \left(\sigma T \sqrt{\frac{\lambda_p}{\Lambda_i}} \right),\ \lambda_p, \sigma, \mu, T > 0;\ i \in \mathbb{N}. \tag{3.36}$$

Further, $\zeta_i(\omega_s)$ are i.i.d random variables from $\mathcal{N}(0,1)$, and the eigenfunctions $e_i(t) = \frac{2}{\sqrt{\left[2T - \beta_i \sin \frac{2T}{\beta_i}\right]}} \sin\left(\frac{t}{\beta_i}\right)$, $\beta_i \triangleq \sqrt{\frac{\Lambda_i}{\lambda_p \sigma^2}}$. ■

Corollary 5. *(**KL expansion of Poisson white noise**) Since* $d\mathcal{P}(\omega_s, t) = \mathrm{PWN}(\omega_s, t)\, dt$, *the KL expansion for Poisson white noise* $\mathrm{PWN}(\omega_s, t)$ *is*

$$\mathrm{PWN}(\omega_s, t) \overset{\text{m.s.}}{=} \sum_{i=1}^{\infty} \sqrt{\Lambda_i} \zeta_i(\omega_s) \frac{\frac{2}{\beta_i}}{\sqrt{\left[2T - \beta_i \sin \frac{2T}{\beta_i}\right]}} \cos\left(\frac{t}{\beta_i}\right), \tag{3.37}$$

where Λ_i, $\zeta_i(\omega_s)$ and β_i are as in Theorem 4.

Proof. The proof follows by taking the derivative of the right-hand-side of (3.35) with respect to t. ■

Remark 3. *Setting the parameters $\lambda_p = 1$, $\mu = 0$, in (3.23), we recover Wiener process as a special case of compound Poisson process. Substituting the same in (3.30), we indeed recover the eigenvalues and eigenfunctions given by (3.17), for the covariance kernel of Wiener process, and consequently, the KL expansion (3.19) for Gaussian white noise.*

3.2.2 Approximating the Langevin Equation

We consider the problem of approximating the joint state PDF evolution of the stochastic flow $x(t)$ that satisfies the Itô SDE (3.1). For notational ease, we drop the "hat" (denoting model dynamics) and "tilde" (denoting extended state space) notation. Dropping the "tilde" means considering only initial condition (x_0) uncertainties, and no uncertainties in parameters p. From what follows, it will be obvious that the proposed algorithm can be applied for the extended state space, by simply augmenting $\dot{p} = \mathbf{0}$ with the actual state dynamics. Hence, there is no loss of generality due to our notational simplification.

Given the Itô SDE (3.1), we write an approximate dynamical system corresponding to its Langevin ODE for the j^{th} state as

$$\dot{x}_N^{(j)}(t) = f^{(j)}(x_N(t), t) + \sum_{k=1}^{n_w} g_x^{(j,k)}(x_N(t), t)\ \mathrm{KL}_N^{(k)}(\omega_s, t)\,, \tag{3.38}$$

where $j = 1, 2, \ldots, n_s$; and $\mathrm{KL}_N^{(k)}(\omega_s, t) = \sum_{i=1}^{N} \sqrt{\Lambda_i}\, \zeta_i^{(k)}(\omega_s)\, e_i(t)$ is the N-term KL expansion of the k^{th} component of the noise vector in the Langevin ODE form;

$k = 1, \ldots, n_w$. Here, $\{\Lambda_i, e_i(t)\}_{i=1}^{\infty}$ is the sequence of eigenvalue-eigenfunction pairs of the covariance function $C(t_1, t_2)$ associated with the additive stationary process noise, which is assumed to be a second order stochastic process. Moreover, $\zeta_i(\omega_s)$ are i.i.d. random variables drawn from the distribution of the noise stochastic process. For example, if $\mathcal{W}(\omega_s, t)$ is *Wiener process* with $C(t_1, t_2) = \sigma^2(t_1 \wedge t_2)$, $t_1, t_2 \in [0, T]$, then $\{\Lambda_i, e_i(t)\}_{i=1}^{\infty}$ is the eigen-pair sequence for *Gaussian white noise* GWN(ω_s, t) given by (3.19) in Corollary 3, and $\zeta_i(\omega_s) \sim \mathcal{N}(0, \sigma^2)$. On the other hand, if the process noise in (3.1) is compound Poisson process $\mathcal{P}(\omega_s, t)$ instead of the Wiener process $\mathcal{W}(\omega_s, t)$, then the KL expansion $\mathrm{KL}_N^{(k)}(\omega_s, t)$ in (3.38) would correspond to that of Poisson White noise, given by (3.37) in Corollary (5). Since this noise KL expansion can be done for any second order stochastic process, not necessarily Gaussian, we summarize the procedure in Table 3.1, for two example cases. We affix subscript N to the flow $x_N(t)$ of the approximate dynamics (3.38), to distinguish it from the sample path $x(t)$ of the original SDE (3.1).

Table 3.1: Noise KL expansion: Examples

Noise in SDE	$C(t_1, t_2)$ for SDE noise	Noise in Langevin ODE	KL expansion in (3.38)
$\mathcal{W}(\omega_s, t)$	$\sigma^2(t_1 \wedge t_2)$	GWN(ω_s, t)	(3.19)
$\mathcal{P}(\omega_s, t)$	$\lambda_p \sigma^2(t_1 \wedge t_2) + (\lambda_p \mu)^2 t_1 t_2$	PWN(ω_s, t)	(3.37)

Next, we augment (3.38) with the characteristic ODE

$$\dot{\xi}_N = -\nabla \cdot (f_N)\, \xi_N, \tag{3.39}$$

where f_N denotes the right-hand-side nonlinearity of (3.38), and $\xi_N\left(x_N(t), t\right)$ denotes the joint state PDF at time t, supported over the state space $x_N(t)$. In other words, (3.39) computes the evolution of joint PDF along the characteristic curves $x_N\left(t\right)$, with the joint initial PDF $\xi_N\left(x_N\left(0\right), 0\right) = \xi\left(x\left(0\right), 0\right) \triangleq \xi_0$. Notice that we do not assume the process noise to be Gaussian. As long as the additive noise has finite second moment, we can write down the approximate dynamical system (3.39) via the noise KL expansion (Table 3.1). The overall formulation is summarized in Fig. 3.1.

It is well-known [104] that as $N \to \infty$, the finite-term noise KL expansion $\mathrm{KL}_N^{(k)}\left(\omega_s, t\right)$ converges uniformly to the k^{th} component of the unstructured process noise in *m.s. sense*. However, to justify our formulation, it remains to answer whether $x_N\left(t\right)$ converges to $x\left(t\right)$, and if yes, then in what sense. For our formulation to make sense, at least $x_N\left(t\right)$ should converge to $x\left(t\right)$ in weak distributional sense, i.e. the PDF $\xi_N\left(x_N(t), t\right)$ should converge to the PDF $\xi\left(x\left(t\right), t\right)$. We next provide numerical evidence to support this. Then we will proceed to rigorously derive the convergence result.

3.2.3 Examples and Numerical Verification

3.2.3.1 Application to Vanderpol's and Duffing Oscillator

The proposed methodology is applied to a noisy Vanderpol's oscillator, given by

$$\ddot{x}\left(t\right) = \left(1 - x^2\left(t\right)\right) \dot{x}\left(t\right) - x\left(t\right) + \mathrm{GWN}\left(\omega_s, t\right), \tag{3.40}$$

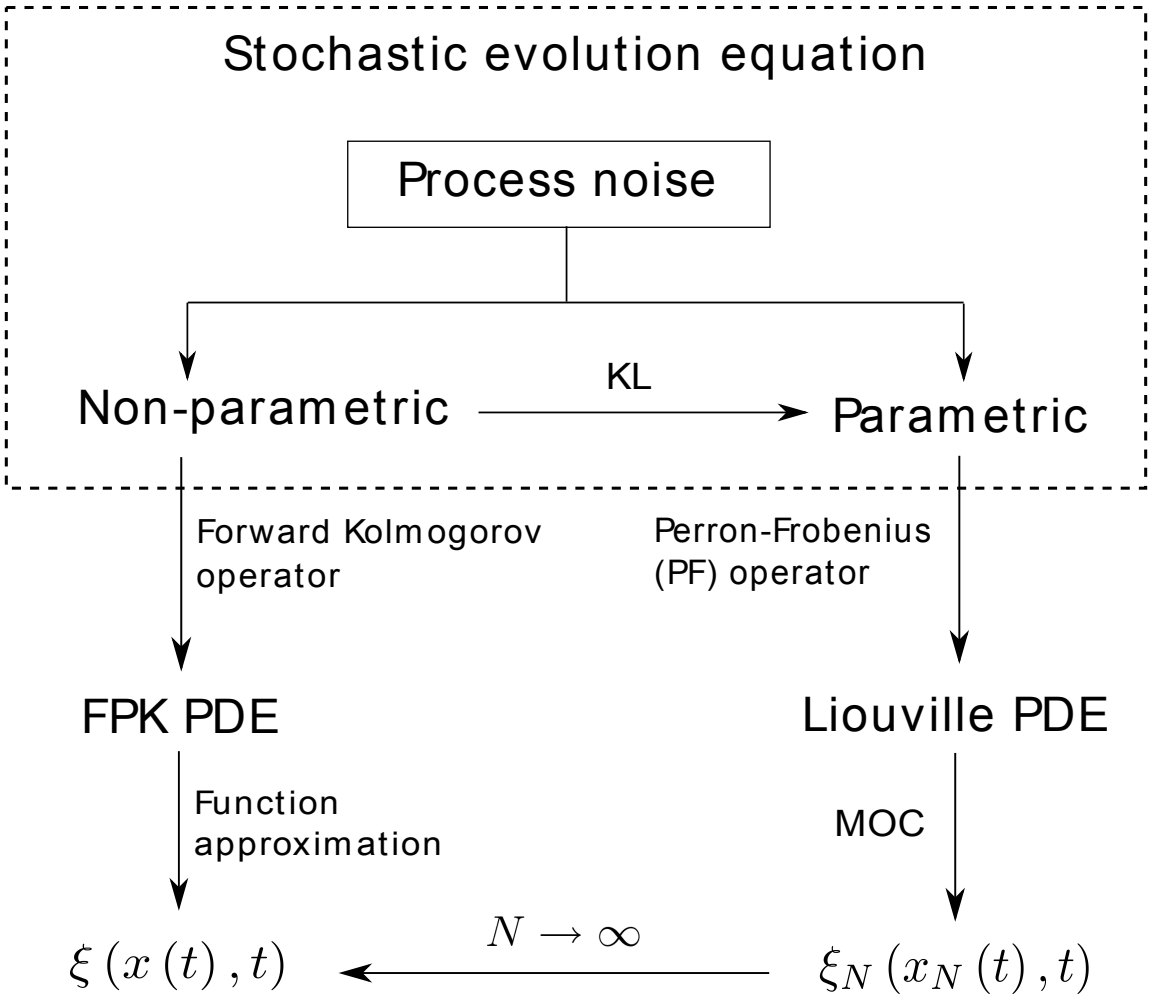


Figure 3.1: Summary of the proposed uncertainty propagation formulation.

and to a noisy Duffing oscillator, with dynamics

$$\ddot{x}\left(t\right) = 10x\left(t\right) - 30x^3\left(t\right) - 10\dot{x}\left(t\right) + \mathrm{GWN}\left(\omega_s, t\right), \tag{3.41}$$

with $\mathrm{GWN}\left(\omega_s, t\right)$ having covariance $2\pi I_2$. The initial state uncertainty, is given by the PDF $\xi_0 = \mathcal{N}\left(\left[0,0\right]^\top, \mathrm{diag}\left(1,1\right)\right)$ for both the systems (3.40) and (3.41).

In state space form, the *approximated* augmented dynamics for the Vanderpol's

oscillator, similar to (3.38) and (3.39), is given by

$$\dot{x}_N^{(1)}(t) = x_N^{(2)}(t), \tag{3.42a}$$

$$\dot{x}_N^{(2)}(t) = \left(1 - \left(x_N^{(1)}(t)\right)^2\right) x_N^{(2)}(t) - x_N^{(1)}(t) + \mathrm{KL}_N(\omega_s, t), \tag{3.42b}$$

$$\dot{\xi}_N = -\left(1 - \left(x_N^{(1)}(t)\right)^2\right) \xi_N, \tag{3.42c}$$

and for the Duffing oscillator, is given by

$$\dot{x}_N^{(1)}(t) = x_N^{(2)}(t), \tag{3.43a}$$

$$\dot{x}_N^{(2)}(t) = 10x_N^{(1)}(t) - 30\left(x_N^{(1)}(t)\right)^3 - 10x_N^{(2)}(t) + \mathrm{KL}_N(\omega_s, t), \tag{3.43b}$$

$$\dot{\xi}_N = 10\xi_N. \tag{3.43c}$$

Next, the initial PDF ξ_0 is sampled with sample size of $\nu = 5000$. For the Vanderpol's oscillator, final time T is taken to be 1 s, for the Duffing oscillator $T = 3$ s. Number of terms in the KL expansion is kept fixed as $N = 7$, for (3.42b) and (3.43b). Fig. 3.2 shows the evolution of joint state PDFs ξ_N with time for the two oscillators. It is observed that, for the Vanderpol's oscillator, the probability mass accumulates along the limit cycle; and for the Duffing oscillator, we get a bimodal PDF at final time. This is in agreement with the qualitative behavior of these systems.

3.2.3.2 Verification of Solution

As mentioned before, it is not apparent whether $x_N(t)$ converges to $x(t)$. Hence, it is important to at least numerically verify the consistency of the solution obtained through the proposed framework. Here, we outline the procedure to verify weak distributional convergence, i.e. the convergence of ξ_N to ξ as $N \to \infty$. We use the Kolmogorov-Smirnov (KS) test to verify the solution of the approximated dynamics

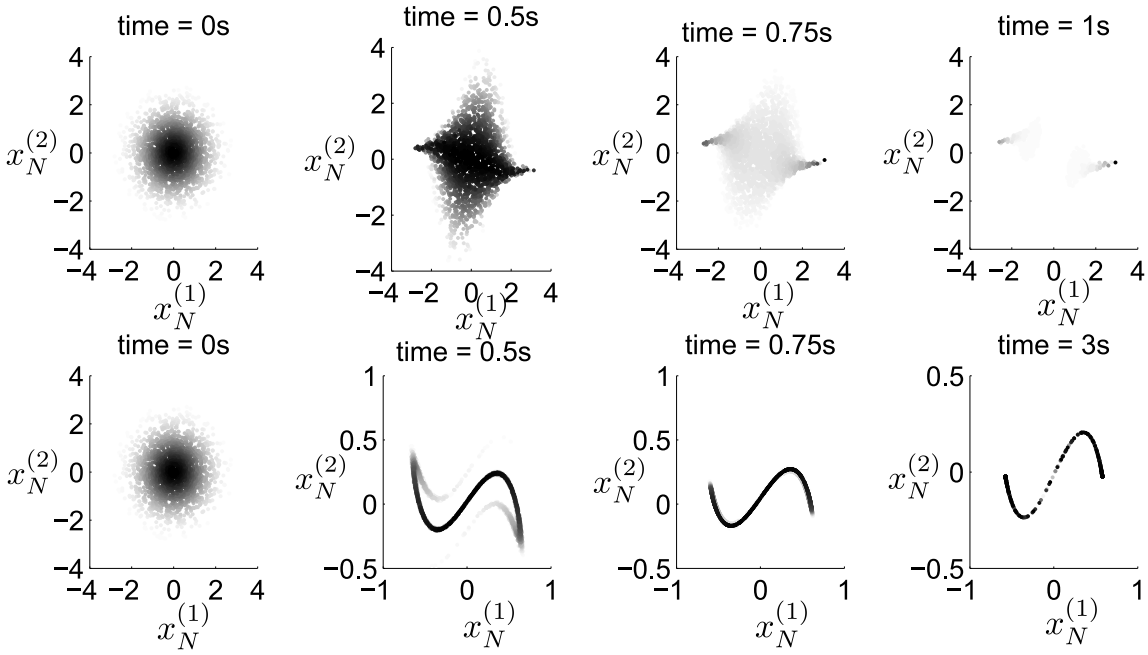


Figure 3.2: Uncertainty Propagation for Vanderpol's oscillator (3.40) in top row, and Duffing oscillator (3.41) in bottom row, obtained via integrating (3.42) and (3.43), respectively. Darker regions have higher joint PDF value than lighter regions.

(3.38).

The Kolmogorov-Smirnov test [107, 108] is a statistical hypothesis testing procedure, used to compare a sample with a reference probability distribution. It quantifies a distance D_ν, between a reference cumulative distribution function (CDF), $F(x)$, and empirical CDF of the sample being tested, $F_\nu(x)$, which is given by

$$D_\nu \triangleq \sup_x |F_\nu(x) - F(x)|, \tag{3.44}$$

where ν refers to the sample size. The null hypothesis is that the sample comes from the reference distribution. Given a significance level α, the null hypothesis is

accepted if

$$\sqrt{\nu} D_\nu \leq K_\alpha, \qquad \text{with} \quad \mathbb{P}\left(K \leq K_\alpha\right) = 1 - \alpha. \tag{3.45}$$

Here K is a random variable that follows Kolmogorov distribution, with CDF

$$\mathbb{P}\left(K \leq x\right) \triangleq F_K\left(x\right) = \frac{\sqrt{2\pi}}{x} \sum_{i=1}^{\infty} \exp\left(-\frac{\left(2i-1\right)^2 \pi^2}{8x^2}\right). \tag{3.46}$$

If (3.45) is satisfied, then the KS test is passed and it is concluded that the given sample comes from the reference PDF. Due to finite number of samples ν, the quantity D_ν is a random variable as each sample will give different D_ν values. The empirical distribution of $\sqrt{\nu} D_\nu$ is given by

$$F_{N_s}\left(x\right) = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{1}_{x < \sqrt{\nu} D_\nu^i}, \tag{3.47}$$

where N_s is the number of sample sets, and D_ν^i is D_ν value for the i^{th} sample. The Dvoretzky-Kiefer-Wolfowitz inequality [109] states that

$$\mathbb{P}\left(\sup_{x \in \mathbb{R}} \left|F_{N_s}(x) - F_K(x)\right| > \epsilon\right) \leq 2 \exp\left(-2N_s\epsilon^2\right), \quad \forall \epsilon > 0, \tag{3.48}$$

where $F_K\left(x\right)$ is the Kolmogorov CDF in (3.46). Hence, as $N_s \to \infty$, the test statistic $\sqrt{\nu} D_\nu$ exponentially converges in distribution to a Kolmogorov random variable with rate $2\exp\left(-2N_s\epsilon^2\right)$.

In our case, the analytical representation of PDF at final time t_f is not known. The only time when we know the PDF *exactly* is at time $t = 0$. Hence to verify our solution, we use the original dynamics (3.1) to back-propagate the probability weighted scattered data obtained by performing MOC on (3.38), till time $t = 0$.

Then, we want to check if the back-propagated sample belongs to the initial PDF ξ_0.

Let the elements sampled from the initial PDF be $x_{0,i}$, $i = 1, \ldots, \nu$; and the elements after propagation of (3.38) till $t = t_f$ be $x_{N,i}(t_f)$. Let the back-propagated sample obtained by propagating $x_{N,i}(t_f)$ via (3.1) during $[t_f, 0]$, be $x^b_{0,i}$. To use the KS test, the sample $x^b_{0,i}$ is compared with ξ_0, for statistical significance. Here, we use a multivariate version [107] of the KS test. The verification methodology is described in algorithm 2.

Algorithm 2 Verification of the proposed uncertainty propagation algorithm

Require: Initial PDF ξ_0, number of sample sets N_s, number of samples ν, significance level α, final time t_f, number of terms in KL expansion N, original dynamics (3.1) in n_s dimensions.

Calculate K_α using (3.45)
for $j = 1 : N_s$ **do** ▷ sample set counter
  Draw ν samples $x^j_{0,i}$ from ξ_0 ▷ using MCMC
  Create null hypothesis H_0: $\xi = \xi_0$ for current sample
  for $i = 1 : \nu$ **do** ▷ sample counter
    Propagate $x^j_{0,i}$ from $[0, t_f]$ using (3.38), to get $x^j_{N,i}(t_f)$ ▷ use N terms in noise KL expansion
    Back-propagate $x^j_{N,i}(t_f)$ from $[t_f, 0]$ using (3.1), to get $x^{b,j}_{0,i}$
    Get $u^j_{0,i} \leftarrow \xi_0\left(x^{b,j}_{0,i}\right)$ ▷ Inverse probability transform, $u^j_{0,i}$ are from $\mathcal{U}([0,1]^{n_s})$
  end for
  Calculate $G^j_\nu(u) \leftarrow \frac{1}{\nu}\sum_{i=1}^{\nu} \mathbb{1}_{u \le u^j_{0,i}}$ ▷ empirical CDF for transformed uniform sample
  Calculate $D^j_\nu \leftarrow \sup_{u \in [0,1]} |G^j_\nu(u) - G(u)|$ ▷ $G(u)$ is uniform CDF in $[0,1]^{n_s}$
  if $D^j_\nu \le \frac{K_\alpha}{\sqrt{\nu}}$ **then**
    Accept H_0 for the j^{th} sample
  else
    Reject H_0 for the j^{th} sample
  end if
end for ▷ Repeat for the next sample set

We illustrate this verification algorithm using the noisy Vanderpol's oscillator (3.42). We take the initial PDF $\xi_0 = \mathcal{N}\left([0,0]^\top, \mathrm{diag}\,(1,1)\right)$ as before, $N_s = 100$ sample sets, each same set with $\nu = 500$ samples, and $N = 21$ terms in noise KL expansion. The noise strength for GWN was taken same as before. At first, we pick a particular sample set, and perform KS test on it. Fig. 3.3 shows the location of original initial samples $\{x_{0,i}\}_{i=1}^{\nu}$ (blue circles), and after back-propagation $\{x_{0,i}^b\}_{i=1}^{\nu}$. It is observed that $\{x_{0,i}^b\}_{i=1}^{\nu}$ is clustered around the origin and sparseness increases as we move away, in agreement with the fact that ξ_0 is standard Gaussian. Plots and a detailed analysis for the KS results for this example can be found at [39].

3.2.4 Asymptotic Consistency of Approximation

In this section, we proceed to rigorously characterize the convergence of $x_N(t)$ to $x(t)$, for all t. We show that the $x_N(t)$ converges to $x(t)$ in mean square sense, which is indeed stronger than distributional convergence, that we hoped for. An illustration of this convergence result is shown in Fig. 3.4. To present the main ideas in a transparent manner, without notational clutter, we restrict to the case when the dimension of the process noise is same as that of the state vector, i.e. $n_s = n_w$, and the diffusion field being identity. It will be apparent from the derivation below that our results extend to the more general case.

Theorem 6. *Let $x\,(\omega_s, t)$ be the solution of the nonlinear Itô SDE*

$$dx(t) = f(x(t), t)dt + d\mathcal{W}(\omega_s, t), \tag{3.49a}$$

$$\Rightarrow \frac{d}{dt}x\,(t) = f\,(x, t) + \wp(\omega_s, t), \tag{3.49b}$$

where $\mathcal{W}(\omega_s, t)$ is any second order stationary stochastic process, and the process $\wp(\omega_s, t)$ is defined as $d\mathcal{W}(\omega_s, t) = \wp\,(\omega_s, t)\;dt$. Further, $f : \mathbb{R}^{n_s} \times [0, t_f] \to \mathbb{R}^{n_s}$

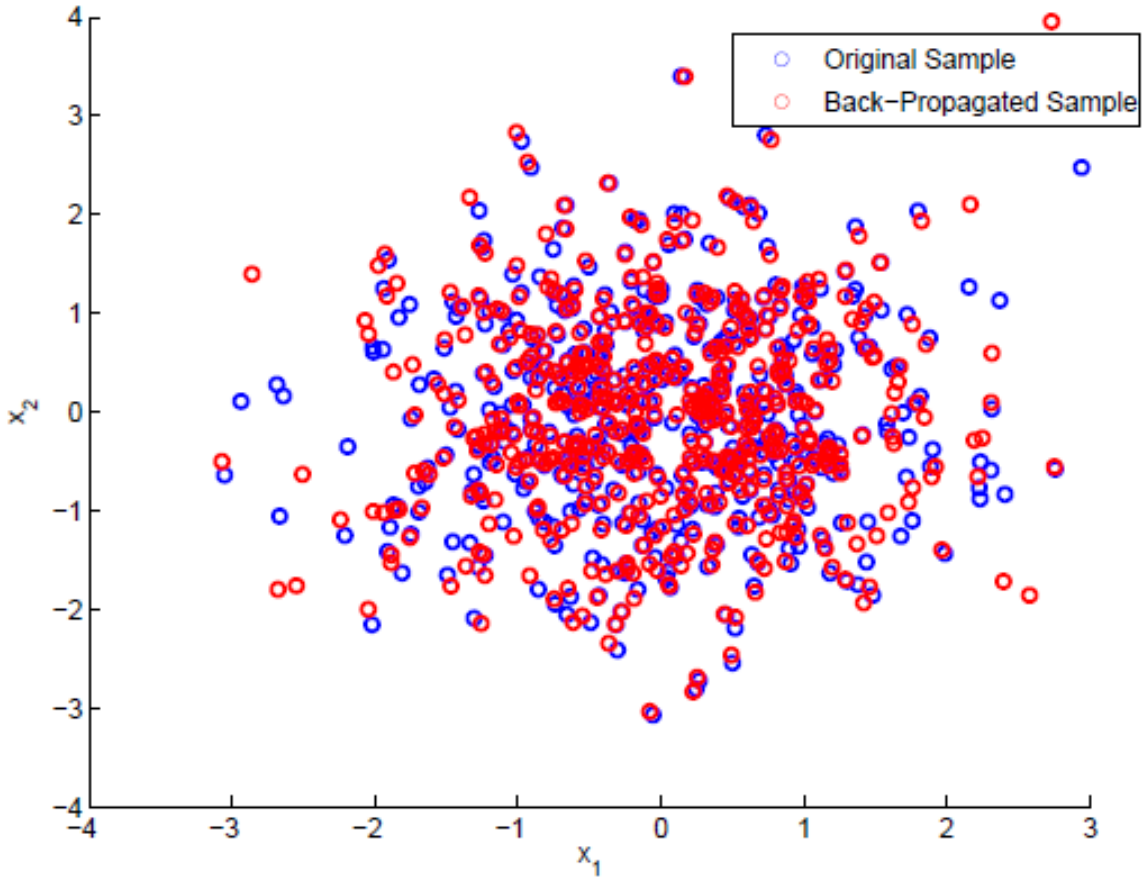


Figure 3.3: Comparison of the original (blue circles) and back-propagated samples (red circles) at $t = 0$, for the noisy Vanderpol oscillator.

satisfies the following:

1. *non-explosion condition:* $\exists\ D \geq 0,\ s.t.\ |f(x,t)| < D(1+|x|)$ *where* $x \in \mathbb{R}^{n_s}$, $t \in [0, t_f]$;

2. *Lipschitz condition:* $\exists\ C \geq 0,\ s.t.\ |f(x,t) - f(\breve{x},t)| < C\,|x - \breve{x}|$, *where* $x, \breve{x} \in \mathbb{R}^{n_s}$, $t \in [0, t_f]$.

Let $x_N(t)$ be solution of the ODE

$$\frac{d}{dt}x_N(t) = f(x_N(t), t) + \wp_N(\omega_s, t), \tag{3.50}$$

where $\wp_N(\omega_s, t)$ is the N-term truncated orthonormal expansion of the L_2 stochastic process $\wp_N(\omega_s, t)$, and $\mathbb{E}\left[\int_0^{t_f} \wp_N(\omega_s, t)dt\right] < \infty$. Then,

$$\lim_{N\to\infty} \mathbb{E}|x(t) - x_N(t)|^2 = 0, \tag{3.51}$$

iff $x_N(t)$ is the KL expansion of $x(t)$.

Proof. ($\Leftarrow$) Given (3.51) holds, we need to show $x_N(\omega_s, t)$ is the KL expansion of $x(\omega_s, t)$. Let $\{\psi_m(t)\}_{m=1}^{\infty}$ be any orthonormal basis. Then $x(\omega_s, t)$ can be written as a convergent sum in $L_2(\Omega_s, \mathscr{F}, \mathbb{P})$, i.e. $x(\omega_s, t) = \sum_{m=1}^{\infty} b_m c_m(\omega_s)\psi_m(t)$.

Let $x_N(\omega_s, t)$ be an N-term m.s. convergent approximation of $x(\omega_s, t)$, and the resulting truncation error equals $\mathscr{E}_N(\omega_s, t) = \sum_{m=N+1}^{\infty} b_m c_m(\omega_s)\psi_m(t)$. Further, projecting $x(\omega_s, t)$ onto the basis $\psi_m(t)$ results $c_m(\omega_s) = \frac{1}{b_m}\int_0^{t_f} x(\omega_s, t)\psi_m(t)dt$. For convergence, the basis $\psi_m(t)$ should minimize $\int_0^{t_f} \mathbb{E}\left[\mathscr{E}_N(\omega_s, t)\right] dt$ subject to the orthonormality constraint $\int_0^{t_f} \psi_m(t)\psi_k(t)dt = \delta_{mk},\ \forall m, k \in \mathbb{N}$.

Introducing b_m^2 as Lagrange multipliers and using the above derived formula for $c_m(\omega_s)$, the first order optimality condition yields $\int_0^{t_f} C_{xx}(t_1, t_2)\psi_m(t_1)dt_1 = b_m^2\psi_m(t_2)$, which is the Fredholm integral equation of second kind for the covariance function of random process $x(\omega_s, t)$. Hence $\{b_m^2, \psi_m(t)\}_{m=1}^{\infty}$ is the eigenvalue-eigenfunction sequence for $C_{xx}(t_1, t_2)$. Thus, the original expansion is indeed a KL expansion. ■

Proof. ($\Rightarrow$) To proceed, we need the following uniqueness conditions on (i) solution

of (3.49a), and (ii) KL expansion of a random process.

Proposition 1 ([19], Chap. 5)**.** *Given, the non-explosion condition and the Lipschitz condition are satisfied for $f(\cdot,\cdot)$ in (3.49a). Let Z be a random variable, independent of the σ-algebra generated by $\wp(\omega_s,t), t\geq 0$, and $\mathbb{E}\left[|Z|^2\right]<\infty$. Then the SDE (3.49a) where $t\in[0,t_f], x(\omega_s,0)=Z$, has a unique t-continuous solution $x(\omega,t)$ adapted to the filtration $\mathscr{F}_t^Z$ generated by Z, and $\mathbb{E}\left[\int_0^{t_f}|x(\omega_s,t)|^2dt\right]<\infty$.*

Proposition 2 ([110], Chap. 2)**.** *The Karhunen-Loève expansion of a random process $x(\omega_s,t)$, given by $x(\omega_s,t)=\sum_{i=1}^{\infty}\sqrt{\Lambda_i}\zeta_i(\omega_s)e_i(t)$, is unique.*

Let us assume that $\breve{x}_N(\omega_s,t)$ is the KL expansion of $x(\omega_s,t)$. Furthermore, if possible, assume that $\breve{x}_N(\omega_s,t)\neq x_N(\omega_s,t)$, which is the solution of (3.50) and converges to the solution of (3.49a) in m.s. sense.

Notice that (3.50) has unique solution as the right-hand-side of (3.50) satisfies Lipschitz condition. This can be proved as follows: for the right-hand-side of (3.50) to satisfy Lipschitz condition, we must have $|f(x,t)+\wp_N(\omega_s,t)-f(\breve{x},t)-\wp_N(\omega_s,t)|\leq C\,|x-\breve{x}|$, which is true since $f(\cdot,\cdot)$ itself satisfies Lipschitz condition. Hence, (3.49a) has unique solution that admits a unique KL expansion. Also according to our assumption, the solution of (3.50) converges to the solution of (3.49a) in m.s. sense. This contradicts our assumption that $\breve{x}_N(\omega_s,t)\neq x_N(\omega_s,t)$, which completes the proof. ■

Remark 4. *Theorem 6 states conditions upon the solutions of approximated and true systems for m.s. convergence to hold, under certain assumptions on the nonlinearities. No condition has been imposed yet on the initial states, which we investigate next.*

Theorem 7. *Given the stochastic dynamical system*

$$dx(t) = \quad f(x(t),t)dt + d\mathcal{W}\left(\omega_s, t\right), \tag{3.52}$$

and its corresponding N-term KL approximation given by

$$dx_N^{(j)}(t) = \quad f^{(j)}(x_N(t),t)\,dt + \sum_{i=1}^{N} \sqrt{\Lambda_i}\zeta_i^{(j)}(\omega_s)\dot{e}_i(t)\,dt, \tag{3.53}$$

where, $\lim\limits_{N\to\infty} \mathbb{E}\left|\mathcal{W}^{(j)}\left(\omega_s,t\right) - \sum\limits_{i=1}^{N}\sqrt{\Lambda_i}\zeta_i^{(j)}(\omega_s)e_i(t)\right|^2 = 0,\ \forall\ j = 1,2,\ldots,n_s.$ *Then,* $\lim\limits_{N\to\infty}\mathbb{E}|x(t)-x_N(t)|^2 = 0$, *if* $x(0) = x_N(0)$.

Proof. Integrating (3.52) and (3.53) and taking the expected value of square of the difference, we obtain

$$\begin{aligned}
&\mathbb{E}|x(t)-x_N(t)|^2 \\
=\ & \mathbb{E}\left[\left|(x(0)-x_N(0)) + \int_0^t \left(f(x,s) - f\left(x_N,s\right)\right)ds + \int_0^t d(\mathcal{W}_s - \sum_{i=1}^{N}\sqrt{\Lambda_i}\zeta_i(\omega_s)e_i(s))\right|^2\right], \\
\leq\ & \underbrace{\mathbb{E}|(x(0)-x_N(0))|^2}_{0=:B\text{ (say)}} + \mathbb{E}\left|\int_0^t (f(x,s)-f(x_N,s))ds\right|^2 + \mathbb{E}\left|\int_0^t d(\mathcal{W}_s - \sum_{i=1}^{N}\sqrt{\Lambda_i}\zeta_i(\omega_s)e_i(s))\right|^2, \\
\leq\ & B + t\mathbb{E}\int_0^t |f(x,s)-f(x_N,s)|^2\,ds + \mathbb{E}\left|\int_0^t d(\mathcal{W}_s - \sum_{i=1}^{N}\sqrt{\Lambda_i}\zeta_i(\omega_s)e_i(s))\right|^2,
\end{aligned} \tag{3.54}$$

where in the last step, we used Chebyshev's integral inequality. Consequently, we have

$$\begin{aligned}
&\lim_{N\to\infty}\mathbb{E}|x(t)-x_N(t)|^2 \leq B + \lim_{N\to\infty} t\mathbb{E}\left[\int_0^t |f(x,s)-f(x_N,s)|^2\,ds\right] \\
&+ \lim_{N\to\infty}\mathbb{E}\left|\int_0^t d(\mathcal{W}_s - \sum_{i=1}^{N}\sqrt{\Lambda_i}\zeta_i(\omega_s)e_i(s))\right|^2.
\end{aligned} \tag{3.55}$$

Using the Lipschitz criterion and property of KL expansion, from (3.55) we get

$$\underbrace{\lim_{N\to\infty} \mathbb{E}|x(t) - x_N(t)|^2}_{v(t)\text{ (say)}} \leq B + tC \int_0^t \lim_{N\to\infty} \mathbb{E}\,|x(s) - x_N(s)|^2\, ds,$$

$$\Rightarrow v(t) \leq B + A\int_0^t v(s)ds \Rightarrow v(t) \leq B\exp(At), \tag{3.56}$$

where the last step follows from Gronwall's inequality, with $tC \leq A, \forall t \in (0, t_f]$. Therefore, $\lim_{N\to\infty} \mathbb{E}|x(t) - x_N(t)|^2 = 0$, since $x(0) = x_N(0) \Rightarrow B = 0$, as per our assumption. ■

Corollary 8. *Suppose* $x_N(0) \neq x\,(0)$. *If* $x_N(0)$ *is the* generalized polynomial chaos *expansion of* $x(0)$, *then* $\lim_{N\to\infty} \mathbb{E}|x(t) - x_N(t)|^2 = 0$.

Proof. In the proof of Theorem 7, for $x(0) \neq x_N(0)$, taking the limit $N \to \infty$ yields

$$\begin{aligned}\lim_{N\to\infty} \mathbb{E}|x(t) - x_N(t)|^2 \leq \lim_{N\to\infty} \mathbb{E}|(x(0) - x_N(0))|^2 + \\ \lim_{N\to\infty} t\mathbb{E}\int_0^t |f(x,s) - f(x_N,s)|^2\, ds + \\ \lim_{N\to\infty} \mathbb{E}\left|\int_0^t d(\mathcal{W}_s - \sum_{i=1}^{N} \sqrt{\Lambda_i}\zeta_i(\omega_s)e_i(s))\right|^2.\end{aligned}$$

Going through the subsequent steps as before, we arrive at

$$\lim_{N\to\infty} \mathbb{E}|x(t) - x_N(t)|^2 = 0, \text{ if } \lim_{N\to\infty} \mathbb{E}|x(0) - x_N(0)|^2 = 0.$$

However, if $x_N(0)$ is the generalized polynomial chaos expansion of $x(0)$, then they asymptotically converge in m.s. sense [110]. Hence $\lim_{N\to\infty} \mathbb{E}|x(0) - x_N(0)|^2 = 0$, which, from the Gronwall's inequality, implies that $\lim_{N\to\infty} \mathbb{E}|x(t) - x_N(t)|^2 = 0$. This completes our proof. ■

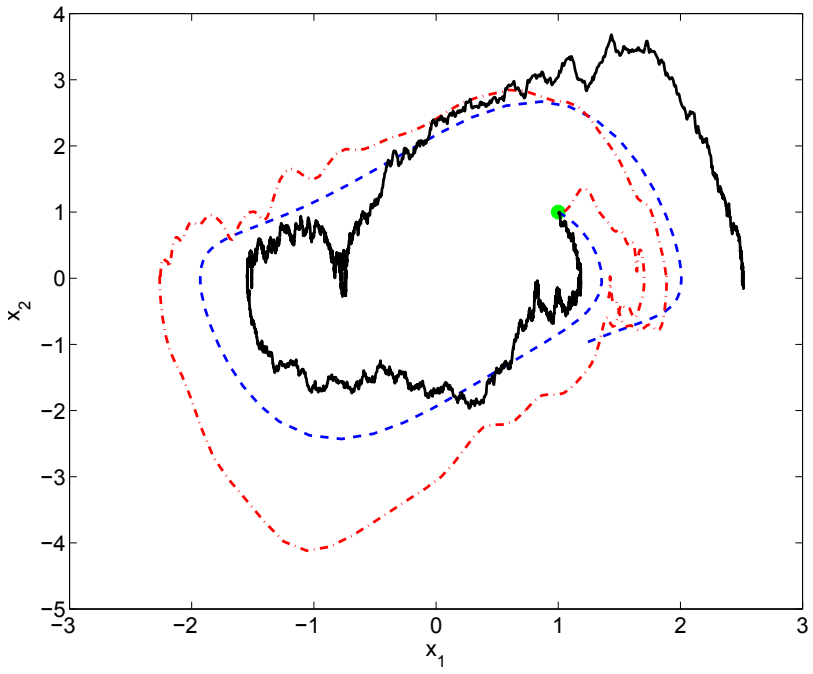


Figure 3.4: This plot illustrates the asymptotic convergence results developed in Section 3.2.4, for the noisy Vanderpol oscillator given by (3.40). Starting from the same initial condition (1; 1), denoted by the filled circle, the dashed, and solid curves show the deterministic (zero noise), and stochastic (SDE sample path with zero-mean additive Gaussian noise of variance 0.25) trajectories, respectively. The dash-dotted curve is the trajectory of the approximated system of the form (3.42) with $N = 100$ terms in the noise KL expansion, starting from the same initial condition, with process noise same as that of the SDE path. As N increases, the dash-dotted curve converges to the solid curve in mean-square sense.

3.2.5 Rate of Convergence of Approximation

The results of the above subsection proves that our proposed algorithm for uncertainty propagation is asymptotically consistent. Next, we want to investigate if we can characterize the rate-of-convergence for our proposed algorithm. This issue is important for numerical implementation perspective. One relevant fact [111] for this purpose is that the rate-of-convergence of KL expansion is governed by the decay rate of the sum of the eigenvalues of the covariance function. We provide an example

to demonstrate this.

3.2.5.1 Example: Rate of Convergence of KL Expansion of Wiener Process

Consider the stochastic process $dx(\omega_s, t) = d\mathcal{W}(\omega_s, t)$. Eigenvalues of the covariance function of $x(\omega_s, t)$ are given by (3.17). Consequently, the truncation error due to N-term KL approximation of $x(\omega_s, t)$, is

$$\begin{aligned}\sum_{i=N+1}^{\infty} \lambda_i &= \sum_{i=1}^{\infty} \lambda_i - \sum_{i=1}^{N} \lambda_i \\ &= \frac{4T}{\pi^2}\left[\frac{\pi^2}{8} - \frac{1}{8}\left(\pi^2 - 2\psi^{(1)}\left(N + \frac{1}{2}\right)\right)\right] = \frac{T}{\pi^2}\psi^{(1)}\left(N + \frac{1}{2}\right), \quad (3.57)\end{aligned}$$

which decays faster than $\mathcal{O}(N^{-1})$, but slower than $\mathcal{O}(N^{-2})$. Here $\psi^{(1)}(.)$ denotes the *trigamma function* [112].

3.2.5.2 Rate of Convergence and Smoothness of the Covariance Function

Recently, it has been shown [113] that one can estimate the convergence rate of KL expansion, by knowing the smoothness of the covariance function, even though the eigenvalues are not available analytically. The main result is that exponential convergence occurs when the covariance function is analytic. If the latter has Sobolev regularity, the KL expansion has algebraic decay. The following example on geometric Brownian motion (GBM) illustrates this.

3.2.5.3 Example: Rate of Convergence of KL Expansion of GBM

Consider the GBM $dx(\omega_s, t) = ax(\omega_s, t)\, dt + bx(\omega_s, t)\, d\mathcal{W}(\omega, t)$, where a, b are deterministic constants, and $\mathcal{W}(\omega, t)$ is the standard Brownian motion. The covariance function for $x(\omega_s, t)$ can be computed from definition as

$$C(t_1, t_2) = (x_0(\omega_s))^2 e^{a(t_1+t_2)}\left(e^{b^2 t_2} - 1\right). \quad (3.58)$$

Since $C(t_1, t_2)$ is analytic, the convergence is exponential.

Remark 5. *In the general nonlinear SDE setting, the covariance function of the state is not known beforehand. What is known here is the dynamics, i.e. the drift and diffusion fields. Thus, one could look for connections between the smoothness of the dynamics with the regularity of $C(t_1, t_2)$. This will be a topic of our future work.*

3.3 Case Study: Nonlinear Estimation with Approximated Prior Dynamics

To demonstrate our "first KL, then MOC" (henceforth KLMOC) algorithm, we now focus on estimation problems, where we propose to compute the prior PDF using our proposed method, followed by Bayesian update to result the posterior PDF. To understand the context of our contribution, we outline some background behind the state-of-the-art for the computation of nonlinear estimation.

State and parameter estimation for nonlinear systems are commonly done using *sequential Monte Carlo* methods, *particle filter* being the most popular amongst them [114]. It is well known [115] that these methods require large number of samples for convergence, leading to higher computational cost. This problem is usually tackled by combining particle filters with *resampling* [116, 117], commonly known as *bootstrap filters* [118]. However, resampling may introduce loss of diversity amongst particles [119]. Several other methods like *regularized particle filter* [120], and *filters with Markov Chain Monte Carlo (MCMC) move step* [121], have been proposed to enhance sample diversity. At the same time, even with resampling, due to the simulation based nature of these filters, the sample size scales exponentially with state dimension [122]. To circumvent this problem, particle filters based on *Rao-Blackwellization* [123] have been proposed to partially solve the estimation problem analytically. However, its application remains limited to systems where the required partition of the state space is possible.

The main novelty in our approach is to recognize the fact that much of the computational burden of particle filter, stems from the Monte Carlo approximation of the prior. Lack of statistically consistent methods for high dimensional uncertainty propagation, has stymied the accurate computation of prior density. Hence it is reasonable to expect that our proposed improvement in the prior PDF computation will also lead to improvement in the estimation performance. We demonstrate these ideas through the following examples.

3.3.1 Simulation Setup

First, we consider two examples for which the estimation problem is exactly solvable and hence the true posterior is known. To demonstrate the performance improvement achieved by KLMOC compared to particle filter, we must show that the posterior computed from our algorithm, is closer to the true posterior, than particle filter. In other words, the "distance" between KLMOC posterior and true posterior, must remain smaller than the "distance" between particle filter posterior and true posterior, *for all times*. The notion of distributional distance used here, is the *quadratic Wasserstein metric of order two* (denoted as ${}_2W_2$), that measures the *difference in shapes* between the two statistical distributions under comparison. We will formally introduce the Wasserstein distance in the next Chapter, and show that it is a metric on the manifold of PDFs, and that its value has the physical interpretation of minimum amount of work needed to morph one PDF to the other. For the time being, the distance ${}_2W_2$ can be thought of as the minimum mean square error between two random vectors whose PDFs under comparison.

3.3.2 Kalman Filter

Let us consider the continuous-discrete Kalman filter with continuous-time state dynamics

$$\dot{x}(t) = -0.05\, I_2\, x(t) + [1 \quad 1]^\top \mathrm{GWN}_x(t), \tag{3.59}$$

and discrete-time measurement model

$$y(k) = [1 \quad 1]\; x(k) + \mathrm{GWN}_y(k), \qquad k \in \mathbb{N}, \tag{3.60}$$

where the process noise $\mathrm{GWN}_x(t)$ is a two dimensional zero mean Gaussian white noise process, with covariance $Q = \frac{1}{8} I_2$. The measurement noise $\mathrm{GWN}_y(k)$ is an one dimensional zero mean Gaussian white noise process, with variance $R = \frac{1}{4}$. We assume the initial joint state PDF to be $\mathcal{N}\left([1 \quad 1]^\top, \mathrm{diag}(1,1)\right)$.

From this initial state PDF, we draw 100 sample sets, each with sample size 500. Then we compute two Wasserstein time histories: ${}_2W_2\left(\xi^+_{\mathrm{Kalman}}(t), \xi^+_{\mathrm{Particle}}(t)\right)$ and ${}_2W_2\left(\xi^+_{\mathrm{Kalman}}(t), \xi^+_{\mathrm{KLMOC}}(t)\right)$, where $\xi^+_{\mathrm{Kalman}}(t)$, $\xi^+_{\mathrm{Particle}}(t)$ and $\xi^+_{\mathrm{KLMOC}}(t)$ denote posteriors at time t, obtained from Kalman filter, particle filter and KLMOC filter, respectively. The Wasserstein distance ${}_2W_2$ between two multivariate Gaussians can be computed in closed form, and will be detailed in Section 4.2. The means and standard deviations of these time histories are shown in Fig. 3.5. This plot shows that the KLMOC filter posterior remains indeed closer to the Kalman posterior, compared to the particle filter posterior.

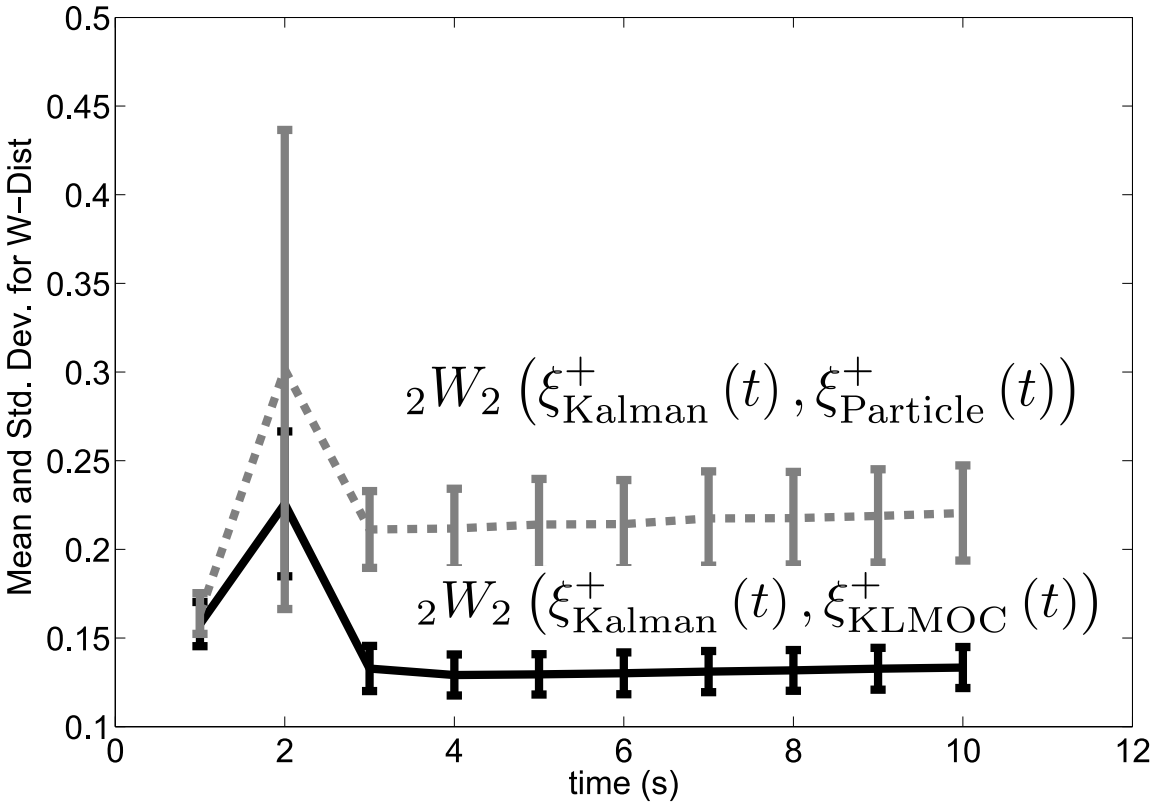


Figure 3.5: Plot of means and standard deviations of the Wasserstein distances of the posteriors from KLMOC filter (*solid line*) and the particle filter (*hyphenated line*) for the Kalman filter. The vertical lines about the means represent $\pm 1\sigma$ limits.

3.3.3 Beneš Filter

Beneš filter is one of the few [124] nonlinear filters which admit a *known finite-dimensional solution* of the nonlinear estimation problem. Here, the nonlinear drift in state dynamics, is assumed to satisfy a Riccati differential equation [125] and the measurement model is taken to be affine in states. We consider the continuous-continuous scalar Beneš filtering problem of the form:

$$dx(t) = \frac{\kappa e^x - e^{-x}}{\kappa e^x + e^{-x}}\, dt + d\mathcal{W}(\omega, t), \tag{3.61}$$

$$dy(t) = x(t)\ dt + d\mathcal{V}(\omega, t), \tag{3.62}$$

with $\kappa = 0.5$ and deterministic initial condition x_0. The process and measurement noise densities are $\mathcal{N}(0, Q)$ and $\mathcal{N}(0, R)$ respectively, with $Q = 1$, $R = 10$. It can be shown [126] that the drift nonlinearity satisfies the necessary Riccati condition and the resulting solution [127] is given by the normalized posterior density

$$\xi\left(x\left(t\right)|\mathcal{Y}_t\right) = \sqrt{\frac{\coth(t)}{2\pi}}\left(\frac{\kappa e^{x} + e^{-x}}{\kappa e^{\mathcal{I}_t(y(\omega_s))} + e^{-\mathcal{I}_t(y(\omega_s))}}\right)\exp\left(-\frac{1}{2}\Gamma\left(t\right)\right), \tag{3.63}$$

where $\mathcal{Y}_t$ is the history (filtration) till time t, and

$$\mathcal{I}_t\left(y\left(\omega_s\right)\right) \triangleq \operatorname{sech}(t)\left[x_0 + \int_0^t \sinh(s) dy\left(\omega_s\right)\right], \tag{3.64}$$

$$\Gamma\left(t\right) \triangleq \tanh(t) + \coth(t)\left(x - \mathcal{I}_t\left(y\left(\omega_s\right)\right)\right)^2. \tag{3.65}$$

for this nonlinear non-Gaussian estimation problem, unlike Kalman filter case, we can not write the Wasserstein distance between the true posterior (3.63) and particle filter/KLMOC posterior, as an analytical expression in terms of the respective sufficient statistics. It will be shown in next Chapter that computing ${}_2W_2$ between two generic multivariate non-Gaussian PDFs, amounts to solving a linear program. At each time, we sample (3.63) using the *Metropolis-Hastings MCMC* technique [71], and solve the LP between the sampled true Beneš posterior and particle filter/KLMOC posterior, to result the Wasserstein trajectories shown in Fig. 3.6. Like the Kalman filter case, as time progresses, KLMOC posterior gets closer, compared to particle filter, to true Beneš posterior.

3.3.4 Nonlinear Estimation for Mars Hypersonic Entry

The KLMOC filtering technique is applied next to estimate states of a hypersonic spacecraft entering the atmosphere of Mars. The entry dynamics is given by the

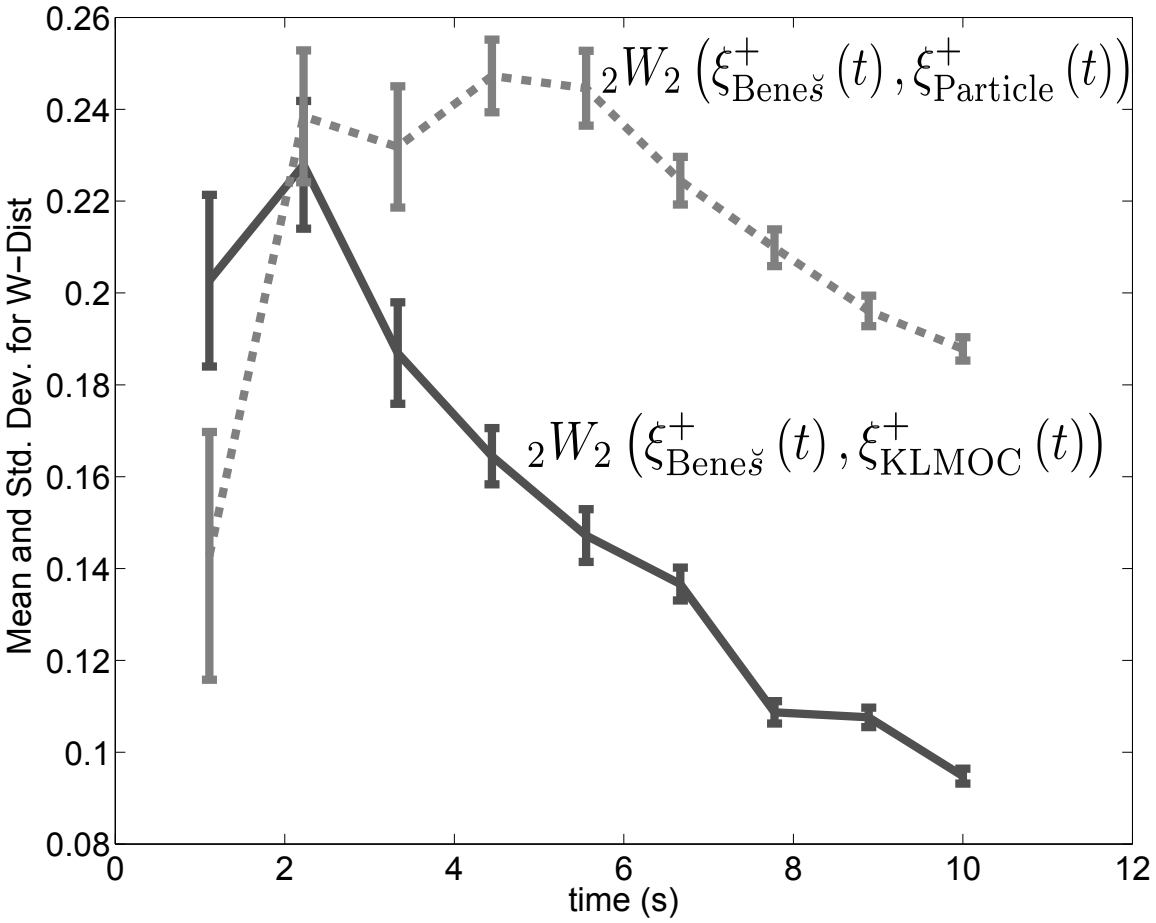


Figure 3.6: Plot of means and standard deviations of the Wasserstein distances of the posteriors from KLPF filter (*solid line*) and the particle filter (*hyphenated line*) for the Beneš filter. The vertical lines about the means represent $\pm 1\sigma$ limits.

"noisy" version of the six state Vinh's equation given in Section 2.2.1, i.e.

$$\dot{h} = V \sin\gamma + \text{GWN}_h, \tag{3.66a}$$

$$\dot{\zeta} = \frac{V \cos\gamma \sin\chi}{(1+h)} + \text{GWN}_\zeta, \tag{3.66b}$$

$$\dot{\lambda} = \frac{V \cos\gamma \cos\chi}{(1+h)\cos\zeta} + \text{GWN}_\lambda, \tag{3.66c}$$

$$\dot{V} = -\frac{\rho R_0}{2B_c} V^2 - \frac{g R_0}{v_c^2} \sin\gamma$$

$$+ \quad \frac{R_0^2\Omega_r^2}{v_c^2}(1+h)\cos\zeta\left(\sin\gamma\cos\zeta - \cos\gamma\sin\zeta\sin\chi\right) + \mathrm{GWN}_V, \tag{3.66d}$$

$$\dot{\gamma} = \frac{\rho R_0}{2B_c}\frac{C_L}{C_D}V\cos\sigma_b + \frac{gR_0}{v_c^2}\cos\gamma\left(\frac{V}{1+h} - \frac{1}{V}\right) + \mathrm{GWN}_\gamma, \tag{3.66e}$$

$$\dot{\chi} = \frac{\rho R_0}{2B_c}\frac{C_L}{C_D}\frac{V\sin\sigma_b}{\cos\gamma} - \frac{V\cos\gamma}{(1+h)}\tan\zeta\cos\chi + \frac{2R_0\Omega_r}{v_c}\left(\tan\gamma\cos\zeta\sin\chi - \sin\zeta\right)$$

$$- \quad \frac{R_0^2\Omega_r^2}{v_c^2}\frac{(1+h)}{V\cos\gamma}\sin\zeta\cos\zeta\cos\chi + \mathrm{GWN}_\chi. \tag{3.66f}$$

The *measurement model* is given by

$$y_1 = \widetilde{q} + \mathrm{GWN}_{y_1}, \tag{3.67a}$$

$$y_2 = H + \mathrm{GWN}_{y_2}, \tag{3.67b}$$

$$y_3 = \zeta + \mathrm{GWN}_{y_3}, \tag{3.67c}$$

$$y_4 = \lambda + \mathrm{GWN}_{y_4}, \tag{3.67d}$$

$$y_5 = \gamma + \mathrm{GWN}_{y_5}, \tag{3.67e}$$

$$y_6 = \chi + \mathrm{GWN}_{y_6}. \tag{3.67f}$$

where $\widetilde{q}$ is the dynamic pressure, and H is the heating rate; and they are given by

$$\widetilde{q} = \frac{1}{2}\rho V^2, \tag{3.68a}$$

$$H = K\rho^{1/2}V^{3.15}. \tag{3.68b}$$

Here $K = 4.47228 \times 10^{-9}$ is the scaled material heating coefficient [128]. Each component of the process noise vector $(\mathrm{GWN}_h, \mathrm{GWN}_\zeta, \mathrm{GWN}_\lambda, \mathrm{GWN}_V, \mathrm{GWN}_\gamma, \mathrm{GWN}_\chi)^\top$ is a zero mean mutually uncorrelated Gaussian white noise, with appropriate units. Same holds true for the 6×1 measurement noise vector, which is also uncorrelated with the process noise vector. The process and measurement noise covariances are

taken as $Q = 3.6 \times 10^{-5} I_6$, and $R = 3.6 \times 10^{-3} I_6$, respectively. The parameter values are as in the simulation set up described in Section 2.2.2.

Starting with an initial state PDF $\mathcal{N}(\mu_0, \Sigma_0)$, with

$$\mu_0 = [54 \text{ Km}, -60^\circ, 30^\circ, 2.4 \text{ Km/s}, -9^\circ, 0.0573^\circ]^\top, \tag{3.69}$$

$$\Sigma_0 = \text{diag}(5.4 \text{ Km}, 3^\circ, 3^\circ, 240 \text{ m/s}, 0.9^\circ, 0.0057^\circ), \tag{3.70}$$

both particle filter and KLMOC filtering schemes are applied to estimate the state vector $x = [h, \zeta, \lambda, V, \gamma, \chi]^\top$. For presenting the results, instead of h in Km, we often plot $r \triangleq (R_0 + h)$ in Km, which denotes the distance of the spacecraft's center-of-mass, measured from the center of the planet Mars. Fig. 3.7 shows the plots for square root of the difference between the respective variance (σ_{x_i}) and *Cramer-Rao lower bound* (CRLB_{x_i}) for each state x_i. The performance of KLMOC filter with sample size 3000, is compared with the same for particle filters with 3000, 20,000 and 50,000 particles. It can be observed that $\sqrt{\sigma_{x_i}^2 - \text{CRLB}_{x_i}}$ for KLMOC filter is lower than that of the particle filters for all the states. This demonstrates that the solution obtained from proposed estimation scheme remains closer to the true minimum variance solution than that obtained from particle filters. Next, we plot the final posterior univariate and bivariate marginals obtained from the two filtering methods, computed using the algorithm given in [38]. Fig. 3.8 and 3.9 respectively compare the univariate and bivariate marginals, obtained from KLMOC estimator and particle filter, both with 3000 samples. It can be observed that the KLMOC estimator is able to reduce variance and capture localization of uncertainty better than the particle filter, with same number of samples. This remains true even when the performance of KLMOC estimator with 3000 samples, is compared with a particle filter with 50,000 samples (Fig. 3.10 and 3.11).

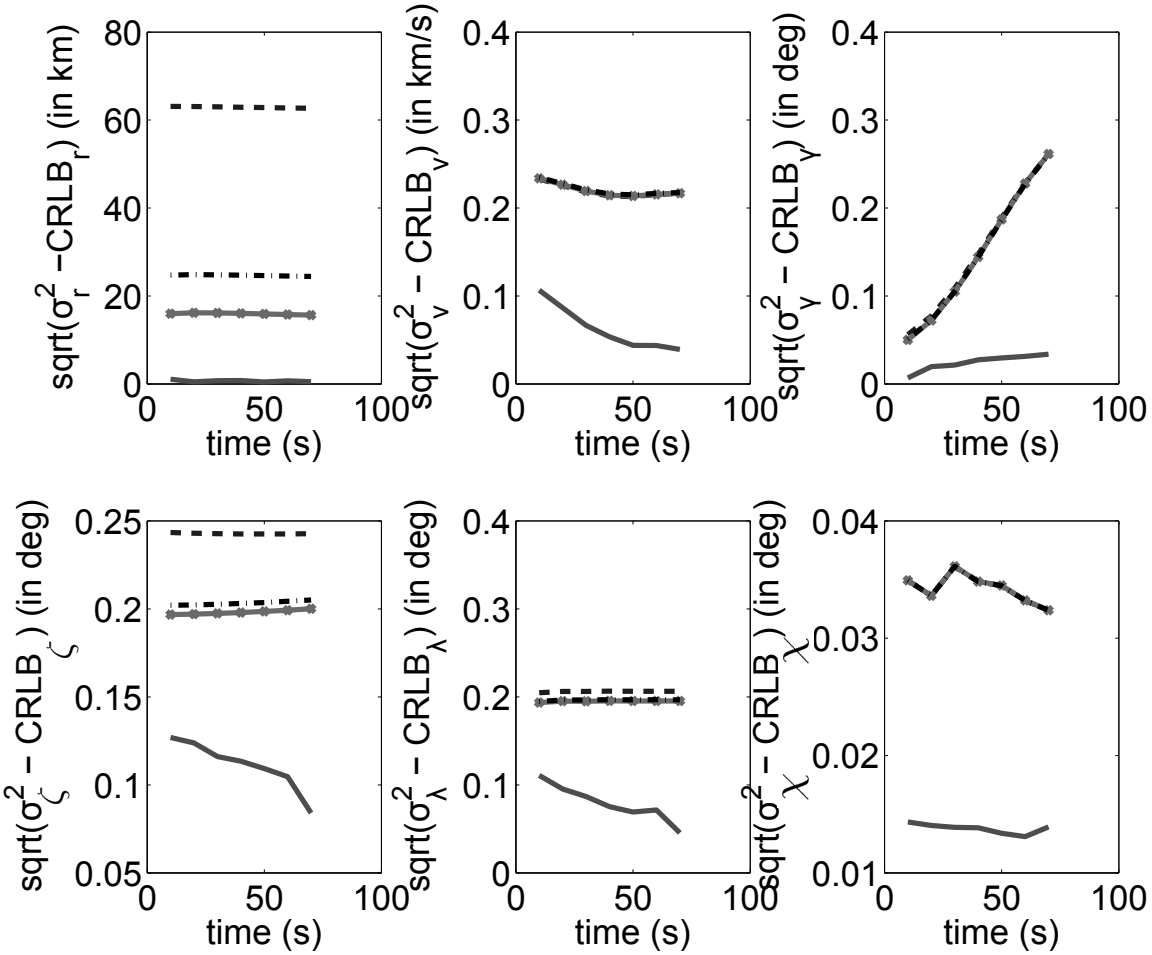


Figure 3.7: Plots for $\sqrt{\sigma_{x_i}^2 - CRLB_{x_i}}$ for states $x_1, \ldots, x_6$. The *solid* line represents KLMOC filter (3000 particles). The *hyphenated*, *hyphen-dotted* and *solid-asterixed* lines represent particle filters with 3000, 20,000 and 50,000 particles, respectively.

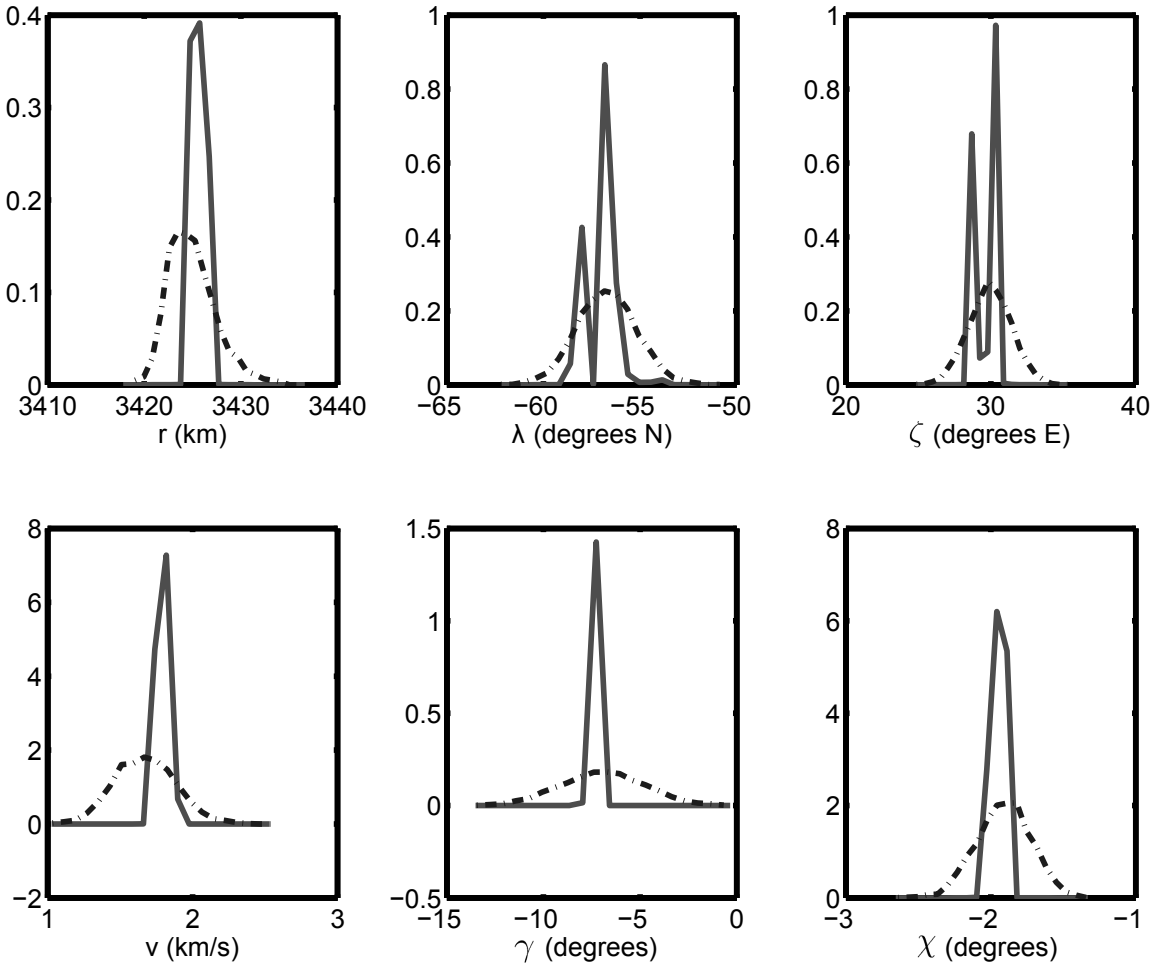
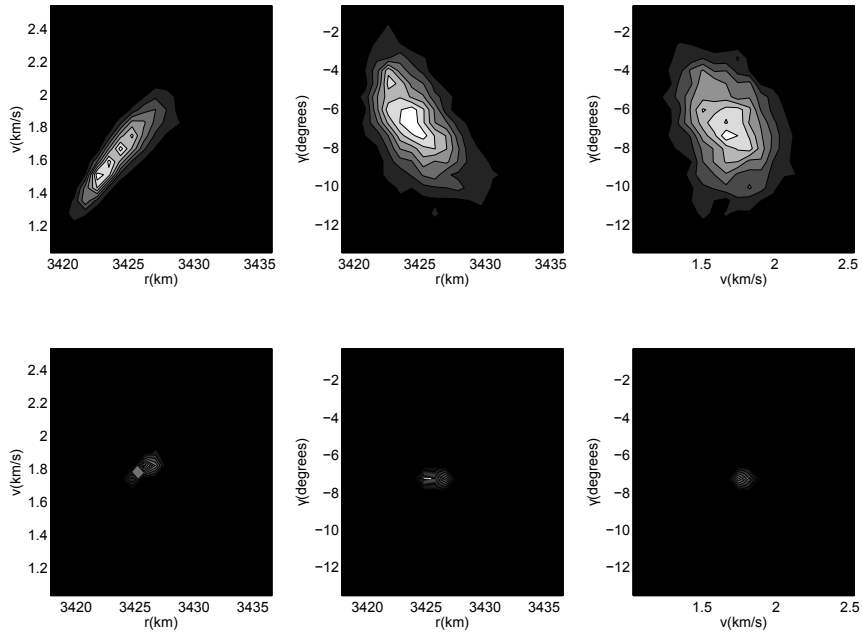
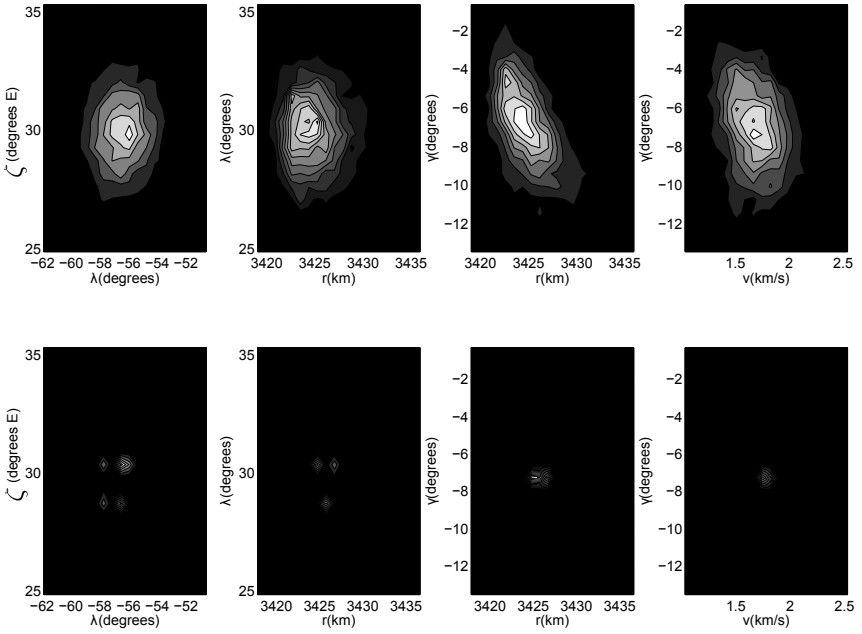


Figure 3.8: Comparison of final posterior univariate marginal PDFs for all states, obtained from KLMOC estimator (*solid line*) and particle filter (*hyphenated line*) with 3000 particles.



(a) Top row : particle filter, bottom row: KLMOC estimator.



(b) Top row : particle filter, bottom row: KLMOC estimator.

Figure 3.9: Plots for the final posterior bivariate marginal PDFs obtained from KLMOC estimator and particle filter with 3000 particles. The *darker (lighter)* regions represent *lower (higher)* PDF values.

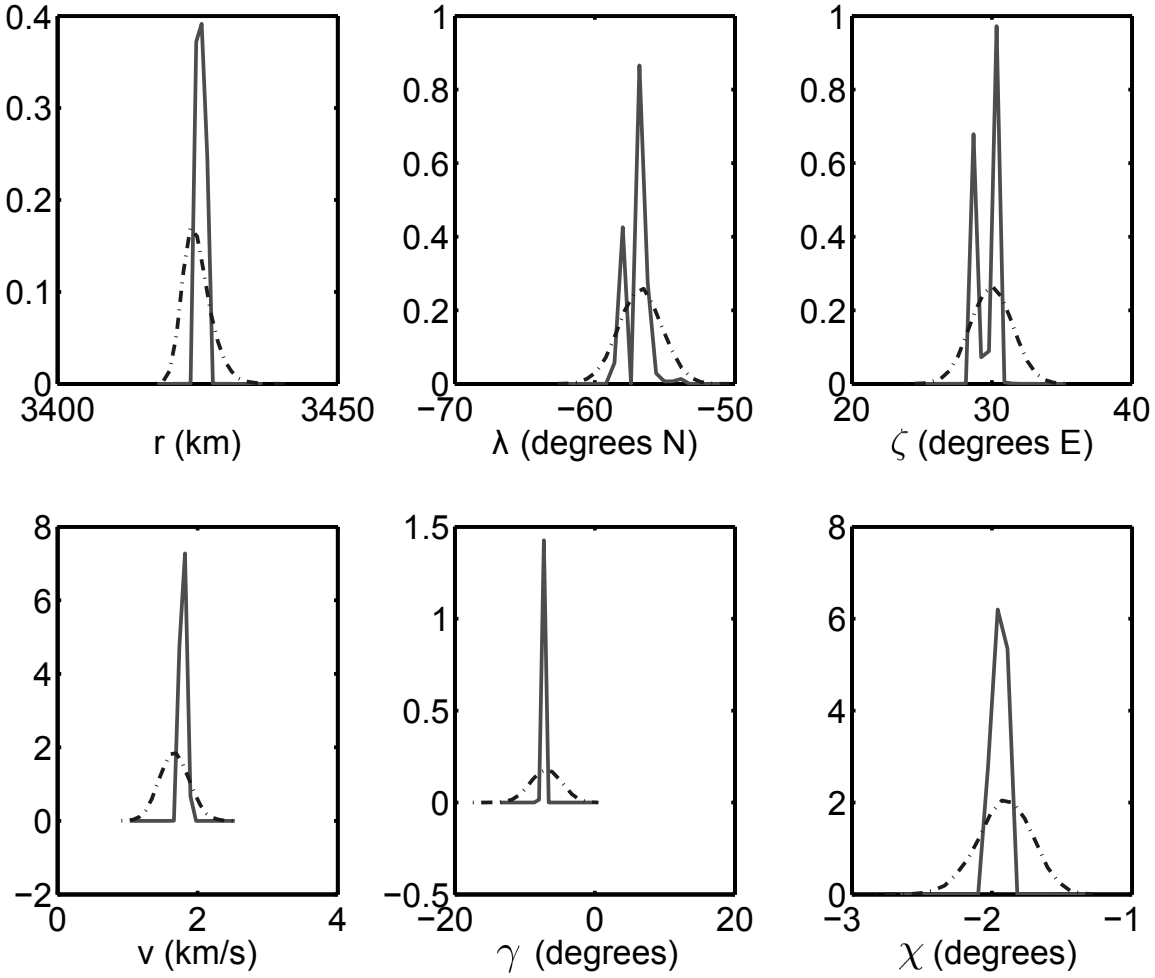
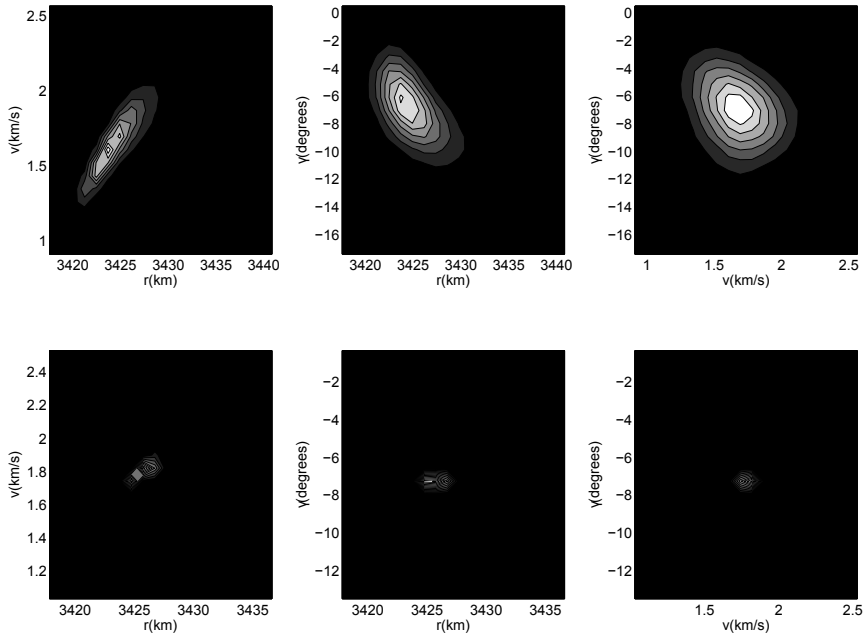
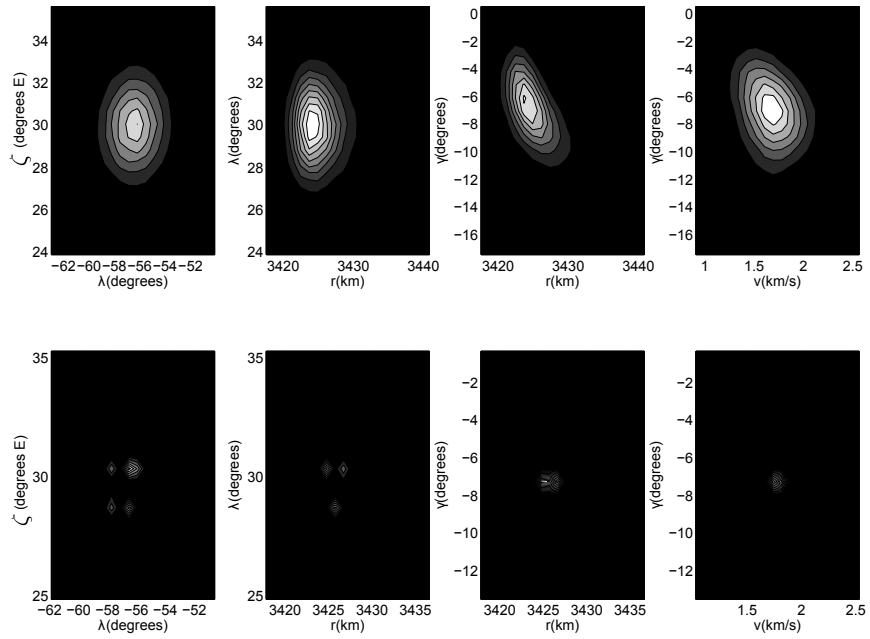


Figure 3.10: Final posterior univariate marginal PDF for all states obtained from KLMOC estimator (*solid line*) with 3000 particles, and particle filter (*hyphenated line*) with 50,000 particles.



(a) Top row: particle filter, bottom row: KLMOC estimator



(b) Top row: particle filter, bottom row: KLMOC estimator

Figure 3.11: Plots for the final posterior bivariate marginal PDFs obtained from KLMOC estimator with 3000 particles, and particle filter with 50,000 particles. The *darker* (*lighter*) regions represent *lower* (*higher*) PDF values.

3.4 Chapter Summary

In this Chapter, an algorithm is proposed for propagating probabilistic uncertainties subject to a stochastic flow. This situation arises when there are modeling uncertainties and/or stochastic forcing, in addition to the uncertainties present in initial conditions and parameters. Unlike the previous Chapter, here we develop an algorithm that approximates the time-varying joint PDF that satisfies the Fokker-Planck or forward Kolmogorov equation. We show that the proposed algorithm not only approximates the joint PDF, but also approximates the sample path of the stochastic flow in mean square sense. The proposed algorithm is a mixed parametric-nonparametric method that leverages the method of characteristics formulation of the previous Chapter. To show that the proposed approximation algorithm performs better than Monte carlo, several case studies in nonlinear estimation are carried out. This leads to a new nonlinear filtering algorithm, which is shown to have better estimation performance than the particle filtering techniques.

4. DISTRIBUTIONAL COMPARISON

The previous two chapters have introduced computational methods for uncertainty propagation in the form of joint PDF, subject to deterministic or stochastic dynamics in continuous time. Given the dynamical system, and associated initial condition, parametric and modeling uncertainties, these methods result time-varying probability weighted scattered data representing model predicted output PDF $\widehat{\eta}\left(\widehat{y},t\right)$, which need to compared with true output PDF $\eta\left(y,t\right)$, at times $\{t_j\}_{j=1}^{\tau}$, when the measurement PDF $\eta\left(y,t_j\right)$ is available. For this purpose, we need a metric to compare the *shapes* of these two PDFs at each time of measurement availability. We argue that the suitable metric in this context, is the *Wasserstein distance*.

4.1 Choice of Metric

Distances on the space of probability distributions [129], can be broadly categorized into two classes, viz. Csisźar's ϕ-divergence [130] and integral probability metrics [131]. The first includes well-known distances like Kullback-Leibler (KL) divergence, Hellinger distance, χ^2 divergence etc. while the latter includes Wasserstein distance, Dudley metric, maximum mean discrepancy. Total variation distance belongs to both of these classes.

4.1.1 Axiomatic Requirements for Model Validation

The choice of a suitable metric depends on application. Following the intuitions of Section 1.3, we list below the *axiomatic requirements*, that a *model validation metric* must satisfy.

R.1 The notion of "distance" must measure the *shape difference* between two instantaneous output PDFs. This is because a good model must emulate similar

concentration of trajectories as observed in the measurement space, i.e. the respective joint PDFs $\eta(y,t)$ and $\widehat{\eta}(\widehat{y},t)$, over the time-varying output supports, must match at times whenever measurements are available. In particular, the distance must be function of *shape difference but not of shape*, i.e. same amount of shape difference must return same magnitude of distance, irrespective of the individual shapes being compared.

R.2 For meaningful validation inference, the choice of distance must be a metric.

R.3 For a given model-data pair, the supports of $\eta(y,t)$ and $\widehat{\eta}(\widehat{y},t)$ may not match at $t = t_j,\ j = 1,\ldots,\tau$. The distance must be well defined and computable under such circumstances.

R.4 The computation of the distance need not require $\eta(y,t)$ and $\widehat{\eta}(\widehat{y},t)$ to be represented by the same number of samples. For the purpose of model validation, this offers practical advantages since experimental data are often expensive to gather. However, model based simulation can harness the computational resources and hence, simulation sample size is often larger than that of experimental data.

R.5 The distance must be asymptotically consistent with respect to finite sample representations of the PDFs under comparison. Namely, in the infinite sample limit, the empirical estimate of the distance must converge to the actual instantaneous value of the distance. For practical computation, this rate-of-convergence is required to be fast with respect to the number of samples.

Next, we introduce the Wasserstein distance on the manifold of PDFs, which will be shown to fulfil the axiomatic requirements listed above.

4.1.2 Wasserstein Metric

Definition 3. *(**Wasserstein distance**) Let the ℓ_p norm between two random output vectors $y \in \mathcal{Y} \subseteq \mathbb{R}^{n_o}$, and $\widehat{y} \in \widehat{\mathcal{Y}} \subseteq \mathbb{R}^{n_o}$, be denoted as $\| y - \widehat{y} \|_p$. Then, the Wasserstein distance of order q, between two PDFs $\eta(y)$ and $\widehat{\eta}(\widehat{y})$, is defined as*

$$ {}_pW_q(\eta, \widehat{\eta}) \triangleq \left[\inf_{\rho \in \mathcal{M}_2(\eta, \widehat{\eta})} \int_{\mathcal{Y} \times \widehat{\mathcal{Y}}} \|y - \widehat{y}\|_p^q \ \rho(y, \widehat{y}) \ dy d\widehat{y} \right]^{\frac{1}{q}}, \tag{4.1} $$

where $\mathcal{M}_2(\eta, \widehat{\eta})$ is the set of all joint PDFs supported on $\mathcal{Y} \times \widehat{\mathcal{Y}}$, having finite second moments, with first marginal as η and second marginal as $\widehat{\eta}$.

Remark 6. *(**Generalizations**) In general, the sets $\mathcal{Y}$ and $\widehat{\mathcal{Y}}$ can be subsets of any complete, separable metric (Polish) space, equipped with a p^{th} order distance metric. Further, (4.1) does not require the distributions under comparison to be absolutely continuous. It remains well defined between output measures μ and $\widehat{\mu}$, even when the corresponding PDFs η and $\widehat{\eta}$ don't exist.*

Remark 7. *(**Choice of** $p = q = 2$) We take Euclidean metric ($p = 2$) as the inter-sample distance between random vectors y and $\widehat{y}$. Further, we set $q = 2$ since it guarantees uniqueness [132] in (4.1), and has the interpretation of minimum effort needed to morph a density shape to other. Also, Jordan, Kinderlehrer and Otto [133] have rigorously demonstrated that uncertainty propagation in a dynamical system can be seen as a gradient flux of free energy with respect to the Wasserstein distance of order $q = 2$.*

The interpretation of ${}_2W_2$ as mass preserving optimal transport between two given shapes, makes it a strong candidate for model validation purpose. Further, it is known [134] that on the set $\mathcal{M}_2$, ${}_2W_2$ defines a metric. Thus, Wasserstein distance meets **R.1** and **R.2**. Also, **R.3** and **R.4** are satisfied since Definition 7.1 does not

require the supports or cardinality of the sample representations of the PDFs to be the same. This will be illustrated further in Section 4.2.4, when we describe the computation of ${}_2W_2$ between two scattered point clouds with probability weights. For **R.5**, convergence of sample Wasserstein estimate to its true deterministic value, will be discussed in Section 4.3.1 (Theorem 10).

4.1.3 Limitations of Pointwise Distances

Commonly used information-theoretic distances like Kullback-Leibler divergence $D_{KL}\left(\eta \parallel \widehat{\eta}\right) \triangleq \mathbb{E}[\log(\eta/\widehat{\eta})]$, its symmetrized version $D_{KL}^{\text{symm}} \triangleq D_{KL}\left(\eta \parallel \widehat{\eta}\right) + D_{KL}\left(\widehat{\eta} \parallel \eta\right)$, are not metrics. On the other hand, Hellinger distance $H\left(\eta, \widehat{\eta}\right) \triangleq \frac{1}{\sqrt{2}} \parallel \sqrt{\eta} - \sqrt{\widehat{\eta}} \parallel_{L_2(\mathbb{R}^{n_o})}$, and the square-root of Jensen-Shannon divergence $JSD\left(\eta, \widehat{\eta}\right) \triangleq \frac{1}{2}\left[D_{KL}\left(\eta \parallel \frac{1}{2}\left(\eta + \widehat{\eta}\right)\right) + D_{KL}\left(\widehat{\eta} \parallel \frac{1}{2}\left(\eta + \widehat{\eta}\right)\right)\right]$ are metrics. However, being pointwise definitions, all of them fail to satisfy **R.3** and **R.4**, resulting computational difficulties for model validation. As for **R.5**, $D_{KL}\left(\eta \parallel \widehat{\eta}\right)$ is known to be asymptotically consistent, but the rate-of-convergence can be arbitrarily slow [135, 136]. Besides these computational problems, we emphasize here that the information theoretic distances may not discriminate shapes in a geometric sense, as desired in **R.1**. We provide two counterexamples below to illustrate this point. The first counterexample highlights that two PDFs with same randomness need not have similar shapes. The second counterexample demonstrates that D_{KL} may depend on the shapes under comparison.

Counterexample 1. *(**Randomness** $\neq$ **shape**) Consider the two parametric family of beta densities* $\mathcal{B}\left(\alpha, \beta\right) \triangleq \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$, $\alpha, \beta > 0$, $x \in [0,1]$, *where* $B\left(\alpha, \beta\right) \triangleq \int_0^1 t^{\alpha-1}\left(1-t\right)^{\beta-1}\, dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, *is the complete beta function, and* $\Gamma\left(z\right)$ *denotes the*

gamma function. The differential entropy for beta family can be computed as [137]

$$\begin{aligned} & H_b(\alpha,\beta) = -\int_0^1 \mathcal{B}(\alpha,\beta)\log\mathcal{B}(\alpha,\beta)\ dx \\ = \ & \log B(\alpha,\beta) - (\alpha-1)\left(\Psi(\alpha)-\Psi(\alpha+\beta)\right) - (\beta-1)\left(\Psi(\beta)-\Psi(\alpha+\beta)\right) \end{aligned} \quad (4.2)$$

where $\Psi(z) \triangleq \frac{d}{dz}\log\Gamma(z)$*, is the digamma function. Since (4.2) remains invariant under* $(\alpha,\beta) \mapsto (\beta,\alpha)$*,* $\alpha \neq \beta$*, hence* $\mathcal{B}(\alpha,\beta)$ *and* $\mathcal{B}(\beta,\alpha)$ *have same entropy, but one is skewed to right and the other to left, as shown in Fig. 4.1. Fig. 4.2 shows the isentropic contours of beta PDFs in* (α,β) *space. Any pair of* ***distinct*** *points chosen on these contours, results two beta PDFs with non-identical shapes, as revealed by Fig. 4.3 and Appendix B, Section B.1.*

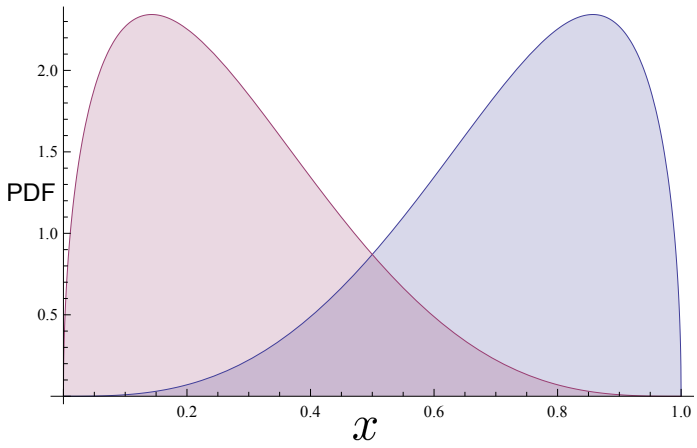


Figure 4.1: The two beta densities $\mathcal{B}\left(4,\frac{3}{2}\right)$ (left-skewed) and $\mathcal{B}\left(\frac{3}{2},4\right)$ (right-skewed) have same entropy or randomness, but have different shapes.

Counterexample 2. *(*$D_{KL} \neq$ ***shape difference****) Consider two* ν*-dimensional homoscedastic Gaussian PDFs* $\mathcal{N}(m_1,\Sigma_1)$ *and* $\mathcal{N}(m_2,\Sigma_2)$*, such that* $\Sigma_1 = \Sigma_2$*. Since*

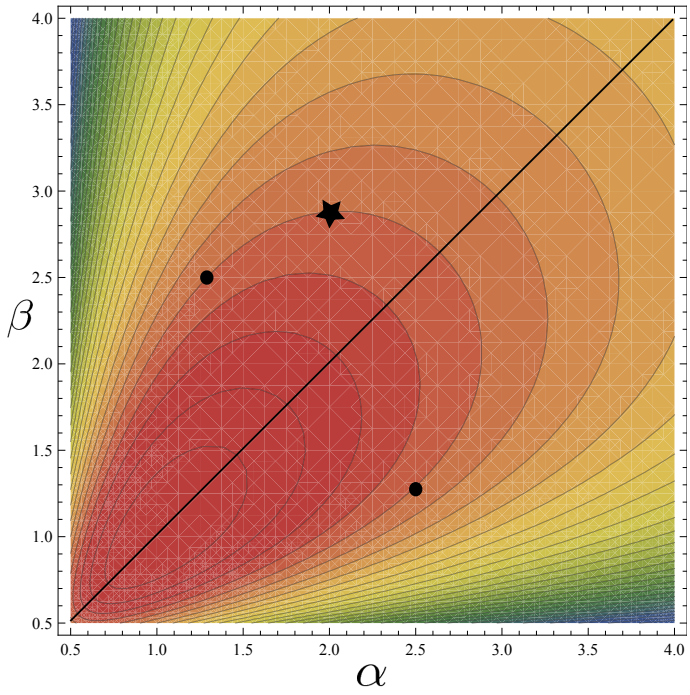


Figure 4.2: Isentropic contours of beta family in (α, β) space. The symmetry of the contours about $\alpha = \beta$ line implies $H_b(\alpha, \beta) = H_b(\beta, \alpha)$. The two solid circles correspond to the symmetric isentropic pair $\mathcal{B}(\alpha, \beta)$ and $\mathcal{B}(\beta, \alpha)$. The star in the figure corresponds to another isentropic beta PDF, that is parametrically asymmetric with the previous two. This plot also shows that uniform distribution $(\alpha = \beta = 1)$ is of maximum entropy.

the only ***difference*** *between the two PDFs is the location of their means, a shape-discriminating distance is expected to be a function of* $\| m_1 - m_2 \|_2$*, and should not depend on the covariance matrix, i.e.* ***shapes of the individual PDFs****.*

In this situation, ${}_2W_2 = \|m_1 - m_2\|_2$ *[138] and* $D_{KL} = \frac{1}{2}(m_2 - m_1)^\top \Sigma_2^{-1}(m_2 - m_1)$ *[139]. If we introduce* $m := m_2 - m_1$*, then* $\frac{D_{KL}}{{}_2W_2} = \frac{\|m\|_2}{2} r$*, where* $r := \frac{m^\top \Sigma_2^{-1} m}{m^\top m}$ *is the Rayleigh quotient corresponding to the positive semi-definite precision matrix* Σ_2^{-1}*. It's known (see e.g., Chapter 7 in [140]) that if we denote* $\mathcal{K} := \{\lambda : \lambda = \sum_{i=1}^{\nu} \alpha_i \lambda_i,\ \sum_{i=1}^{\nu} \alpha_i = 1,\ \alpha_i \geqslant 0,\ \forall i = 1, 2, \ldots, \nu\}$ *as the convex hull of the eigenvalues*

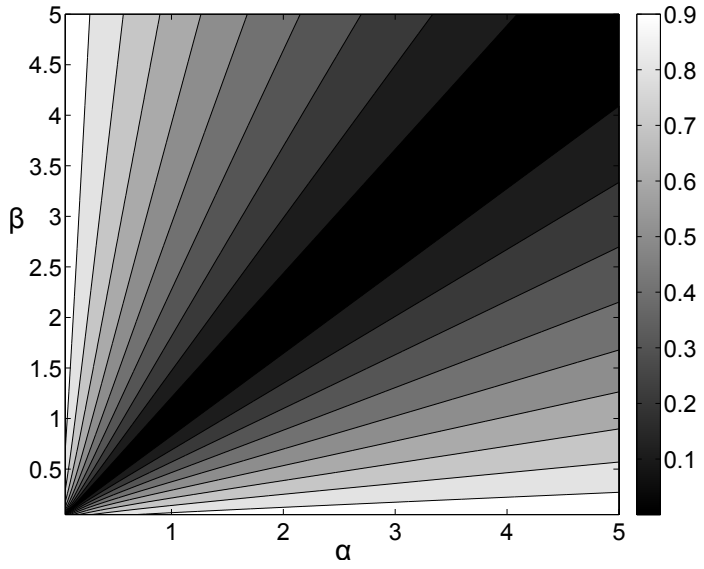


Figure 4.3: Iso-Wasserstein contours of ${}_2W_2\left(\mathcal{B}\left(\alpha,\beta\right),\mathcal{B}\left(\beta,\alpha\right)\right)$ in (α,β) space. Since ${}_2W_2$ is a metric, it has symmetry about $\alpha=\beta$ line, and vanishes only along this line.

of the precision matrix Σ_2^{-1}*, then* $r\left(m\right)\in\mathcal{K}$*. In particular,*

$$r_{\min}=\lambda_{\min}\left(\Sigma_2^{-1}\right)=\frac{1}{\lambda_{\min}\left(\Sigma_2\right)}>0,$$
$$r_{\max}=\lambda_{\max}\left(\Sigma_2^{-1}\right)=\frac{1}{\lambda_{\max}\left(\Sigma_2\right)}>0,$$

and these extrema are attained when $m:=m_2-m_1$ *respectively coincides with the minimum and maximum eigenvector of* Σ_2^{-1}*. Thus the spectrum of* Σ_2^{-1} *governs the magnitude of the ratio* $\frac{D_{KL}}{{}_2W_2}$*, even when* $\|m\|_2$ *is kept fixed. In particular, the ratio assumes unity iff* $r=\frac{2}{\|m\|_2}\Rightarrow\Sigma_2^{-1}=\frac{2}{\|m\|_2}\,I_\nu\Rightarrow\Sigma_1=\Sigma_2=\frac{\|m\|_2}{2}\,I_\nu$.

Further discussions on the inadequacy of D_{KL} *for capturing shape characteristics and the utility of Wasserstein distance for the same, can be found in [141, 142].*

4.2 Wasserstein Gap Between Dynamical Systems

4.2.1 Single Output Systems

Proposition 3. *[143] At time $t > 0$, let $F(y,t)$ and $\widehat{F}(\widehat{y},t)$ be the cumulative distribution functions (CDFs) corresponding to the univariate PDFs $\eta(y,t)$ and $\widehat{\eta}(\widehat{y},t)$, respectively. Then*

$$ {}_2W_2(t) \;=\; \sqrt{\int_0^1 \left(F^{-1}(\varsigma,t) - \widehat{F}^{-1}(\varsigma,t)\right)^2 d\varsigma}, \tag{4.3}$$

and the optimizer in (4.1) is the PDF $\rho^\star(y,\widehat{y},t)$ corresponding to the CDF $F^\star(y,\widehat{y},t) \triangleq \min\left(F(y,t), \widehat{F}(\widehat{y},t)\right)$.

4.2.2 Linear Gaussian Systems

Proposition 4. *Consider stable, observable LTI system pairs in continuous and discrete time:*

$$dx_i(t) = A_i x_i(t)dt + G_i d\mathcal{W}_i(t), \quad y_i(t) = C_i x_i(t), \tag{4.4}$$

$$x_i(k+1) = A_i x_i(k) + G_i \mathcal{W}_i(k), \quad y_i(k) = C_i x_i(k), \tag{4.5}$$

where $i = 1,2$. $\mathcal{W}_i(t)$ are Wiener processes with auto-covariances $Q_i(t_1 \wedge t_2)$, $t_1, t_2 > 0$, and $\mathcal{W}_i(k)$ are Gaussian white noises with covariances $Q_i(k)$. If the initial PDF $\xi_0 = \mathcal{N}(\mu_0, \Sigma_0)$, then the Wasserstein distance between output PDFs $\eta_i = \mathcal{N}(\mu_{y_i}, \Sigma_{y_i})$, is given by [138]

$$ {}_2W_2 = \sqrt{\| \mu_{y_1} - \mu_{y_2} \|_2^2 + \mathrm{tr}\left(\Sigma_{y_1} + \Sigma_{y_2} - 2\left[\sqrt{\Sigma_{y_1}}\Sigma_{y_2}\sqrt{\Sigma_{y_1}}\right]^{\frac{1}{2}}\right)}, \tag{4.6}$$

where $\mu_{y_i} = C_i\mu_{x_i}$, $\Sigma_{y_i} = C_i\Sigma_{x_i}C_i^\top$. *For the continuous-time case,*

$$\dot{\mu}_{x_i}(t) = A_i\mu_{x_i}(t), \tag{4.7}$$

$$\dot{\Sigma}_{x_i}(t) = A_i\Sigma_{x_i}(t) + \Sigma_{x_i}(t)A_i^\top + G_iQ_iG_i^\top, \tag{4.8}$$

and for the discrete-time case,

$$\mu_{x_i}(k+1) = A_i\mu_{x_i}(k), \tag{4.9}$$

$$\Sigma_{x_i}(k+1) = A_i\Sigma_{x_i}(k)A_i^\top + G_iQ_iG_i^\top, \tag{4.10}$$

to be solved with $\mu_{x_i}(0) = \mu_0$, *and* $\Sigma_{x_i}(0) = \Sigma_0$. *Deterministic results are recovered from above by setting the diffusion matrix* $G_i = 0$.

4.2.3 Asymptotic Wasserstein Distance

In Table 4.1, we have listed asymptotic Wasserstein distances between different pairs of stable dynamical systems. The asymptotic ${}_2W_2$ between two deterministic linear systems (**first row**) is zero since the origin being unique equilibria for both systems, Dirac delta is the stationary density for both. For a pair of deterministic affine systems (**second row**), asymptotic ${}_2W_2$ is simply the ℓ_2 norm between their respective fixed points. This holds true even for a pair of nonlinear systems, each having a **unique** globally asymptotically stable equilibrium. For the stochastic linear case (**third row**), $\Sigma_{y\infty} = C\Sigma_{x\infty}C^\top$, and $\widehat{\Sigma}_{\widehat{y}\infty} = \widehat{C}\widehat{\Sigma}_{\widehat{x}\infty}\widehat{C}^\top$; where $\Sigma_{x\infty}, \widehat{\Sigma}_{\widehat{x}\infty}$ respectively solve $A\Sigma_{x\infty} + \Sigma_{x\infty}A^\top + BQB^\top = 0$, and $\widehat{A}\widehat{\Sigma}_{\widehat{x}\infty} + \widehat{\Sigma}_{\widehat{x}\infty}\widehat{A}^\top + \widehat{B}\widehat{Q}\widehat{B}^\top = 0$. Q and $\widehat{Q}$ are process noise covariances associated with Wiener processes $\beta(t)$ and $\widehat{\beta}(t)$. For the **fourth** and **fifth row**, the set of stable equilibria for the true and model nonlinear system, are given by $\{y_i^\star\}_{i=1}^{n^\star}$ and $\{\widehat{y}_i^\star\}_{i=1}^{\widehat{n}^\star}$, respectively. Further, we assume that the nonlinear systems have no invariant sets other than these stable equilibria. In

such cases, the stationary densities are convex sum of Dirac delta densities, located at these equilibria. The weights for this convex sum, denoted as $m_i^\star$ and $\widehat{m}_i^\star$, depend on the initial PDF ξ_0. In particular, if we denote $\mathcal{R}_i$ as the **region-of-attraction** of the i^{th} equilibrium, then (see Appendix B, Section B.2)

$$m_i^\star = \int_{supp(\xi_0)\cap\mathcal{R}_i} \xi_0\left(x_0\right)\, dx_0 \;\in\; [0,1]\,. \tag{4.11}$$

To further illustrate this idea, a numerical example corresponding to the **fourth row** in Table 4.1, will be provided in Section 4.5.

4.2.4 Computing Transient Wasserstein Distance for Multivariate Output Systems

Computing Wasserstein distance from (4.1) calls for solving *Monge-Kantorovich optimal transportation plan* [144]. In this formulation, the difference in shape between two statistical distributions is quantified by the minimum amount of work required to convert a shape to the other. The ensuing optimization, often known as *Hitchcock-Koopmans problem* [145, 146, 147], can be cast as a linear program (LP), as described next.

Consider a complete, weighted, directed bipartite graph $K_{m,n}\left(U \cup V, E\right)$ with $\#\left(U\right) = m$ and $\#\left(V\right) = n$. If $u_i \in U, i = 1,\ldots,m$, and $v_j \in V, j = 1,\ldots,n$, then the edge weight $c_{ij} := \parallel u_i - v_j \parallel_{\ell_2}^2$ denotes the cost of transporting unit mass from vertex u_i to v_j. Then, according to (4.1), computing ${}_2W_2^2$ translates to

$$\text{minimize} \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\,\varphi_{ij} \tag{4.12}$$

subject to the constraints

$$\sum_{j=1}^{n} \varphi_{ij} = \alpha_i, \qquad \forall\; u_i \in U, \tag{C1}$$

Table 4.1: List of asymptotic Wasserstein distance, defined as ${}_2W_2\left(\eta_\infty, \widehat{\eta}_\infty\right)$, or various stable dynamical system pairs. Here η_∞ and $\widehat{\eta}_\infty$ are the stationary PDFs of the true and model dynamics, respectively.

Systems	Dynamics	Stationary PDFs	Asymptotic ${}_2W_2$
Deterministic linear pair	$\dot{x}(t) = Ax(t),\ y(t) = Cx(t),$ $\dot{\widehat{x}}(t) = \widehat{A}\widehat{x}(t),\ \widehat{y}(t) = \widehat{C}\widehat{x}(t)$	$\eta_\infty = \delta\left(y\right)$ $\widehat{\eta}_\infty = \delta\left(\widehat{y}\right)$	0
Deterministic affine pair	$\dot{x}(t) = Ax(t) + b,\ y(t) = Cx(t) + d,$ $\dot{\widehat{x}}(t) = \widehat{A}\widehat{x}(t) + \widehat{b},\ \widehat{y}(t) = \widehat{C}\widehat{x}(t) + \widehat{d}$	$\eta_\infty = \delta\left(y + CA^{-1}b - d\right)$ $\widehat{\eta}_\infty = \delta\left(\widehat{y} + \widehat{C}\widehat{A}^{-1}\widehat{b} - \widehat{d}\right)$	$\left\| \left(d - \widehat{d}\right) - \left(CA^{-1}b - \widehat{C}\widehat{A}^{-1}\widehat{b}\right) \right\|_2$
Stochastic linear pair	$dx(t) = Ax(t)dt + Gd\mathcal{W}(t),\ y(t) = Cx(t),$ $d\widehat{x}(t) = \widehat{A}\widehat{x}(t)dt + \widehat{G}d\widehat{\mathcal{W}}(t),\ \widehat{y}(t) = \widehat{C}\widehat{x}(t)$	$\eta_\infty = \mathcal{N}\left(0, \Sigma_{y\infty}\right)$ $\widehat{\eta}_\infty = \mathcal{N}\left(0, \widehat{\Sigma}_{\widehat{y}\infty}\right)$	$\left(\mathrm{tr}\left(\Sigma_{y\infty} + \widehat{\Sigma}_{\widehat{y}\infty} - 2\left[\Sigma_{y\infty}^{\frac{1}{2}}\widehat{\Sigma}_{\widehat{y}\infty}\Sigma_{y\infty}^{\frac{1}{2}}\right]^{\frac{1}{2}}\right)\right)^{\frac{1}{2}}$
Deterministic nonlinear and deterministic linear	$\dot{x}(t) = f\left(x(t)\right),\ y(t) = h\left(x(t)\right),$ $\dot{\widehat{x}}(t) = \widehat{A}\widehat{x}(t),\ \widehat{y}(t) = \widehat{C}\widehat{x}(t)$	$\eta_\infty = \sum_{i=1}^{n^\star} m_i^\star \delta\left(y - y_i^\star\right)$ $\widehat{\eta}_\infty = \delta\left(\widehat{y}\right)$	$\left(\sum_{i=1}^{n^\star} \left\|y_i^\star\right\|_2^2 m_i^\star \delta\left(y - y_i^\star\right)\right)^{\frac{1}{2}}$
Deterministic nonlinear pair	$\dot{x}(t) = f\left(x(t)\right),\ y(t) = h\left(x(t)\right),$ $\dot{\widehat{x}}(t) = \widehat{f}\left(\widehat{x}(t)\right),\ \widehat{y}(t) = \widehat{h}\left(\widehat{x}(t)\right)$	$\eta_\infty = \sum_{i=1}^{n^\star} m_i^\star \delta\left(y - y_i^\star\right)$ $\widehat{\eta}_\infty = \sum_{i=1}^{\widehat{n}^\star} \widehat{m}_i^\star \delta\left(\widehat{y} - \widehat{y}_i^\star\right)$	Monge-Kantorovich optimal transport LP (4.12), (C1)–(C3)

$$\sum_{i=1}^{m} \varphi_{ij} = \beta_j, \qquad \forall\ v_j \in V, \tag{C2}$$

$$\varphi_{ij} \geqslant 0, \qquad \forall\ (u_i, v_j) \in U \times V. \tag{C3}$$

The objective of (4.12) is to come up with an optimal mass transportation policy $\varphi_{ij} := \varphi(u_i \to v_j)$ associated with cost c_{ij}. Clearly, in addition to constraints (C1)–(C3), (4.12) must respect the necessary feasibility condition

$$\sum_{i=1}^{m} \alpha_i = \sum_{j=1}^{n} \beta_j \tag{C0}$$

denoting the conservation of mass. In our context of measuring the shape difference between two PDFs, we treat the joint probability mass function (PMF) vectors α_i and β_j to be the marginals of some unknown joint PMF φ_{ij} supported over the product space $U \times V$. Since determining joint PMF with given marginals is not unique, (4.12) strives to find that particular joint PMF which minimizes the total cost for transporting the probability mass while respecting the normality condition. Notice that the finite-dimensional LP (4.12) is a direct discretization of the Wasserstein definition (4.1), and it is known [148] that the solution of (4.12) is asymptotically consistent with that of the infinite dimensional LP (4.1).

4.3 Computational Complexity for ${}_2W_2$

4.3.1 Sample Complexity

For a desired accuracy of Wasserstein distance computation, we want to specify the bounds for number of samples $m = n$, for a given initial PDF. Since the finite sample estimate of Wasserstein distance is a random variable, we need to answer how large should n be, in order to guarantee that the empirical estimate of Wasserstein distance obtained by solving the LP (4.12), (C1)–(C3) with $m = n$, is close to the

true deterministic value of (4.1) in probability. In other words, given $\epsilon, \delta \in (0,1)$, we want to estimate a lower bound of $m = n$ as a function of ϵ and δ, such that

$$\mathbb{P}\left(\left|{}_2W_2\left(\eta_m^j(y), \widehat{\eta}_n^j(\widehat{y})\right) - {}_2W_2\left(\eta^j(y), \widehat{\eta}^j(\widehat{y})\right)\right| < \epsilon\right) > 1-\delta, \forall j = 1, \ldots, \tau. \quad (4.13)$$

Similar consistency and sample complexity results are available in the literature (see Corollary 9(i) and Corollary 12(i) in [149]) for Wasserstein distance of order $q = 1$. From Hölder's inequality, $W_{q_2} > W_{q_1}$ for $q_2 > q_1$, and hence that sample complexity bound, in general, does not hold for $q = 2$.To proceed, we need the following results.

Lemma 1. *Given random variables X, Y, Z, such that $X \leqslant Y + Z$, then for $\epsilon > 0$, we have*

$$\mathbb{P}(X > \epsilon) \leqslant \mathbb{P}(Y + Z > \epsilon) \leqslant \mathbb{P}\left(Y > \frac{\epsilon}{2}\right) + \mathbb{P}\left(Z > \frac{\epsilon}{2}\right). \quad (4.14)$$

Proof. Since $X(\omega_s) \leqslant Y(\omega_s) + Z(\omega_s)$, $\forall\, \omega_s \in \Omega_s$, hence we have $\{\omega_s : X(\omega_s) > \epsilon\} \subseteq \{\omega_s : Y(\omega_s) + Z(\omega_s) > \epsilon\} \subseteq \{\omega_s : Y(\omega_s) > \frac{\epsilon}{2}\} \cup \{\omega_s : Z(\omega_s) > \frac{\epsilon}{2}\}$, $\forall\, \omega_s \in \Omega_s$. Thus, we get $\mathbb{P}(X > \epsilon) \leqslant \mathbb{P}\left(\{Y > \frac{\epsilon}{2}\} \cup \{Z > \frac{\epsilon}{2}\}\right) \leqslant \mathbb{P}\left(Y > \frac{\epsilon}{2}\right) + \mathbb{P}\left(Z > \frac{\epsilon}{2}\right)$, from Boole-Bonferroni inequality (see e.g., Appendix C in [150]). ■

Definition 4. *(**Transportation cost inequality**)[151] A probability measure μ is said to satisfy the L_p-*transportation cost inequality *(TCI) of order q, if there exists some constant $C > 0$ such that for any probability measure ν, ${}_pW_q(\mu, \nu) \leqslant \sqrt{2CD_{KL}(\nu \parallel \mu)}$. In short, we write $\mu \in T_q(C)$. In particular, for $\mu \sim \mathcal{N}(m_{\kappa\times 1}, \Sigma_{\kappa\times\kappa})$, we have [152] $\mu \in T_2(\lambda_{max}(\Sigma))$.*

Theorem 9. *(**Rate-of-convergence of empirical measure in Wasserstein metric**)(Theorem 5.3 in [153]) For a probability measure $\rho \in T_q(\mathscr{C})$, $1 \leqslant q \leqslant 2$,*

and its n-sample estimate ρ_n, we have

$$\mathbb{P}\left({}_pW_q\left(\rho,\rho_n\right)>\theta\right)\leqslant K_\theta\,\exp\left(-\frac{n\theta^2}{8\mathscr{C}}\right),\quad \theta>0, \tag{4.15}$$

and $\log K_\theta := \dfrac{1}{\mathscr{C}}\inf_{\mu}\#\left(\operatorname{supp}\mu\right)\left(\operatorname{diam}\left(\operatorname{supp}\mu\right)\right)^2$. *The optimization takes place over all probability measures μ of finite support, such that ${}_pW_q\left(\rho,\mu\right)\leqslant\theta/4$.*

We now make few notational simplifications. In this subsection, we denote $\eta^j\left(y\right)$ and $\widehat{\eta}^j\left(y\right)$ by η and $\widehat{\eta}$, and their finite sample representations by η_m and $\widehat{\eta}_n$, respectively. Then we have the following result.

Theorem 10. ***(Rate-of-convergence of empirical Wasserstein estimate)*** *For true densities η and $\widehat{\eta}$, let corresponding empirical densities be η_m and $\widehat{\eta}_n$, evaluated at respective uniform sampling of cardinality m and n. Let $\mathscr{C}_1$, $\mathscr{C}_2$, be the TCI constants for η and $\widehat{\eta}$, respectively and fix $\epsilon>0$. Then*

$$\mathbb{P}\left(\left|\,{}_2W_2\left(\eta_m,\widehat{\eta}_n\right)\;-\;{}_2W_2\left(\eta,\widehat{\eta}\right)\,\right|>\epsilon\right)\leqslant K_1\,\exp\left(-\frac{m\epsilon^2}{32\mathscr{C}_1}\right)+K_2\,\exp\left(-\frac{n\epsilon^2}{32\mathscr{C}_2}\right) \tag{4.16}$$

Proof. Since Wasserstein distance is a metric, from triangle inequality

$$\begin{aligned}{}_2W_2\left(\eta_m,\widehat{\eta}_n\right)\leqslant\;{}_2W_2\left(\eta_m,\eta\right)+\;{}_2W_2\left(\widehat{\eta}_n,\eta\right)\leqslant\;{}_2W_2\left(\eta_m,\eta\right)\;+\;{}_2W_2\left(\widehat{\eta}_n,\widehat{\eta}\right)\;+\;{}_2W_2\left(\eta,\widehat{\eta}\right)\\ \Rightarrow\;{}_2W_2\left(\eta_m,\widehat{\eta}_n\right)\;-\;{}_2W_2\left(\eta,\widehat{\eta}\right)\;\leqslant\;{}_2W_2\left(\eta_m,\eta\right)\;+\;{}_2W_2\left(\widehat{\eta}_n,\widehat{\eta}\right).\end{aligned}$$

Combining the above with Lemma 1, we have

$$\begin{aligned}\mathbb{P}\left(\left|\,{}_2W_2\left(\eta_m,\widehat{\eta}_n\right)-{}_2\,W_2\left(\eta,\widehat{\eta}\right)\,\right|>\epsilon\right)\;&\leqslant\;\mathbb{P}\left({}_2W_2\left(\eta_m,\eta\right)>\frac{\epsilon}{2}\right)\\ &+\;\mathbb{P}\left({}_2W_2\left(\widehat{\eta}_n,\widehat{\eta}\right)>\frac{\epsilon}{2}\right),\end{aligned} \tag{4.17}$$

where each term in the right-hand-side of (4.17) can be separately upper-bounded using Theorem 1 with $\theta \mapsto \frac{\epsilon}{2}$. Hence the result. ■

Remark 8. *At a fixed time, K_1, K_2, $\mathscr{C}_1$ and $\mathscr{C}_2$ are constants in a given model validation problem, i.e. for a given pair of experimental data and proposed model. However, values of these constants depend on true and model dynamics. In particular, the TCI constants $\mathscr{C}_1$ and $\mathscr{C}_2$ depend on the dynamics via respective PDF evolution operators. The constants K_1 and K_2 depend on η and $\widehat{\eta}$, which in turn depend on the dynamics.*

4.3.2 Runtime Complexity

The LP formulation (4.12), (C1)–(C3), requires solving for mn unknowns subject to $(m + n + mn)$ constraints. For $m = n$, it can be shown that [154, 155] the runtime complexity for solving the LP is $\mathcal{O}\left(n_o\, n^{2.5} \log \nu\right)$. Notice that the output dimension n_o enters only through the cost c_{ij} in (4.12) and hence affects the computational time linearly.

In actual simulations, we found the runtime of the LP (4.12) to be sensitive on how the constraints were implemented. Suppose, we put (4.12) in standard form

$$\text{minimize } \widetilde{c}^{\top}\widetilde{\varphi}, \qquad \text{subject to } A\widetilde{\varphi} = b, \quad \widetilde{\varphi} \geqslant 0, \tag{4.18}$$

where $\widetilde{c}_{mn\times 1} \triangleq \text{vec}\,(c)$, $\widetilde{\varphi}_{mn\times 1} \triangleq \text{vec}\,(\varphi)$, $b_{(m+n)\times 1} := [\alpha_{m\times 1}, \beta_{n\times 1}]^{\top}$. If we let $e_n := \underbrace{[1, 1, \ldots, 1]}_{n\,\text{times}}^{\top}$, then the implementation $A_{(m+n)\times mn} = \begin{bmatrix} e_n^{\top} \otimes I_m \\ I_n \otimes e_m^{\top} \end{bmatrix}$ was found to achieve fast offline construction of the constraint matrix.

4.3.3 Storage Complexity

For $m = n$, the constraint matrix A in (4.18), is a binary matrix of size $2n \times n^2$, whose each row has n ones. Consequently, there are total $2n^2$ ones in the constraint matrix and the remaining $2n^2\,(n-1)$ elements are zero. Hence at any fixed time, the sparse representation of the constraint matrix needs # non-zero elements $\times 3 = 6n^2$ storage. The PMF vectors are, in general, fully populated. In addition, we need to store the model and true sample coordinates, each of them being a n_o-tuple. Hence at any fixed time, constructing cost matrix requires storing $2n_o n$ values. Thus total storage complexity at any given snapshot, is $2n\,(3n + n_o + 1) = \mathcal{O}\,(n^2)$, assuming $n > n_o$. However, if the sparsity of constraint matrix is not exploited by the solver, then storage complexity rises to $2n\,(n^2 + n_o + 1) = \mathcal{O}\,(n^3)$. For example, if we take $n = 1000$ samples and use double precision arithmetic, then solving the LP at each time requires either megabytes or gigabytes of storage, depending on whether or not sparse representation is utilized by the solver[1]. For $m \neq n$, it is easy to verify that the sparse storage complexity is $(6mn + (m+n)\,n_o + m + n)$, and the non-sparse storage complexity is $(m+n)\,(mn + n_0 + 1)$.

4.4 Construction of Validation Certificates

4.4.1 Probabilistically Robust Model Validation

Often in practice, the exact initial density is not known to facilitate our model validation framework; instead a class of densities may be known. For example, it may be known that the initial density is symmetric unimodal but its exact shape (e.g. normal, semi-circular etc.) may not be known. Even when the distribution-type is known (e.g. normal), it is often difficult to pinpoint the parameter values describing the initial density function. To account such scenarios, consider a random

[1]We used MOSEK (available at `www.mosek.com`) as the LP solver.

variable $\Delta : \Omega \to E$, that induces a probability triplet $(\Omega, \mathcal{F}, \mathbb{P})$ on the space of initial densities. Here $E \subset \Omega$ and $\#(E) = 1$. The random variable Δ picks up an initial density from the collection of admissible initial densities $\Omega := \{\xi_0^{(1)}(\widetilde{x}), \xi_0^{(2)}(\widetilde{x}), \ldots\}$ according to the law of Δ. For example, if we know $\xi_0 \sim \mathcal{N}(\mu_0, \sigma_0^2)$ with a given joint distribution over the (μ_0, σ_0^2) space, then in our model validation framework, one sample from this space will return one distance measure between the instantaneous output PDFs. How many such (μ_0, σ_0^2) samples are necessary to guarantee the robustness of the model validation oracle? The Chernoff bound provides such an estimate for finite sample complexity.

At time step t_k, let the *validation probability* be $p(\gamma_k) \triangleq \mathbb{P}({}_2W_2(\eta_k(y), \widehat{\eta}_k(\widehat{y})) \leqslant \gamma_k)$. Here $\gamma_k \in \mathbb{R}^+$ is the prescribed instantaneous tolerance level. If the model validation is performed by drawing N samples from Ω, then the *empirical validation probability* is $\widehat{p}_N(\gamma_k) \triangleq \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \mathbb{1}_{V_k^{(i)}}$ where $V_k^{(i)} \triangleq \{\widehat{\eta}_k^{(i)}(\widehat{y}) : {}_2W_2\left(\eta_k^{(i)}(y), \widehat{\eta}_k^{(i)}(\widehat{y})\right) \leqslant \gamma_k\}$. Consider $\epsilon, \delta \in (0,1)$ as the desired accuracy and confidence, respectively.

Lemma 2. *(**Chernoff bound**)[156] For any $\epsilon, \delta \in (0,1)$, if $N \geqslant N_{\text{ch}} := \dfrac{1}{2\epsilon^2}\log\dfrac{2}{\delta}$, then $\mathbb{P}(|p(\gamma_k) - \widehat{p}_N(\gamma_k)| < \epsilon) > 1 - \delta$.*

The above lemma allows us to construct *probabilistically robust validation certificate* (PRVC) $\widehat{p}_N(\gamma_k)$ through the Algorithm 3 stated next. The PRVC vector, with ϵ accuracy, returns the probability that the model is valid at time t_k, in the sense that the instantaneous output PDFs are no distant than the required tolerance level γ_k. Lemma 2 lets the user control the accuracy ϵ and the confidence δ, with which the preceding statement can be made. Thus the framework enables us to compute a provably correct validation certificate on the face of uncertainty with finite sample complexity.

Algorithm 3 Construct PRVC

Require: $\epsilon, \delta \in (0,1)$, T, ν, law of Δ, experimental data $\{\eta_k(y)\}_{k=1}^{\tau}$, model, tolerance vector $\{\gamma_k\}_{k=1}^{\tau}$

$N \leftarrow N_{\text{ch}}(\epsilon, \delta)$ ▷ Using lemma 2
Draw random functions $\xi_0^{(1)}(\widetilde{x}), \xi_0^{(2)}(\widetilde{x}), \ldots, \xi_0^{(N)}(\widetilde{x})$ according to the law of Δ
for $k = 1$ to τ **do** ▷ Index for time step
  for $i = 1$ to N **do** ▷ Index for initial density
    for $j = 1$ to ν **do** ▷ Samples drawn from $\xi_0^{(i)}(\widetilde{x})$
      Propagate states using dynamics
      Propagate measurements
    end for
    Propagate $\widehat{\xi}_k^{(i)}\left(\widehat{\widetilde{x}}\right)$ ▷ Use (2.9) or (3.38)
    Compute $\widehat{\eta}_k^{(i)}(\widehat{y})$
    Compute ${}_2W_2\left(\eta_k^{(i)}(y), \widehat{\eta}_k^{(i)}(\widehat{y})\right)$ ▷ Distributional comparison by solving LP (4.12) subject to (C0)–(C3)
    sum $\leftarrow 0$ ▷ Initialize
    if ${}_2W_2\left(\eta_k^{(i)}(y), \widehat{\eta}_k^{(i)}(\widehat{y})\right) \leqslant \gamma_k$ **then**
      sum $\leftarrow$ sum $+ 1$
    end if
  end for
  $\widehat{p}_N(\gamma_k) \leftarrow \dfrac{\text{sum}}{N}$ ▷ Construct PRVC vector
end for

4.4.2 Probabilistically Worst Case Model Validation

Following [157, 158, 159], one can also define a probabilistic notion of the worst-case model validation performance as $\gamma_k^{\text{wc}} := \sup_{\Delta} {}_2W_2(\eta_k(y), \widehat{\eta}_k(\widehat{y}))$, and its empirical estimate $\widehat{\gamma}_k^N := \max_{i=1,\ldots,N} {}_2W_2\left(\eta_k^{(i)}(y), \widehat{\eta}_k^{(i)}(\widehat{y})\right)$. The sample complexity for probabilistically worst-case model validation is given by the lemma below.

Lemma 3. *(**Worst-case bound**) (p. 128 in [156]) For any $\epsilon, \delta \in (0,1)$, if $N \geqslant N_{wc} := \dfrac{\log \frac{1}{\delta}}{\log \frac{1}{1-\epsilon}}$, then $\mathbb{P}\left(\mathbb{P}\left({}_2W_2(\eta_k(y), \widehat{\eta}_k(\widehat{y})) \leqslant \widehat{\gamma}_k^N\right) \geqslant 1-\epsilon\right) > 1-\delta$.*

Notice that in general, there is no guarantee that the empirical estimate $\widehat{\gamma}_k^N$

Algorithm 4 Construct PWVC

Require: $\epsilon, \delta \in (0,1)$, τ, ν, law of Δ, experimental data $\{\eta_k(y)\}_{k=1}^{\tau}$, model

$N \leftarrow N_{\text{wc}}(\epsilon, \delta)$ ▷ Using lemma 3

Draw N random functions $\xi_0^{(1)}(\widetilde{x}), \xi_0^{(2)}(\widetilde{x}), \ldots, \xi_0^{(N)}(\widetilde{x})$ according to the law of Δ ▷ Use MCMC

for $k = 1$ to τ **do** ▷ Index for time step

  for $i = 1$ to N **do** ▷ Index for initial density

    for $j = 1$ to ν **do** ▷ Index for samples in the extended state space, drawn from $\xi_0^{(i)}(\widetilde{x})$

      Propagate states using dynamics

      Propagate measurements

    end for

    Propagate $\widehat{\xi}_k^{(i)}\left(\widehat{\widetilde{x}}\right)$ ▷ Use (2.9) or (3.38)

    Compute $\widehat{\eta}_k^{(i)}(\widehat{y})$ ▷ Algebraic transformation

    Compute ${}_2W_2\left(\eta_k^{(i)}(y), \widehat{\eta}_k^{(i)}(\widehat{y})\right)$ ▷ Distributional comparison by solving LP (4.12) subject to (C0)–(C3)

    $\widehat{\gamma}_k^N \leftarrow \max_{i=1,\ldots,N} {}_2W_2\left(\eta_k^{(i)}(y), \widehat{\eta}_k^{(i)}(\widehat{y})\right)$ ▷ Empirically estimate worst-case performance

  end for

end for

is close to the true worst-case performance γ_k^{wc}. Also, the performance bound is obtained *a posteriori* while the robust validation framework accounted for *a priori* tolerance levels. The corresponding *probabilistically worst-case validation certificate* (PWVC) $\widehat{\gamma}_k^N$ can be computed from Algorithm 4, given below. In summary, the algorithm, with high probability $(1-\epsilon)$, only ensures that the output PDFs are at most $\widehat{\gamma}_k^N$ far. The preceding statement can be made with probability at least $1-\delta$.

4.5 Examples

4.5.1 Validating Deterministic Flow Model

Consider the following nonlinear dynamical system

$$\ddot{x} = -ax - b\sin 2x - c\dot{x}, \quad a = 0.1, b = 0.5, c = 1. \tag{4.19}$$

The system has five fixed points $P_0 = (0,0)$, $P_1^{\pm} = (\pm 1.7495, 0)$, $P_2^{\pm} = (\pm 2.8396, 0)$, which can be solved by noting the abscissa values of the points of intersection of two curves $f(x) = b\sin 2x$ and $g(x) = -ax$, as shown in Fig 4.4. From linear analysis, it is easy to verify that P_0 and $P_2^{\pm}$ are stable foci while $P_1^{\pm}$ are saddles (Fig. 4.5).

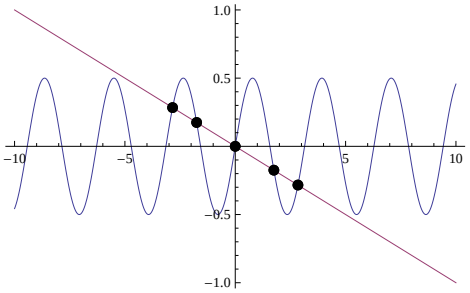


Figure 4.4: Points of intersection of the curve $f(x) = b\sin 2x$ and the line $g(x) = -ax$.

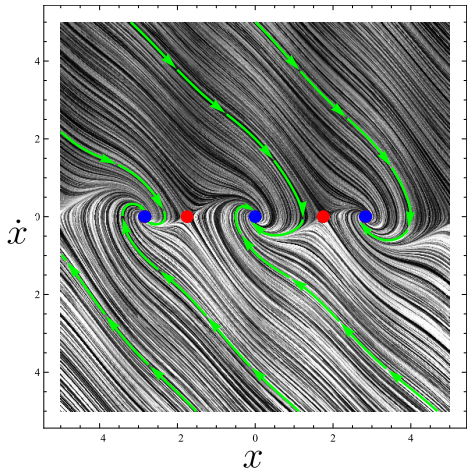


Figure 4.5: Phase portrait of the vector field (4.19) with three stable and two saddle fixed points.

To illustrate our model validation framework, let's assume that 'true data' is generated by the dynamics (4.19). However, this true dynamics is unknown to the

modeler, whose proposed model is a linearization of (4.19) about the origin. We emphasize here that the purpose of (4.19) is only to create the synthetic data and to demonstrate the proof-of-concept. In a realistic model validation, the data arrives from experimental measurements, not from another model. For simplicity, we take the outputs same as states for both true and model dynamics.

Starting from the bivariate uniform distribution $\mathcal{U}([-\pi,\pi]\times[-\pi,\pi]) =: \xi_0$, we evolve the respective joint PDFs $\xi = \eta$ and $\widehat{\xi} = \widehat{\eta}$, through true and model dynamics via MOC implementation of Liouville equation discussed in Chapter 2. The distributional shape discrepancy is captured via the Wasserstein gap $({}_2W_2(\eta,\widehat{\eta}))$ between these instantaneous joint PDFs, as shown in Fig. 4.6 (*solid line*), computed by solving the LP (4.12), (C1)–(C3). As the individual joint PDFs converge toward their respective stationary densities, the slope of the Wasserstein time-history decreases progressively. Fig. 4.7 shows the Wasserstein gap trajectories when ξ_0 is taken to be $\mathcal{N}(0,\sigma_0^2 I_2)$, instead of uniform. In this case, we observe that larger initial dispersion causes larger Wasserstein gap. Suppose the user-specified tolerance level $\{\gamma_j\}_{j=1}^{40}$ is 0.8 for first 10 instances and 0.6 for next 30 instances of measurement availability, as shown by the shaded area in Fig. 4.7. Given the set of admissible initial densities $\{\xi_0^{(1)},\ldots,\xi_0^{(9)}\}$ with $\xi_0^{(i)} := \mathcal{N}(0,\sigma_{0i}^2 I_2)$, $i = 1,\ldots,9$, we can compute the PRVC vector, shown as the dashed line in Fig. 4.7, to be

$$\left[\underbrace{1,\ldots,1}_{3\text{ times}},0.89,\underbrace{0.78,\ldots,0.78}_{5\text{ times}},0.67,\underbrace{0.56,\ldots,0.56}_{30\text{ times}}\right]^\top.$$

4.5.2 Validating Stochastic Flow Model

Here we assume the true data to be generated by (4.19) with additive white noise having autocorrelation $Q\delta(t_1-t_2)$, $t_1,t_2 \geqslant 0$. Letting $x_1 = x$ and $x_2 = \dot{x}$, the

associated Itô SDE can be written in state-space form similar to (3.1)

$$\begin{Bmatrix} dx_1 \\ dx_2 \end{Bmatrix} = \begin{Bmatrix} x_2 \\ -ax_1 - b\sin 2x_1 - cx_2 \end{Bmatrix} dt + \begin{Bmatrix} 0 \\ 1 \end{Bmatrix} d\mathcal{W}, \tag{4.20}$$

where $\mathcal{W}(t)$ is a Wiener process with autocorrelation $Q(t_1 \wedge t_2)$. The stationary Fokker-Planck equation for (4.20) can be solved in closed form (Appendix B, Section B.3)

$$\eta_\infty(x_1, x_2) \propto \exp\left(-\frac{c}{2Q}\left(ax_1^2 + x_2^2 - b\cos 2x_1\right)\right), \tag{4.21}$$

and one can verify that peaks of (4.21) appear at the fixed points of the nonlinear drift.

Let the proposed model be the linearization of (4.20) about the origin. It is well-known [93] that the stationary density of a linear SDE of the form $d\widehat{x} = A\widehat{x}dt + GdW$, is given by

$$\widehat{\eta}_\infty(\widehat{x}) = \mathcal{N}(\mathbf{0}, \Sigma_\infty) = \frac{\exp\left(-\frac{1}{2}\ \widehat{x}^\top \Sigma_\infty^{-1} \widehat{x}\right)}{\sqrt{(2\pi)^2 \det(\Sigma_\infty)}}, \tag{4.22}$$

provided A is Hurwitz and (A, g) is a controllable pair. The steady-state covariance matrix Σ_∞ solves $A\Sigma_\infty + \Sigma_\infty A^\top + GQG^\top = 0$. For the linearized version of (4.20), $A = \begin{bmatrix} 0 & 1 \\ (-a-2b) & -c \end{bmatrix}$ and $G = \begin{Bmatrix} 0 \\ 1 \end{Bmatrix}$ satisfy the aforementioned conditions and the stationary density is obtained from (4.22).

Taking the initial density same as in the deterministic case, we propagated the joint PDFs for (4.20) and the linear SDE using the KLMOC method described in Chapter 3. The *dashed line* in Fig. 4.6 shows the Wasserstein trajectory for this case.

The dash-dotted line in Fig. 4.6 shows the asymptotic Wasserstein gap between the respective stationary densities (4.21) and (4.22). Due to randomized sampling, all stochastic computations are in probabilistically approximate sense [160].

4.5.3 Comparison with Prajna's Barrier Certificate Based Model Falsification

Consider the nonlinear model validation problem stated as Example 4 in [7], where the model is $\dot{x} = -px^3$, with parameter $p \in \mathcal{P} = [0.5, 2]$. The measurement data are interval-valued sets $\mathcal{X}_0 = [0.85, 0.95]$ at $t = 0$, and $\mathcal{X}_T = [0.55, 0.65]$ at $t = T = 4$. A barrier certificate of the form $B(x,t) = B_1(x) + tB_2(x)$ was found in [7] through sum of squares (SOS) optimization [161] where $B_1(x) = 8.35x + 10.40x^2 - 21.50x^3 + 9.86x^4$, and $B_2(x) = -1.78 + 6.58x - 4.12x^2 - 1.19x^3 + 1.54x^4$. The model was thereby invalidated by the existence of such certificate, i.e. the model $\dot{x} = -px^3$, with parameter $p \in \mathcal{P}$ was shown to be inconsistent with measurements $\{\mathcal{X}_0, \mathcal{X}_T, T\}$.

To tackle this problem in our model validation framework, consider the spatio-temporal evolution of the joint PDF $\xi(x,p,t)$ over the extended state space $\widetilde{x} = [x \;\; p]^\top$, with initial support $\widetilde{\mathcal{X}}_0 := \mathcal{X}_0 \times \mathcal{P}$, under the action of the extended vector field $\widetilde{f}(x,p) = [-px^3 \quad 0]^\top$. Our objective then, is to prove that for $T = 4$, the PDF $\xi_T(x_T, p, T) = \mathcal{U}(x_T, p) = 1/\mathrm{vol}\left(\widetilde{\mathcal{X}}_T\right)$ is not finite-time reachable from $\xi_0(x_0, p) = \mathcal{U}(x_0, p) = 1/\mathrm{vol}\left(\widetilde{\mathcal{X}}_0\right)$, subject to the proposed model dynamics on the extended state space.

Theorem 11. *The two-point boundary value problem*

$$\begin{aligned}
\frac{\partial \xi}{\partial t} + \nabla_{\widetilde{x}} \cdot \left(\widetilde{f}(x,p)\,\xi\right) = \frac{\partial \xi}{\partial t} + \nabla_x \cdot \left(-px^3\xi\right) = 0,\\
\xi(x(0),p,0) = \xi_0(x_0,p) = \mathcal{U}(x_0,p) = 1/\mathrm{vol}\left(\widetilde{\mathcal{X}}_0\right),\\
\xi(x(T),p,T) = \xi_T(x_T,p,T) = \mathcal{U}(x_T,p) = 1/\mathrm{vol}\left(\widetilde{\mathcal{X}}_T\right),
\end{aligned}$$

has no solution for $\xi(x,p,t)$*, such that* $\int_{\widetilde{\mathcal{X}}(t)} \xi(x,p,t)\, dx\, dp = 1$, $\forall t \in (0,T)$.

Proof. MOC ODE (2.9) corresponding to the Liouville PDE $\frac{\partial \xi}{\partial t} + \nabla_{\widetilde{x}} \cdot \left(\widetilde{f}(x,p)\,\xi\right) = 0$, yields a solution of the form

$$\xi(x,p,t) = \xi_0(x_0,p) \exp\left(-\int_0^t \nabla_{\widetilde{x}} \cdot \left(\widetilde{f}(x(\tau),p)\right) d\tau\right). \tag{4.23}$$

For the model dynamics $\dot{x} = -px^3$, we have $\nabla_{\widetilde{x}} \cdot \left(\widetilde{f}(x(\tau),p)\right) = -3p(x(\tau))^2$ and $\frac{1}{x^2} = \frac{1}{x_0^2} + 2pt$. Consequently (4.23) results

$$\begin{aligned} \xi(x,p,t) &= \xi_0(x_0,p)\left(1+2x_0^2pt\right)^{3/2} \\ &= \frac{1}{(1-2x^2pt)^{3/2}}\,\xi_0\left(\pm\frac{x}{\sqrt{1-2x^2pt}},p\right). \end{aligned} \tag{4.24}$$

In particular, for $\xi_0(x_0,p) = 1/\mathrm{vol}\left(\widetilde{\mathcal{X}}_0\right)$, $\xi_T(x_T,p,T) = 1/\mathrm{vol}\left(\widetilde{\mathcal{X}}_T\right)$, and $T = 4$, (4.24) requires us to satisfy

$$\left(1-8x_T^2p\right) = \left(\frac{\mathrm{vol}\left(\widetilde{X}_T\right)}{\mathrm{vol}\left(\widetilde{X}_0\right)}\right)^{2/3} > 0 \Rightarrow 1 > 8x_T^2p. \tag{4.25}$$

Since $8x_T^2p$ is an increasing function in both $x_T \in \mathcal{X}_T$ and $p \in \mathcal{P}$, we need at least $1 > 8(x_T)_{\min}^2 p_{\min} = 8 \times (0.55)^2 \times 0.5 = 1.21$, which is incorrect. Thus the PDF $\xi_T(x_T,p,T) \sim \mathcal{U}(x_T,p)$ is not finite-time reachable from $\xi_0(x_0,p) \sim \mathcal{U}(x_0,p)$ for $T = 4$, via the proposed model dynamics. Hence our measure-theoretic formulation recovers Prajna's invalidation result [7] as a special case. ■

Remark 9. ***(Relaxation of set-based invalidation)*** *Instead of binary (in)validation oracle, we can now measure the "degree of validation" by computing the Wasserstein*

distance ${}_2W_2\left(\frac{1}{\left(1-2x_T^2pT\right)^{3/2}}\frac{1}{vol\left(\widetilde{\mathcal{X}}_0\right)},\ \frac{1}{vol\left(\widetilde{\mathcal{X}}_T\right)}\right)$ *between the model predicted and experimentally measured joint PDFs. More importantly, it dispenses off the conservatism in barrier certificate based model validation by showing that the goodness of a model depends on the measures over the same pair of supports* $\widetilde{\mathcal{X}}_0$ *and* $\widetilde{\mathcal{X}}_T$*, than on the supports themselves. Indeed, given a joint PDF* $\xi\left(x_T,p,T\right)$ *supported over* $\widetilde{\mathcal{X}}_T$ *at* $T=4$*, from (4.24) we can explicitly compute the initial PDF* $\xi_0\left(x_0,p\right)$ *supported over* $\widetilde{\mathcal{X}}_0$ *that, under the proposed model dynamics, yields the prescribed PDF, i.e.*

$$\xi_0\left(x_0,p\right)=\frac{1}{\left(1+8x_0^2p\right)^{3/2}}\,\xi\left(\pm\frac{x_0}{\sqrt{1+8x_0^2p}},p,4\right). \tag{4.26}$$

In other words, if the measurements find the initial density given by (4.26) and final density $\xi\left(x_T,p,T\right)$ *at* $T=4$*, then the Wasserstein distance at* $T=4$ *will be zero, thereby perfectly validating the model. This reinstates the importance of considering the* ***reachability of densities*** *over sets than* ***reachability of sets****, for model validation.*

4.5.4 Comparison with Rantzer's Density Function Based Model Falsification

Similar to barrier certificates, Rantzer's density functions [162] can provide deductive invalidation guarantees (see Theorem 1 in [163]) by constructing a scalar function via convex program. Various applications of these two approaches for temporal verification problems have been reported [164]. It is interesting to note that the main idea of Rantzer's density function stems from an integral form of Liouville equation, given by (see Lemma A.1 in [162])

$$\int_{\mathcal{X}_T}\xi\,dx\ -\ \int_{\mathcal{X}_0}\xi\,dx=\int_0^T\int_{\phi_t(\mathcal{X}_0)}\nabla_x\cdot\left(\xi f\right)dx\,dt, \tag{4.27}$$

where the initial set $\mathcal{X}_0$ gets mapped to the set $\mathcal{X}_T$ at time $t = T$, under the action of the flow $\phi_t(\cdot)$ associated with the nonlinear dynamics $\dot{x} = f(x)$. The convex relaxation proposed for invalidation/safety verification (Theorem 1 in [163]), strives to construct an artificial "density" $\xi = \xi_r(x,t)$ satisfying three conditions, viz. (i) $\xi_r(x,0) > 0$, $\forall\, x \in \mathcal{X}_0$, (ii) $\xi_r(x,T) \leqslant 0$, $\forall\, x \in \mathcal{X}_T$, and (iii) $\nabla_x \cdot (\xi_r f) \geqslant 0$, $\forall x \in \phi_t(\mathcal{X}_0), t \in (0,T)$. From (4.27), such a construction results a "sign-based invalidation", and is only **sufficient** unless a Slater-like condition [165] is satisfied. On the other hand, the "validation in probability" framework proposed in this paper, relies on Liouville PDE-based exact arithmetic computation of ξ, and is a direct simulation-based non-deductive formulation. In this approach, model invalidation equals violation of (4.27), not just the sign-mismatch of its left-hand and right-hand side, and hence is **necessary and sufficient**. As shown in this subsection, for Liouville-integrable nonlinear vector fields (not necessarily semi-algebraic), our framework can recover the deductive falsification inference while bypassing the **additional conservatism** due to SOS-based computation.

4.6 Chapter Summary

In this Chapter, we considered the following question: for the purpose of model validation, what should be the metric for comparing observed and model predicted PDFs? To address this issue, we first laid down the axiomatic requirements that the validation metric must inherit. Then we showed that various pointwise notions of distances, defined on the space of probability densities, fail to satisfy these requirements. Next, we propose to use quadratic Wasserstein distance of order 2, which is an integral notion of distance on the space of probability densities, as a V&V metric. We discuss the computation of the metric in our general context of comparing dynamical systems, and in particular for the case, when the model predicted PDFs

are computed using the algorithms developed in previous two Chapters. The latter amounts to solving a linear program at each instance of measurement availability, whose computational complexity is discussed in detail. Next, we describe the construction of validation certificates to account the case when the initial PDF is not known exactly. Several examples are given to demonstrate the ideas, and to compare the proposed framework with existing nonlinear model falsification techniques.

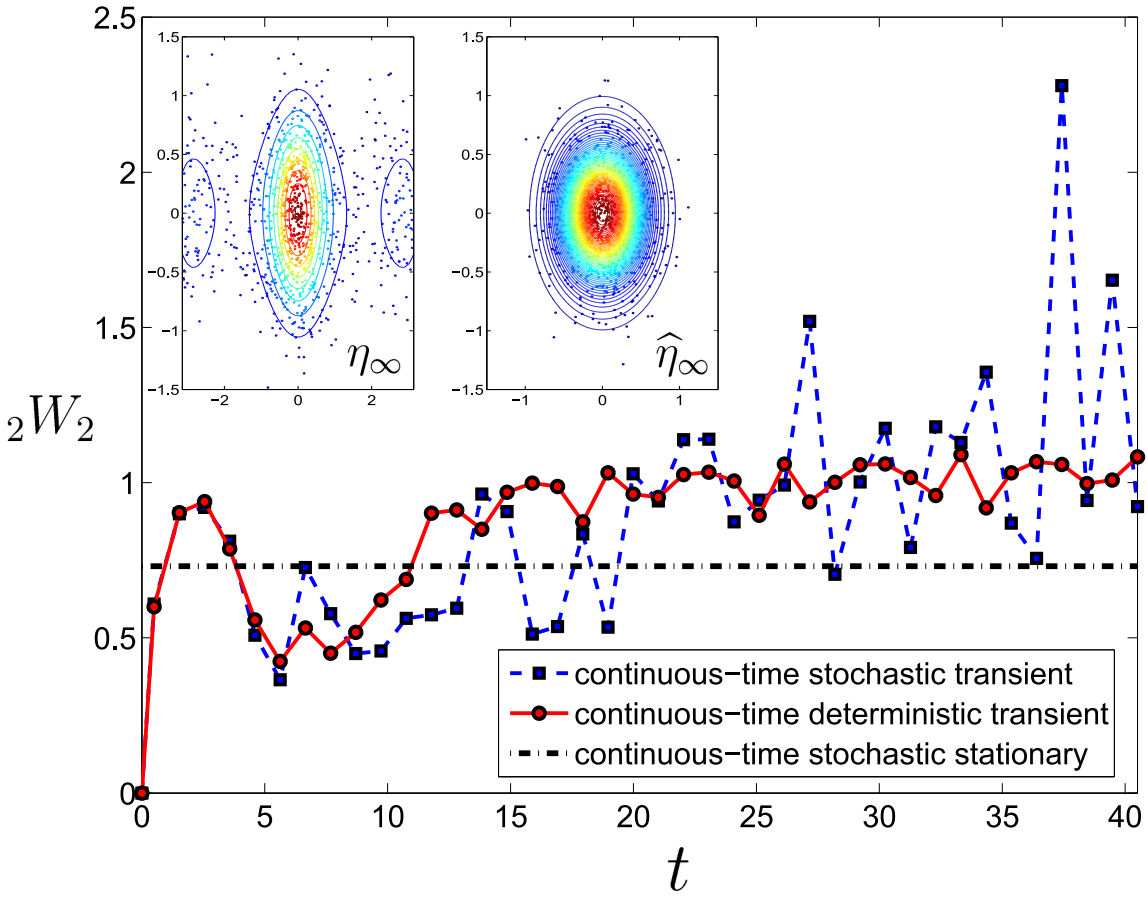


Figure 4.6: Starting with $\xi_0 = \mathcal{U}([-\pi, \pi] \times [-\pi, \pi])$, the *solid line* shows time history of ${}_2W_2$ measured between the joint state PDFs for (4.19) and its linearization about the origin. The *dashed line* shows the same between (4.20) and its linearization about the origin. The *dash-dotted line* shows the stationary ${}_2W_2$ between known η_∞ and $\widehat{\eta}_\infty$ (*contours in the inset plot*), given by (4.21) and (4.22) respectively, and is computed by solving the optimal transport LP between their MCMC samples (*scattered points in the inset plot*). All computations were done with 1000 Halton samples [2] from ξ_0 and 50 eigenfunctions in noise KL expansion.

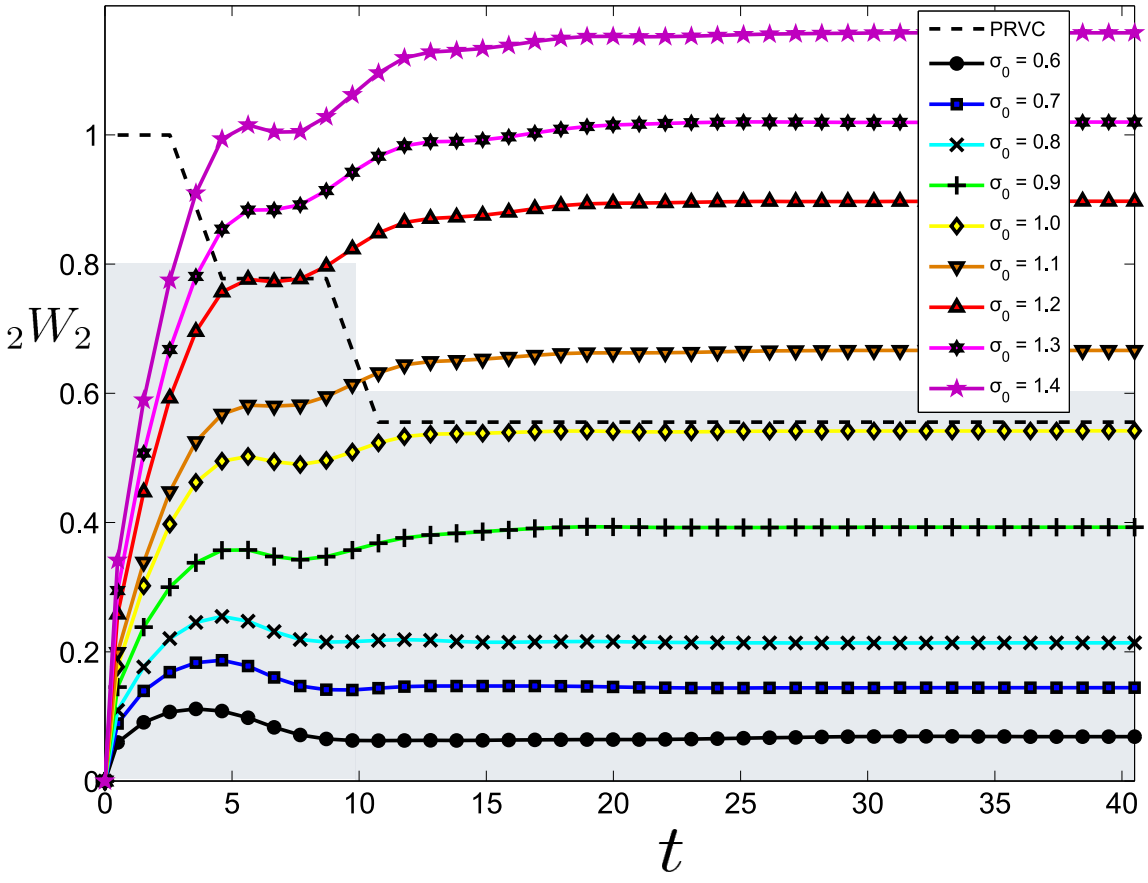


Figure 4.7: Starting with $\xi_0 = \mathcal{N}(0, \sigma_0^2 I_2)$, transient Wasserstein time histories, measured between the joint state PDFs for (4.19) and its linearization about the origin. In this case, increasing σ_0 increases ${}_2W_2$ *at all times*. Further, notice that ${}_2W_2$ trajectories with larger σ_0, converges to higher asymptotic values. This is due to the fact that the stationary density of (4.19) is of the form $\eta_\infty(y) = \sum_{i=1}^{5} m_i^\star \delta(y - y_i^\star)$, and hence depends on ξ_0, as explained in Section 4.2.3 and fourth row of Table 4.1. The *shaded area* shows user-specified tolerance level $\{\gamma_j\}_{j=1}^{40}$, from which PRVC (*dashed line*) can be computed. In this case, PWVC is simply the ${}_2W_2$ trajectory corresponding to $\sigma_0 = 1.4$.

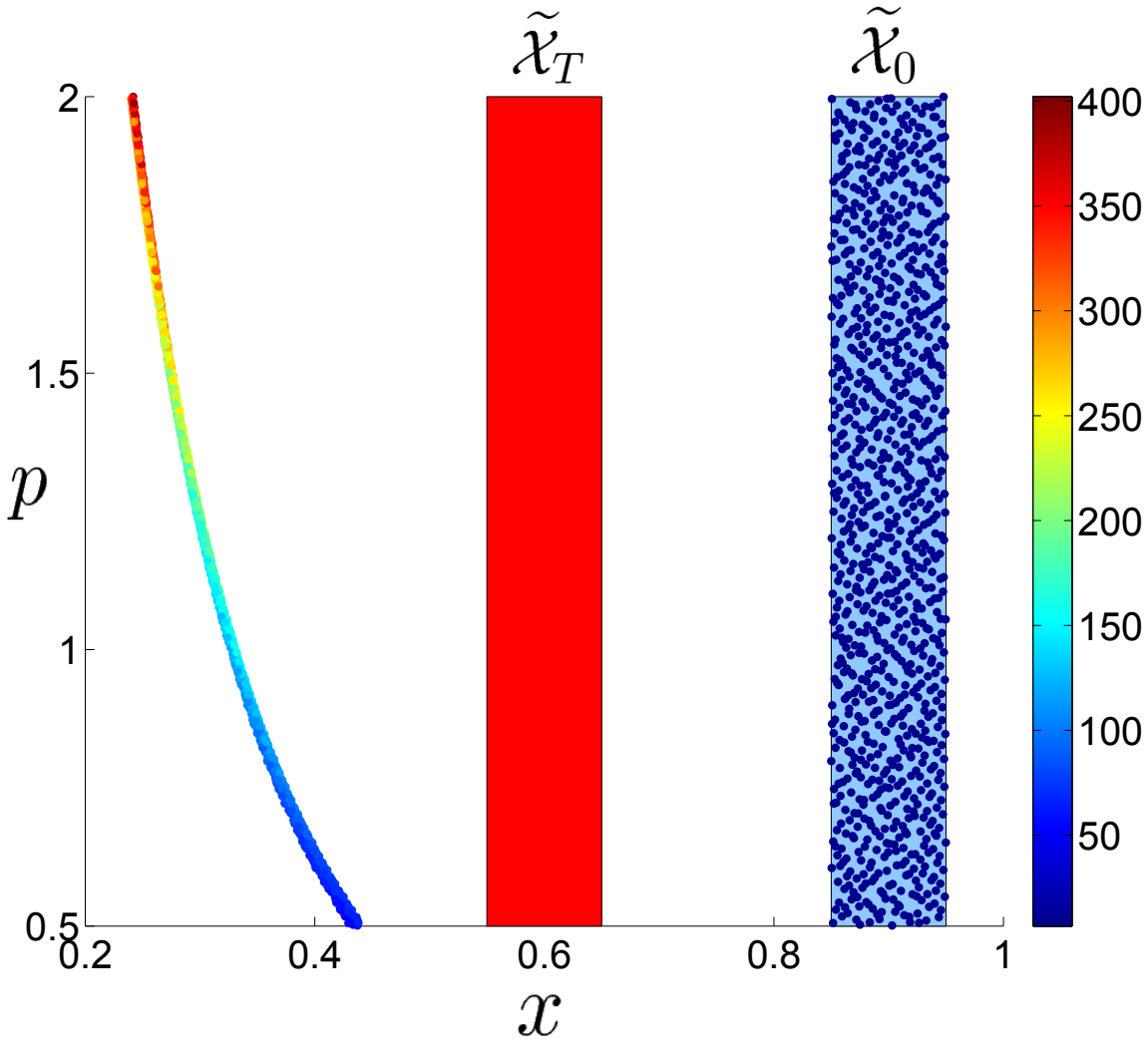


Figure 4.8: This plot illustrates how Prajna's barrier certificate-based invalidation result can be recovered in our probabilistic model validation framework. To show $\widetilde{\mathcal{X}}_T$ is not reachable from the set $\widetilde{\mathcal{X}}_0$ in time $T = 4$, we sample $\widetilde{\mathcal{X}}_0$ uniformly and propagate that uniform ensemble subject to the proposed model dynamics till $T = 4$. The samples are gray-scale color coded (white = high probability, black = low probability) according to the value of the joint PDF at that location. Here, the model is invalidated since the pair of joint PDFs at initial and final time, does not satisfy the Liouville transport PDE corresponding to the model dynamics, as proved in Theorem 11.

5. CASE STUDY: F-16 CONTROLLER ROBUSTNESS VERIFICATION

In this chapter, we will apply the probabilistic V&V formulation developed so far, to the problem of controller verification. This application set up is a slight modification of the model validation problem in the sense that the open loop dynamics is now well-known owing to a good understanding of the first principle physics; however, given a candidate controller, the closed loop performance remains to be verified against uncertainties present in the system. Thus, the problem is essentially that of ensuring probabilistic robustness of a controller.

5.1 Background on Probabilistic Robustness

In recent times, the notion of probabilistic robustness [166, 167, 168, 169, 170, 156, 171], has emerged as an attractive alternative to classical worst-case robust control framework. Probabilistic robustness formulation offers a promising alternative to address these challenges. Instead of the interval-valued structured uncertainty descriptions, it adopts a risk-aware perspective to analyze robustness, and hence, explicitly accounts the distributional information associated with unstructured uncertainty. Furthermore, significant progress have been made in the design and analysis of randomized algorithms [172, 173] for computations related to probabilistic robustness. These recent developments are providing impetus to a transition from "worst-case" to "distributional robustness" [174, 175].

In order to fully leverage the potential of distributional robustness, the associated computation must be scalable and of high accuracy. However, numerical implementation of most probabilistic methods rely on Monte Carlo like realization-based algorithms, leading to high computational cost for implementing them to nonlinear systems. In particular, the accuracy of robustness computation depends on the nu-

merical accuracy of histogram-based piecewise constant approximation of the PDF that evolves spatio-temporally over the joint state and parameter space, under the action of closed-loop nonlinear dynamics. Nonlinearities at trajectory level cause non-Gaussianity at PDF level, even when the initial uncertainty is Gaussian. Thus, in Monte Carlo approach, at any given time, a high-dimensional nonlinear system requires a dense grid to sufficiently resolve the non-Gaussian PDF, incurring the 'curse of dimensionality' [94].

This is a serious bottleneck in applications like flight control software certification [176], where the closed loop dynamics is nonlinear, and linear robustness analysis supported with Monte Carlo, remains the state-of-the art. Lack of nonlinear robustness analysis tools, coupled with the increasing complexity of flight control algorithms, have caused loss of several F/A-18 aircrafts due to nonlinear "falling leaf mode" [177], that went undetectable [178] by linear robustness analysis algorithms. On the other hand, accuracy of sum of squares optimization-based *deterministic* nonlinear robustness analysis [176, 177] depends on the quality of semi-algebraic approximation, and is still computationally expensive for large-scale nonlinear systems. Thus, there is a need for controller robustness verification methods, that does not make any structural assumption on nonlinearity, and allows scalable computation while accommodating stochastic uncertainty.

Rest of this Chapter is structured as follows. In next Section, we describe the nonlinear open-loop dynamics of F-16 aircraft in longitudinal flight, followed by the synthesis of linear quadratic regulator (LQR) and gain-scheduled linear quadratic regulator (gsLQR) – the two controllers whose state regulation performances are being compared. The proposed framework is detailed next and consists of closed-loop uncertainty propagation and optimal transport to trim. Numerical results illustrating the proposed method, are presented in the end.

5.2 F-16 Flight Dynamics

5.2.1 Longitudinal Equations of Motion

The longitudinal equations of motion for F-16 considered here, follows the model given in [179, 180, 181], with the exception that we restrict the maneuver to a constant altitude ($h = 10,000$ ft) flight. Further, the north position state equation is dropped since no other longitudinal states depend on it. This results a reduced *four state, two input model* with $x = (\theta, V, \alpha, q)^\top$, $u = (T, \delta_e)^\top$, given by

$$\begin{aligned}
\dot{\theta} &= q, && (5.1a)\\
\dot{V} &= \frac{1}{m}\cos\alpha\left[T - mg\sin\theta + \overline{q}S\left(C_X + \frac{\overline{c}}{2V}C_{X_q}q\right)\right] \\
&+ \frac{1}{m}\sin\alpha\left[mg\cos\theta + \overline{q}S\left(C_Z + \frac{\overline{c}}{2V}C_{Z_q}q\right)\right], && (5.1b)\\
\dot{\alpha} &= q - \frac{\sin\alpha}{mV}\left[T - mg\sin\theta + \overline{q}S\left(C_X + \frac{\overline{c}}{2V}C_{X_q}q\right)\right] \\
&+ \frac{\cos\alpha}{mV}\left[mg\cos\theta + \overline{q}S\left(C_Z + \frac{\overline{c}}{2V}C_{Z_q}q\right)\right], && (5.1c)\\
\dot{q} &= \frac{\overline{q}S\overline{c}}{J_{yy}}\left[C_m + \frac{\overline{c}}{2V}C_{m_q}q + \frac{\left(x_{\text{cg}}^{\text{ref}} - x_{\text{cg}}\right)}{\overline{c}}\left(C_Z + \frac{\overline{c}}{2V}C_{Z_q}q\right)\right]. && (5.1d)
\end{aligned}$$

The state variables are second Euler angle θ (deg), total velocity V (ft/s), angle-of-attack α (deg), and pitch rate q (deg/s), respectively. The control variables are thrust T (lb), and elevator deflection angle δ_e (deg). Table 5.1 lists the parameters involved in (5.1). Furthermore, the dynamic pressure $\overline{q} = \frac{1}{2}\rho(h)V^2$, where the atmospheric density $\rho(h) = \rho_0\left(1 - 0.703 \times 10^{-5}h\right)^{4.14} = 1.8 \times 10^{-3}$ slugs/ft^3 remains fixed.

5.2.2 Aerodynamic Coefficients

The aerodynamic force and moment coefficients C_X, C_Z, and C_m are functions of α and δ_e, expressed as look-up table from wind tunnel test data [179, 180, 181].

Table 5.1: Parameters in (5.1)

Description of parameters	Values with dimensions
Mass of the aircraft	$m = 636.94$ slugs
Acceleration due to gravity	$g = 32.17$ ft/s^2
Wing planform area	$S = 300$ ft^2
Mean aerodynamic chord	$\overline{c} = 11.32$ ft
Reference x-position of c.g.	$x_{\text{cg}}^{\text{ref}} = 0.35\,\overline{c}$ ft
True x-position of c.g.	$x_{\text{cg}} = 0.30\,\overline{c}$ ft
Pitch moment-of-inertia	$J_{yy} = 55,814$ slug-ft^2
Nominal atmospheric density	$\rho_0 = 2.377 \times 10^{-3}$ slugs/ft^3

Similarly, the stability derivatives C_{X_q}, C_{Z_q}, and C_{m_q} are look-up table functions of α. We refer the readers to above references for details.

5.3 Controller Synthesis

We consider two controllers: LQR and gsLQR, as shown in Fig. 5.1 and 5.2 respectively, with the common objective of regulating the state to its trim value. Here, w denotes the actuator disturbance. Both controllers minimize the infinite-horizon cost functional

$$\mathcal{J} = \int_0^\infty \left(x(t)^\top Q\, x(t) + u(t)^\top R\, u(t) \right)\, dt, \tag{5.2}$$

with $Q = \text{diag}\,(100, 0.25, 100, 10^{-4})$, and $R = \text{diag}\,(10^{-6}, 625)$. The control saturation shown in the block diagrams, is modeled as

$$1000\,\text{lb} \leqslant T \leqslant 28,000\,\text{lb}, \quad -25^\circ \leqslant \delta_e \leqslant +25^\circ. \tag{5.3}$$

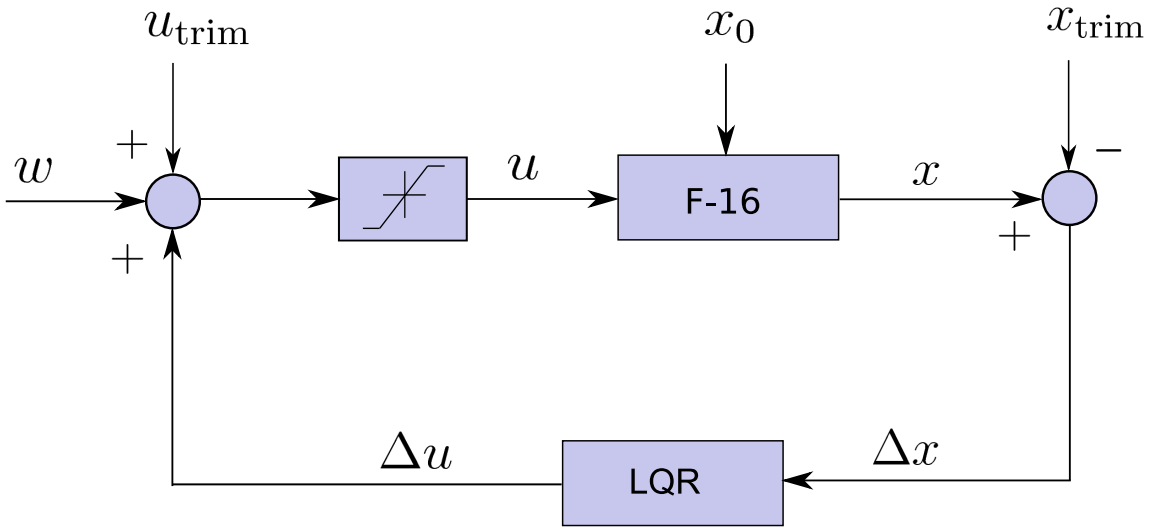


Figure 5.1: Block diagram for LQR closed-loop system.

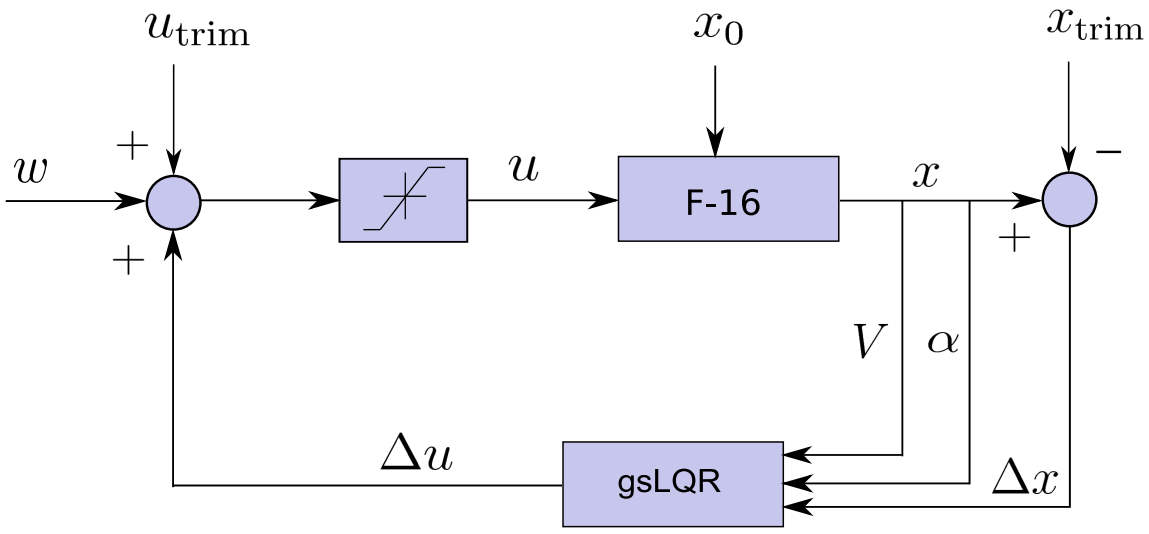


Figure 5.2: Block diagram for gsLQR closed-loop system.

5.3.1 LQR Synthesis

The nonlinear open loop plant model was linearized about $x_{\text{trim}}, u_{\text{trim}}$, using simulink `linmod` command. The trim conditions were computed via the nonlinear optimization package SNOPT [182], and are given by

$$x_{\text{trim}} = \left(2.8190\,\text{deg}, 407.8942\,\text{ft/s}, 6.1650\,\text{deg}, 6.8463 \times 10^{-4}\,\text{deg/s}\right)^{\top}, \quad (5.4)$$

$$u_{\text{trim}} = \left(1000\,\text{lb}, -2.9737\,\text{deg}\right)^{\top}. \quad (5.5)$$

The LQR gain matrix K, computed for this linearized model, was found to be

$$K = \begin{bmatrix} 7144.9 & -400.58 & -1355.8 & 2002.8 \\ 0.7419 & -0.0113 & -0.2053 & 0.3221 \end{bmatrix}. \quad (5.6)$$

As observed in Fig. 5.3 (a), both open-loop and LQR closed-loop linear systems are stable.

5.3.2 Gain-scheduled LQR Synthesis

As shown in Fig. 5.2 (b), V and α are taken as the scheduling states. We generate 100 grid points in the box

$$100\,\text{ft/s} \leqslant V \leqslant 1000\,\text{ft/s}, \quad -10^{\circ} \leqslant \alpha \leqslant +45^{\circ}, \quad (5.7)$$

and compute trim conditions $\{x^j_{\text{trim}}, u^j_{\text{trim}}\}_{j=1}^{100}$, using SNOPT, for each of these grid points. Next, we synthesize a sequence of LQR gains $\{K_j\}_{j=1}^{100}$, corresponding to the linearized dynamics about each trim. For the closed-loop nonlinear system, the gain matrices at other state vectors are linearly interpolated over $\{K_j\}_{j=1}^{100}$. As shown in Fig. 5.3 (b), depending on the choice of the trim conditions corresponding to the

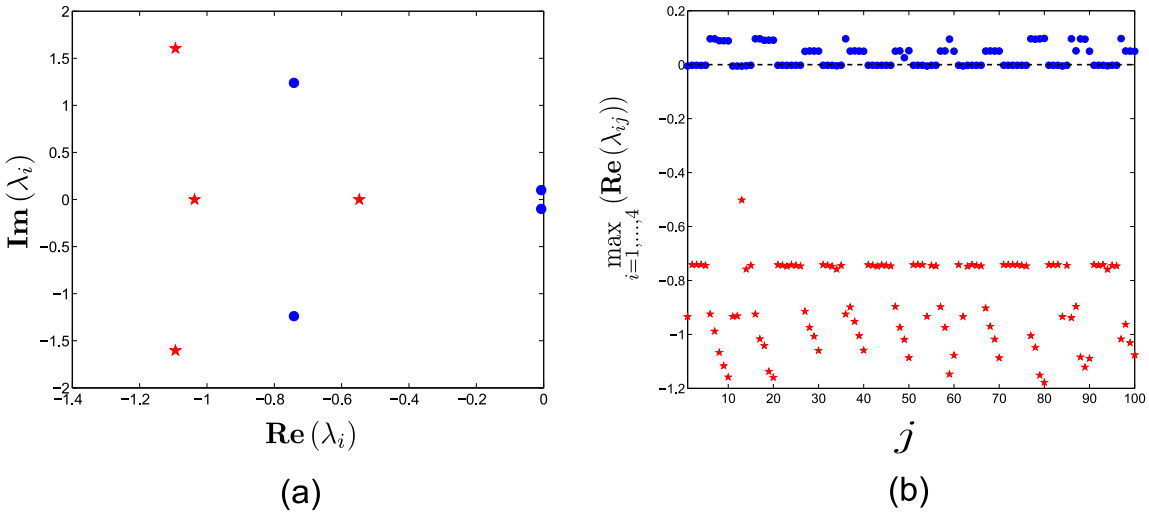


Figure 5.3: (a) The open-loop (*circles*) and LQR closed-loop (*stars*) eigenvalues shown in the complex plane, for the linearized model. (b) For gsLQR synthesis, maximum of the real parts of open-loop (*circles*) and closed-loop (*stars*) eigenvalues for each of the $j = 1, \ldots, 100$ linearizations are plotted. Depending on the trim condition, some open-loop linearized plants can be unstable but all closed-loop synthesis are stable.

grid-points in scheduling subspace, some open-loop linearized plants are unstable but all closed-loop synthesis are stable.

5.4 Optimal Transport to Trim

5.4.1 Probabilistic Robustness of State Regulation Performance As Optimal Transport

We assume that the uncertainties in initial conditions (x_0) and parameters (p) are described by the initial joint PDF $\xi_0(x_0, p)$, and this PDF is known for the purpose of performance analysis. For $t > 0$, under the action of the closed-loop dynamics, ξ_0 evolves over the *extended state space*, defined as the joint space of states and parameters, to yield the instantaneous joint PDF $\xi(x(t), p, t)$. We use (2.9) for MOC computation of $\xi(x(t), p, t)$. Notice that the divergence computation in

(2.9) can be done analytically *offline* for our case of LQR and gsLQR closed-loop systems, provided we obtain function approximations for aerodynamic coefficients. However, there are two drawbacks for such offline computation of the divergence. **First**, the accuracy of the computation will depend on the quality of function approximations for aerodynamic coefficients. **Second**, for nonlinear controllers like MPC [183], which numerically realize the state feedback, analytical computation for closed-loop divergence is not possible. For these reasons, we implement an alternative *online* computation of divergence in this paper. Using the Simulink® command `linmod`, we linearize the *closed-loop systems* about each characteristics, and obtain the instantaneous divergence as the trace of the time-varying Jacobian matrix. Then, the uncertainty propagation follows the generic structure of Algorithm 1 in Chapter 2.

To provide a quantitative comparison for LQR and gsLQR controllers' performance, we need a notion of "distance" between the respective time-varying state PDFs and the desired state PDF. Since the controllers strive to bring the state trajectory ensemble to x_{trim}, hence we take $\xi^*(x_{\text{trim}})$, a Dirac delta distribution at x_{trim}, as our desired joint PDF. For this purpose, we use the Wasserstein metric ${}_2W_2$ introduced in Chapter 4. Recall that the Wasserstein distance computation from its definition leads to solving an infinite dimensional LP associated with the Monge-Kantorovich optimal transport. As we have seen in Section 4.3, the discrete version of this problem is a large but finite-dimensional LP, which though consistent, is not easy to implement in generic solvers. For example, an MOC with 1000 samples leads to an LP with one million decision variables, even after sparseness is exploited. This issue, in general, leads to numerical accuracy versus storage capacity trade off. Next, we show that for state regulation performance considered in this Chapter, this computational burden can be reduced significantly.

Table 5.2: Admissible state perturbation limits

x_{pert}	Interval
$\theta_{\text{pert}} \in \left[\theta_{\text{pert}}^{\min}, \theta_{\text{pert}}^{\max}\right]$	$[-35^\circ, +35^\circ]$
$V_{\text{pert}} \in \left[V_{\text{pert}}^{\min}, V_{\text{pert}}^{\max}\right]$	$[-65$ ft/s, $+65$ ft/s$]$
$\alpha_{\text{pert}} \in \left[\alpha_{\text{pert}}^{\min}, \alpha_{\text{pert}}^{\max}\right]$	$[-20^\circ, +50^\circ]$
$q_{\text{pert}} \in \left[q_{\text{pert}}^{\min}, q_{\text{pert}}^{\max}\right]$	$[-70$ deg/s, $+70$ deg/s$]$

5.4.2 Reduction of Storage Complexity

For our purpose of computing $W\left(\xi\left(x\left(t\right),t\right),\xi^{*}\left(x_{\text{trim}}\right)\right)$, the storage complexity can be reduced by leveraging the fact that $\xi^{*}\left(x_{\text{trim}}\right)$ is a stationary Dirac distribution. Hence, it suffices to represent the joint probability mass function (PMF) of $\xi^{*}\left(x_{\text{trim}}\right)$ as a *single sample* located at x_{trim} with PMF value unity. This trivializes the optimal transport problem, since

$$W\left(t\right) \triangleq W\left(\xi\left(x\left(t\right),t\right),\xi^{*}\left(x_{\text{trim}}\right)\right) = \sqrt{\sum_{i=1}^{n} \parallel x_i\left(t\right) - x_{\text{trim}} \parallel_2^2 \ \gamma_i}, \tag{5.8}$$

where $\gamma_i \geqslant 0$ denotes the joint PMF value at sample $x_i\left(t\right)$, $i = 1, \ldots, n$. Consequently, the storage complexity reduces to $(nn_s + n + n_s)$, which is *linear* in number of samples n.

5.5 Numerical Results

5.5.1 Robustness Against Initial Condition Uncertainties

5.5.1.1 Stochastic Initial Condition Uncertainty

We first consider analyzing the controller robustness subject to initial condition uncertainties. For this purpose, we let the initial condition x_0 to be a stochastic perturbation from x_{trim}, i.e. $x_0 = x_{\text{trim}} + x_{\text{pert}}$, where x_{pert} is a random vector with prob-

ability density $\xi_{\text{pert}} = \mathcal{U}\left(\left[\theta_{\text{pert}}^{\min}, \theta_{\text{pert}}^{\max}\right] \times \left[V_{\text{pert}}^{\min}, V_{\text{pert}}^{\max}\right] \times \left[\alpha_{\text{pert}}^{\min}, \alpha_{\text{pert}}^{\max}\right] \times \left[q_{\text{pert}}^{\min}, q_{\text{pert}}^{\max}\right]\right)$, where the perturbation range for each state, is listed in Table 5.2. Consequently, x_0 has a joint PDF $\xi_0(x_0)$. For this analysis, we assume no actuator disturbance.

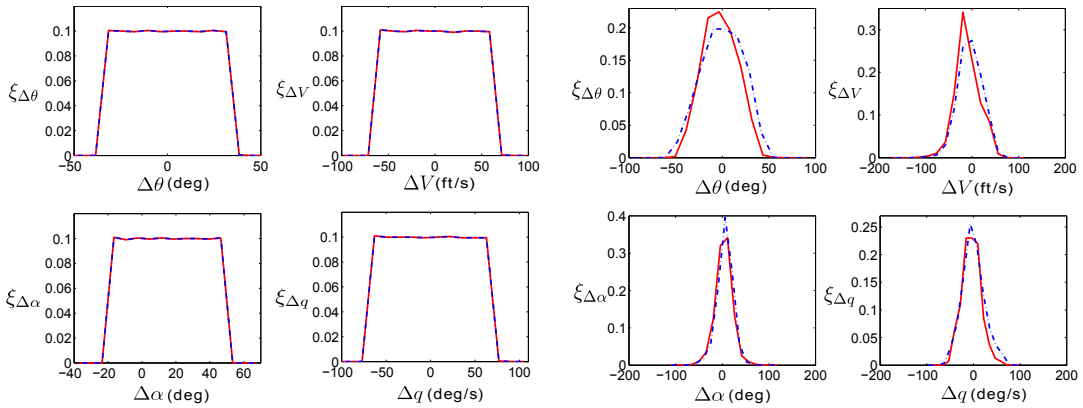
5.5.1.2 Simulation Set Up

We generated pseudo-random Halton sequence [2] in $\left[\theta_{\text{pert}}^{\min}, \theta_{\text{pert}}^{\max}\right] \times \left[V_{\text{pert}}^{\min}, V_{\text{pert}}^{\max}\right] \times \left[\alpha_{\text{pert}}^{\min}, \alpha_{\text{pert}}^{\max}\right] \times \left[q_{\text{pert}}^{\min}, q_{\text{pert}}^{\max}\right]$, to sample the uniform distribution ξ_{pert}, and hence ξ_0 supported on the four dimensional state space. With 2000 Halton samples for ξ_0, we propagate joint state PDFs for both LQR and gsLQR closed-loop dynamics via MOC ODE (2.9), from $t = 0$ to 20 seconds, using fourth-order Runge-Kutta integrator with fixed step-size $\Delta t = 0.01$ s.

We observed that the `linmod` computation for evaluating time-varying divergence along each trajectory, takes the most of computational time. To take advantage of the fact that computation along characteristics are independent of each other, all simulations were performed using 12 cores with NVIDIA® Tesla GPUs in MATLAB® environment. It was noticed that with LQR closed-loop dynamics, the computational time for single sample from $t = 0$ to 20 s, is approx. 90 seconds. With sequential `for`-loops over 2000 samples, this scales to 50 hours of runtime. The same for gsLQR scales to 72 hours of runtime. In parallel implementation on Tesla, MATLAB® `parfor`-loops were used to reduce these runtimes to 4.5 hours (for LQR) and 6 hours (for gsLQR), respectively.

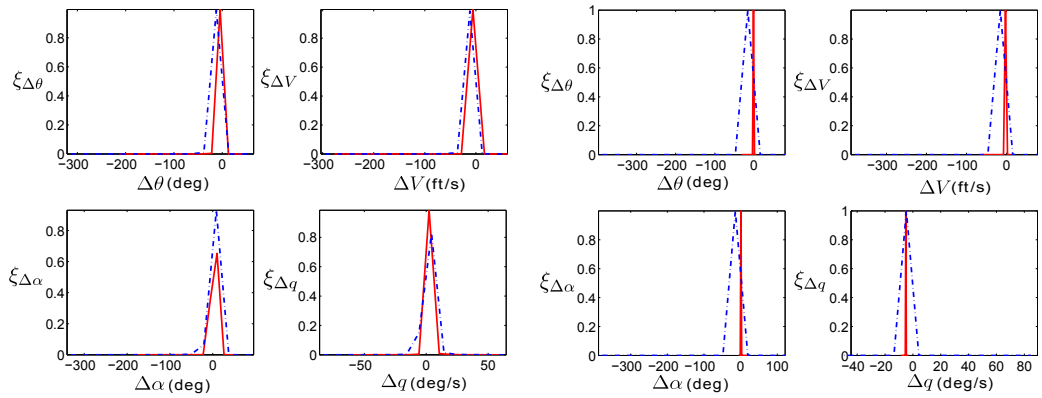
5.5.1.3 Density Based Qualitative Analysis

Fig. 5.4 shows the evolution of univariate *marginal error PDFs*. All marginal computations were performed using algorithms given in [38]. Since ξ_0 and its marginals were uniform, Fig. 5.4(a) shows similar trend for small t, and there seems no visible difference between LQR and gsLQR performance. As t increases, both LQR and



(a) Snapshot at $t = 0.01$ second.

(b) Snapshot at $t = 1$ second.



(c) Snapshot at $t = 5$ seconds.

(d) Snapshot at $t = 20$ seconds.

Figure 5.4: Snapshots of univariate marginal error PDFs for each state, with LQR *(blue, dashed)* and gsLQR *(red, solid)* closed loop dynamics.

gsLQR error PDFs shrink about zero. By $t = 20$ s (Fig. 5.4(d)), both LQR and gsLQR controllers make the respective state marginals $\xi^j(t)$, $j = 1, \ldots, 4$, converge to the Dirac distribution at x^j_{trim}, although the rate of convergence of gsLQR error marginals is faster than the same for LQR.

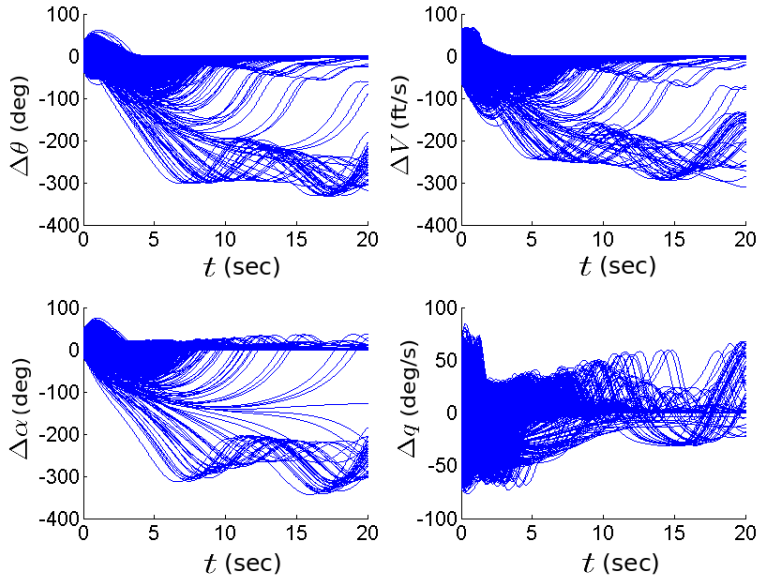
Thus, Fig. 5.4 *qualitatively* show that both LQR and gsLQR exhibit comparable *immediate* and *asymptotic* performance, as far as robustness against initial condition uncertainty is concerned. However, there are some visible mismatches in Fig. 5.4(b) and 5.4(c), that suggests the need for a careful quantitative investigation of the

transient performance.

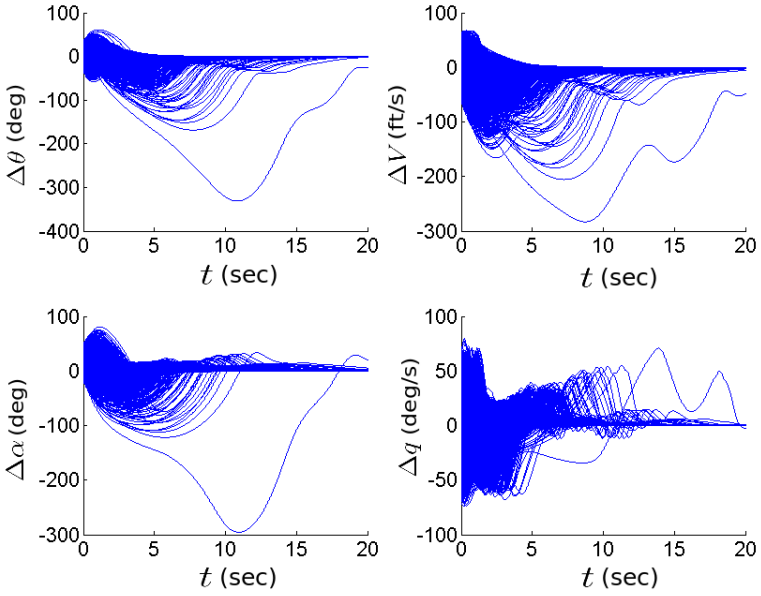
The insights obtained from Fig. 5.4 can be verified against the MC simulations (Fig. 5.5). Compared to LQR, the MC simulations reveal faster regulation performance for gsLQR, and hence corroborate the faster rate of convergence of gsLQR error marginals observed in Fig. 5.4. From Fig. 5.5, it is interesting to observe that by $t = 20$ s, some of the LQR trajectories do not converge to trim while all gsLQR trajectories do. For risk aware control design, it is natural to ask: how probable is this event, i.e. can we probabilistically assess the severity of the loss of performance for LQR? To address this question, in Fig. 5.6, we plot the time evolution of the peak value of LQR joint state PDF, and compare that with the joint state PDF values along the LQR closed-loop trajectories that don't converge to x_{trim} by 20 s. Fig. 5.6 reveals that the probabilities that the LQR trajectories don't converge, remain *at least* an order of magnitude less than the peak value of the LQR joint PDF. In other words, the performance degradation for LQR controller, as observed in Fig. 5.5(a), is a low-probability event. This conclusion can be further verified from Fig. 5.7, which shows that for gsLQR controller, both maximum and minimum probability trajectories achieve satisfactory regulation performance by $t = 20$ s. However, for LQR controller, although the maximum probability trajectory achieves regulation performance as good as the corresponding gsLQR case, the minimum probability LQR trajectory results in poor regulation. Furthermore, even for the maximum probability trajectories (Fig. 5.7, *top row*), there are transient performance mismatch between LQR and gsLQR, for approximately 3–8 s.

5.5.1.4 Optimal Transport Based Quantitative Analysis

From a systems-control perspective, instead of performing an *elaborate qualitative* statistical analysis as above, one would like to have a *concise and quantitative*



(a) State error vs. time for LQR controller



(b) State error vs. time for gsLQR controller

Figure 5.5: MC state error ($\Delta x^j(t) \triangleq x^j(t) - x^j_{\text{trim}},\ j = 1, \ldots, 4$) trajectories for LQR and gsLQR closed-loop dynamics.

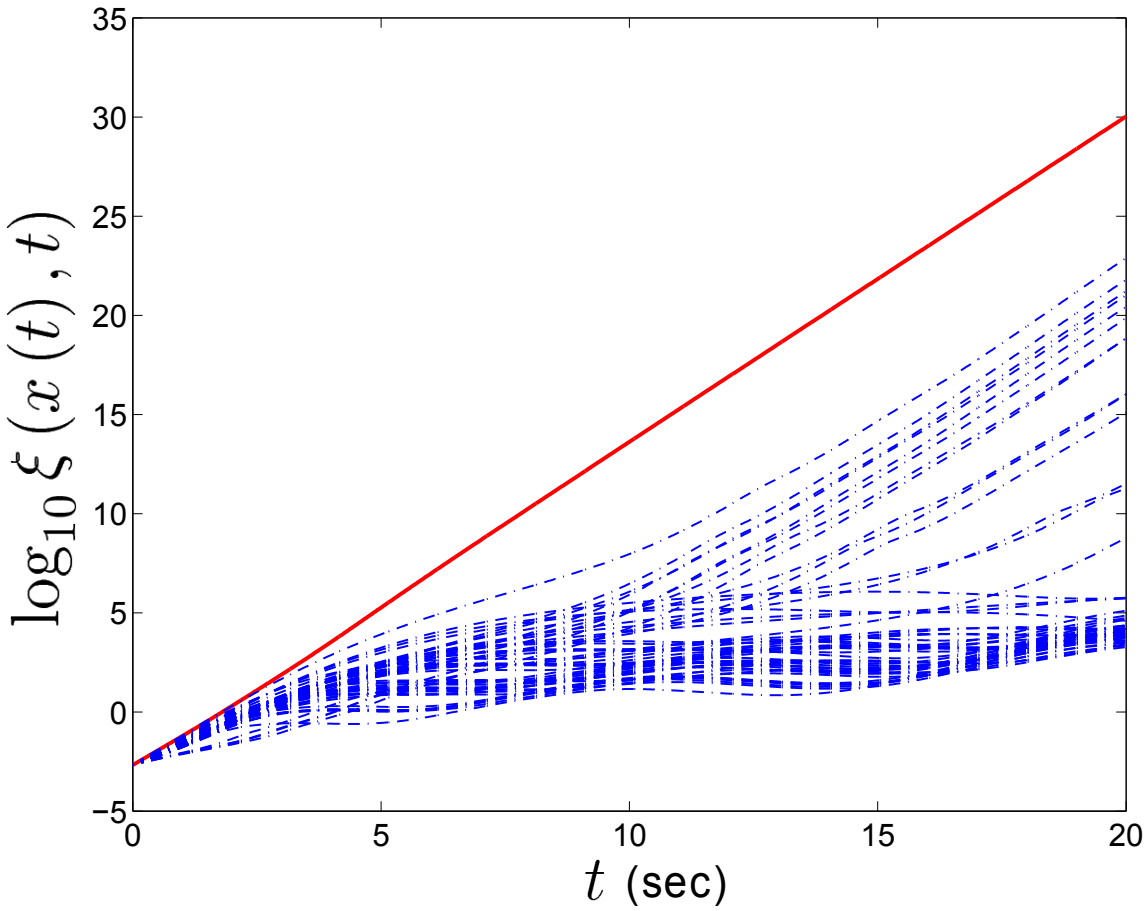


Figure 5.6: Time evolution of *maximum value* of joint PDF $\xi_{\mathrm{LQR}}\left(x(t), t\right)$ (*red solid*) and $\xi_{\mathrm{LQR}}\left(x(t), t\right)$ *along the diverging trajectories* (*blue dashed*), as seen in Fig. 5.5(a). The plots are in *log-linear* scale.

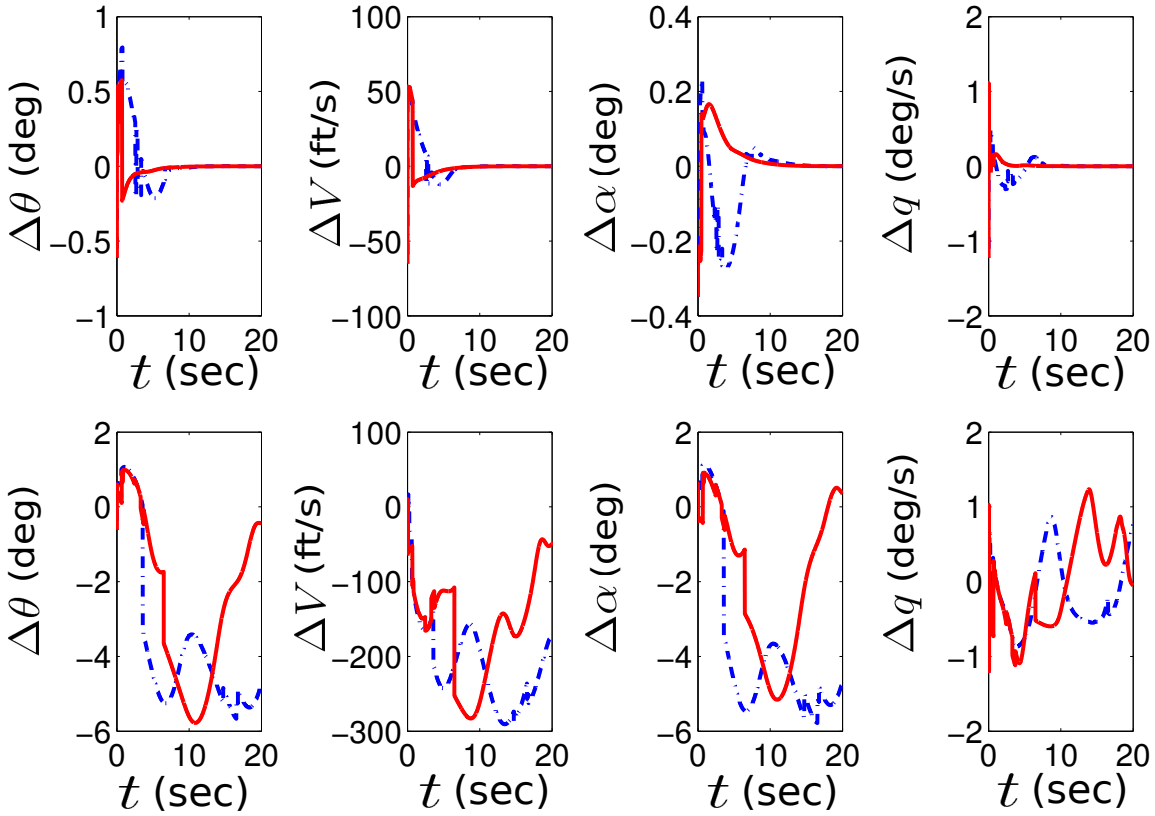


Figure 5.7: Time evolution of the most likely (*top row*) and least likely (*bottom row*) state errors for LQR (*blue dashed*) and gsLQR (*red solid*) closed-loop dynamics.

robustness analysis tool, enabling the inferences of the previous subsection. We now illustrate that the Wasserstein distance introduced in Section 4, serves this need.

In this formulation, a controller is said to have better regulation performance if it makes the closed-loop state PDF converge faster to the Dirac distribution located at x_{trim}. In other words, for a better controller, at all times, the distance between the closed-loop state PDF and the Dirac distribution, as measured in ${}_2W_2$, must remain smaller than the same for the other controller. In this Chapter, for notational ease, we drop the subscripts from the Wasserstein distance notation ${}_2W_2$, and denote it simply by W. Thus, we compute the time-evolution of the two Wasserstein distances:

$$W_{\text{LQR}}(t) \triangleq W\left(\xi_{\text{LQR}}(x(t),t), \xi^*(x_{\text{trim}})\right), \tag{5.9}$$

$$W_{\text{gsLQR}}(t) \triangleq W\left(\xi_{\text{gsLQR}}(x(t),t), \xi^*(x_{\text{trim}})\right). \tag{5.10}$$

The schematic of this computation is shown in Fig. 5.8. Fig. 5.9 indeed confirms

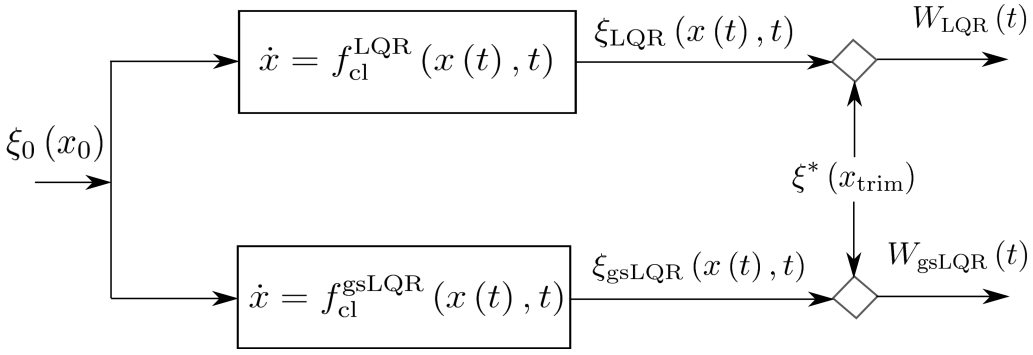


Figure 5.8: Schematic of probabilistic robustness comparison for controllers based on Wasserstein metric. The "diamond" denotes the Wasserstein computation by solving the Monge-Kantorovich optimal transport. The internal details of LQR and gsLQR closed-loop dynamics blocks are as in Fig. 5.1 and Fig. 5.2.

the qualitative trends, observed in the density based statistical analysis mentioned before, that LQR and gsLQR exhibit comparable immediate and asymptotic performance. Furthermore, Fig. 5.9 shows that for $t = 3 - 8$ seconds, W_{LQR} stays higher than W_{gsLQR}, meaning the gsLQR joint PDF $\xi_{\text{gsLQR}}(x(t), t)$ is closer to $\xi^*(x_{\text{trim}})$, compared to the LQR joint PDF $\xi_{\text{LQR}}(x(t), t)$. This corroborates well with the transient mismatch observed in Fig. 5.4(c). As time progresses, both W_{LQR} and W_{gsLQR} converge to zero, meaning the convergence of both LQR and gsLQR closed-loop joint state PDFs to the Dirac distribution at x_{trim}.

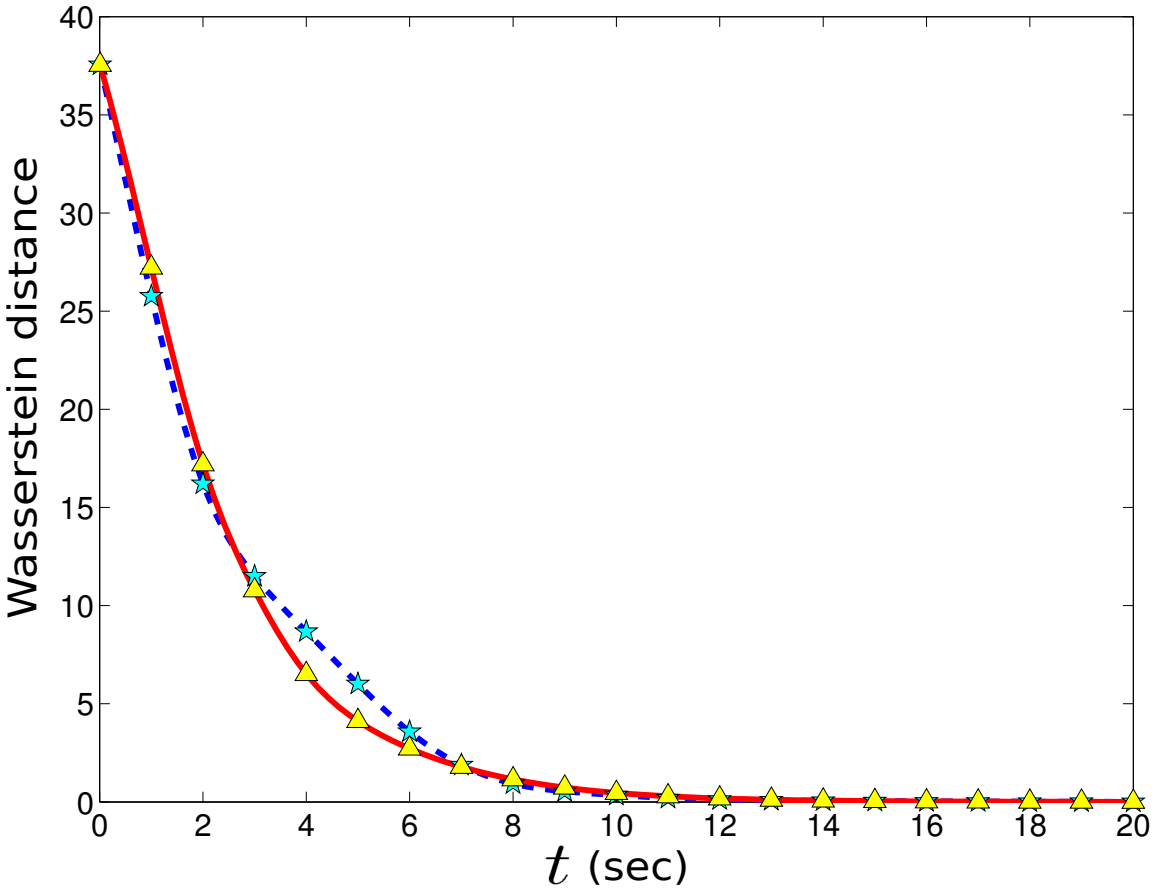


Figure 5.9: Comparison of time histories of $W(\xi_{\text{LQR}}(t), \xi^*)$ *(blue dashed, star)* and $W(\xi_{\text{gsLQR}}(t), \xi^*)$ *(red solid, triangle).*

Remark 10. *At this point, we highlight a subtle distinction between the two approaches of probabilistic robustness analysis presented above: (1)* density based qualitative analysis*, and (2)* the optimal transport based quantitative analysis using Wasserstein distance*. For density based qualitative analysis, controller performance assessment was done using Fig. 5.4 that compares the asymptotic convergence of the univariate marginal state PDFs. However, this analysis is only* sufficient *since convergence of marginals does not* necessarily *imply convergence of joints. Conversely, the optimal transport based quantitative analysis is* necessary and sufficient *since Fig. 5.9 compares the Wasserstein distance between the joint PDFs. We refer the readers to Appendix C (Theorem 26) for a precise statement and proof.*

Further, since $W_{LQR}(t) \to 0$ for large t, we can affirmatively say that the divergent LQR trajectories are indeed of low-probability, as hinted by Fig. 5.6 and 5.7. Otherwise, $W_{LQR}(t)$ would show a steady-state error. Thus, the optimal transport theoretic Wasserstein distance is shown to be an effective way of comparing the robustness of controllers.

5.5.2 Robustness Against Parametric Uncertainties

5.5.2.1 Deterministic Initial Condition with Stochastic Parametric Uncertainty

Instead of the stochastic initial condition uncertainties described in above, we now consider uncertainties in three parameters: mass of the aircraft (m), true x-position of c.g. (x_{cg}), and pitch moment-of-inertia (J_{yy}). The uncertainties in these geometric parameters can be attributed to the variable rate of fuel consumption depending on the flight conditions. For the simulation purpose, we assume that each of these three parameters has $\pm\Delta\%$ *uniform* uncertainties about their nominal values listed in Table 5.1. To verify the controller robustness, we vary the parametric uncertainty range by allowing $\Delta = 0.5, 2.5, 5, 7.5$ and 15. As before, we set the

actuator disturbance $w = 0$.

5.5.2.2 Simulation Set Up

We let the initial condition be a deterministic vector: $x_0 = x_{\text{trim}} + x_{\text{pert}}$, where $x_{\text{pert}} = [1.1803\,\text{rad},\ 5.1058\,\text{ft/s},\ 2.8370\,\text{rad},\ 10^{-4}\,\text{rad/s}]^\top$. We keep the rest of the simulation set up same as in the previous case. Notice that since $\dot{p} = 0$, the characteristic ODE for joint PDF evolution remains the same. However, the state trajectories, along which the characteristic ODE needs to be integrated, now depend on the realizations of the random vector p.

5.5.2.3 Density Based Qualitative Analysis

Due to parametric uncertainties in $p \triangleq [m,\ x_{\text{cg}},\ J_{yy}]^\top$, we now have $n_s = 4, n_p = 3$, and hence the joint PDF evolves over the extended state space $\widetilde{x}(t) \triangleq [x(t),\ p]^\top \in \mathbb{R}^7$. Since we assumed x_0 to be deterministic, both initial and asymptotic joint PDFs ξ_0 and ξ_∞ are degenerate distributions, supported over the three dimensional parametric subspace of the extended state space in $\mathbb{R}^7$. In other words, $\xi_0 = \xi_p(p)\,\delta(x - x_0)$, and $\xi_\infty = \xi_p(p)\,\delta(x - x_{\text{trim}})$, i.e. the PDFs ξ_0 and ξ_∞ differ only by a translation of magnitude $\| x_0 - x_{\text{trim}} \|_2 = \| x_{\text{pert}} \|_2$. However, for any intermediate time $t \in (0, \infty)$, the joint PDF $\xi(\widetilde{x}(t), t)$ has a support obtained by nonlinear transformation of the initial support. This is illustrated graphically in Fig. 5.10.

The MC simulations in Fig. 5.11 show that both LQR and gsLQR have similar asymptotic performance, however, the transient overshoot for LQR is much larger than the same for gsLQR. Hence, the transient performance for gsLQR seems to be more robust against parametric uncertainties. Similar trends were observed for other values of Δ.

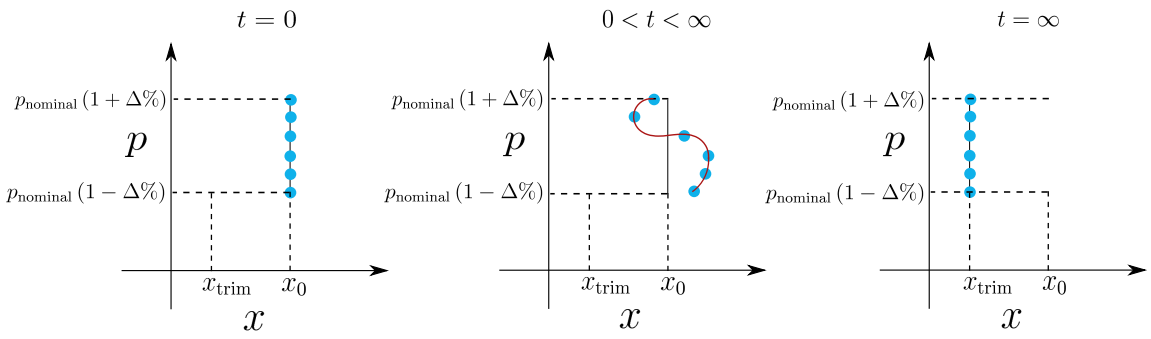
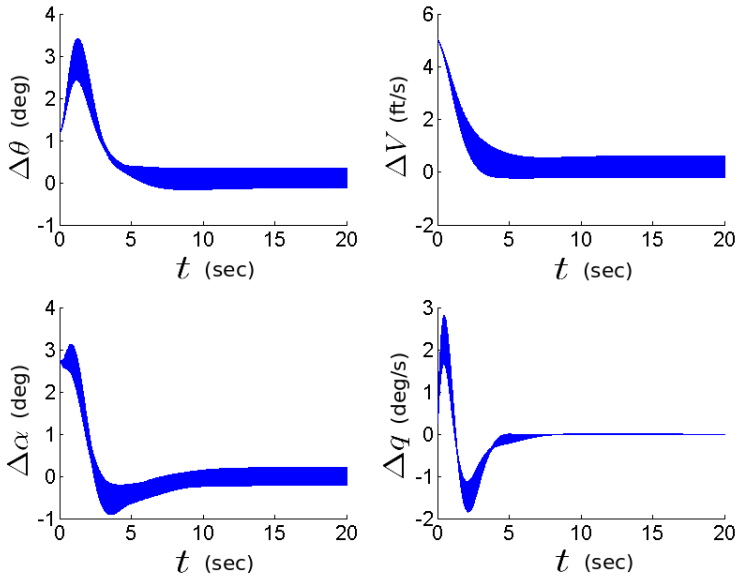
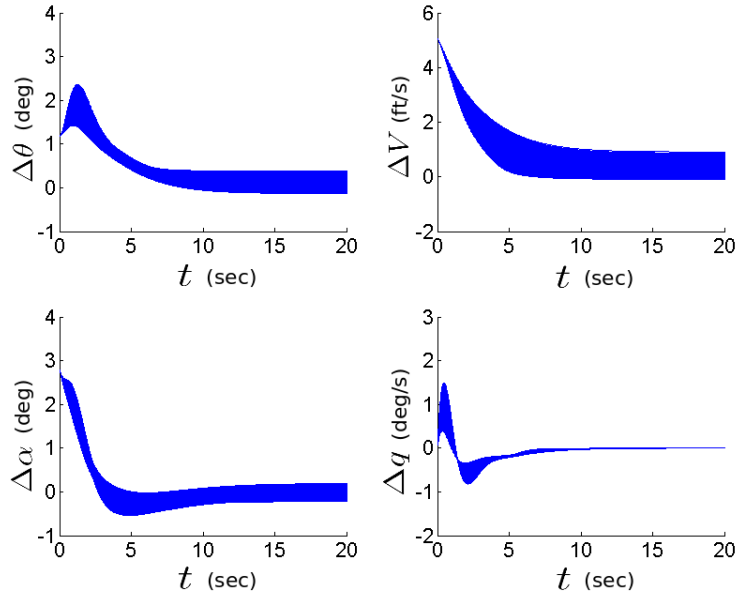


Figure 5.10: A schematic of how the support of a joint PDF evolves in the extended state space under parametric uncertainty. For ease of understanding, we illustrate here a case for one state x and one parameter p. Since x_0 is deterministic but p is random, the initial joint PDF ξ_0 is simply the univariate parametric PDF $\xi_p(p)$ translated to $x = x_0$. Consequently, ξ_0 is supported on a straight line segment (one dimensional subspace) in the two dimensional extended state space, as shown in the *left figure*. For $0 < t < \infty$, due to state dynamics, the samples (denoted as *circles*) on that line segment move in the horizontal (x) direction while keeping the respective ordinate (p) value constant, resulting the instantaneous support to be a curve (*middle figure*). If the system achieves regulation, then $\lim_{t\to\infty} x(t) = x_{\text{trim}}$, $\forall p$ in the parametric uncertainty set, resulting the asymptotic joint PDF ξ_∞ to be supported on a straight line segment (*right figure*) at $x = x_{\text{trim}}$.



(a) State error vs. time for LQR controller



(b) State error vs. time for gsLQR controller

Figure 5.11: MC state error ($\Delta x^j(t) \triangleq x^j(t) - x^j_{\text{trim}},\ j = 1, \ldots, 4$) trajectories for LQR and gsLQR closed-loop dynamics, with $\pm 2.5\%$ uniform uncertainties in $p = [m,\ x_{\text{cg}},\ J_{yy}]^\top$, i.e. $p = p_{\text{nominal}}(1 \pm \Delta\%)$, where $\Delta = 2.5$, and p_{nominal} values are listed in Table 5.1.

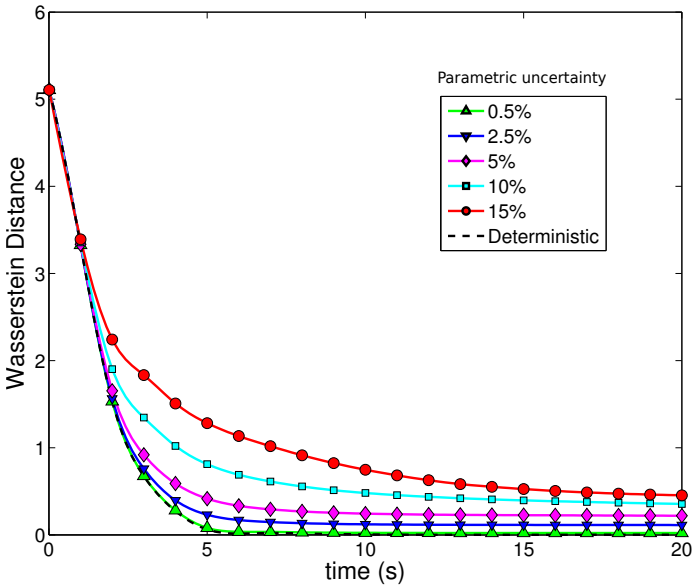
5.5.2.4 Optimal Transport Based Quantitative Analysis

Here, we solve the LP (4.12) with cost

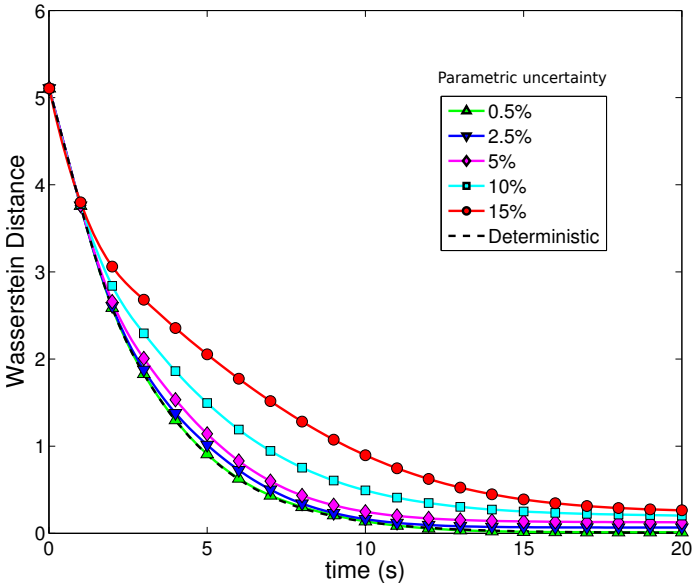
$$c_{ij} = n \sum_{i=1}^{n} \parallel x_i(t) - x_{\text{trim}} \parallel_2^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n_p=3} \left(p_k\left(i\right) - p_k\left(j\right) \right)^2, \tag{5.11}$$

with ς_i being the joint PMF value at the i^{th} sample location $\widetilde{x}_i(t) = [x_i(t),\ p(i)]^\top$, and $\widehat{\varsigma}_j$ being the trim joint PMF value at the j^{th} sample location $[x_{\text{trim}},\ p(j)]^\top$. Fig. 5.12(a) and 5.12(b) show $W\left(t\right)$ vs. t under parametric uncertainty for LQR and gsLQR, respectively. For both the controllers, the plots confirm that larger parametric uncertainty results in larger transport efforts at all times, causing higher value of W. In both cases, the deterministic (no uncertainty) W curves (*dashed lines* in Fig. 5.12(a) and 5.12(b)) almost coincide with those of $\pm 0.5\%$ parametric uncertainties. Notice that in the deterministic case, W is simply the Euclidian distance of the current state from trim, i.e. convergence in W reduces to the classical ℓ_2 convergence of a signal.

It is interesting to compare the LQR and gsLQR performance against parametric uncertainty for each fixed Δ. For $0-3$ s, the rate-of-convergence for $W(t)$ is faster for LQR, implying probabilistically faster regulation. However, the LQR W curves tend to flatten out after 3 s, thus slowing down its joint PDF's rate-of-convergence to ξ_∞. On the other hand, gsLQR W curves exhibit somewhat opposite trend. The initial regulation performance for gsLQR is slower than that of LQR, but gsLQR achieves better *asymptotic* performance by bringing the probability mass closer to ξ_∞ than the LQR case, resulting smaller values of W. Further, one may notice that for large ($\pm 15\%$) parametric uncertainties, the W curve for LQR shows a mild bump around 3 s, corresponding to the significant transient overshoot observed in Fig. 5.11(a).



(a) Time evolution of Wasserstein distance for LQR, with varying levels of Δ.



(b) Time evolution of Wasserstein distance for gsLQR, with varying levels of Δ.

Figure 5.12: Time evolution of Wasserstein distances for LQR and gsLQR, with parametric uncertainties.

This can be contrasted with the corresponding W curve for gsLQR, that does not show any prominent effect of transient overshoot at that time. The observation is consistent with the MC simulation results in Fig. 5.11(b). Thus, we can conclude that gsLQR is more robust than LQR, against parametric uncertainties.

5.5.3 Robustness Against Actuator Disturbance

5.5.3.1 Stochastic Initial Condition Uncertainty with Actuator Disturbance

Here, in addition to the initial condition uncertainties described in Section 5.5.1, we consider actuator disturbance in elevator. Our objective is to analyze how the additional disturbance in actuator affects the regulation performance of the controllers.

5.5.3.2 Simulation Set Up

We let the initial condition uncertainties to be described as in Table 5.2, and consequently the initial joint PDF is uniform. Further, we assume that the elevator is subjected to a periodic disturbance of the form $w(t) = 6.5 \sin(\Omega t)$. The simulation results of Section 5.1.1 corresponds to the special case when the forcing angular frequency $\Omega = 0$. To investigate how $\Omega > 0$ alters the system response, we first perform frequency-domain analysis of the LQR closed-loop system, linearized about x_{trim}. Fig. 5.13 shows the variation in singular value magnitude (in dB) with respect to frequency (rad/s), for the transfer array from disturbance $w(t)$ to states $x(t)$. This frequency-response plot shows that the *peak frequency* is $\omega^\star \approx 2$ rad/s.

5.5.3.3 Density Based Qualitative Analysis

To compare the LQR and gsLQR performance under peak frequency excitation (as per linearized LQR analysis), we set $\Omega = \omega^\star = 2$ rad/s, and evolve the initial uniform joint PDF over the LQR and gsLQR closed-loop state space. Notice that the LQR closed-loop dynamics is *nonlinear*, and the extent to which the linear analysis

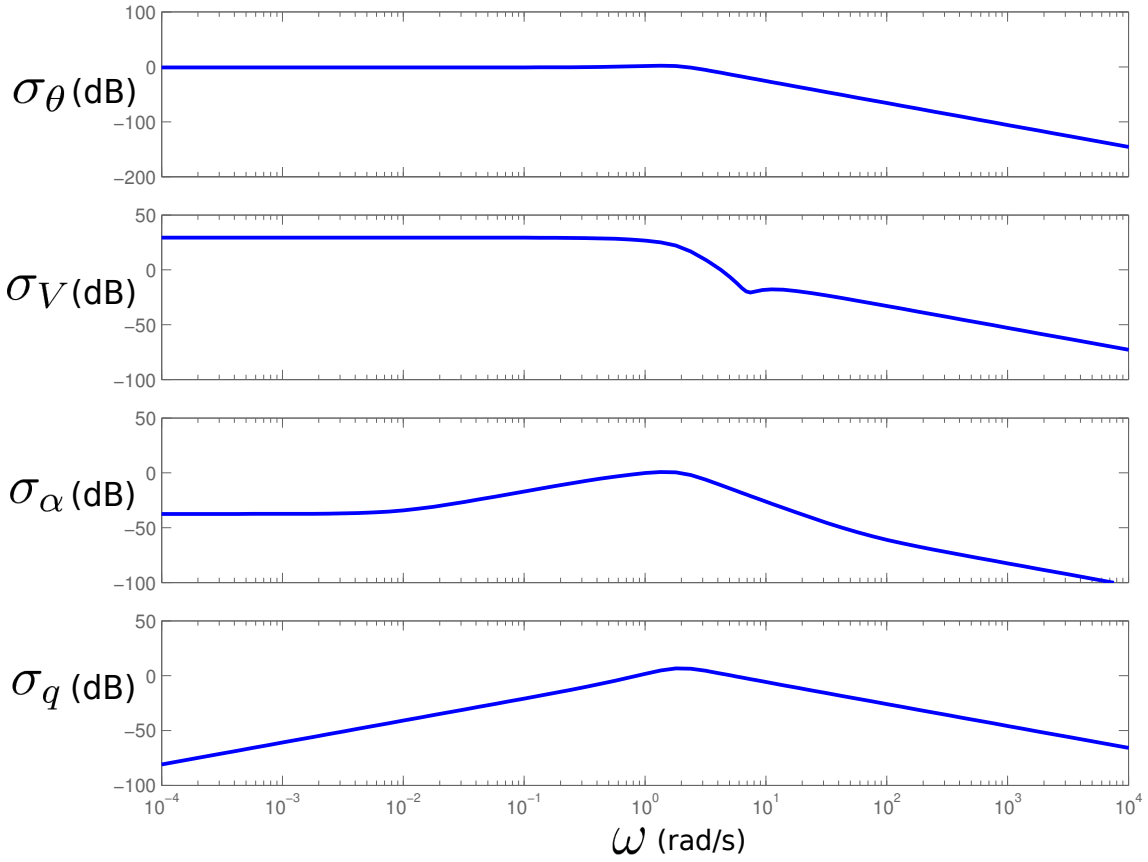


Figure 5.13: Singular values for the LQR closed-loop dynamics, linearized about x_{trim}, computed from the 4×1 transfer array corresponding to the disturbance to states.

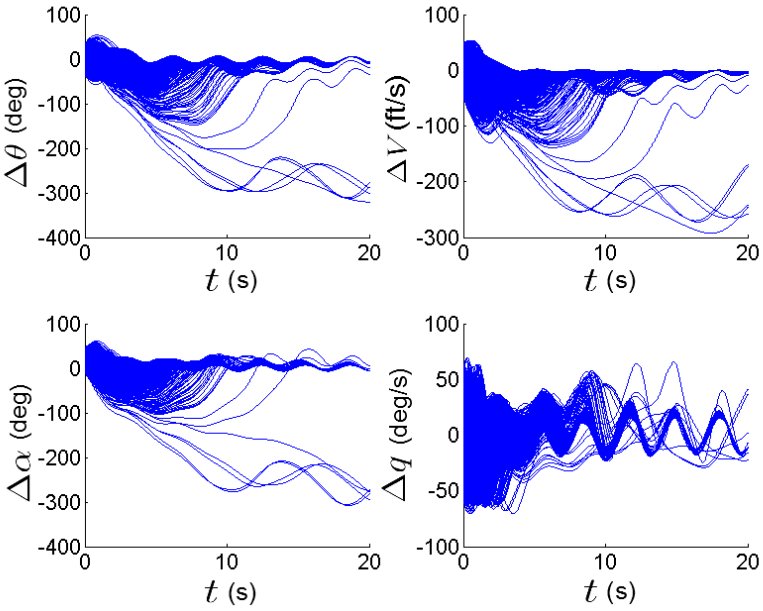
would be valid, depends on the robustness of regulation performance. Fig. 5.14(a) shows the LQR state error trajectories from the MC simulation. It can be observed that after $t = 10$ s, most of the LQR trajectories exhibit constant frequency oscillation with $\omega = 2$ rad/s. This trend is even more prominent for the gsLQR trajectories in Fig. 5.14(b), which seem to settle to the constant frequency oscillation quicker than the LQR case.

5.5.3.4 Optimal Transport Based Quantitative Analysis

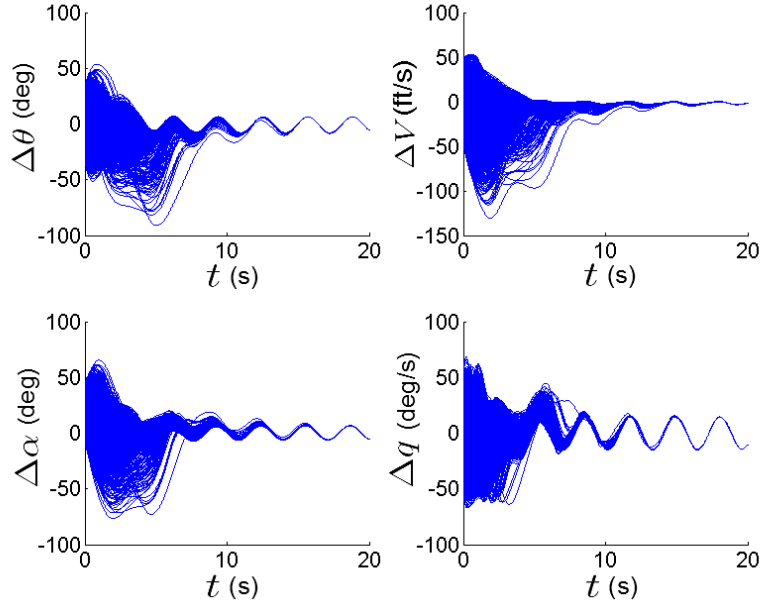
We now investigate the effect of elevator disturbance $w(t) = 6.5 \sin(2t)$ and initial condition uncertainties, via the optimal transport framework. In this case, the computation of Wasserstein distance is of the form (5.8).

For the LQR closed-loop system, Fig. 5.15(a) compares the Wasserstein distances for no actuator disturbance, i.e. $\Omega = 0$ rad/s (*circles*), actuator disturbances with $\Omega = 2$ rad/s (*solid line*) and $\Omega = 100$ rad/s (*dashed line*), respectively. It can be seen that the Wasserstein curves for $\Omega = 0$ rad/s and $\Omega = 100$ rad/s are indistinguishable, meaning the LQR closed-loop *nonlinear* system rejects high frequency elevator disturbance, similar to the *linearized* closed-loop system, as observed in Fig. 5.13. For $\Omega = 2$ rad/s, the Wasserstein curve reflects the effect of closed-loop nonlinearity in joint PDF evolution till approximately $t = 10$ s. For $t > 10$ s, we observe that the LQR Wasserstein curve itself settles to an oscillation with $\omega = 2$ rad/s. This is due to the fact that by $t = 10$ s, the joint probability mass comes so close to x_{trim}, that the linearization about x_{trim} becomes a valid approximation of the closed-loop nonlinear dynamics. This observation is consistent with the MC simulations in Fig. 5.14(a).

For the gsLQR closed-loop system, Fig. 5.15(b) compares the Wasserstein distances with $\Omega = 0, 2, 100$ rad/s. It is interesting to observe that, similar to the LQR

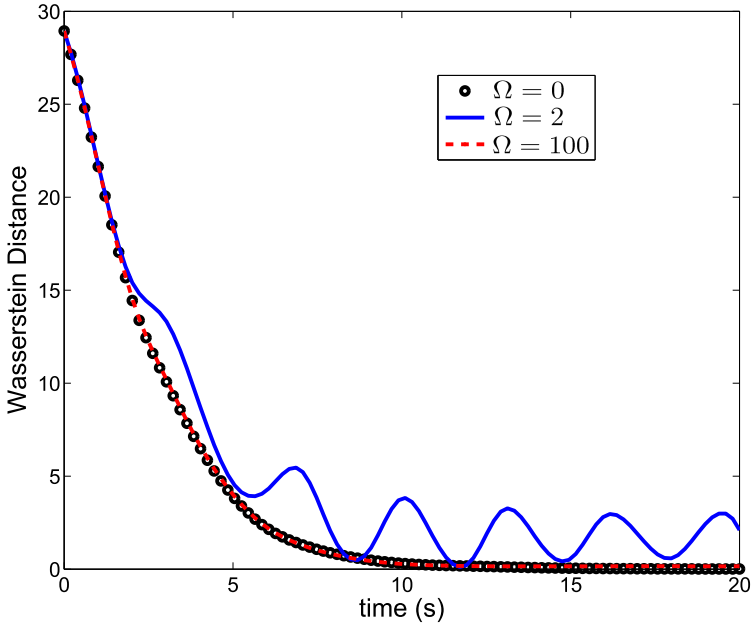


(a) State error vs. time for LQR controller

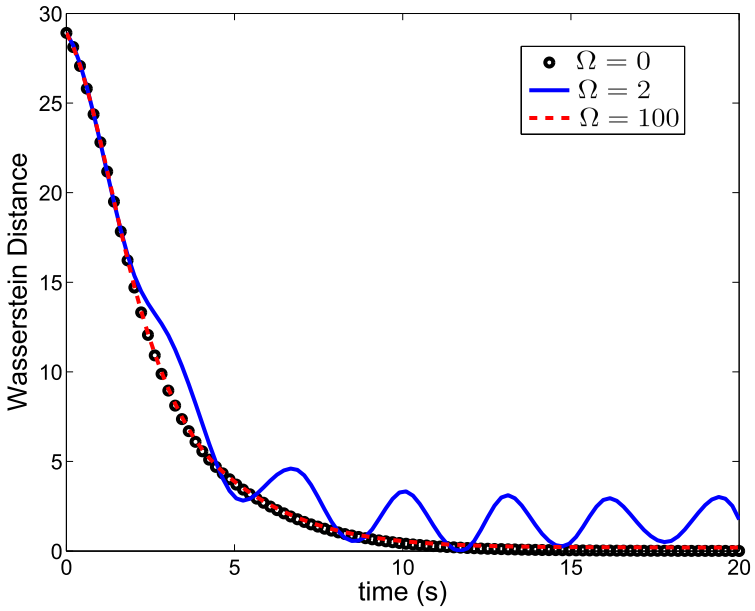


(b) State error vs. time for gsLQR controller

Figure 5.14: MC state error ($\Delta x^j(t) \triangleq x^j(t) - x^j_{\text{trim}}$, $j = 1, \ldots, 4$) trajectories for LQR and gsLQR closed-loop dynamics, with periodic disturbance $w(t) = 6.5 \sin(2t)$ in the elevator, and initial condition uncertainties.



(a) Time evolution of Wasserstein distance for LQR, with elevator disturbance $w(t) = 6.5 \sin(\Omega t)$.



(b) Time evolution of Wasserstein distance for gsLQR, with elevator disturbance $w(t) = 6.5 \sin(\Omega t)$.

Figure 5.15: Time evolution of Wasserstein distance for LQR and gsLQR, with initial condition uncertainties and elevator disturbance.

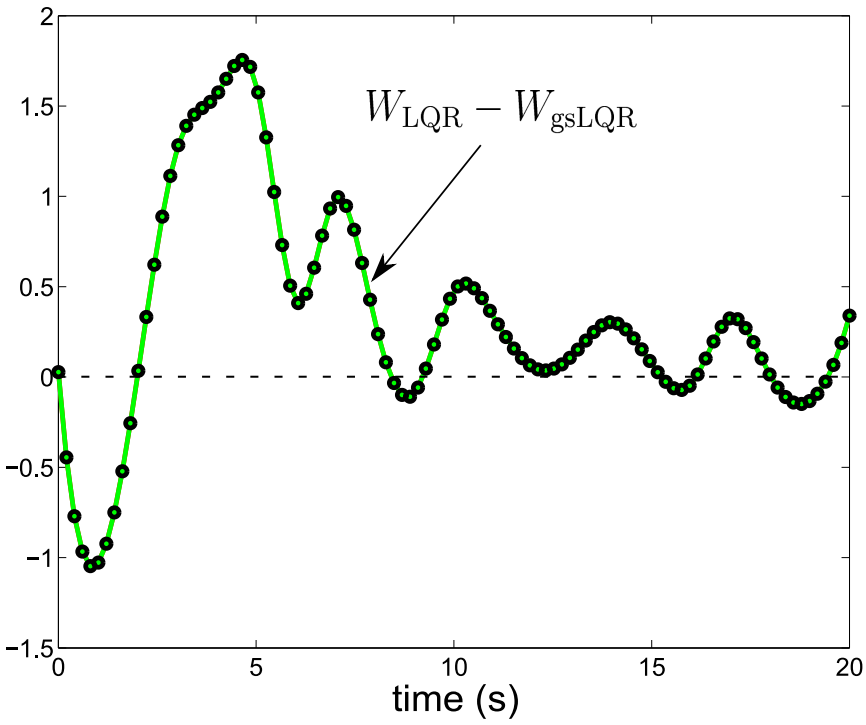


Figure 5.16: Time history of the difference between W_{LQR} and W_{gsLQR}, with initial condition uncertainties and elevator disturbance with $\Omega = 2$ rad/s.

case, gsLQR closed loop system rejects the high frequency elevator disturbance, and hence the Wasserstein curves for $\Omega = 0$ rad/s and $\Omega = 100$ rad/s look almost identical. Further, beyond $t = 10$ s, the gsLQR closed-loop response is similar to the LQR case, and hence the respective Wasserstein curves have similar trends. However, if we compare the LQR and gsLQR Wasserstein curves for $\Omega = 2$ rad/s, then we observe that gsLQR transient performance is slightly more robust than LQR, resulting lower values of Wasserstein distance for approximately $3 - 5$ seconds. This transient performance difference between LQR and gsLQR, can also be seen in Fig. 5.16 that shows the time evolution of $W_{\text{LQR}} - W_{\text{gsLQR}}$.

5.6 Chapter Summary

This Chapter is a case study to demonstrate the efficacy of the proposed density based V&V method. The case study concerns the controller performance verification for F-16 longitudinal flight dynamics. The objective is to ascertain which, among two given controllers, one LQR and another gain-scheduled LQR, has better regulation performance in the presence of probabilistic uncertainties in initial conditions, parameters and actuator disturbance. Since the open-loop flight dynamics is nonlinear, this amounts to verifying the controllers' nonlinear robustness of the state regulation performance. It is shown that the *quantitative* results obtained by the proposed V&V framework, are in unison with the *qualitative* predictions made by Monte Carlo and the uncertainty propagation algorithm developed in Chapter 2.

6. SYSTEMS THEORETIC RESULTS FOR PROBABILISTIC V&V

In the previous chapters, we have developed a density based model validation framework in time domain. The computational aspects have been covered in detail, and several examples and case studies have been worked out to demonstrate its applicability for nonlinear dynamical systems. In this chapter, we focus on investigating the systems-theoretic interpretations of the proposed V&V framework. In particular, we would like to understand the meaning of two *linear systems* being close in Wasserstein sense. This chapter provides two results in this direction. First, we consider a deterministic LTI system and its approximation (another deterministic LTI system), both being excited by Gaussian white noise (GWN). Then, we show that the asymptotic Wasserstein gap can be characterized in terms of the average gains of the systems under comparison. Next, we consider the linear model discrimination problem, i.e. finding the initial PDF which maximizes the Wasserstein gap between the systems under comparison. We provide geometric intuitions for the solutions of both these problems.

6.1 Wasserstein Gap Between Deterministic Linear Systems Driven By Gaussian White Noise

6.1.1 Frequency Domain Formulae for Asymptotic $_2W_2$

In this chapter, the asymptotic Wasserstein distance $\lim_{t\to\infty} {}_2W_2(t)$ is denoted as W_∞. We next derive formulae for W_∞ for the single input single output (SISO), multiple input single output (MISO), and multiple input multiple output (MIMO) LTI systems.

6.1.1.1 SISO Case

Theorem 12. *Consider two stable LTI systems with transfer functions G_i, $i = 1, 2$, both excited by stationary Gaussian input $u(t) \sim \mathcal{N}(\mu_u, \sigma_u^2)$, with power spectral density (PSD) $S_u(\omega)$. Then the asymptotic Wasserstein gap W_∞ between them, is given by*

$$W_\infty(G_1, G_2) = \sqrt{\mu_u^2 \left(G_1(0) - G_2(0)\right)^2 + \left(\sigma_1 - \sigma_2\right)^2}, \tag{6.1}$$

with $\sigma_i = \sqrt{\int_{-\infty}^{+\infty} |G_i(j\omega)|^2 S_u(\omega)\, d\omega - \mu_u^2 G_i^2(0)}$, $i = 1, 2$.

Proof. Let $y_i(t)$, $i = 1, 2$, denote the sample paths of the respective stationary outputs, with densities $\mathcal{N}(\mu_i, \sigma_i^2)$. If $R_{y_i}(\tau)$ are the respective autocorrelations with τ being the corresponding correlation window, then stationarity implies

$$\begin{aligned} & R_{y_i}(\tau) = \mathbb{E}\left[y_i(t+\tau)\, y_i(t)\right] \Rightarrow R_{y_i}(0) = \mathbb{E}\left[\left(y_i(t)\right)^2\right] \\ \Rightarrow \quad & \int_{-\infty}^{+\infty} S_{y_i}(\omega)\, d\omega = \int_{-\infty}^{+\infty} |G_i(j\omega)|^2 S_u(\omega)\, d\omega = \mu_i^2 + \sigma_i^2, \end{aligned} \tag{6.2}$$

where $S_{y_i}(\omega)$ are the auto-PSDs, defined as the inverse Fourier transform of the respective auto-correlations, and we have used the well-known [184] SISO PSD relation $S_{y_i}(\omega) = |G_i(j\omega)|^2 S_u(\omega)$. On the other hand,

$$\mu_i = \mathbb{E}\left[y_i(t)\right] = \mathbb{E}\left[\int_{-\infty}^{+\infty} h_i(\tau)\, u(t-\tau)\, d\tau\right],$$

where $h_i(.)$ are the respective impulse response functions. By changing the order of integrals, and using the fact that $\mathbb{E}\left[u(t-\tau)\right] = \mathbb{E}\left[u(t)\right]$ (due to stationarity), we get

$$\mu_i = \mu_u \int_{-\infty}^{+\infty} h_i(\tau)\, e^{-j.0.\tau} d\tau = \mu_u G_i(0). \tag{6.3}$$

Notice that (6.2) and (6.3) yield four equations in four unknowns μ_i, σ_i, $i = 1, 2$. Since Wasserstein distance between two univariate Gaussians $\mathcal{N}(\mu_i, \sigma_i^2)$ is given by [138]

$$W_\infty = \sqrt{(\mu_1 - \mu_2)^2 + (\sigma_1 - \sigma_2)^2}, \tag{6.4}$$

the result follows from (6.2) and (6.3). The input PSD $S_u(\omega)$ is a function of σ_u, depending on the temporal properties of the stochastic process $u(t)$. ■

Corollary 13. *If $u(t)$ is Gaussian* white *noise, then $\mu_u = 0$, $S_u(\omega) = \sigma_u^2 = $ constant. Consequently, W_∞* is the difference between the respective $\mathcal{H}_2$ norms*, up to scaling by the strength of input disturbance, i.e.*

$$W_\infty = \sqrt{2\pi}\sigma_u \Big| \, ||G_1(j\omega)||_2 - ||G_2(j\omega)||_2 \, \Big|. \tag{6.5}$$

The proof follows immediately from the definition of SISO $\mathcal{H}_2$ norm: $||G_i||_2 := \sqrt{\frac{1}{2\pi}\int_{-\infty}^{+\infty} G_i^*(j\omega)\, G_i(j\omega)\, d\omega}$, $i = 1, 2$. Notice that, in the definitions of Fourier transform pairs auto-correlation and PSD, the factor $\frac{1}{\sqrt{2\pi}}$ is usually omitted in the signal processing community [184, 104], and we have adopted the same convention in proving Theorem 12. Thus, by scaling the variance of the input noise, one can normalize the factor $\sqrt{2\pi}\sigma_u$ in (6.5), a condition we will assume in most derivations without loss of generality.

In the remaining of this section, we will derive the results assuming the input to be Gaussian white noise. Given the input auto and cross-PSDs, how to handle more general cases, will become apparent from the proofs.

6.1.1.2 MISO Case

Theorem 14. *Consider two stable LTI systems with m inputs and single output, having transfer arrays G and $\widehat{G}$, each being a row vector of size $1 \times m$. If both the systems are excited by Gaussian white noise vector $u(t) \sim \mathcal{N}(0_{m\times 1}, diag(\sigma_u^2))$, then the asymptotic Wasserstein gap W_∞ between them, is given by the scaled difference between respective $\mathcal{H}_2$ norms:*

$$W_\infty\left(G, \widehat{G}\right) = \sqrt{2\pi}\sigma_u \Big|\, ||G(j\omega)||_2 - ||\widehat{G}(j\omega)||_2 \,\Big|. \tag{6.6}$$

Proof. Like the SISO proof, we still have $\int_{-\infty}^{+\infty} S_y(\omega)\, d\omega = \sigma^2$, since $\mu = \widehat{\mu} = 0$, due to whiteness of the input. This equation holds for both pairs $(S_y(\omega), \sigma^2)$ and $(S_{\widehat{y}}(\omega), \widehat{\sigma}^2)$. For the general correlated stationary input, the MISO PSD relation is known [184] to be

$$S_y(\omega) = \sum_{i=1}^{m}\sum_{k=1}^{m} G_i^*(j\omega)\, G_k(j\omega) \quad \underbrace{S_{u_iu_k}(\omega)}_{\text{input PSD matrix}} \quad d\omega. \tag{6.7}$$

Now, for white input vector, each dimension is an independent white noise process, implying the dimensions are mutually uncorrelated. Hence, $R_{u_iu_k}(\tau), S_{u_iu_k}(\omega) = 0$, $\forall i \neq k$. Thus, for Gaussian white vector $u(t) \sim \mathcal{N}(0_{m\times 1}, \text{diag}(\sigma_i^2))$, (6.4) results

$$W = \left| \sqrt{\int_{-\infty}^{+\infty} \sum_{i=1}^{m} |G_i(j\omega)|^2 \sigma_{u_i}^2 d\omega} - \sqrt{\int_{-\infty}^{+\infty} \sum_{i=1}^{m} |\widehat{G}_i(j\omega)|^2 \sigma_{u_i}^2 d\omega} \right|$$

which reduces to (6.6) for *spherical* Gaussian case, since the $\mathcal{H}_2$ norm for multivariate case, is defined as $||G||_2 := \sqrt{\dfrac{1}{2\pi}\displaystyle\int_{-\infty}^{+\infty} \text{tr}\left(G^{\text{H}}(j\omega)\, G(j\omega)\right) d\omega}$. ∎

6.1.1.3 MIMO Case

Theorem 15. *Consider two stable LTI systems with m inputs and p outputs, having transfer matrices G and $\widehat{G}$. If both the systems are excited by Gaussian white noise vector $u(t) \sim \mathcal{N}(0_{m\times 1}, diag(\sigma_u^2))$, then the asymptotic Wasserstein gap W_∞ between them, is given by*

$$\begin{aligned} & W_\infty\left(G, \widehat{G}\right) \\ = \; & \sqrt{2\pi}\sigma_u \left(||G(j\omega)||_2^2 + ||\widehat{G}(j\omega)||_2^2 - 2\,\mathrm{tr}\left[\left(\frac{1}{2\pi}\int_{-\infty}^{+\infty} G^{\mathrm{H}}(j\omega)\, G(j\omega)\, d\omega \right)^{1/2} \right.\right. \\ & \left.\left. \left(\frac{1}{2\pi}\int_{-\infty}^{+\infty} \widehat{G}^{\mathrm{H}}(j\omega)\, \widehat{G}(j\omega)\, d\omega \right) \left(\frac{1}{2\pi}\int_{-\infty}^{+\infty} G^{\mathrm{H}}(j\omega)\, G(j\omega)\, d\omega \right)^{1/2} \right]^{1/2} \right)^{1/2}. \end{aligned} \tag{6.8}$$

Proof. In this case, the output correlation matrices of size $p \times p$ satisfy $\mathbf{R}_y(0) = \int_{-\infty}^{+\infty} \mathbf{S}_y(\omega)\, d\omega = \Sigma_y$. Similar equation holds for the "hat" system.

Let the $(i,k)^{\text{th}}$ elements $G_{ik}, \widehat{G}_{ik}$, of the transfer matrices relate the response from i^{th} input to the respective k^{th} output. Then the MIMO PSD relation [185] becomes

$$\mathbf{S}_y(\omega) \;=\; G^{\mathrm{H}}(j\omega) \quad \underbrace{\mathbf{S}_u(\omega)}_{\text{Input PSD matrix}} \quad G(j\omega), \tag{6.9}$$

where for Gaussian white input, we have

$$\mathbf{S}_u(\omega) = \sigma_u^2 I_{m\times m} \Rightarrow \mathbf{S}_y(\omega) = \sigma_u^2 G^{\mathrm{H}}(j\omega)\, G(j\omega). \tag{6.10}$$

Thus, $\Sigma_y = \sigma_u^2 \int_{-\infty}^{+\infty} G^{\mathrm{H}}(j\omega)\, G(j\omega)\, d\omega$. Hence, $\mathrm{tr}(\Sigma_y) = 2\pi\sigma_u^2 ||G||_2^2$; $\mathrm{tr}(\Sigma_{\widehat{y}}) = 2\pi\sigma_u^2 ||\widehat{G}||_2^2$. Since $\mu_y = \mu_{\widehat{y}} = 0$, the Wasserstein distance between $y(t) \sim \mathcal{N}(\mu_y, \Sigma_y)$

and $\widehat{y}(t) \sim \mathcal{N}(\mu_{\widehat{y}}, \Sigma_{\widehat{y}})$ is given by [138]

$$W = \sqrt{\operatorname{tr}(\Sigma_y) + \operatorname{tr}(\Sigma_{\widehat{y}}) - 2\operatorname{tr}\left[\sqrt{\Sigma_y}\Sigma_{\widehat{y}}\sqrt{\Sigma_y}\right]^{1/2}}. \tag{6.11}$$

Substituting for Σ_y and $\Sigma_{\widehat{y}}$ in (6.11), in terms of G and $\widehat{G}$, we arrive at (6.8). ■

6.1.2 Bounds for MIMO Case

Following results provide simpler and easier-to-interpret bounds for W_∞ in the MIMO case.

6.1.2.1 Lower Bound

Lemma 4. *Given stable transfer matrices G and $\widehat{G}$, the asymptotic Wasserstein distance for MIMO case, is lower bounded by the corresponding expression for SISO or MISO case, i.e.* $\left|||G||_2 - ||\widehat{G}||_2\right| \leqslant W_\infty$.

Proof. Since the stationary covariances Σ_y and $\Sigma_{\widehat{y}}$ are positive semi-definite, they satisfy (p. 527, Fact 8.12.20, [186])

$$\operatorname{tr}\left(\sqrt{\Sigma_y}\Sigma_{\widehat{y}}\sqrt{\Sigma_y}\right)^{1/2} \leqslant \sqrt{\operatorname{tr}(\Sigma_y)\operatorname{tr}(\Sigma_{\widehat{y}})}. \tag{6.12}$$

Since $\operatorname{tr}(\Sigma_y) = ||G||_2^2$ and $\operatorname{tr}(\Sigma_{\widehat{y}}) = ||\widehat{G}||_2^2$, we get

$$\underbrace{\operatorname{tr}(\Sigma_y) + \operatorname{tr}(\Sigma_{\widehat{y}}) - 2\operatorname{tr}\left(\sqrt{\Sigma_y}\Sigma_{\widehat{y}}\sqrt{\Sigma_y}\right)^{1/2}}_{W_\infty^2} \geqslant \left(||G||_2 - ||\widehat{G}||_2\right)^2. \tag{6.13}$$

Hence the result. ■

6.1.2.2 Upper Bound

Lemma 5. *If Σ_y and $\Sigma_{\widehat{y}}$ are the stationary output covariance matrices corresponding to stable transfer matrices G and $\widehat{G}$ respectively, then we have the following upper*

bound for MIMO Wasserstein distance: $W \leqslant \left|\left|\sqrt{\Sigma_y} - \sqrt{\Sigma_{\widehat{y}}}\right|\right|_F$.

Proof. It is known (Fact 8.19.21, [186]) that for $0 \leqslant p \leqslant 1$, we have

$$\operatorname{tr}\left(\Sigma_y^p \Sigma_{\widehat{y}}^p\right) \leqslant \operatorname{tr}\left(\Sigma_{\widehat{y}}^{1/2} \Sigma_y \Sigma_{\widehat{y}}^{1/2}\right)^p. \tag{6.14}$$

Setting $p = \frac{1}{2}$ results

$$\operatorname{tr}\left(\sqrt{\Sigma_y}\sqrt{\Sigma_{\widehat{y}}}\right) \leqslant \left(\sqrt{\Sigma_{\widehat{y}}}\Sigma_y\sqrt{\Sigma_{\widehat{y}}}\right)^{\frac{1}{2}} = \left(\sqrt{\Sigma_y}\Sigma_{\widehat{y}}\sqrt{\Sigma_y}\right)^{\frac{1}{2}}, \tag{6.15}$$

where the last equality follows from the symmetry of Wasserstein distance, and can be separately proved by noting that $\operatorname{tr}\left(\sqrt{MM^\top}\right) = \operatorname{tr}\left(\sqrt{M^\top M}\right)$ for $M = \sqrt{\Sigma_y}\sqrt{\Sigma_{\widehat{y}}}$. Since $\operatorname{tr}(\Sigma_y) = ||G||_2^2$ and $\operatorname{tr}(\Sigma_{\widehat{y}}) = ||\widehat{G}||_2^2$, we obtain

$$W \leqslant \underbrace{\left(\operatorname{tr}(\Sigma_y) + \operatorname{tr}(\Sigma_{\widehat{y}}) - 2\operatorname{tr}\left(\sqrt{\Sigma_y}\sqrt{\Sigma_{\widehat{y}}}\right)\right)^{1/2}}_{||\sqrt{\Sigma_y} - \sqrt{\Sigma_{\widehat{y}}}||_F}. \tag{6.16}$$

■

6.1.3 Sensitivity of W_∞ in Frequency Domain

One may interpret the frequency domain formulae for W derived above, as the difference between average gains of the two systems. Following are some observations regarding the same.

6.1.3.1 Scaling

W_∞ is sensitive to scaling. For example, if we set $\widehat{G} = kG$, where k is some non-zero scaling constant, then (6.5), (6.6) and (6.8) yields $W_\infty = |1 - k|\ ||G||_2$ (assuming $\sigma_u = 1/\sqrt{2\pi}$). Thus a linear relative amplification between two stable LTI systems, results a linear amplification of the Wasserstein gap. This can be

contrasted with some recent works [30, 29] in the literature, on defining gap between dynamical systems, where a gap was shown to be either insensitive [30] to scaling, or a nonlinear function [29] of the scaling constant. As pointed out in [187], which one is a desirable property depends on the application.

6.1.3.2 Minimum versus Non-minimum Phase Systems

The frequency domain expressions for W_∞ depend only on the *magnitudes* of transfer functions. Thus, if we compare stable transfer functions of the form $G_\pm = \frac{\prod_{i=1}^{n_z}(s \pm z_i)}{\prod_{k=1}^{n_p}(s+p_k)}$ with $n_p > n_z$, and $\mathrm{Re}\,(p_k) > 0\ \forall k$, then $W_\infty(G_+, G_-) = 0$. This result is intuitively consistent (see the discussion at p. 1592 in [29]). However, the non-minimum phase zeros are related to $\mathcal{H}_2$ norms of the respective transfer functions via *Poisson-Jensen half-plane formula* (Appendix C.8.2 in [188]).

6.1.3.3 SISO Invariance Properties

Being a function of magnitudes only, the asymptotic Wasserstein distance, like the chordal metric [189], remains invariant under complex conjugate transformation, i.e. $W_\infty\left(G, \widehat{G}\right) = W\left(G^*, \widehat{G}^*\right)$. However, unlike chordal metric, $W_\infty\left(G, \widehat{G}\right) \neq W_\infty\left(\frac{1}{G}, \frac{1}{\widehat{G}}\right)$, in general.

6.1.4 Geometric Interpretation of the SISO Formula and Comparison with ν-gap Metric

The ν-gap metric [190, 191, 192], was introduced as an important tool for linear model validation with good robustness properties. It is natural to ask how the Wasserstein distance, proposed in time domain [41, 45] for both linear and nonlinear systems, relate with ν-gap. Like the SISO ν-gap, we look for a geometric interpretation of the SISO formula (6.5), which may be helpful for comparison between the metrics.

6.1.4.1 SISO ν-gap and W_∞

Given two transfer functions G_1 and G_2, let $G_i = N_i M_i^{-1} = \widetilde{M}_i^{-1}\widetilde{N}_i$, $i = 1, 2$, denote the normalized right and left coprime factorizations [193], respectively. Further, let $\Gamma_i(s) := \{N_i(s) \quad M_i(s)\}^\top$, and $\widetilde{\Gamma}_i(s) := \{\widetilde{M}_i(s) \quad -\widetilde{N}_i(s)\}$. If $\text{wno}(\det(\Gamma_2^*(j\omega)\Gamma(j\omega))) \neq 0$, then the SISO ν-gap metric δ_ν is given by

$$\delta_\nu = \sup_{\omega \in \mathbb{R}\cup\{\infty\}} \frac{|G_1(j\omega) - G_2(j\omega)|}{\sqrt{1+|G_1(j\omega)|^2}\sqrt{1+|G_2(j\omega)|^2}}, \tag{6.17}$$

and lies between 0 to 1. When the winding number condition is not satisfied, then $\delta_\nu := 1$. Geometrically, ν-gap measures the largest chordal distance $\kappa(\omega)$ between the Nyquist plots of G_1 and G_2, projected on the Riemann sphere (Fig. 6.1). On the other hand, (6.5) can be geometrically interpreted as the difference between the lengths of the r.m.s. distances to the Nyquist plots of G_1 and G_2, measured from the origin (Fig. 6.1).

One difficulty in directly comparing W and δ_ν is that (6.17) is normalized, but (6.5) is not. Hence we can *either* compare (6.5) with the "un-normalized equivalent" of (6.17), *or* we can normalize (6.5) and then compare with (6.17). In the latter case, as of yet, it is not clear what should be the intrinsic normalization factor. However, as we show next, the geometric insight will guide us to answer both.

6.1.4.2 Comparison on the Complex Plane

Let $\kappa^{\text{proj}}(\omega)$ be the projection of $\kappa(\omega)$ to the extended complex plane. For given transfer functions G and $\widehat{G}$, $\sup_\omega \kappa^{\text{proj}}(\omega)$ is the largest pointwise distance between the two Nyquist plots, and can be taken as an "un-normalized analogue" of the largest normalized chord length δ_ν.

Theorem 16. *Given two stable LTI transfer functions G and $\widehat{G}$, the difference*

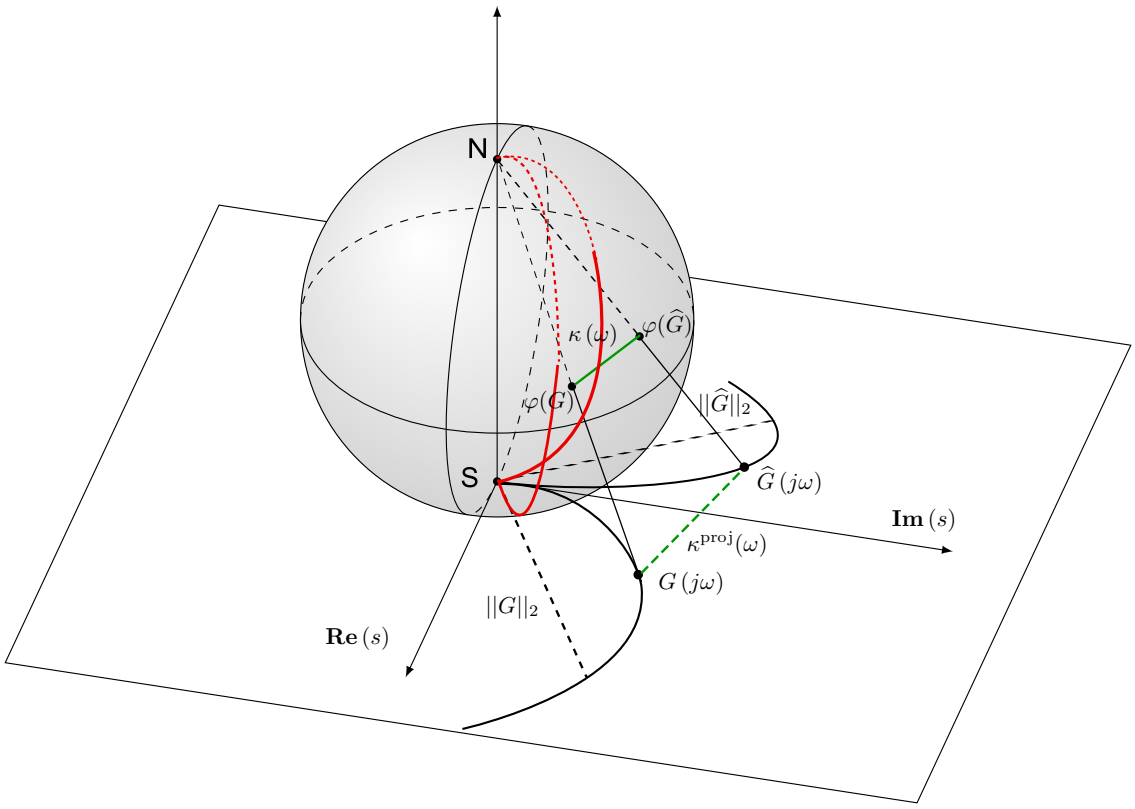


Figure 6.1: The stereographic projection $\varphi : \mathbb{C} \cup \{\infty\} \mapsto \mathcal{S}$, where $\mathcal{S}$ is the Riemann sphere of unit diameter, with south pole at the origin of the extended complex plane. $\kappa^{\mathrm{proj}}(\omega)$ (*dotted green*) is the projection of the chordal distance $\kappa(\omega)$ (*solid green*) to the plane. The *dotted black* lines on the plane denote the $\mathcal{H}_2$ norms, which measure the r.m.s. distances of the respective Nyquist plots (*solid black*) from the origin. The *solid red* arcs show the projections of the two $\mathcal{H}_2$ norms on the Riemann sphere. The projected W is the difference between these arc-lengths. The *dotted red* arcs show that values of the respective $\mathcal{H}_2$ norm projections can at most be $\frac{\pi}{2}$, resulting a normalization for projected W.

between their r.m.s. lengths, is upper bounded by the maximum projected chordal length, i.e. $\sup_{\omega} \kappa^{\text{proj}}(\omega) \geqslant W$.

Proof. The stereographic projection of a point $(x, y) := x + jy$ on the plane, to the point (ξ, η, ζ) on Riemann sphere, is given by [194]

$$\xi = \frac{x}{1 + x^2 + y^2}, \quad \eta = \frac{y}{1 + x^2 + y^2}, \quad \zeta = \frac{x^2 + y^2}{1 + x^2 + y^2}. \tag{6.18}$$

From Fig. 6.1, $NG = \sqrt{1 + |G|^2}$, $N\widehat{G} = \sqrt{1 + |\widehat{G}|^2}$. Since $\frac{N\varphi(G)}{NG} = \frac{1 - \zeta}{1} = \frac{1}{1 + |G|^2}$, we have $N\varphi(G) = (1 + |G|^2)^{-1/2}$, and $N\varphi(\widehat{G}) = \left(1 + |\widehat{G}|^2\right)^{-1/2}$. Further, $N\varphi(G)\, NG = N\varphi(\widehat{G})\, N\widehat{G} = 1$ implies that triangles $N\varphi(G)\varphi(\widehat{G})$ and $NG\widehat{G}$ are similar. Consequently, $\frac{\kappa(\omega)}{\kappa^{\text{proj}}(\omega)} = \frac{N\varphi(\widehat{G})}{NG} = \frac{1}{(1 + |G|^2)^{1/2} \left(1 + |\widehat{G}|^2\right)^{1/2}}$. Thus, $\kappa^{\text{proj}}(\omega) = |G - \widehat{G}|$. Now, notice that

$$\sup_{\omega} \kappa^{\text{proj}}(\omega) = ||G - \widehat{G}||_{\infty} \geqslant ||G - \widehat{G}||_2 \geqslant |\, ||G||_2 - ||\widehat{G}||_2\, |,$$

where the last step is the reverse triangle inequality. This completes the proof. ■

6.1.4.3 Comparison on the Riemann Sphere

To make a stereographic projection of Wasserstein distance onto the Riemann sphere, we first consider projections of the individual r.m.s. lengths given by the $\mathcal{H}_2$ norms of G and $\widehat{G}$. The following lemma is relevant in this regard.

Lemma 6. *(p. 40, Theorem 2.5.1, [195]) Under stereographic projection, circlines in the complex plane get projected to circles on the Riemann sphere and vice versa. For straight lines on the plane, the corresponding circles on Riemann sphere pass through the north pole.*

Corollary 17. *The stereographic projection of $\mathcal{H}_2$ norm of any LTI transfer function can be of length at most $\frac{\pi}{2}$.*

Proof. From Lemma 6, all straight lines on the complex plane, passing through origin, must go through both north and south poles, i.e. will be meridians on Riemann sphere. Points on such straight lines, situated infinite extent away from the origin, under stereographic projection, approach the north pole from both sides of the Riemann sphere. Thus, a ray on the complex plane, with fixed end at the origin, projects to half meridian of circumference at most $\frac{\pi}{2}$. This completes the proof. Notice that, half-meridian arc length on the Riemann sphere will be exactly $\frac{\pi}{2}$ iff $\mathcal{H}_2$ norm is infinity, *either* due to unstable transfer function *or* due to non-zero feed-through. ■

Theorem 18. *Given two LTI transfer functions G and $\widehat{G}$, the normalized Wasserstein distance $\overline{W}_{\mathcal{S}}$ on the Riemann sphere $\mathcal{S}$, is given by*

$$\overline{W}_{\mathcal{S}}\left(G, \widehat{G}\right) = \frac{2}{\pi}\Big|\ \arctan||G||_2 - \arctan||\widehat{G}||_2\ \Big|. \tag{6.19}$$

Proof. If the $\mathcal{H}_2$ norm of the transfer function is finite, the stereographic projection maps a line segment on complex plane (Fig. 6.1) to an arc of the half-meridian. The infinitesimal lengths on the plane and on the sphere relate [195] by

$$\frac{dl_{\mathcal{S}}}{dl_{\mathbb{C}\cup\infty}} = \frac{1}{1+x^2+y^2}. \tag{6.20}$$

Taking infinitesimal elements dr and $d\widehat{r}$ along the $\mathcal{H}_2$ norms (Fig. 6.1), (6.20) yields the half-meridial arc lengths

$$s \ = \ \int_0^{||G||_2} \frac{dr}{1+r^2} = \arctan||G||_2, \tag{6.21}$$

$$\widehat{s} \;=\; \int_0^{||\widehat{G}||_2} \frac{d\widehat{r}}{1+\widehat{r}^2} = \arctan||\widehat{G}||_2. \tag{6.22}$$

Since $0 \leqslant ||G||_2, ||\widehat{G}||_2 \leqslant \infty \Rightarrow 0 \leqslant s, \widehat{s} \leqslant \frac{\pi}{2}$, therefore $|s-\widehat{s}| \leqslant \frac{\pi}{2}$. Assuming the scaling $\sqrt{2\pi}\sigma_u = 1$ in (6.5), we get the projected Wasserstein distance $W_{\mathcal{S}} = |s-\widehat{s}| = \big|\ \arctan||G||_2 - \arctan||\widehat{G}||_2\ \big|$, which can be normalized by $\frac{\pi}{2}$ to result $\overline{W}_{\mathcal{S}}$. Notice that, either of the $\mathcal{H}_2$ norms can be infinity. ■

The following theorem provides an indirect comparison between $\overline{W}_{\mathcal{S}}$ and δ_ν. It presents only sufficiency condition.

Theorem 19. *Given stable LTI transfer functions G and $\widehat{G}$, let P and $\widehat{P}$ be the points on the respective Nyquist plots corresponding to their $\mathcal{H}_2$ norms. Let $\gamma_2 := \angle NPS$, $\widehat{\gamma}_2 := \angle N\widehat{P}S$. Similarly, define angles $\gamma(\omega)$ and $\widehat{\gamma}(\omega)$, for generic points $G(j\omega)$ and $\widehat{G}(j\omega)$. If $\big|\ ||\cos\gamma\sin\widehat{\gamma}||_2 - ||\sin\gamma\cos\widehat{\gamma}||_2\ \big| \geqslant \frac{2}{\pi}\big|\gamma_2 - \widehat{\gamma}_2\big|$, then $\delta_\nu \geqslant \overline{W}_{\mathcal{S}}$.*

Proof. We observe that

$$\begin{aligned}
\delta_\nu = \sup_\omega \frac{|G-\widehat{G}|}{\sqrt{1+|G|^2}\sqrt{1+|\widehat{G}|^2}} &= \Big|\Big|\frac{|G-\widehat{G}|}{\sqrt{1+|G|^2}\sqrt{1+|\widehat{G}|^2}}\Big|\Big|_\infty, \\
&\geqslant \Big|\Big|\frac{|G-\widehat{G}|}{\sqrt{1+|G|^2}\sqrt{1+|\widehat{G}|^2}}\Big|\Big|_2, \\
\geqslant \Big|\ \Big|\Big|\frac{|G|}{\sqrt{1+|G|^2}\sqrt{1+|\widehat{G}|^2}}\Big|\Big|_2 &- \Big|\Big|\frac{|\widehat{G}|}{\sqrt{1+|G|^2}\sqrt{1+|\widehat{G}|^2}}\Big|\Big|_2\ \Big|,
\end{aligned} \tag{15}$$

where the last step follows from triangle difference inequality for $\mathcal{H}_2$ norms.

Notice that, γ_2 and $\widehat{\gamma}_2$ are the angles subtended by the respective $\mathcal{H}_2$ norms with its stereographic projections. Clearly, $0 \leqslant \gamma_2, \widehat{\gamma}_2 \leqslant \frac{\pi}{2}$, and we can rewrite (6.19) as $\frac{2}{\pi}\big|\gamma_2 - \widehat{\gamma}_2\big|$. Likewise, we define the running angles γ and $\widehat{\gamma}$ as functions of ω,

associated with points $G(j\omega)$ and $\widehat{G}(j\omega)$ (see Fig. 6.1). Further, notice that

$$\cos\gamma = \frac{|G|}{\sqrt{1+|G|^2}}, \quad \sin\widehat{\gamma} = \frac{1}{\sqrt{1+|\widehat{G}|^2}}, \tag{6.23}$$

$$\cos\widehat{\gamma} = \frac{|\widehat{G}|}{\sqrt{1+|\widehat{G}|^2}}, \quad \sin\gamma = \frac{1}{\sqrt{1+|G|^2}}, \tag{6.24}$$

and consequently, (6.23) can be written as the *difference between the r.m.s. values* of $\cos\gamma\sin\widehat{\gamma}$ and $\sin\gamma\cos\widehat{\gamma}$. Hence the result. ■

6.2 Linear Model Discrimination in Wasserstein Metric

In predictive modeling applications like systems biology, an important problem is of *model discrimination* [196, 197], where one looks for an initial PDF that *maximizes* the gap between two models, which seem to exhibit comparable performance. The motivation for model discrimination comes from the fact that the presence of uncertainties "mask" the difference between a family of completing models *or* between model prediction and physical observation. Thus, the idea is similar to optimal input design for system identification.

The main result of this section is that the initial PDF that maximizes Wasserstein distance, depends on the model and true dynamics. In particular, we show that for a linear dynamics pair, the gap is oblivious beyond the first two moments of ξ_0. We restrict ourselves to scalar dynamics for this analysis.

6.2.1 Tools for Analysis

Definition 5. *(**Quantile function**) Consider the probability space $(\Omega_y, \mathscr{F}, \mathbb{P})$ for the output random variable Y. Let $y := Y(\omega_y)$, for $\omega_y \in \Omega_y$. The quantile function*

$Q_y : \Omega_y \mapsto [0,1]$, *is defined as the generalized inverse of the CDF for* Y, *i.e.*

$$Q_y(\varsigma) := \inf\left(y \in \Omega_y : \varsigma \leq \mathbb{P}(Y \leq y)\right). \tag{6.25}$$

Here $\varsigma \in [0,1]$ *denotes probability mass.*

Proposition 5. *(**Quantile transport PDE**)[198] Consider the scalar SDE* $dx(t) = f(x)\ dt + g(x)\ d\beta$, *where* β *is the standard Wiener process. Then the* ***quantile Fokker-Planck equation*** *(QFPE), given by*

$$\partial_t Q = f(Q,t) - \frac{1}{2}\partial_Q\left(g(Q,t)\right)^2 + \frac{1}{2}\left(g(Q,t)\right)^2 \frac{\partial_{\varsigma\varsigma} Q}{\left(\partial_\varsigma Q\right)^2}, \tag{6.26}$$

describes the transport of quantile function $Q(\varsigma, t)$ *for the process* $x(t)$.

Proposition 6. *(**Quantile transformation rule**)[199] For an algebraic map* $y = h(x)$, *we have*

$$Q_y(\varsigma) = \begin{cases} h \circ Q_x(\varsigma) & \text{if } h(\cdot) \text{ is non-decreasing,} \\ h \circ Q_x(1-\varsigma) & \text{if } h(\cdot) \text{ is non-increasing.} \end{cases} \tag{6.27}$$

Next, we work out some specific results by imposing structural assumptions on the true and model dynamics.

6.2.2 Discriminating Deterministic Linear Systems

Let the dynamics of the two systems be

$$\dot{x_i} = a_i x, \quad y_i = c_i x,\ a_i < 0, c_i > 0, \qquad i = 1,2. \tag{6.28}$$

Theorem 20. *For any initial density* $\xi_0(x_0)$, *the Wasserstein gap between the sys-*

tems in (6.28), is given by

$$ {}_2W_2(t) = \sqrt{m_{20}} \left| c_1 e^{a_1 t} - c_2 e^{a_2 t} \right|, \tag{6.29} $$

where $m_{20} = \mu_0^2 + \sigma_0^2$*, is the second raw moment of* $\xi_0(x_0)$*, while* μ_0 *and* σ_0 *are its mean and standard deviation, respectively.*

Proof. For (6.28), $Q_{y_i} = c_i Q_{x_i}$, and the QFPE reduces to a linear PDE $\partial_t Q_{x_i} = a_i Q_{x_i}$, yielding $Q_{x_i}(\varsigma, t) = Q_0(\varsigma) e^{a_i t}$, where Q_0 is the initial quantile function corresponding to ξ_0. Thus, we have

$$ \begin{aligned} \left({}_2W_2(t) \right)^2 &= \int_0^1 \left(Q_{y_1}(\varsigma, t) - Q_{y_2}(\varsigma, t) \right)^2 d\varsigma \\ &= \left(c_1 e^{a_1 t} - c_2 e^{a_2 t} \right)^2 \int_0^1 \left(Q_0(\varsigma) \right)^2 d\varsigma. \end{aligned} \tag{6.30} $$

Since the quantile function maps probability to the sample space, hence $x_0 = Q_0(\varsigma)$, and $d\varsigma = \xi_0(x_0)\, dx_0$. Consequently, we can rewrite (6.30) as

$$ \left({}_2W_2(t) \right)^2 = \left(c_1 e^{a_1 t} - c_2 e^{a_2 t} \right)^2 \underbrace{\int_{-\infty}^{\infty} x_0^2\, \xi_0(x_0)\, dx_0}_{m_{20}}. \tag{6.31} $$

Taking square root to both sides, we obtain the result. It's straightforward to check that $m_{20} = \mu_0^2 + \sigma_0^2$, relating the central moments with m_{20}. ■

Remark 11. *(*${}_2W_2$ ***has limited dependence on*** ξ_0*) The above result shows that the Wasserstein gap between scalar linear systems, depends on the initial density up to mean and variance. Any other aspect (skewness, kurtosis etc.) of* ξ_0*, even when it's non-Gaussian, has no effect on* ${}_2W_2(t)$*. The next example demonstrates that our result: "the initial PDF with maximum second raw moment, maximizes Wasserstein*

distance" (Fig. 6.2), may be counterintuitive in some situations.

Example 1. *(**Uniform initial PDF may not maximize** ${}_2W_2$) For (6.28), let the set of admissible initial PDFs be* $S_0 := \{\xi_0 : \mathrm{supp}\,(\xi_0) = [a,b], \xi_0(x_0) = \frac{1}{(b-a)^{\alpha+\beta-1}B(\alpha,\beta)}(x_0-a)^{\alpha-1}(b-x_0)^{\beta-1}, \alpha>0, \beta>0\}$*, i.e. the set of all scaled beta PDFs supported on* $[a,b]$*. One can readily compute that* $\mu_0 = \frac{\alpha b+\beta a}{\alpha+\beta}$*, and* $\sigma_0^2 = \frac{\alpha\beta(b-a)^2}{(\alpha+\beta)^2(\alpha+\beta+1)}$*. For* $\alpha=\beta=1$*,* $\xi_0 = \mathcal{U}([a,b])$*, and for* $\alpha=\beta=\frac{1}{2}$*,* $\xi_0 = \mathcal{A}([a,b])$*. Thus, we have*

$$m_{20}(\mathcal{U}[a,b]) = \frac{1}{3}\left(a^2+b^2+ab\right), \tag{6.32}$$

$$m_{20}(\mathcal{A}[a,b]) = \frac{1}{8}\left(3a^2+3b^2+2ab\right), \tag{6.33}$$

and hence $m_{20}(\mathcal{A}[a,b]) > m_{20}(\mathcal{U}[a,b])$*,* $\forall\, b>a$*. From Theorem 20,* ${}_2W_2(t)$ *trajectory for uniform initial PDF, stays below the same for arcsine initial PDF, as shown in Fig. 6.3.*

Remark 12. *(**Discrete-time linear systems**) Consider the true and model maps* $x_i^{(k+1)} = a_i x_i^{(k)}$*,* $y_i^{(k)} = c_i x_i^{(k)}$*,* $i=1,2$*, where* $k \in \mathbb{N}\cup\{0\}$*, denotes the discrete time index. From linear recursion, one can obtain a result similar to (6.29):* $W(k) = \sqrt{m_{20}}\left|c_1 a_1^k - c_2 a_2^k\right|$*.*

Remark 13. *(**Linear Gaussian systems**) For the linear Gaussian case, one can verify (6.29) without resorting to the QFPE. To see this, notice that if* $\xi_0(x_0) = \mathcal{N}(\mu_0,\sigma_0^2)$*, then the state PDFs evolve as* $\xi_{x_i}(x_i,t) = \mathcal{N}\left(\mu_{x_i}(t), \sigma_{x_i}^2(t)\right)$*, where* $\mu_{x_i}(t)$ *and* $\sigma_{x_i}^2(t)$ *satisfy their respective state and Lyapunov equations, which, in the scalar case, can be solved in closed form. Since* $\eta_{y_i}(y_i,t) = \mathcal{N}\left(c_i\mu_{x_i}(t), c_i^2\sigma_{x_i}^2(t)\right)$*, and* ${}_2W_2$ *between two Gaussian PDFs is known [138] to be* $\sqrt{(\mu_{y_1}-\mu_{y_2})^2+(\sigma_{y_1}-\sigma_{y_2})^2}$*, the result follows.*

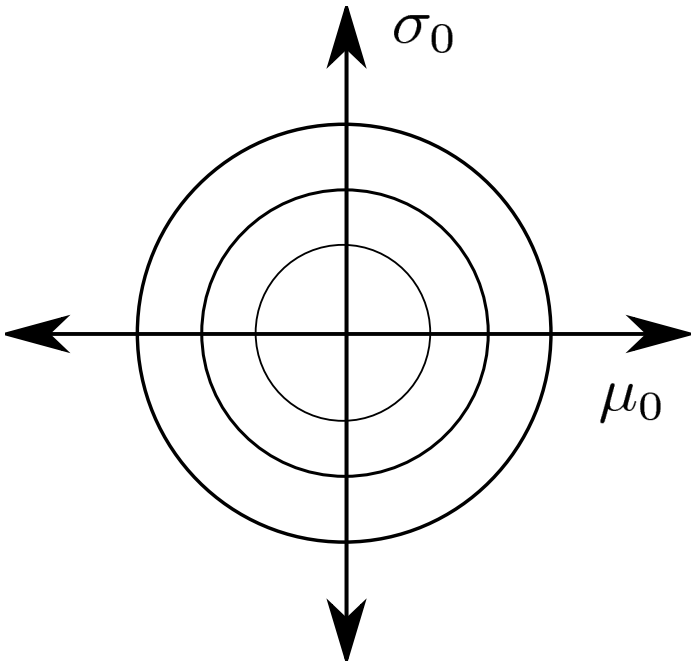


Figure 6.2: The results of Section 6.2 can be summarized through a graphical algorithm illustrated above. For scalar linear systems, given a set of admissible initial PDFs over state space, we construct concentric circles centered at origin, over the two dimensional (μ_0, σ_0) subspace of the (infinite-dimensional) moment space. From (6.29), ξ_0 corresponding to the circle with largest radius, maximizes ${}_2W_2(t)$, $\forall t > 0$. For affine systems, (6.34) implies a similar construction in (μ_0, σ_0) subspace, with circles centered at $\left(-\frac{q(t)}{p(t)}, 0\right)$. The direction of this translation along μ_0 axis, depends on parameters (a_i, b_i, c_i, d_i), $i = 1, 2$, of the systems under comparison.

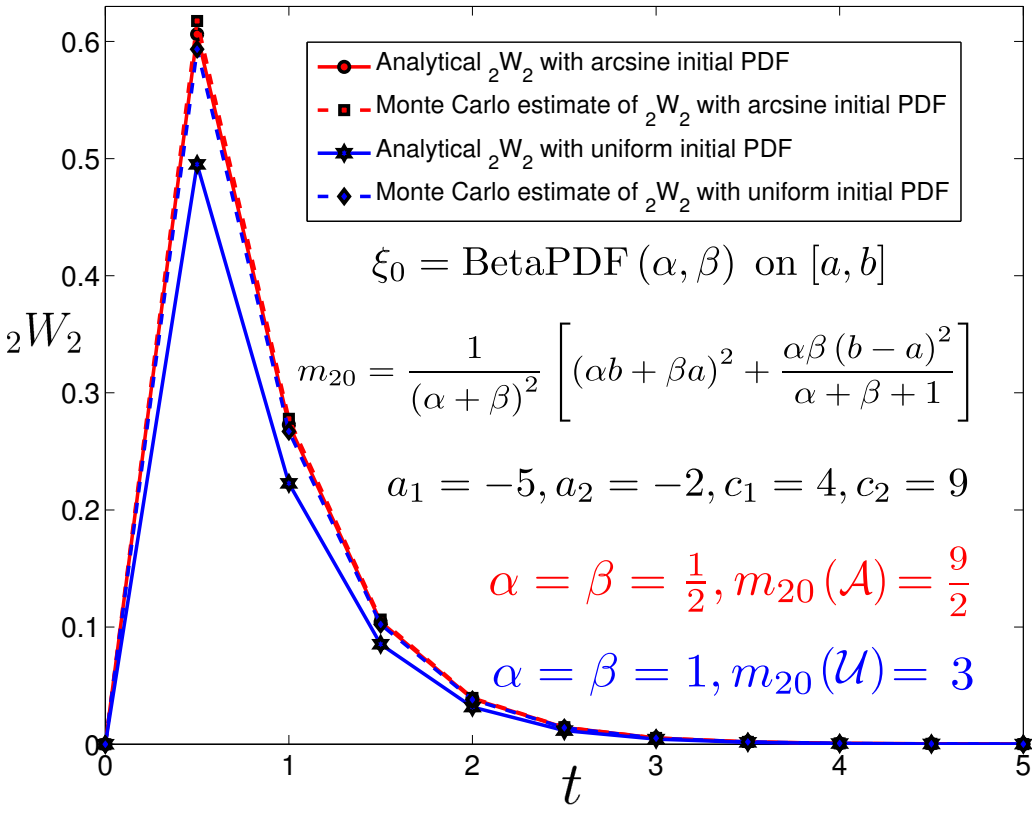


Figure 6.3: Wasserstein time histories between linear system pair (6.28) with ξ_0 as $\mathcal{A}([a,b])$ and $\mathcal{U}([a,b])$, respectively. Here $a = -3$, $b = 3$, and we set sampling interval $\Delta t_k = 0.5$. We observe that the Wasserstein gap for $\xi_0 = \mathcal{A}([a,b])$ remains above the same for $\xi_0 = \mathcal{U}([a,b])$, as predicted by Theorem 20. The *solid lines* are direct computation from (6.29), while the *dashed lines* are Monte Carlo estimates of ${}_2W_2$ using (4.3).

Remark 14. *(**Affine dynamics**) Instead of (6.28), if the dynamics are given by* $\dot{x}_i = a_i x + b_i$, $y_i = c_i x + d_i$, $i = 1, 2$, *then by variable substitution, one can derive that* $Q_{x_i}(\varsigma, t) = Q_0(\varsigma)\, e^{a_i t} + \frac{b_i}{a_i}\left(e^{a_i t} - 1\right)$. *Hence, we get*

$$ {}_2W_2(t) = \sqrt{(p(t))^2\, m_{20} + 2p(t)\, q(t)\, m_{10} + (q(t))^2}, \tag{6.34}$$

where $m_{10} = \mu_0$, $p(t) := (c_1 e^{a_1 t} - c_2 e^{a_2 t})$, *and* $q(t) := \frac{b_1 c_1}{a_1}\left(e^{a_1 t} - 1\right) - \frac{b_2 c_2}{a_2}\left(e^{a_2 t} - 1\right) +$

$(d_1 - d_2)$.

6.2.3 Discriminating Stochastic Linear Systems

Consider two stochastic dynamical systems with linear drift and constant diffusion coefficients, given by

$$dx_i = a_i x\, dt + b_i\, d\beta, \quad y_i = c_i x, \qquad i = 1, 2, \tag{6.35}$$

where β is the standard Wiener process.

Theorem 21. *For any initial density $\xi_0(x_0)$, the Wasserstein gap ${}_2W_2(t)$ between the systems in (6.35), is given by*

$${}_2W_2(t) = \sqrt{(p(t))^2 m_{20} + 2p(t)\, r(t)\, s(F_0) + (r(t))^2}, \tag{6.36}$$

where $r(t) := \dfrac{|b_1|c_1}{\sqrt{2a_1}}\sqrt{e^{2a_1 t} - 1} - \dfrac{|b_2|c_2}{\sqrt{2a_2}}\sqrt{e^{2a_2 t} - 1}$, *and*

$$s(F_0) := \sqrt{2}\, \mathbb{E}\left[x_0\, \mathrm{erf}^{-1}(2F_0(x_0) - 1)\right], \tag{6.37}$$

F_0 being the CDF of x_0.

Proof. For systems (6.35), quantile functions for the states evolve as (p. 102, [198])

$$Q_{x_i}(\varsigma, t) = Q_0(\varsigma)\, e^{a_i t} + |b_i| Q_N(\varsigma) \sqrt{\frac{e^{2a_i t} - 1}{2a_i}}, \tag{6.38}$$

where $Q_N(\varsigma) := \sqrt{2}\, \mathrm{erf}^{-1}(2\varsigma - 1)$, is the standard normal quantile. Thus, the Wasserstein distance becomes

$$({}_2W_2(t))^2 \;=\; \int_0^1 (c_1 Q_{x_1}(\varsigma, t) - c_2 Q_{x_2}(\varsigma, t))^2\, d\varsigma$$

$$
\begin{aligned}
&= \left(p\left(t\right)\right)^2 \int_0^1 \left(Q_0\left(\varsigma\right)\right)^2 \, d\varsigma \\
&+ 2p\left(t\right) r\left(t\right) \int_0^1 Q_0\left(\varsigma\right) Q_N\left(\varsigma\right) \, d\varsigma \\
&+ \left(r\left(t\right)\right)^2 \int_0^1 \left(Q_N\left(\varsigma\right)\right)^2 \, d\varsigma. \qquad (6.39)
\end{aligned}
$$

Notice that the first and third integrals are m_{20} and 1, respectively. Since $\varsigma = F_0\left(x_0\right)$, the second integral becomes

$$
\begin{aligned}
&\int_{-\infty}^{\infty} x_0 \, F_N^{-1} \circ F_0\left(x_0\right) \, \rho_0\left(x_0\right) dx_0 \\
&= \sqrt{2}\, \mathbb{E}\left[x_0 \operatorname{erf}^{-1}\left(2F_0\left(x_0\right) - 1\right)\right] = s\left(F_0\right). \qquad (6.40)
\end{aligned}
$$

This completes the proof. ■

Remark 15. *(**Gaussian case**) Consider the special case when $\xi_0\left(x_0\right) = \mathcal{N}\left(\mu_0, \sigma_0^2\right)$. Then $Q_0\left(\varsigma\right) = \mu_0 + \sigma_0 Q_N\left(\varsigma\right)$, and hence the second integral equals σ_0. Thus, if the initial density is normal, then*

$$
{}_2W_2\left(t\right) = \sqrt{\left(p\left(t\right)\right)^2 m_{20} + 2p\left(t\right) r\left(t\right) \sigma_0 + \left(r\left(t\right)\right)^2}, \qquad (6.41)
$$

a function of μ_0 and σ_0, which can be verified otherwise by solving the mean and variance propagation equations.

6.3 Chapter Summary

In this Chapter, we provided two systems-theoretic results pertaining to the proposed V&V framework. First, we showed that the asymptotic Wasserstein gap between two deterministic single output linear systems, excited by Gaussian white noise, is the difference between their average gains, up to a scaling by the strength of the noise excitation. This gives an intuitive systems-theoretic interpretation of

Wasserstein gap. Similar formula was derived for the multi input, multi output case. The single output asymptotic Wasserstein gap was shown to admit an intrinsic normalization, thus allowing a geometric comparison between the proposed Wasserstein gap and the conventional ν-gap metric. The second result concerns linear model discrimination in Wasserstein metric. It was shown that given a set of initial PDFs, the one that maximizes the Wasserstein gap between two scalar linear systems, has maximum second raw moment.

7. PROBABILISTIC MODEL REFINEMENT

The previous chapters have introduced a probabilistic V&V framework based on the theory of optimal transport. Given a model and experimentally measured data, we have discussed how to compute the probability that the proposed model is valid at any given time. In this chapter, we focus on the next step that a modeler can take after such V&V inference has been obtained, i.e. *model refinement*.

The motivation of model refinement comes from the fact that most physics based models perform well against some, but not all temporal segments of the observed data. For example, a model may be able to make excellent asymptotic prediction, as substantiated by the observed data, but there may be significant prediction vs. observation mismatch in the transient regime. The hard invalidation methods available in the literature simply discards such a model. In contrast, we propose here that probabilistic V&V information can be utilized in a constructive manner to refine the model. The purpose of this chapter is to lay the foundation of probabilistic model refinement in the optimal transport framework we have set so far. After introducing the main ideas, we will show that how the proposed refinement formulation can be used for finite-time control of PDFs, a problem that is of independent practical interest. To embark on this pursuit, we revisit the optimal transport problem and review its equivalent formulations.

7.1 Background on Optimal Transport

7.1.1 Primal Formulation

The optimal transport theory originated in 1781 when Gaspard Monge considered [200] the problem of moving a pile of soil from an excavation to another site that entails minimum work. This idea went mostly unnoticed for 160 years until

Leonid Kantorovich provided a modern treatment [201] of this subject in 1942 (the English translation [202] appeared in 1958), which eventually led to the Nobel prize in economics in 1975. In the theory of Monge-Kantorovich optimal transport, one defines a distance, called *Wasserstein distance*, between two *given* PDFs ρ and $\widehat{\rho}$, that measures the *shape difference* between them.

The Wasserstein distance and the associated infinite dimensional LP, as we introduced in Section 4.1 and 4.2, are the primal formulation for the Monge-Kantorovich optimal transport problem. We restate this definition here for clarity of exposition and notational consistency.

Definition 6. *(**Wasserstein distance**) The L_2 Wasserstein distance of order 2 (henceforth referred simply as* Wasserstein distance *W), between two d-dimensional random vectors $y \sim \rho$, and $\widehat{y} \sim \widehat{\rho}$, is defined as*

$$W\left(\rho, \widehat{\rho}\right) \triangleq \left(\inf_{\varrho \in \mathcal{P}_2(\rho, \widehat{\rho})} \underbrace{\mathbb{E}\left[\| \, y - \widehat{y} \, \|^2_{\ell_2\left(\mathbb{R}^d\right)} \right]}_{J_1(\varrho)} \right)^{\frac{1}{2}}, \tag{7.1}$$

where the $\mathbb{E}\left[\cdot\right]$ is taken with respect to the joint PDF $\varrho\left(y, \widehat{y}\right)$ that makes the cost function achieve the infimum. The symbol $\mathcal{P}_2\left(\rho, \widehat{\rho}\right)$ denotes the set of all joint PDFs supported over $\mathbb{R}^{2d}$, having finite second moments, whose first marginal is ρ, and second marginal is $\widehat{\rho}$.

Computation for our V&V formulation so far has relied exclusively on this formulation of optimal transport. Next, we list equivalent formulations of optimal transport, which may be more amenable from model refinement perspective.

7.1.2 Spatial Variational Formulation for Optimal Transport Map

Instead of solving the infinite dimensional LP (7.1) , one could directly solve for the the optimal transport map $\beta : \mathbb{R}^d \mapsto \mathbb{R}^d$, that satisfies $y = \beta(\widehat{y})$, by solving

$$\inf_{\beta(\cdot)} \underbrace{\int_{\mathbb{R}^d} \| \beta(\widehat{y}) - \widehat{y} \|^2_{\ell_2(\mathbb{R}^d)} \ \widehat{\rho}(\widehat{y}) \ d\widehat{y}}_{J_2(\beta)}, \text{ subject to } \rho = \beta \sharp \widehat{\rho}. \tag{7.2}$$

Remark 16. *Since there are infinite ways to morph $\widehat{\rho}$ to ρ, (7.2) looks for an* optimal push-forward map *$\beta^\star(\cdot)$ that would require minimum amount of transport effort among all possible push-forward maps $\beta(\cdot)$. Then the map $\beta^\star(\cdot)$ characterizes the optimal transport.*

Remark 17. *In a seminal paper [203], Brenier proved the existence and uniqueness of $\beta^\star(\cdot)$. Further, his* polar factorization theorem *[203] proved that the unique vector function $\beta^\star(\cdot)$ can be written as a gradient of a scalar function, i.e. $\beta^\star = \nabla\psi$. Furthermore, the scalar function ψ is convex. The optimal transport map $\beta^\star$ is also known as the* Brenier map*.*

Remark 18. *Although the cost function in (7.2) is quadratic in $\beta(\cdot)$, the push-forward constraint is nonlinear and non-convex in $\beta(\cdot)$. Thus, a direct numerical optimization is not straight-forward. As shown in the second row of Table 7.1, [204] used the fact that $\beta^\star(\cdot)$ is curl-free, to formulate a* regularized *sequential quadratic program (SQP) to solve (7.2) as*

$$\inf_{\beta(\cdot)} \widetilde{J}_2(\beta), \quad \textit{subject to} \quad c(\beta) = 0, \tag{7.3}$$

where $\widetilde{J}_2(\beta) \triangleq J_2(\beta) + \alpha \int_{\mathbb{R}^d} \| \nabla \times \beta \|^2_{\ell_2(\mathbb{R}^d)} \ d\widehat{y}$, and $\alpha > 0$ is a regularization parameter. The constraint $c(\beta) \triangleq |det(\nabla\beta)| \, \rho \circ \beta(\widehat{y}) - \widehat{\rho}(\widehat{y})$.

7.1.3 Spatial PDE Formulation for Optimal Transport Map

From Remark 17, we can substitute $\beta = \nabla\psi$ in the push-forward constraint $c(\beta) = 0$. Then it follows that ψ must solve

$$|\det(\text{Hess}(\psi(\widehat{y})))|\, \rho(\nabla\psi(\widehat{y})) = \widehat{\rho}(\widehat{y}). \tag{7.4}$$

This is a second order, nonlinear, stationary, elliptic PDE, known as the *Monge-Ampère equation*, to be solved for ψ as a function of $\widehat{y}$. In principle, if we can solve (7.4), then $\nabla\psi$ would yield the optimal transport map. However, as mentioned in the third row of Table 7.1, numerically solving the PDE (7.4) remains a research challenge.

7.1.4 Benamou-Brenier Space-time Variational Formulation

Benamou and Brenier proposed [207] a *dynamic* reformulation of the *static* optimization problem (7.2) by introducing a synthetic notion of time, which we denote as $s \in [0, \tau]$. Their main result is that the spatial optimization problem (7.2), is equivalent to solving the following space-time optimization problem:

$$W^2 = \tau \inf_{(\varphi, v)} \int_{\mathbb{R}^d} \int_0^\tau \varphi(\widehat{y}, s) \parallel v(\widehat{y}, s) \parallel^2_{\ell_2(\mathbb{R}^d)} \, d\widehat{y}\, ds, \tag{7.5}$$

$$\text{subject to } \frac{\partial\varphi}{\partial s} + \nabla\cdot(\varphi v) = 0,\ \varphi(\cdot, 0) = \widehat{\rho},\ \varphi(\cdot, \tau) = \rho. \tag{7.6}$$

Remark 19. *It is important to understand the meaning of solving the optimization problem (7.5)-(7.6). Notice that the spatial and temporal integrals in the cost function can be interchanged. Thus, if we fix s, then the cost is the instantaneous kinetic energy of the ensemble during transport, where each sample moves according to the deterministic ODE $\frac{d}{ds}\widehat{y} = v(\widehat{y}(s), s)$, corresponding to the Liouville PDE in Chapter*

Table 7.1: Different formulations of optimal transport, as described in Section 7.1

Mathematical formulation	Problem type	Numerical method	Solver
Primal formulation in ϱ	Infinite dimensional LP (7.1)	Direct discretization	Large scale LP solver (e.g. MOSEK® used in [41, 45])
Variational Formulation for optimal transport map $\beta(\cdot)$	Quadratic cost with nonlinear non-convex constraints (7.2)	Regularized SQP (optimality not guaranteed)	"Discretize-then-optimize" solver in [204]
PDE Formulation for optimal transport map $\beta(\cdot)$	Monge-Ampère PDE (7.4) (Second order nonlinear elliptic PDE)	Not well studied	See review article [205] for current research status
Brenier-Benamou space-time variational formulation in (v, φ)	Non-smooth convex optimization problem (7.5)-(7.6)	Proximal operator splitting	Staggered grid discretization in [206]

2: $\dfrac{\partial \varphi}{\partial s} + \nabla \cdot (\varphi v) = 0$*, appearing in the constraint. Hence, the cost function in (7.5) is equal to the total kinetic energy up to time* τ*. Consequently,* W^2 *equals total work done during the transport process. The optimization is over a pair of vector field* v *and joint PDF* φ*, and is convex in both.*

Remark 20. *It can be shown [132] that the minimizing vector field* $v^\star(\widehat{y}, s)$ *in the above optimization problem, is a* pressureless potential flow*. In other words,* $v^\star(\widehat{y}, s) = \nabla \phi(\widehat{y}, s)$*, where the scalar function* $\phi(\widehat{y}, s)$ *solves the Hamilton-Jacobi equation*

$$\frac{\partial \phi}{\partial s} + \frac{1}{2} \parallel \nabla \phi \parallel^2_{\ell_2\left(\mathbb{R}^d\right)} = 0. \tag{7.7}$$

Remark 21. *In p. 384 of [207], using Legendre transform, (7.5)-(7.6) was further converted to a saddle point optimization problem, which was numerically solved using the augmented Lagrangian technique [208]. Recently, an improved numerical method to solve (7.5)-(7.6) has been proposed [206] via proximal operator splitting. We will use this technique later in this chapter for numerical simulations.*

7.1.5 Wasserstein Geodesics on the Manifold of PDFs

One merit of the Benamou-Brenier approach described above is that it constructs the transportation path, which is a geodesic connecting the source and target PDFs, and yields the intermediate PDFs satisfying McCann's displacement interpolation [3]. In particular, the following results [132] hold.

1. Without loss of generality, let the synthetic time $s \in [0, 1]$, i.e. in the notation of Section 7.1.4, set $\tau = 1$. Then the Benamou-Brenier vector field constructs

the geodesic curve between $(\widehat{\rho}, \rho) : s \in [0,1] \mapsto \varphi(s)$. Recall that the PDF $\widehat{\rho}$ is the source PDF, and ρ is the target PDF. In other words, φ has the variational characterization

$$\varphi(s) = \underset{\varphi}{\operatorname{argmin}} (1-s) W^2(\widehat{\rho}, \varphi) + s W^2(\rho, \varphi), \tag{7.8}$$

and it lies on the geodesic curve connecting $\widehat{\rho}$ and ρ. The Wasserstein distance $W(\widehat{\rho}, \rho)$ is the length of this geodesic curve on the manifold of PDFs.

2. As a corollary of the above result, the intermediate optimal transport map β_s that satisfies $\varphi(s) = \beta_s \sharp \widehat{\rho}$, is obtained via *linear interpolation* between the identity map Id and β, i.e.

$$\beta_s = (1-s)\,\mathrm{Id} + s\beta. \tag{7.9}$$

Also, the intermediate Wasserstein distance is obtained via *linear interpolation*:

$$W(\widehat{\rho}, \varphi(s)) = s\, W(\widehat{\rho}, \rho), \tag{7.10}$$

$$W(\rho, \varphi(s)) = (1-s)\, W(\widehat{\rho}, \rho). \tag{7.11}$$

However, the intermediate PDF is obtained via *nonlinear (displacement) interpolation*:

$$\varphi(s) = \beta_s \sharp \widehat{\rho} = [(1-s)\,\mathrm{Id} + s\beta] \sharp \widehat{\rho}, \tag{7.12}$$

$$= \beta_{1-s} \sharp \rho = [s\,\mathrm{Id} + (1-s)\beta] \sharp \rho. \tag{7.13}$$

7.2 Model Refinement Through Brenier Map

In this section, we consider refining a baseline model against experimental data. We first formulate the model refinement problem as that of finding the optimal transport map introduced in Section 7.1.2.

7.2.1 Problem Formulation

We formulate the model refinement problem (Fig. 7.1) as the natural successor of the density based model validation formulation presented in the previous chapters.

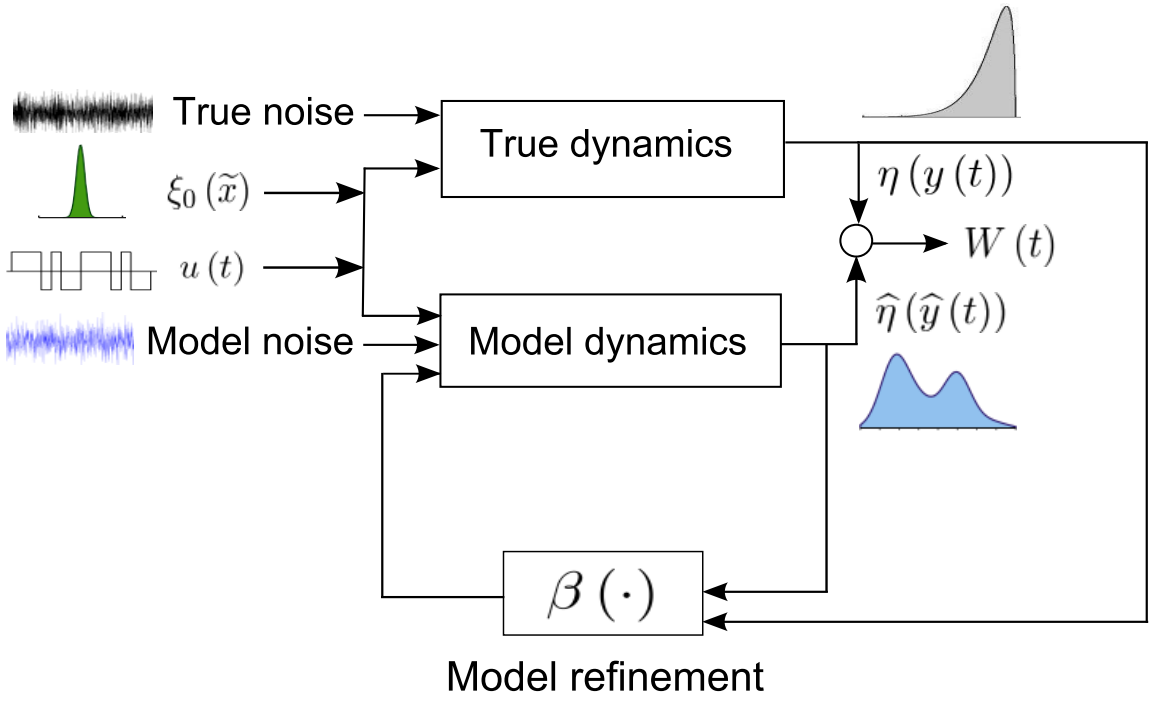


Figure 7.1: The block diagram for proposed model refinement formulation. Here $\xi_0(\widetilde{x})$ refers to the joint PDF supported on the space of initial conditions and parameters, $u(t)$ is an open-loop control command, and the true and model dynamics can be affected by different noises.

In the validation problem, the model predicted output PDF $\widehat{\eta}$ is compared with

the experimentally observed output PDF η, at each instance of measurement availability t_j, $j = 0, 1, \ldots, M$, and an inference is made by looking at the prediction-observation gap quantified via $W(t_j)$. The key insight behind our refinement formulation is that usually there is no specific requirement on the structure of the refined model, as long as we can make the refined dynamics track the observed output PDFs. This provides us the freedom to formulate the model refinement problem over the model's output map while keeping the model's state equation intact. This has two implications: (i) the refinement algorithm will involve the output dimension n_o, typically less than the state dimension, and (ii) both state and output modeling errors would be accounted by updating the model's output map. To make the ideas precise, we give the model refinement problem statement for a model whose output map is given by $\widehat{y} = \widehat{h}(\widehat{x})$, where $\widehat{x}$ and $\widehat{y}$ are model-predicted state and output vectors.

At $t = t_j$, let us introduce $\widehat{y_j^-} \triangleq \widehat{y}_j$, and denote $\widehat{h}^-(\cdot) \triangleq \widehat{h}(\cdot)$. We want to find the *Brenier map* $\beta_j(\cdot)$ for updating the predicted output, i.e. $\widehat{y}_j^+ = \beta_j\left(\widehat{y_j^-}\right)$, where $\widehat{y}_j^+ \sim \eta_j$ and $\widehat{y_j^-} \sim \widehat{\eta}_j$. In other words, find $\beta_j(\cdot)$ such that $\eta_j = \beta_j \sharp\, \widehat{\eta}_j$. Clearly, this problem is underdetermined since there are many ways to morph $\widehat{\eta}_j$ to η_j. Then we must look for an *optimal push-forward map* $\beta_j^\star(\cdot)$ that would require minimum amount of transport effort among all possible push-forward maps $\beta_j(\cdot)$, i.e. we solve (7.2). Once $\beta_j^\star(\cdot)$ has been found, the refined model is given by augmenting the model's state equation with the new output map:

$$\widehat{y}_j = \beta_j^\star \circ \widehat{h}(\widehat{x}). \tag{7.14}$$

7.2.2 Refining Linear Model against Gaussian Measurements

In this subsection, we work with an example where a discrete time linear *model* is given. Further the true data is generated by evolving an initial Gaussian joint

PDF ξ_0, subject to a *true* (and hence unknown to the modeler) discrete time linear system.

Suppose, the true data being generated by the discrete-time LTI system $x_{j+1} = Ax_j$, $y_j = Cx_j$, that is unknown to the modeler. The proposed model is $\widehat{x}_{j+1} = \widehat{A}\widehat{x}_j$, $\widehat{y}_j = \widehat{C}\widehat{x}_j$. In this linear Gaussian case, both the spatial variational formulation (Section 7.1.2), and the space-time variation formulation (Section 7.1.4 and 7.1.5) are computationally tractable. Hence, we demonstrate both for this problem. In our case, the result from the first formulation, can be summarized in the following proposition.

Proposition 7. *Consider the true dynamics*

$$x_{j+1} = Ax_j, \quad y_j = Cx_j, \tag{7.15}$$

and model dynamics

$$\widehat{x}_{j+1} = \widehat{A}\widehat{x}_j, \quad \widehat{y}_j = \widehat{C}\widehat{x}_j, \tag{7.16}$$

and an initial PDF $\xi_0 = \mathcal{N}(\mu_0, \Sigma_0)$. *Suppose that the true data at time* t_j *is generated as* $y(t_j) \sim \mathcal{N}(\mu_j, \Sigma_j)$ *under the action of (7.15). From (7.16), the model predicted output PDF is given by* $\widehat{y}(t_j) \sim \mathcal{N}\left(\widehat{\mu}_j, \widehat{\Sigma}_j\right)$. *Then, the refined model is*

$$\widehat{x}_{j+1} = \widehat{A}\widehat{x}_j, \quad \widehat{y}_j = T_j\widehat{C}\widehat{x}_j + \tau_j, \tag{7.17}$$

where the matrix-vector pair (T_j, τ_j) *is given by*

$$T_j = \Sigma_j^{1/2}\left(\Sigma_j^{1/2}\widehat{\Sigma}_j\Sigma_j^{1/2}\right)^{-1/2}\Sigma_j^{1/2}, \tag{7.18}$$

$$\tau_j = \mu_j - \widehat{\mu}_j. \tag{7.19}$$

Proof. The proof can be adapted from [209, 210] by noting that

$$\mu_j = CA^j\mu_0, \qquad \Sigma_j = CA^j\Sigma_0\left(CA^j\right)^\top, \tag{7.20}$$

$$\widehat{\mu}_j = \widehat{C}\widehat{A}^j\mu_0, \qquad \widehat{\Sigma}_j = \widehat{C}\widehat{A}^j\Sigma_0\left(\widehat{C}\widehat{A}^j\right)^\top. \tag{7.21}$$

We skip the details here, and direct the interested readers to above cited references. ■

While the above result is obtained by solving the spatial variational formulation (Section 7.1.2) for Brenier map, it does not provide the refinement path. Next, we show that the space-time variational formulation (Section 7.1.4 and 7.1.5) allows the same.

To give a concrete example, we take the Schur-Cohn stable matrices A and $\widehat{A}$ to be

$$A = \begin{bmatrix} 0.4 & -0.1 \\ 2 & 0.6 \end{bmatrix}, \quad \widehat{A} = \begin{bmatrix} 0.2 & -0.7 \\ -0.7 & 0.1 \end{bmatrix}, \tag{7.22}$$

and the output matrices as

$$C = \begin{bmatrix} -1 & 0.03 \\ -0.2 & 0.8 \end{bmatrix}, \quad \widehat{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{7.23}$$

Starting from the initial Gaussian state PDF $\xi_0 = \mathcal{N}(\mu_0, P_0)$ with $\mu_0 = \{1, 3\}^\top$, $P_0 = \begin{bmatrix} 10 & 6 \\ 6 & 7 \end{bmatrix}$, we refine the model at three instances of measurement availability: $j = 1, 2$, and 3. The results for the model refinement algorithm are shown in Fig. 7.2.

To illustrate how the results of Section 7.1.5 are applied in this particular refinement problem, we provide the following Theorem.

Theorem 22. ***(s^{th} synthetic time PDF at j^{th} physical time)*** *Let $s \in [0,1]$ and consider the above linear Gaussian refinement problem with initial PDF $\mathcal{N}(\mu_0, P_0)$. At the j^{th} instance of measurement availability, the intermediate PDF during refinement is a Gaussian PDF $\mathcal{N}(\mu_{\widehat{y}\to y}(s), \Sigma_{\widehat{y}\to y}(s))$ where*

$$\mu_{\widehat{y}\to y}(s) = \left[(1-s)\ \widehat{C}\widehat{A}^j\ +\ s\,CA^j\right]\mu_0, \tag{7.24}$$

$$\Sigma_{\widehat{y}\to y}(s) = [(1-s)\ I\ +\ s\,\Gamma(j)]\left(\left(\widehat{C}\widehat{A}^j\right)P_0\left(\widehat{C}\widehat{A}^j\right)^\top\right)[(1-s)\ I\ +\ s\,\Gamma(j)]^\top, \tag{7.25}$$

where

$$\Gamma(j) \triangleq \sqrt{(CA^j)\,P_0\,(CA^j)^\top}\left(\sqrt{(CA^j)\,P_0\,(CA^j)^\top}\left(\widehat{C}\widehat{A}^j\right)P_0\left(\widehat{C}\widehat{A}^j\right)^\top \sqrt{(CA^j)\,P_0\,(CA^j)^\top}\right)^{-1/2}\sqrt{(CA^j)\,P_0\,(CA^j)^\top}. \tag{7.26}$$

Proof. We know that $\mu_y(j) = C\mu_x(j) = CA^j\mu_0$, and similarly, $\widehat{\mu}_{\widehat{y}}(j) = \widehat{C}\widehat{A}^j\mu_0$. On the other hand, we have $\Sigma_y(j) = C\Sigma_x(j)\,C^\top = CA^jP_0\,(A^j)^\top C^\top$, and similarly, $\widehat{\Sigma}_{\widehat{y}}(j) = \left(\widehat{C}\widehat{A}^j\right)P_0\left(\widehat{C}\widehat{A}^j\right)^\top$.

From (7.12), we get $\mu_{\widehat{y}\to y}(s) = (1-s)\,\widehat{\mu}_{\widehat{y}}(j) + s\mu_y(j) = \left[(1-s)\ \widehat{C}\widehat{A}^j\ +\ s\,CA^j\right]\mu_0$. Similarly, $\Sigma_{\widehat{y}\to y}(s) = [(1-s)\ I\ +\ s\,\Gamma(j)]\,\widehat{\Sigma}_{\widehat{y}}(j)\,[(1-s)\ I\ +\ s\,\Gamma(j)]$, where $\Gamma(j) = \sqrt{\Sigma_y(j)}\left(\sqrt{\Sigma_y(j)}\,\widehat{\Sigma}_{\widehat{y}}(j)\,\sqrt{\Sigma_y(j)}\right)^{-\frac{1}{2}}\sqrt{\Sigma_y(j)}$ (from (7.18)). Substituting the covariance matrix formulae in terms of the respective system matrices, the result follows. ■

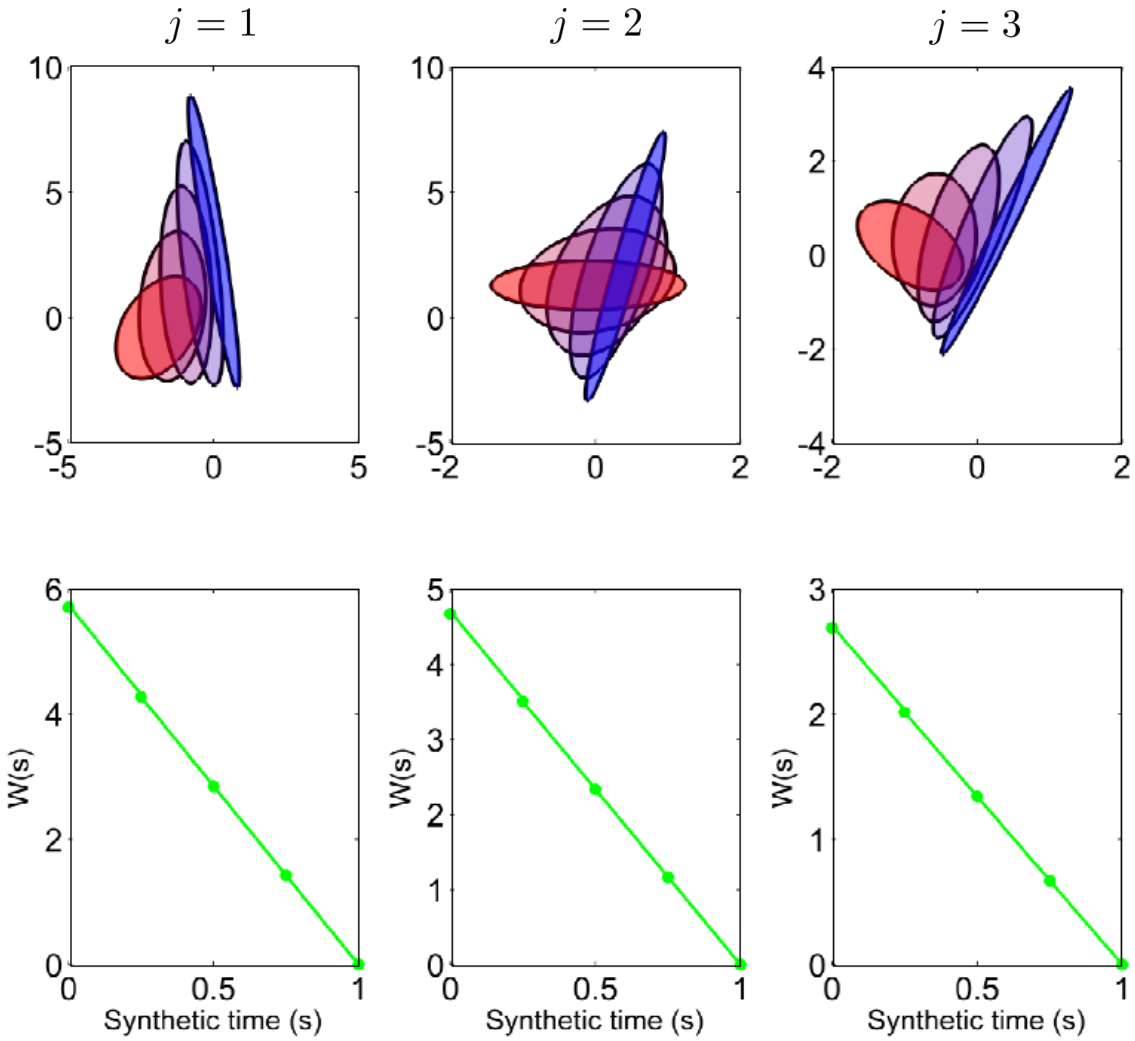


Figure 7.2: Shown here is the refinement process of the linear model $\left(\widehat{A}, \widehat{C}\right)$ at times $j = 1, 2$ and 3, so that the model predicted output PDFs match with the true PDF, generated by (A, C) at each j. Since both processes are Gaussian, the *top row* shows 1-σ ellipses of the respective normal PDFs (*red = model predicted*, *blue = true*). At every fixed j, we also plot intermediate Gaussians generated during the refinement process. The color of these intermediate 1-σ ellipses are interpolated *from red to blue*, to show the notion of synthetic time $s \in [0, 1]$, as the physical time index j remains zero-order hold. This also shows that the Gaussian-to-Gaussian refinement happens via Gaussians, i.e. the set of Gaussian PDFs is geodesically convex. The *bottom row* shows that although the PDFs over synthetic time gets nonlinearly interpolated (McCann's *displacement interpolation* [3]), the Wasserstein distance $W(s)$ gets linearly interpolated, as predicted by (7.10).

7.2.3 Nonlinear Model Refinement via Proximal Operator Splitting

Nonlinear models result non-Gaussian PDF evolution, and hence explicitly solving the Brenier map from the spatial variation formulation of Section 7.1.2 is rarely possible. Even numerical computation of the same is difficult due to the reasons discussed in Remark 18, Section 7.1.2. A comparatively tractable computation results if we resort think of the model refinement problem in the space-time variational context of Section 7.1.4, as follows.

We think of the Brenier map $\beta^{\star}(\cdot)$ in (7.14), as the flow of some synthetic ODE. Computing $\beta^{\star}(\cdot)$, then amounts to computing the ODE over simulation time s, whose Liouville PDE would satisfy the two point boundary PDF constraints: model predicted PDF at simulation time $s = 0$, and observed PDF at simulation time $s = 1$. This way, model refinement transcribes to solving the optimization problem (7.5)-(7.6), which was [206] shown to be amenable via proximal operator splitting techniques. We refer the readers to [206] for details of this computation. We will use this technique in next section.

7.3 Case Studies

We provide here two case studies. The first shows how the model refinement ideas can be applied to solve PDF tracking problem. This is illustrated for linear Gaussian case, using the spatial variational formulation of Section 7.1.2. The next case study concerns reduced order data driven modeling, using the space-time variational formulation of Section 7.1.4, solved by proximal operator splitting techniques.

7.3.1 Feedback Control for Linear Gaussian PDF Tracking

Consider a linear system $x_{j+1} = Ax_j + Bu_j,\ x_j \in \mathbb{R}^d, u_j \in \mathbb{R}^m$, with a sequence of Gaussian PDFs $\eta_j = \mathcal{N}(\mu_j, \Sigma_j),\ j = 0, 1, \ldots, M$. The objective is to

find state feedback $u_j^\star \triangleq u^\star(x_j)$ over each time interval $\Delta t_j \triangleq [t_j, t_{j+1})$, such that $x_j \sim \eta_j = \mathcal{N}(\mu_j, \Sigma_j)$, while guaranteeing minimal transportation cost. We transcribe the problem of finding optimal control $u_j^\star$ to that of finding the optimal transport map (a.k.a. *Brenier map*) $\beta_j^\star : x_j \mapsto x_{j+1}$, where

$$\beta_j^\star \triangleq \beta^\star(x_j) = \underset{\beta(.)}{\operatorname{argmin}} \int_{\mathbb{R}^d} \| \beta(x_j) - x_j \|^2_{\ell_2(\mathbb{R}^d)} \, \eta_j \, dx_j, \tag{7.27}$$

subject to the constraints (C1) $x_j \sim \eta_j$, (C2) $\beta(x_j) \sim \eta_{j+1}$, and (C3) $\eta_{j+1} = \beta \sharp \eta_j$. Then we have the following result.

Theorem 23. *Consider the discrete-time Gaussian PDF control problem under LTI structure, and let* $\eta_j = \mathcal{N}(\mu_j, \Sigma_j)$, *where* $\ker(\Sigma_j) \cap Im(\Sigma_{j+1}) = \{0\}$. *Further, let* $\mathcal{S}_{j+1}$ *be given by the discrete-time LTI structure:* $x_{j+1} = Ax_j + Bu_j$, $\forall j = 0, 1, \ldots, M$. *Then the state feedback* $u_j^\star \triangleq u^\star(x_j)$ *that minimizes the transportation cost (7.2), has the following properties.*

1. *The optimal state feedback, if exists, must be affine.*

2. *Optimal state feedback* $u_j^\star$ *exists iff* $(\Gamma_j - A), \gamma_j \in ker\left(I - BB^\dagger\right)$, *where*

$$\Gamma_j = \sqrt{\Sigma_{j+1}} \left(\sqrt{\Sigma_{j+1}} \, \Sigma_j \, \sqrt{\Sigma_{j+1}} \right)^{-\frac{1}{2}} \sqrt{\Sigma_{j+1}}, \tag{7.28}$$

$$\gamma_j = \mu_{j+1} - \mu_j. \tag{7.29}$$

3. *If exists, then the optimal state feedback is given by the pair* (K_j, κ_j), *i.e.* $u_j^\star = K_j x_j + \kappa_j$, *where* $K_j = B^\dagger(\Gamma_j - A) - \left(I - BB^\dagger\right) R$, *and* $\kappa_j = B^\dagger \gamma_j - \left(I - BB^\dagger\right) r$, *for* arbitrary *real matrix-vector pair* (R, r) *of appropriate dimen-*

sions.

4. *If B is full rank, then the optimal state feedback is unique, and is given by $K_j = B^{-1}(\Gamma_j - A)$, $\kappa_j = B^{-1}\gamma_j$.*

Proof. Given, $\ker(\Sigma_j) \cap \mathrm{Im}(\Sigma_{j+1}) = \{0\}$, we know [209, 210] that $\beta_j^\star$ satisfying

$$\mathcal{N}(\mu_{j+1}, \Sigma_{j+1}) = \beta_j^\star \sharp \mathcal{N}(\mu_j, \Sigma_j), \tag{7.30}$$

is a unique affine transformation $z \mapsto \Gamma_j z + \gamma_j$. Since the optimal transport is $x_{j+1} = \Gamma_j x_j + \gamma_j$, the *optimal controller, if exists, must be affine*, i.e. of the form $u_j^\star = K_j x_j + \kappa_j$, where K_j and κ_j solve the *linear matrix equations*

$$A + BK_j = \Gamma_j, \qquad B\kappa_j = \gamma_j. \tag{7.31}$$

Now, from Lemma 2.4 in [211], there exists K_j solving the equation $BK_j = (\Gamma_j - A)$ iff

$$\begin{aligned} BB^\dagger(\Gamma_j - A) = (\Gamma_j - A) &\Leftrightarrow \left(I - BB^\dagger\right)(\Gamma_j - A) = 0 \\ &\Leftrightarrow (\Gamma_j - A) \in \ker\left(I - BB^\dagger\right). \end{aligned} \tag{7.32}$$

On the other hand, the matrix-vector equation $B\kappa_j = \gamma_j$ admits solution iff

$$BB^\dagger\gamma_j = \gamma_j \Leftrightarrow \gamma_j \in \ker\left(I - BB^\dagger\right). \tag{7.33}$$

When $(\Gamma_j - A), \gamma_j \in \ker\left(I - BB^\dagger\right)$, then the (non-unique) solution is given by [211]: $K_j = B^\dagger(\Gamma_j - A) - \left(I - BB^\dagger\right)R$, and $\kappa_j = B^\dagger\gamma_j - \left(I - BB^\dagger\right)r$, for *arbitrary* real

matrix-vector pair (R, r) of appropriate dimensions. If B is full rank, then B^{-1} exists and $\left(I - BB^{\dagger}\right) = 0$, resulting the unique solution. ■

Remark 22. *It is easy to see that the above Theorem generalizes when the LTI pair* (A_j, B_j) *are different for different horizons.*

7.3.2 Data Driven Reduced Order Modeling: Duffing Oscillator

In this subsection, we consider the problem of interpolating observed distributional data by identifying dynamical models over each finite horizon, in the absence of *a priori* structural knowledge (unlike the previous case study) about the models. The only choice the modeler can make is to decide whether a discrete-time or continuous-time model is apt. Once this choice is made, a *deterministic* trajectory-level model is desired that satisfies the two point boundary value problem in the output PDF level, at the beginning and end of the horizon length. Notice that we restrict ourselves to derive a deterministic flow or map, even though the observed PDFs may have been generated by a true but unknown state dynamics governed by PDE or SDE. In this sense, the modeling problem can be thought of as a sequence of finite-horizon distributional realization problems.

Consider the case when the true dynamics is given by the Duffing oscillator

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = -\alpha x_1^3 - \beta x_1 - \delta x_2,\, y = \{x_1,\, x_2\}^{\top}, \tag{7.34}$$

where $\alpha = 1$, $\beta = -1$, $\delta = 0.5$. One can verify that for these values of the parameters α, β, δ, the dynamics (7.34) has three equilibria: $(0, 0)$, $\left(\pm\sqrt{\frac{-\beta}{\alpha}}, 0\right)$. Linear stability analysis tells that the origin is a saddle node while the remaining two equilibria are stable foci. We use (7.34) only to generate synthetic data and assume that the knowledge of this true vector field is unavailable to the modeler.

To generate the true distributional PDFs, we assume that the initial joint state PDF $\xi_0(x_0) = \mathcal{U}\left([-2,2]^2\right)$. We generate 500 samples from this uniform PDF, and evaluate them at ξ_0. Starting from these samples, we evolve the joint state PDF $\xi(x_1(t), x_2(t), t)$ subject to (7.34) by solving the Liouville PDE $\frac{\partial \xi}{\partial t} + \nabla \cdot (\xi f) = 0$, where $f(x_1, x_2)$ is the Duffing vector field. We perform this uncertainty propagation by solving the method-of-characteristics ODE corresponding to the Liouville PDE, detailed in Chapter 2. This procedure results scattered colored data (Fig. 7.3) at every time t_j, $j = 1, 2, \ldots, 10$, where the location of the samples are determined from the dynamics while the color value at a sample location indicates the *exact* (unlike Monte Carlo histograms) joint PDF value at that sample location, at that time. Since $y = \{x_1,\, x_2\}^\top$, hence Fig. 7.3 depicts the sequence $\{t_j, \eta_j\}_{j=1}^{10}$.

Let Fig. 7.3 be the distributional data observed by the modeler. A continuous-time model is sought over each horizon: $t \in [t_j, t_{j+1})$. To solve this problem, we employ the Benamou-Brenier space-time optimization formulation of Section 7.1.4, resulting a vector field $v_j(x_1(t), x_2(t), t)$ per horizon, which solves the two point Liouville boundary value problem (guaranteeing end-point PDF matches) while incurring minimum amount of work over each $[t_j, t_{j+1})$. For this purpose, we take the two end point scattered data representation of η_j and η_{j+1}, and interpolate the data over a regular grid, followed by Douglas-Rachford proximal operator splitting algorithm [206] to solve the ensuing non-smooth convex optimization (7.5)-(7.6), resulting the vector field $v_j(x_1(t), x_2(t), t)$. Fig. 7.4 and 7.5 show the gridded observed PDFs and the intermediate PDF reconstructions for $(t_1, \eta_1) \to (t_2, \eta_2)$, and $(t_8, \eta_8) \to (t_9, \eta_9)$, respectively, superimposed with their respective Benamou-Brenier vector fields (*black arrows*). In Fig. 7.6, we compare the PDF transportation paths for $t \in [t_1, t_2)$ in W, for the true Duffing dynamics (7.34) and the optimal transport dynamics. In view of Remark 20, this plot shows that unlike the Brenier-Benamou *gradient* vector

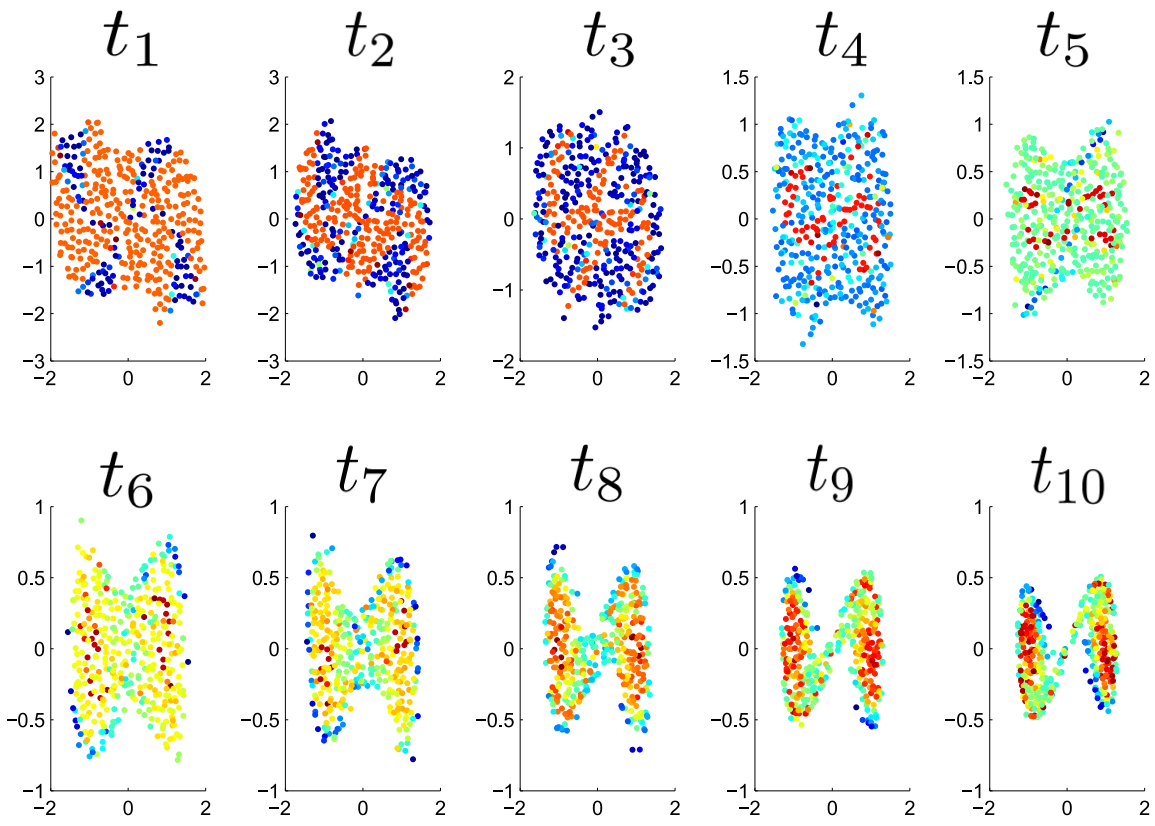


Figure 7.3: The distributional scattered data $\{t_j, \eta_j\}_{j=1}^{10}$ obtained by solving Liouville PDE for Duffing dynamics. The color value indicates the magnitude (*red = high, blue = low*) of the joint PDF η_j. In our simulation, $t_j = \frac{j}{2}$, where $j = 1, 2, \ldots, 10$.

field, (7.34) does not result into geodesic PDF transport. This is not surprising, since $\nabla \times f(x_1, x_2) = (-3\alpha x_1^2 - \beta - 1)\,\widehat{\mathbf{k}} = -3x_1^2\,\widehat{\mathbf{k}}$, i.e. Duffing vector field has non-zero vorticity everywhere except $x_1 = 0$, thus causing a clockwise rotational flow that requires more transportation effort than what could be achieved by a gradient flow.

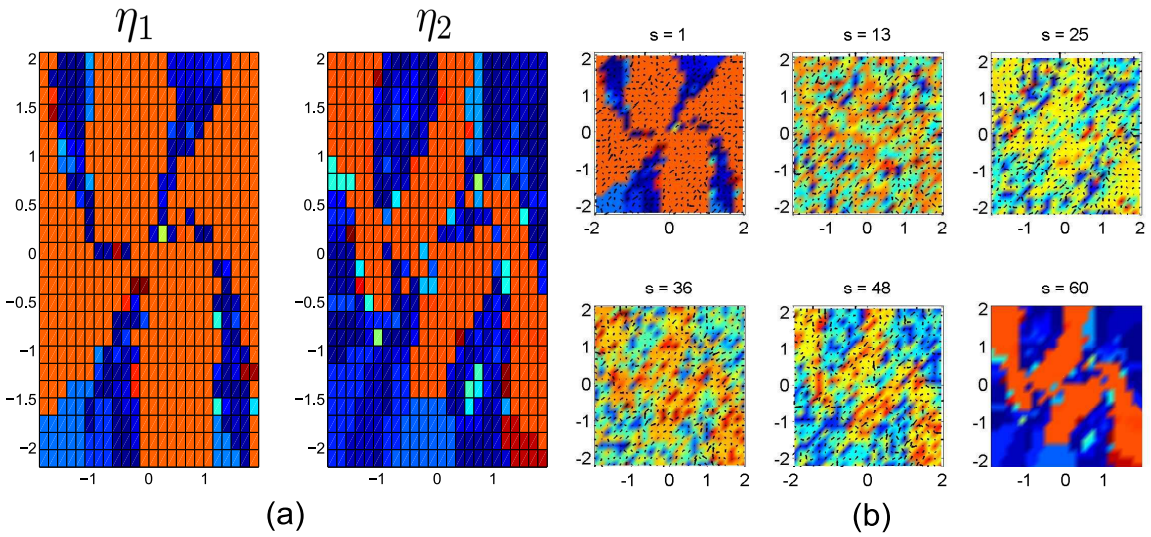


Figure 7.4: (a) The gridded PDFs η_1 and η_2; (b) The background color (*red = high, blue = low*) shows optimal transport reconstructions for PDF $\eta(t)$, $t \in [t_1, t_2)$, superimposed with Benamou-Brenier vector field $v_1^\star$ (*black arrows*). The interval $[t_1, t_2)$ was subdivided into 60 divisions, denoted by the index s above, i.e. $s = 0 \Leftrightarrow t_1$, $s = 60 \Leftrightarrow t_2$. Notice that the vector field vanishes at t_2.

7.4 Chapter Summary

In this Chapter, we showed that the optimal transport map can be used for automatically refining a proposed model. Such model refinement is consistent with the V&V formulation proposed in the previous Chapters in the sense that the model refinement at every time occurs in a way to drive the instantaneous Wasserstein gap to zero. Several equivalent formulations are pointed out and it is shown that

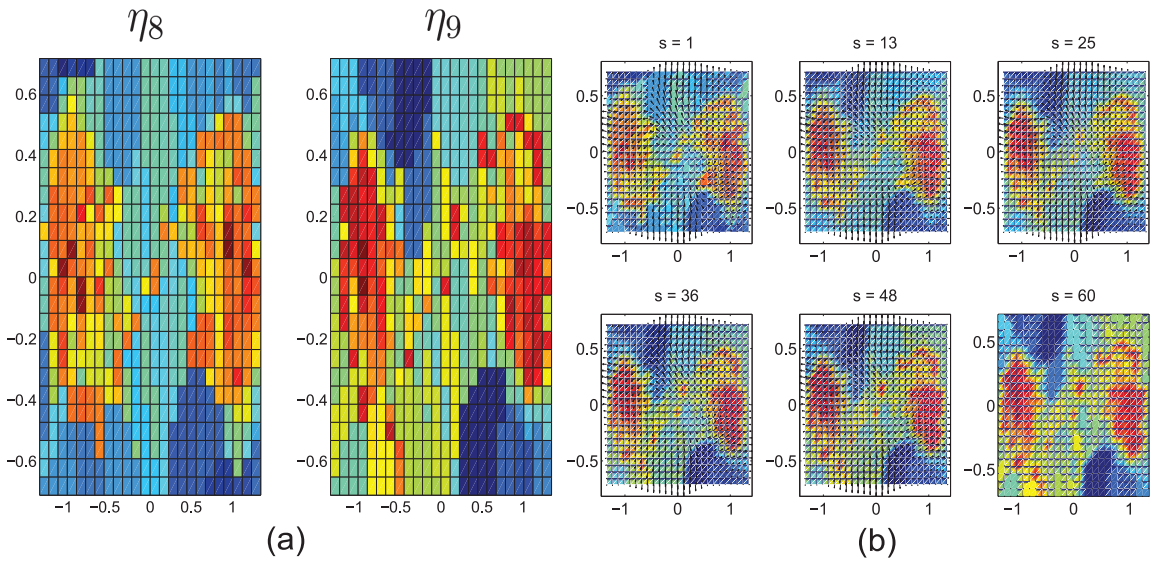


Figure 7.5: (a) The gridded PDFs η_8 and η_9; (b) The background color (*red = high, blue = low*) shows optimal transport reconstructions for PDF $\eta(t)$, $t \in [t_8, t_9)$, superimposed with Benamou-Brenier vector field $v_8^\star$ (*black arrows*). Like Fig. 7.4, $s = 0 \Leftrightarrow t_8$, $s = 60 \Leftrightarrow t_9$. Again, the vector field vanishes at t_9.

closed form solutions can be obtained for refining linear model against Gaussian measurements. For nonlinear model refinement, a non-smooth convex optimization framework was found to be more amenable for numerical implementation. Two case studies were provided: one showing how these ideas can enable feedback control for linear Gaussian PDF tracking, and the other showing data driven nonlinear reduced order modeling.

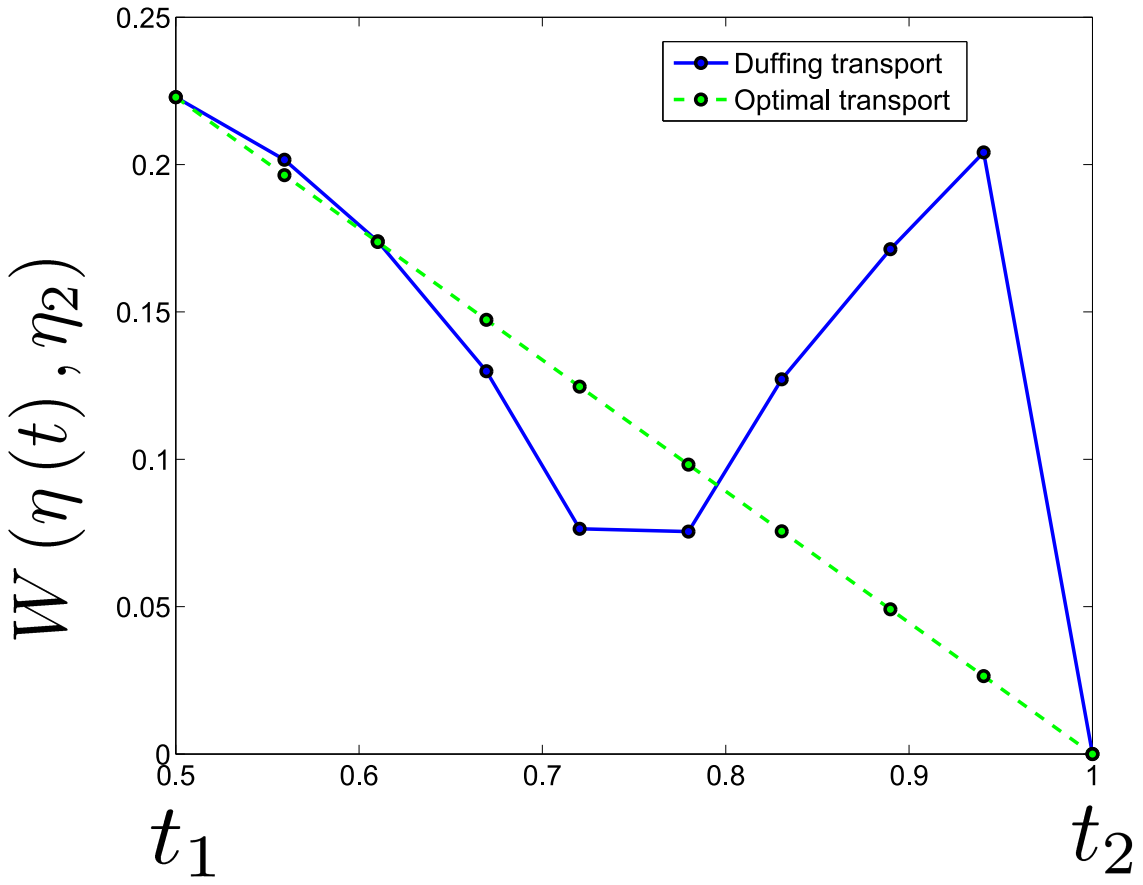


Figure 7.6: Comparison of optimal transport (Benamou-Brenier dynamics) with Duffing transport (true dynamics) for $t \in [t_1, t_2)$.

8. CONCLUSIONS

8.1 Summary of Contributions

In the 2010 U.S. Air Force Report on Technology Horizons [212], it was clearly pointed out that "It is possible to develop systems having high levels of autonomy, but *it is the lack of suitable V&V methods* that prevents all but relatively low levels of autonomy from being certified for use." Indeed, the convergence of communication, control and computation in all vistas of the 21^{st} century lifestyle, has brought the question of systems safety and performance reliability at the forefront of technological challenges. As systems are getting more interconnected and the underlying estimation-control software, and the communication protocols are getting increasingly sophisticated, the need for a "systems level V&V science" is becoming acute.

In this dissertation, we have presented a unifying V&V framework that can be used for performance certification of both deterministic and stochastic dynamical systems with uncertainties. The motivation for a probabilistic V&V methodology was outlined in Chapter 1, and the scope of this work in the context of existing systems-control literature was pointed out. The proposed framework is built on novel methods of uncertainty propagation subject to nonlinear dynamics, and hence Chapter 2 and 3 were devoted in developing new algorithms for uncertainty propagation. Since this is a topic of independent research interest, various case studies were presented to illustrate the practical utility of these uncertainty propagation algorithms, in the context of risk assessment and nonlinear estimation. In Chapter 4, based on axiomatic requirements for model validation, we argued that the Wasserstein metric related to optimal mass transport, is a natural notion of distance between the observed and model predicted concentration or "density" of trajectories, and hence

between the plant and model under comparison. It was shown that the resulting computation amounts to solving an LP at each time of measurement availability, and was demonstrated to fit naturally with the meshless uncertainty propagation algorithms developed in earlier chapters. Illustrative numerical examples were given, and ideas from the analysis of randomized algorithms were used to give probabilistic guarantees on the V&V inference. Chapter 5 provided a case study on controller robustness verification for F-16 aircraft. In Chapter 6, we focused on comparing linear systems in the proposed V&V framework. We showed that closeness of two linear systems in Wasserstein sense, means closeness in their average gains. Also, it was found that among a set of given initial PDFs, the one that best discriminates two linear systems, has maximum second raw moment. Chapter 7 extended the V&V framework to enable automatic model refinement and it was shown that the same ideas can be used for finite time PDF tracking in minimum effort sense.

8.2 Scope for Future Research

Below we outline the research directions that can be pursued based on the results of this dissertation.

8.2.1 Compositionality in Probabilistic V&V

The ideas presented in this dissertation, are envisioned for *component level* V&V applications in the presence of uncertainties. The final goal is to enable "trust in autonomy" through certifiable V&V. This brings forth the notion of compositionality in V&V. The key idea here is to perform V&V only at the component level, but such exercise should come with the guarantee of system performance when connected with other systems. The need for this compositionality has been long recognized [212]: "In effect, the number of possible input states that such systems can be presented with is so large that not only is it impossible to test all of them directly, it is not even

possible to test more than an insignificantly small fraction of them. Development of such systems is thus inherently unverifiable by todays methods, and as a result their operation in all but comparatively trivial applications is uncertifiable."

Some specific research questions in this regard, are listed below.

1. When does open loop validation (in Wasserstein sense) implies closed loop validation? In particular, it will be of interest to investigate the following scenario. Suppose a model has been (open loop) validated with γ level of Wasserstein closeness. Then, if a controller is designed for that model with some notion of "goodness", such as stabilizing controller, does it imply that the same notion of "goodness" gets translated when the designed controller is commissioned to the actual plant? A positive answer to this question, even for a given class of models/controllers, would mean that there will be no need for repeating the V&V exercise for the final closed-loop implementation. In other words, the performance would be "certifiably correct by construction". Different notions of "goodness" for the controller can be investigated, e.g. $\mathcal{H}_2/\mathcal{H}_\infty$ controllers, passivity etc. Also, different notions of interconnections can be investigated, for a given choice of controller.

2. Another direction is to generalize the systems-theoretic results of Section 6.1, for different input classes, e.g. bounded energy input signal $u(t)$ having finite $L_2[0,T]$ norm. In particular, we would like to understand the connection between finite-time incremental stochastic L_2 gain and the magnitude of Wasserstein distance. It is envisaged that such a pursuit would reveal a connection between dissipativity, small gain theorem, and the phrase "${}_2W_2(t) \leq \gamma$". In effect, this would help in answering "what does the value of Wasserstein distance mean for a nonlinear system?"

8.2.2 Nonlinear Model Reduction via Optimal Transport

A promising way to extend ideas of Section 7 is to use Brenier-Benamou space-time variational formulation to derive *gradient nonlinear models* from data, which are naturally reduced order (ODE over output space). We plan to compare such optimal transport theoretic nonlinear model reduction with popular *linear* model reduction algorithms, such as POD and DMD. Also, it would be interesting to consider such nonlinear model reduction in the context of preserving desirable quantities, such as passivity, stability, controllability and observability, which are currently accounted through "balanced" variants of POD and DMD. On a similar vein, one could design *nonlinear* controllers for the piecewise gradient models derived from data via Brenier-Benamou algorithm, then guarantee the closed-loop performance. The fact that reduced order models are piecewise gradient, could lead to controller design using piecewise Lyapunov functions.

8.2.3 Connections with Formal Verification

A direction which we have not explored in this dissertation, is how to bring together formal V&V methods such as the linear temporal logic specifications, with the probabilistic V&V ideas presented here. Connections with ideas from computer science community, such as probabilistic model checking and bisimulation, could be investigated.

8.2.4 Application to Fault Detection, Isolation, and Reconfiguration

Another potential research direction could be applying the optimal transport based V&V formulation for density based fault detection, isolation and reconfiguration. The detection could be based on the Wasserstein distance computation, while the reconfiguration strategy could be thought of as Brenier map. In most applica-

tions, fault scenarios are enumerable (e.g. by constructing a fault tree), and if the faults are assumed to occur randomly, due to execution level uncertainties, then the fault isolation could be interpreted as the mode estimation problem for a stochastic jump system.

REFERENCES

[1] R. D. Braun and R. M. Manning, "Mars exploration entry, descent, and landing challenges," *Journal of Spacecraft and Rockets*, vol. 44, no. 2, pp. 310–323, 2007.

[2] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, vol. 63. SIAM, Philadelphia, 1992.

[3] R. J. McCann, "A convexity principle for interacting gases," *Advances in Mathematics*, vol. 128, no. 1, pp. 153–179, 1997.

[4] K. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge.* Routledge, London and New York, 1963.

[5] R. S. Smith and J. C. Doyle, "Model validation: a connection between robust control and identification," *Automatic Control, IEEE Transactions on*, vol. 37, no. 7, pp. 942–952, 1992.

[6] K. Poolla, P. Khargonekar, A. Tikku, J. Krause, and K. Nagpal, "A time-domain approach to model validation," *Automatic Control, IEEE Transactions on*, vol. 39, no. 5, pp. 951–959, 1994.

[7] S. Prajna, "Barrier certificates for nonlinear model validation," *Automatica*, vol. 42, no. 1, pp. 117–126, 2006.

[8] P. B. Brugarolas and M. G. Safonov, "A data driven approach to learning dynamical systems (invited paper)," in *IEEE Conference on Decision and Control*, vol. 4, pp. 4670–4675, IEEE; 1998, 2002.

[9] D. Sornette, A. Davis, K. Ide, K. Vixie, V. Pisarenko, and J. Kamm, "Algorithm for model validation: Theory and applications," *Proceedings of the National Academy of Sciences*, vol. 104, no. 16, pp. 6562–6567, 2007.

[10] Committee on Mathematical Foundations of Verification, Validation, and Uncertainty Quantification and Board on Mathematical Sciences and Their Applications and Division on Engineering and Physical Sciences and National Research Council (US), *Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification*. National Academies Press, Washington D.C., 2012.

[11] C. Baier and J.-P. Katoen, *Principles of Model Checking*. MIT Press, Cambridge, Massachusetts, 2008.

[12] F. Ciesinski and M. Größer, "On probabilistic computation tree logic," in *Validation of Stochastic Systems*, pp. 147–188, Springer, 2004.

[13] R. Smith and G. E. Dullerud, "Continuous-time control model validation using finite experimental data," *Automatic Control, IEEE Transactions on*, vol. 41, no. 8, pp. 1094–1105, 1996.

[14] J. Chen and S. Wang, "Validation of linear fractional uncertain models: Solutions via matrix inequalities," *Automatic Control, IEEE Transactions on*, vol. 41, no. 6, pp. 844–849, 1996.

[15] B. Wahlberg and L. Ljung, "Hard frequency-domain model error bounds from least-squares like identification techniques," *Automatic Control, IEEE Transactions on*, vol. 37, no. 7, pp. 900–912, 1992.

[16] D. Xu, Z. Ren, G. Gu, and J. Chen, "LFT uncertain model validation with time-and frequency-domain measurements," *Automatic Control, IEEE Transactions on*, vol. 44, no. 7, pp. 1435–1441, 1999.

[17] S. L. V. Campbell and S. Campbell, *Singular Systems of Differential Equations*, vol. 1. Pitman, London, 1980.

[18] A. Megretski and A. Rantzer, "System analysis via integral quadratic constraints," *Automatic Control, IEEE Transactions on*, vol. 42, no. 6, pp. 819–830, 1997.

[19] B. Øksendal, *Stochastic Differential Equations*. Springer, Heidelberg, Germany, 2003.

[20] A. J. Van Der Schaft, J. M. Schumacher, A. J. van der Schaft, and A. J. van der Schaft, *An Introduction to Hybrid Dynamical Systems*, vol. 251. Springer, London, 2000.

[21] L. Ljung, "System identification: theory for the user," *Information and System Sciences Series*, vol. 7632, 1987.

[22] L. Ljung and L. Guo, "The role of model validation for assessing the size of the unmodeled dynamics," *Automatic Control, IEEE Transactions on*, vol. 42, no. 9, pp. 1230–1239, 1997.

[23] R. G. Ghanem, A. Doostan, and J. Red-Horse, "A probabilistic construction of model validation," *Computer Methods in Applied Mechanics and Engineering*, vol. 197, no. 29, pp. 2585–2595, 2008.

[24] M. Gevers, X. Bombois, B. Codrons, G. Scorletti, and B. Anderson, "Model validation for control and controller validation in a prediction error identification framework – part i: theory," *Automatica*, vol. 39, no. 3, pp. 403–415, 2003.

[25] L. Lee and K. Poolla, "On statistical model validation," *Journal of Dynamic Systems, Measurement, and Control*, vol. 118, no. 2, pp. 226–236, 1996.

[26] J. Van Schuppen, "Stochastic realization problems," *Three Decades of Mathematical System Theory, Lecture Notes in Control and Information Sciences,*

H. Nijmeijer, J.M. Schumacher (eds.), vol. 135, pp. 480–523, 1989.

[27] V. Ugrinovskii, "Risk-sensitivity conditions for stochastic uncertain model validation," *Automatica*, vol. 45, no. 11, pp. 2651–2658, 2009.

[28] A. Lasota and M. C. Mackey, *Chaos, fractals, and noise: stochastic aspects of dynamics*, vol. 97. Springer, New York, 1994.

[29] S. Yu and P. G. Mehta, "The Kullback–Leibler rate pseudo-metric for comparing dynamical systems," *Automatic Control, IEEE Transactions on*, vol. 55, no. 7, pp. 1585–1598, 2010.

[30] T. T. Georgiou, "Distances and riemannian metrics for spectral density functions," *Signal Processing, IEEE Transactions on*, vol. 55, no. 8, pp. 3995–4003, 2007.

[31] H. Wang, "Robust control of the output probability density functions for multivariable stochastic systems with guaranteed stability," *Automatic Control, IEEE Transactions on*, vol. 44, no. 11, pp. 2103–2107, 1999.

[32] H. Wang, H. Baki, and P. Kabore, "Control of bounded dynamic stochastic distributions using square root models: an applicability study in papermaking systems," *Transactions of the Institute of Measurement and Control*, vol. 23, no. 1, pp. 51–68, 2001.

[33] J.-S. Li and N. Khaneja, "Ensemble control of bloch equations," *Automatic Control, IEEE Transactions on*, vol. 54, no. 3, pp. 528–536, 2009.

[34] E. Brown, J. Moehlis, and P. Holmes, "On the phase reduction and response dynamics of neural oscillator populations," *Neural Computation*, vol. 16, no. 4, pp. 673–715, 2004.

[35] S. A. Wadoo and P. Kachroo, "Feedback control of crowd evacuation in one dimension," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 11, no. 1, pp. 182–193, 2010.

[36] P. A. Parrilo, *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, 2000.

[37] A. Halder and R. Bhattacharya, "Beyond Monte Carlo: A computational framework for uncertainty propagation in planetary entry, descent and landing," in *AIAA Guidance, Navigation and Control Conference, Toronto, ON*, 2010.

[38] A. Halder and R. Bhattacharya, "Dispersion analysis in hypersonic flight during planetary entry using stochastic Liouville equation," *Journal of Guidance, Control, and Dynamics*, vol. 34, no. 2, pp. 459–474, 2011.

[39] P. Dutta, A. Halder, and R. Bhattacharya, "Uncertainty quantification for stochastic nonlinear systems using Perron-Frobenius operator and Karhunen-Loève expansion," in *Control Applications (CCA), 2012 IEEE International Conference on*, pp. 1449–1454, IEEE, 2012.

[40] P. Dutta, A. Halder, and R. Bhattacharya, "Nonlinear filtering with transfer operator," in *American Control Conference (ACC), 2013*, pp. 3069–3074, IEEE, 2013.

[41] A. Halder and R. Bhattacharya, "Model validation: A probabilistic formulation," in *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pp. 1692–1697, IEEE, 2011.

[42] A. Halder and R. Bhattacharya, "Probabilistic model validation for uncertain nonlinear systems (under review)," *Automatica*, 2014.

[43] A. Halder, K. Lee, and R. Bhattacharya, "Probabilistic robustness analysis of F-16 controller performance: An optimal transport approach," in *American Control Conference (ACC), 2013*, pp. 5562–5567, IEEE, 2013.

[44] A. Halder and R. Bhattacharya, "A probabilistic method for nonlinear robustness analysis of F-16 controllers (under review)," *Journal of Guidance, Control, and Dynamics*, 2014.

[45] A. Halder and R. Bhattacharya, "Further results on probabilistic model validation in Wasserstein metric," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pp. 5542–5547, IEEE, 2012.

[46] A. Halder and R. Bhattacharya, "Frequency domain model validation in Wasserstein metric," in *American Control Conference (ACC), 2013*, pp. 5845–5850, IEEE, 2013.

[47] A. Halder and R. Bhattacharya, "Geodesic density tracking with applications to data driven modeling (invited paper)," in *American Control Conference (ACC), 2014*, IEEE, 2014.

[48] D. A. Spencer and R. D. Braun, "Mars Pathfinder atmospheric entry-trajectory design and dispersion analysis," *Journal of Spacecraft and Rockets*, vol. 33, no. 5, pp. 670–676, 1996.

[49] S. A. Striepe, D. Way, A. Dwyer, and J. Balaram, "Mars Science Laboratory simulations for entry, descent, and landing," *Journal of Spacecraft and Rockets*, vol. 43, no. 2, pp. 311–323, 2006.

[50] M. Ehrendorfer, "The Liouville equation and its potential usefulness for the prediction of forecast skill. part i: Theory," *Monthly Weather Review*, vol. 122, no. 4, pp. 703–713, 1994.

[51] M. Ehrendorfer, "The Liouville equation and its potential usefulness for the prediction of forecast skill. part ii: Applications," *Monthly Weather Review*, vol. 122, no. 4, pp. 714–728, 1994.

[52] O. Knio and O. Le Maitre, "Uncertainty propagation in CFD using polynomial chaos decomposition," *Fluid Dynamics Research*, vol. 38, no. 9, pp. 616–640, 2006.

[53] J. Ching, K. A. Porter, and J. L. Beck, "Propagating uncertainties for loss estimation in performance-based earthquake engineering using moment matching," *Structure and Infrastructure Engineering*, vol. 5, no. 3, pp. 245–262, 2009.

[54] F. J. Regan and S. M. Anandakrishnan, *Dynamics of Atmospheric Re-entry*. AIAA, Washington, D.C., 1993.

[55] P. D. Spanos and R. Ghanem, "Stochastic finite element expansion for random media," *Journal of Engineering Mechanics*, vol. 115, no. 5, pp. 1035–1053, 1989.

[56] D. Xiu and J. S. Hesthaven, "High-order collocation methods for differential equations with random inputs," *SIAM Journal on Scientific Computing*, vol. 27, no. 3, pp. 1118–1139, 2005.

[57] B. J. Debusschere, H. N. Najm, P. P. Pébay, O. M. Knio, R. G. Ghanem, and O. P. Le Maître, "Numerical challenges in the use of polynomial chaos representations for stochastic processes," *SIAM Journal on Scientific Computing*, vol. 26, no. 2, pp. 698–719, 2004.

[58] D. L. Marchisio and R. O. Fox, "Solution of population balance equations using the direct quadrature method of moments," *Journal of Aerosol Science*, vol. 36, no. 1, pp. 43–73, 2005.

[59] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*, vol. 1. American Mathematical Society, New York, 1943.

[60] A. Mura and R. Borghi, "Introducing a new partial PDF approach for turbulent combustion modeling," *Combustion and Flame*, vol. 136, no. 3, pp. 377–382, 2004.

[61] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, vol. 26. Chapman and Hall, London, 1986.

[62] C. Pantano and B. Shotorban, "Least-squares dynamic approximation method for evolution of uncertainty in initial conditions of dynamical systems," *Physical Review E*, vol. 76, no. 6, pp. 066705–1 – 066705–13, 2007.

[63] B. Epstein, *Partial Differential Equations – An Introduction.* McGraw-Hill, New York, 1962.

[64] M. Delgado, "Classroom note: The Lagrange–Charpit method," *SIAM review*, vol. 39, no. 2, pp. 298–304, 1997.

[65] O. Babelon, D. Bernard, and M. Talon, *Introduction to Classical Integrable Systems.* Cambridge University Press, Cambridge, 2003.

[66] P. Gaspard, *Chaos, Scattering and Statistical Mechanics*, vol. 9. Cambridge University Press, Cambridge, 2005.

[67] M. Ehrendorfer, "The Liouville equation and atmospheric predictability," *Predictability of Weather and Climate. Ed. Tim Palmer and Renate Hagedorn*, pp. 59–98, 2006.

[68] V. I. Arnol'd, *Mathematical Methods of Classical Mechanics*, vol. 60. Springer, New York, 1989.

[69] L. Devroye, *Non-uniform Random Variate Generation*. Springer, New York, 1986.

[70] P. Diaconis, "The Markov Chain Monte Carlo revolution," *Bulletin of the American Mathematical Society*, vol. 46, no. 2, pp. 179–205, 2009.

[71] S. Chib and E. Greenberg, "Understanding the Metropolis-Hastings algorithm," *The American Statistician*, vol. 49, no. 4, pp. 327–335, 1995.

[72] C. Andrieu, N. De Freitas, A. Doucet, and M. I. Jordan, "An introduction to MCMC for machine learning," *Machine Learning*, vol. 50, no. 1-2, pp. 5–43, 2003.

[73] P. N. Desai, R. D. Braun, R. W. Powell, W. C. Engelund, and P. V. Tartabini, "Six-degree-of-freedom entry dispersion analysis for the METEOR recovery module," *Journal of Spacecraft and Rockets*, vol. 34, no. 3, pp. 334–340, 1997.

[74] P. N. Desai, R. A. Mitcheltree, and F. M. Cheatwood, "Entry dispersion analysis for the Stardust comet sample return capsule," *Journal of Spacecraft and Rockets*, vol. 36, no. 3, pp. 463–469, 1999.

[75] R. D. Braun, R. A. Mitcheltree, and F. M. Cheatwood, "Mars microprobe entry-to-impact analysis," *Journal of Spacecraft and Rockets*, vol. 36, no. 3, pp. 412–420, 1999.

[76] S. A. Striepe, E. M. Queen, R. W. Powell, R. D. Braun, F. Cheatwood, J. T. Aguirre, L. A. Sachi, and D. T. Lyons, "An atmospheric guidance algorithm testbed for the Mars Surveyor Program 2001 Orbiter and Lander," *AIAA Paper*, pp. 98–4569, 1998.

[77] P. N. Desai and P. C. Knocke, "Mars Exploration Rovers entry, descent, and landing trajectory analysis," *The Journal of the Astronautical Sciences*, vol. 55, no. 3, pp. 311–323, 2007.

[78] W. M. Congdon, "Ablation model validation and analytical sensitivity study for the Mars Pathfinder heat shield," in *AIAA, Thermophysics Conference, 30 th, San Diego, CA*, 1995.

[79] R. Powell *et al.*, "Program to optimize simulated trajectories (POST) utilization manual, volume ii, version 5.2," *NASA Langley Research Center, Hampton, VA*, 1997.

[80] J. Balaram, R. Austin, P. Banerjee, T. Bentley, D. Henriquez, B. Martin, E. McMahon, and G. Sohl, "DSENDS – a high-fidelity dynamics and spacecraft simulator for entry, descent and surface landing," in *Aerospace Conference Proceedings, 2002. IEEE*, vol. 7, pp. 7–3343, IEEE, 2002.

[81] N. Vinh, A. Busemann, and R. Culp, *Hypersonic & Planetary Entry Flight Mechanics*. The University of Michigan Press, Ann Arbor, 1980.

[82] M. Noton, *Spacecraft Navigation and Guidance*. Springer Verlag, New York, 1998.

[83] J. H. Halton, "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals," *Numerische Mathematik*, vol. 2, no. 1, pp. 84–90, 1960.

[84] P. Diaconis and L. Saloff-Coste, "What do we know about the Metropolis algorithm?," in *Proceedings of the twenty-seventh annual ACM symposium on Theory of computing*, pp. 112–129, ACM, 1995.

[85] R. J. Glauber, "Time-dependent statistics of the Ising model," *Journal of Mathematical Physics*, vol. 4, no. 2, pp. 294–307, 2004.

[86] V. F. Turchin, "On the computation of multidimensional integrals by the Monte-Carlo method," *Theory of Probability & Its Applications*, vol. 16, no. 4, pp. 720–724, 1971.

[87] P. Diaconis, K. Khare, L. Saloff-Coste, *et al.*, "Gibbs sampling, exponential families and orthogonal polynomials," *Statistical Science*, vol. 23, no. 2, pp. 151–178, 2008.

[88] M. K. Cowles and B. P. Carlin, "Markov chain Monte Carlo convergence diagnostics: a comparative review," *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 883–904, 1996.

[89] G. O. Roberts, J. S. Rosenthal, *et al.*, "General state space Markov chains and MCMC algorithms," *Probability Surveys*, vol. 1, pp. 20–71, 2004.

[90] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. Cambridge University Press, New York, 2009.

[91] A. Halder and R. Bhattacharya, "Performance bounds for dispersion analysis: A comparison between Monte Carlo and Perron-Frobenius operator approach," *Preprint*, 2011.

[92] C. W. Gardiner, *Handbook of Stochastic Methods*, vol. 3. Springer, Berlin, 1985.

[93] D. Liberzon and R. W. Brockett, "Nonlinear feedback systems perturbed by noise: steady-state probability distributions and optimal control," *Automatic Control, IEEE Transactions on*, vol. 45, no. 6, pp. 1116–1130, 2000.

[94] R. E. Bellman, *Dynamic Programming*. Princeton University Press, Princeton, New Jersey, 1957.

[95] D. Liberzon and R. W. Brockett, "Spectral analysis of Fokker–Planck and related operators arising from linear stochastic differential equations," *SIAM Journal on Control and Optimization*, vol. 38, no. 5, pp. 1453–1467, 2000.

[96] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*. vol. 18 of Springer Series in Synergetics, Springer, Berlin, 1989.

[97] K. Karhunen, "Über lineare methoden in der wahrscheinlichkeitsrechnung," *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys*, vol. 37, pp. 1–79, 1947.

[98] S. Lall, J. Marsden, and S. Glavaski, "A subspace approach to balanced truncation for model reduction of nonlinear control systems," *International Journal of Robust and Nonlinear Control*, vol. 12, no. 6, pp. 519–535, 2002.

[99] D. Kosambi, "Statistics in function space," *J. Indian Math. Soc*, vol. 7, no. 1, pp. 76–88, 1943.

[100] M. Kac and A. Siegert, "An explicit representation of a stationary Gaussian process," *The Annals of Mathematical Statistics*, vol. 18, no. 3, pp. 438–442, 1947.

[101] M. Loéve, *Fonctions Aléatoires du Second Ordre*. Supplement to P. Lévy, Processus Stochastic et Mouvement Brownien, Gauthier-Villars, Paris, 1948.

[102] J. Lumley, *Stochastic Tools in Turbulence*. Academic Press, New York, 1970.

[103] N. Cressie, "Statistics for spatial data," *Terra Nova*, vol. 4, no. 5, pp. 613–617, 1992.

[104] M. Grigoriu, *Stochastic Calculus: Applications in Science and Engineering*. Birkhäuser, Boston, 2002.

[105] A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*. Prentice Hall, New Jersey, 1994.

[106] M. Grigoriu, "White noise processes," *Journal of Engineering Mechanics*, vol. 113, no. 5, pp. 757–765, 1987.

[107] A. Justel, D. Peña, and R. Zamar, "A multivariate Kolmogorov-Smirnov test of goodness of fit," *Statistics & Probability Letters*, vol. 35, no. 3, pp. 251–259, 1997.

[108] R. H. Lopes, I. Reid, and P. R. Hobson, "The two-dimensional Kolmogorov-Smirnov test," in *XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research, Nikhef, Amsterdam, the Netherlands, April 23-27, 2007*, 2007.

[109] A. Dvoretzky, J. Kiefer, and J. Wolfowitz, "Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator," *The Annals of Mathematical Statistics*, pp. 642–669, 1956.

[110] R. G. Ghanem and P. D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, vol. 41. Springer, New York, 1991.

[111] P. Frauenfelder, C. Schwab, and R. A. Todor, "Finite elements for elliptic problems with stochastic coefficients," *Computer Methods in Applied Mechanics and Engineering*, vol. 194, no. 2, pp. 205–228, 2005.

[112] M. Abramowitz, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables. Edited by Milton Abramowitz and Irene A. Stegun.* US Government Printing Office, Washington D.C., 1967.

[113] C. Schwab and R. A. Todor, "Karhunen–Loève approximation of random fields by generalized fast multipole methods," *Journal of Computational Physics*,

vol. 217, no. 1, pp. 100–122, 2006.

[114] M. Sanjeev Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *Signal Processing, IEEE Transactions on*, vol. 50, no. 2, pp. 174–188, 2002.

[115] F. Daum and J. Huang, "Curse of dimensionality and particle filters," in *Proceedings of the 2003 IEEE Aerospace Conference*, vol. 4, pp. 4_1979–4_1993, IEEE, 2003.

[116] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.

[117] C. Manohar and D. Roy, "Monte Carlo filters for identification of nonlinear structural dynamical systems," *Sadhana*, vol. 31, no. 4, pp. 399–427, 2006.

[118] N. J. Gordon, D. J. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-gaussian Bayesian state estimation," in *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, pp. 107–113, IET, 1993.

[119] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, Boston, 2004.

[120] N. Oudjane and C. Musso, "Progressive correction for regularized particle filters," in *Information Fusion (FUSION 2000) , Proceedings of the Third International Conference on*, vol. 2, pp. THB2–10, IEEE, 2000.

[121] W. R. Gilks and C. Berzuini, "Following a moving target – monte carlo inference for dynamic Bayesian models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 1, pp. 127–146, 2001.

[122] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson, "Obstacles to high-dimensional particle filtering.," *Monthly Weather Review*, vol. 136, no. 12, 2008.

[123] G. Casella and C. P. Robert, "Rao-Blackwellisation of sampling schemes," *Biometrika*, vol. 83, no. 1, pp. 81–94, 1996.

[124] F. E. Daum, "Exact finite-dimensional nonlinear filters," *Automatic Control, IEEE Transactions on*, vol. 31, no. 7, pp. 616–622, 1986.

[125] V. Beneš, "Exact finite-dimensional filters for certain diffusions with nonlinear drift," *Stochastics: An International Journal of Probability and Stochastic Processes*, vol. 5, no. 1-2, pp. 65–92, 1981.

[126] A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering*, vol. 3. Springer, New York, 2009.

[127] D. Crisan, "A direct computation of the Beneš filter conditional density," *Stochastics: An International Journal of Probability and Stochastic Processes*, vol. 55, no. 1-2, pp. 47–54, 1995.

[128] K. P. Bollino, I. M. Ross, and D. D. Doman, "Optimal nonlinear feedback guidance for reentry vehicles," in *Proceeding of AIAA Guidance, Navigation, and Control Conference and Exhibit*, pp. 563–582, 2006.

[129] A. L. Gibbs and F. E. Su, "On choosing and bounding probability metrics," *International Statistical Review*, vol. 70, no. 3, pp. 419–435, 2002.

[130] I. Csiszár, "Information-type measures of difference of probability distributions and indirect observations," *Studia Sci. Math. Hungar.*, vol. 2, pp. 299–318, 1967.

[131] A. Müller, "Integral probability metrics and their generating classes of functions," *Advances in Applied Probability*, pp. 429–443, 1997.

[132] C. Villani, *Topics in optimal transportation*. No. 58, American Mathematical Society, Providence, Rhode Island, 2003.

[133] R. Jordan, D. Kinderlehrer, and F. Otto, "The variational formulation of the Fokker–Planck equation," *SIAM Journal on Mathematical Analysis*, vol. 29, no. 1, pp. 1–17, 1998.

[134] S. T. Rachev, *Probability Metrics and the Stability of Stochastic Models*, vol. 334. Wiley, New York, 1991.

[135] Q. Wang, S. R. Kulkarni, and S. Verdú, "Divergence estimation of continuous distributions based on data-dependent partitions," *Information Theory, IEEE Transactions on*, vol. 51, no. 9, pp. 3064–3074, 2005.

[136] X. Nguyen, M. J. Wainwright, and M. I. Jordan, "Estimating divergence functionals and the likelihood ratio by convex risk minimization," *Information Theory, IEEE Transactions on*, vol. 56, no. 11, pp. 5847–5861, 2010.

[137] A. V. Lazo and P. Rathie, "On the entropy of continuous probability distributions (corresp.)," *Information Theory, IEEE Transactions on*, vol. 24, no. 1, pp. 120–122, 1978.

[138] C. R. Givens and R. M. Shortt, "A class of Wasserstein metrics for probability distributions.," *The Michigan Mathematical Journal*, vol. 31, no. 2, pp. 231–240, 1984.

[139] R. Kulhavy, *Recursive Nonlinear Estimation: A Geometric Approach*. Springer, Berlin, 1996.

[140] A. Poznyak, *Advanced Mathematical Tools for Automatic Control Engineers: Deterministic Systems (vol. 1)*. Elselvier Science, Amsterdam, 2008.

[141] B.-W. Hong, S. Soatto, K. Ni, and T. Chan, "The scale of a texture and its application to segmentation," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1–8, IEEE, 2008.

[142] K. Ni, X. Bresson, T. Chan, and S. Esedoglu, "Local histogram based segmentation using the Wasserstein distance," *International Journal of Computer Vision*, vol. 84, no. 1, pp. 97–111, 2009.

[143] S. Vallender, "Calculation of the Wasserstein distance between probability distributions on the line," *Theory of Probability & Its Applications*, vol. 18, no. 4, pp. 784–786, 1974.

[144] S. T. Rachev, "The Monge-Kantorovich mass transference problem and its stochastic applications," *Theory of Probability & Its Applications*, vol. 29, no. 4, pp. 647–676, 1985.

[145] F. L. Hitchcock, "The distribution of a product from several sources to numerous localities," *J. Math. Phys*, vol. 20, no. 2, pp. 224–230, 1941.

[146] T. C. Koopmans, "Optimum utilization of the transportation system," *Econometrica: Journal of the Econometric Society*, pp. 136–146, 1949.

[147] T. C. Koopmans, "Analysis of production as an efficient combination of activities," *Activity Analysis of Production and Allocation*, vol. 13, pp. 33–37, 1951.

[148] B. K. Sriperumbudur, K. Fukumizu, A. Gretton, B. Schölkopf, G. R. Lanckriet, *et al.*, "On the empirical estimation of integral probability metrics," *Electronic Journal of Statistics*, vol. 6, pp. 1550–1599, 2012.

[149] B. K. Sriperumbudur, K. Fukumizu, A. Gretton, B. Schölkopf, and G. R. Lanckriet, "On integral probability metrics, ϕ-divergences and binary classifi-

cation," *arXiv preprint arXiv:0901.2698*, 2009.

[150] R. Motwani and P. Raghavan, *Randomized Algorithms*. Cambridge University Press, Cambridge, 1995.

[151] M. Talagrand, "Transportation cost for gaussian and other product measures," *Geometric & Functional Analysis GAFA*, vol. 6, no. 3, pp. 587–600, 1996.

[152] H. Djellout, A. Guillin, L. Wu, *et al.*, "Transportation cost-information inequalities and applications to random dynamical systems and diffusions," *Annals of Probability*, vol. 32, no. 3, pp. 2702–2732, 2004.

[153] E. Boissard and T. le Gouic, ""Exact" deviations in Wasserstein distance for empirical and occupation measures," *arXiv preprint arXiv:1103.3188v1*, 2011.

[154] R. E. Burkard, M. Dell'Amico, S. Martello, *et al.*, *Assignment Problems, Revised Reprint*. SIAM, Philadelphia, 2009.

[155] J. Rabin, G. Peyré, J. Delon, and M. Bernot, "Wasserstein barycenter and its application to texture mixing," in *Scale Space and Variational Methods in Computer Vision*, pp. 435–446, Springer, 2012.

[156] R. Tempo, G. Calafiore, and F. Dabbene, *Randomized Algorithms for Analysis and Control of Uncertain Systems*. Springer, London, 2004.

[157] P. Khargonekar and A. Tikku, "Randomized algorithms for robust control analysis and synthesis have polynomial complexity," in *Decision and Control, 1996., Proceedings of the 35th IEEE Conference on*, vol. 3, pp. 3470–3475, IEEE, 1996.

[158] R. Tempo, E.-W. Bai, and F. Dabbene, "Probabilistic robustness analysis: Explicit bounds for the minimum number of samples," in *Decision and Control,*

1996., Proceedings of the 35th IEEE Conference on, vol. 3, pp. 3424–3428, IEEE, 1996.

[159] X. Chen and K. Zhou, "Order statistics and probabilistic robust control," *Systems & Control Letters*, vol. 35, no. 3, pp. 175–182, 1998.

[160] M. Vidyasagar, "Randomized algorithms for robust controller synthesis using statistical learning theory," *Automatica*, vol. 37, no. 10, pp. 1515–1528, 2001.

[161] S. Prajna, A. Papachristodoulou, and P. A. Parrilo, "Introducing SOSTOOLS: A general purpose sum of squares programming solver," in *Decision and Control, 2002, Proceedings of the 41st IEEE Conference on*, vol. 1, pp. 741–746, IEEE, 2002.

[162] A. Rantzer, "A dual to lyapunov's stability theorem," *Systems & Control Letters*, vol. 42, no. 3, pp. 161–168, 2001.

[163] A. Rantzer and S. Prajna, "On analysis and synthesis of safe control laws," in *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Allerton Conference, 2004.

[164] S. Prajna and A. Rantzer, "Convex programs for temporal verification of nonlinear dynamical systems," *SIAM Journal on Control and Optimization*, vol. 46, no. 3, pp. 999–1021, 2007.

[165] S. Prajna and A. Rantzer, "On the necessity of barrier certificates," in *Proceedings of the IFAC World Congress*, 2005.

[166] B. Barmish, "Probabilistic robustness: A new line of research," in *Control'96, UKACC International Conference on (Conf. Publ. No. 427)*, vol. 1, pp. 181–181, IET, 1996.

[167] G. C. Calafiore, F. Dabbene, and R. Tempo, "Randomized algorithms for probabilistic robustness with real and complex structured uncertainty," *Automatic Control, IEEE Transactions on*, vol. 45, no. 12, pp. 2218–2235, 2000.

[168] B. Polyak and R. Tempo, "Probabilistic robust design with linear quadratic regulators," *Systems & Control Letters*, vol. 43, no. 5, pp. 343–353, 2001.

[169] Q. Wang and R. F. Stengel, "Robust control of nonlinear systems with parametric uncertainty," *Automatica*, vol. 38, no. 9, pp. 1591–1599, 2002.

[170] Y. Fujisaki, F. Dabbene, and R. Tempo, "Probabilistic design of LPV control systems," *Automatica*, vol. 39, no. 8, pp. 1323–1337, 2003.

[171] X. Chen, J. L. Aravena, and K. Zhou, "Risk analysis in robust control–making the case for probabilistic robust control," in *American Control Conference, 2005. Proceedings of the 2005*, pp. 1533–1538, IEEE, 2005.

[172] G. Calafiore, F. Dabbene, and R. Tempo, "A survey of randomized algorithms for control synthesis and performance verification," *Journal of Complexity*, vol. 23, no. 3, pp. 301–316, 2007.

[173] G. C. Calafiore, F. Dabbene, and R. Tempo, "Research on probabilistic methods for control system design," *Automatica*, vol. 47, no. 7, pp. 1279–1293, 2011.

[174] C. M. Lagoa and B. R. Barmish, "Distributionally robust monte carlo simulation: a tutorial survey," in *Proceedings of the IFAC world congress*, pp. 1–12, 2002.

[175] Z. K. Nagy and R. D. Braatz, "Worst-case and distributional robustness analysis of finite-time control trajectories for nonlinear distributed parameter systems," *Control Systems Technology, IEEE Transactions on*, vol. 11, no. 5, pp. 694–704, 2003.

[176] A. Chakraborty, P. Seiler, and G. J. Balas, "Nonlinear region of attraction analysis for flight control verification and validation," *Control Engineering Practice*, vol. 19, no. 4, pp. 335–345, 2011.

[177] P. Seiler, G. J. Balas, and A. K. Packard, "Assessment of aircraft flight controllers using nonlinear robustness analysis techniques," in *Optimization Based Clearance of Flight Control Laws*, pp. 369–397, Springer, 2012.

[178] A. Chakraborty, P. Seiler, and G. J. Balas, "Susceptibility of F/A-18 flight controllers to the falling-leaf mode: Linear analysis," *Journal of Guidance, Control, and Dynamics*, vol. 34, no. 1, pp. 57–72, 2011.

[179] L. T. Nguyen, M. E. Ogburn, W. P. Gilbert, K. S. Kibler, P. W. Brown, and P. L. Deal, "Simulator study of stall/post-stall characteristics of a fighter airplane with relaxed longitudinal static stability (F-16)," *NASA Technical Paper 1538*, pp. 1–223, 1979.

[180] F. L. Lewis and B. L. Stevens, *Aircraft Control and Simulation*. John Wiley& Sons, Hoboken, New Jersey, 1992.

[181] R. Bhattacharya, G. J. Balas, M. A. Kaya, and A. Packard, "Nonlinear receding horizon control of an F-16 aircraft," *Journal of Guidance, Control, and Dynamics*, vol. 25, no. 5, pp. 924–931, 2002.

[182] P. Gill, W. Murray, and M. Saunders, "User's guide for SNOPT version 7: Software for large-scale nonlinear programming," *Available on-line:* `http://www.ccom.ucsd.edu/~peg/papers/sndoc7.pdf`, 2006.

[183] E. F. Camacho and C. B. Alba, *Model Predictive Control*. Springer, London, 2003.

[184] D. E. Newland, *An Introduction to Random Vibrations, Spectral & Wavelet Analysis*. Courier Dover Publications, Mineola, New York, 2012.

[185] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, vol. 729. John Wiley & Sons, Hoboken, New Jersey, 2011.

[186] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton University Press, Princeton, New Jersey, 2009.

[187] X. Jiang, L. Ning, and T. T. Georgiou, "Distances and riemannian metrics for multivariate spectral densities," *Automatic Control, IEEE Transactions on*, vol. 57, no. 7, pp. 1723–1735, 2012.

[188] G. C. Goodwin, S. F. Graebe, and M. E. Salgado, *Control System Design*, vol. 240. Prentice Hall, New Jersey, 2001.

[189] J. R. Partington, *Linear Operators and Linear systems: An Analytical Approach to Control Theory*. No. 60, Cambridge University Press, Cambridge, 2004.

[190] G. Vinnicombe, "The robustness of feedback systems with bounded complexity controllers," *Automatic Control, IEEE Transactions on*, vol. 41, no. 6, pp. 795–803, 1996.

[191] G. Vinnicombe, *Uncertainty and Feedback: $\mathcal{H}_\infty$ Loop-shaping and the ν-gap Metric*. Imperial College Press, London, 2001.

[192] J. H. Steele and G. Vinnicombe, "Closed-loop time-domain model validation in the ν-gap metric," in *Decision and Control, 2001. Proceedings of the 40th IEEE Conference on*, vol. 5, pp. 4332–4337, IEEE, 2001.

[193] M. Vidyasagar, *Control System Synthesis: A Factorization Approach*. Morgan & Claypool Publishers, California, 2011.

[194] B. Dubrovin, A. Fomenko, and S. Novikov, *Modern Geometry – Methods and Applications. Part I. Graduate Texts in Mathematics, No. 93.* Springer Verlag, New York, 1985.

[195] E. Hille, *Analytic Function Theory*, vol. 1. American Mathematical Society, Providence, Rhode Island, 1959.

[196] D. Georgiev and E. Klavins, "Model discrimination of polynomial systems via stochastic inputs," in *Decision and Control, 47th IEEE Conference on*, pp. 3323–3329, IEEE, 2008.

[197] A. Kremling, S. Fischer, K. Gadkar, F. J. Doyle, T. Sauter, E. Bullinger, F. Allgöwer, and E. D. Gilles, "A benchmark for methods in reverse engineering and model discrimination: problem formulation and solutions," *Genome Research*, vol. 14, no. 9, pp. 1773–1785, 2004.

[198] G. Steinbrecher and W. T. Shaw, "Quantile mechanics," *European Journal of Applied Mathematics*, vol. 19, no. 2, pp. 87–112, 2008.

[199] W. Gilchrist, *Statistical Modelling with Quantile Functions.* CRC Press, Boca Raton, Florida, 2000.

[200] G. Monge, "Mémoire sur la théorie des déblais et des remblais," *Histoire de l'Académie Royale des Sciences de Paris*, pp. 666–704, 1781.

[201] L. Kantorovich, "On the translocation of masses," *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS*, vol. XXXVII, no. 7-8, pp. 199–201, 1942.

[202] L. Kantorovitch, "On the translocation of masses," *Management Science*, vol. 5, no. 1, pp. 1–4, 1958.

[203] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions," *Communications on Pure and Applied Mathematics*, vol. 44, no. 4, pp. 375–417, 1991.

[204] E. Haber, T. Rehman, and A. Tannenbaum, "An efficient numerical method for the solution of the L_2 optimal mass transfer problem," *SIAM Journal on Scientific Computing*, vol. 32, no. 1, pp. 197–211, 2010.

[205] X. Feng, R. Glowinski, and M. Neilan, "Recent developments in numerical methods for fully nonlinear second order partial differential equations," *SIAM Review*, vol. 55, no. 2, pp. 205–267, 2013.

[206] N. Papadakis, G. Peyré, and E. Oudet, "Optimal transport with proximal splitting," *SIAM Journal on Imaging Sciences*, vol. 7, no. 1, pp. 212–238, 2014.

[207] J.-D. Benamou and Y. Brenier, "A computational fluid mechanics solution to the monge-kantorovich mass transfer problem," *Numerische Mathematik*, vol. 84, no. 3, pp. 375–393, 2000.

[208] M. Fortin and R. Glowinski, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-value Problems*. Elsevier, Amsterdam, 2000.

[209] I. Olkin and F. Pukelsheim, "The distance between two random vectors with given dispersion matrices," *Linear Algebra and its Applications*, vol. 48, pp. 257–263, 1982.

[210] M. Knott and C. Smith, "On the optimal mapping of distributions," *Journal of Optimization Theory and Applications*, vol. 43, no. 1, pp. 39–49, 1984.

[211] A. Ohara and T. Kitamori, "Geometric structures of stable state feedback systems," *Automatic Control, IEEE Transactions on*, vol. 38, no. 10, pp. 1579–1583, 1993.

[212] W. Dahm, "Technology horizons: a vision for Air Force Science & Technology during 2010-2030," *USAF HQ Report*, 2010.

[213] V. I. Klyatskin, *Dynamics of Stochastic Systems*. Elsevier, Amsterdam, 2005.

APPENDIX A

STATISTICS OF LTI SYSTEM

A.1 Deterministic LTI Flow

Consider the deterministic LTI system with open-loop control signal $u(t)$, given by

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad x(0) = x_0, \tag{A.1a}$$

$$y(t) = Cx(t) + Du(t), \tag{A.1b}$$

where $x \in \mathbb{R}^{n_x}$, $u \in \mathbb{R}^{n_u}$ and $y \in \mathbb{R}^{n_y}$. One can solve for the state flow $x(t)$ by solving the ODE (A.1a) via computing the *homogeneous solution* and *particular integral*, and then adding the two. This results the state flow

$$x(t) = \underbrace{e^{At}x_0}_{\text{homogeneous solution}} + \underbrace{\int_0^t e^{A(t-\tau)}Bu(\tau)\ d\tau}_{\text{particular integral}}, \tag{A.2}$$

which together with (A.1b) yields the output flow $y(t)$ as

$$y(t) = Ce^{At}x_0 + Ce^{At}\int_0^t e^{-A\tau}Bu(\tau)\ d\tau + Du(t). \tag{A.3}$$

A.1.1 Mean Flow

Taking $\mathbb{E}[\cdot]$ to both sides of the state flow (A.2), we get the flow for the mean of the state vector:

$$\mu_x(t) = e^{At}\mu_{x_0} + \int_0^t e^{A(t-\tau)}Bu(\tau)\ d\tau, \tag{A.4}$$

which shows that even if we choose μ_0 to be zero, $\mu_x(t)$ need not be zero at $t > 0$, due to the control forcing. Thus, the open-loop control signal $u(t)$ contributes to the *average drift* of the state ensemble. Alternatively, this result could be derived by directly taking $\mathbb{E}[\cdot]$ to both sides of (A.1a), thus resulting an ODE for state mean:

$$\dot{\mu}_x(t) = A\mu_x(t) + Bu(t), \qquad \mu(0) = \mathbb{E}[x(0)] = \mathbb{E}[x_0] = \mu_0, \tag{A.5}$$

and then solving the linear ODE (A.5) akin to (A.1a).

Now, taking $\mathbb{E}[\cdot]$ to both sides of (A.1b) and substituting (A.4), we get the flow for the mean of the output vector:

$$\mu_y(t) = C\mu_x(t) + Du(t) = Ce^{At}\mu_{x_0} + Ce^{At}\int_0^t e^{-A\tau}Bu(\tau)\ d\tau + Du(t), \tag{A.6}$$

which we could also obtain by taking $\mathbb{E}[\cdot]$ to both sides of (A.3).

A.1.2 Covariance Flow

The state covariance $P_x(t)$ is given by

$$\begin{aligned} & P_x(t) \\ \triangleq\ & \mathbb{E}\left[(x(t) - \mu_x(t))(x(t) - \mu_x(t))^\top\right], && \text{(A.7)} \\ =\ & \mathbb{E}\left[e^{At}(x_0 - \mu_0)(x_0 - \mu_0)^\top\left(e^{At}\right)^\top\right], \text{ (substituting (A.2) and (A.4))} && \text{(A.8)} \end{aligned}$$

$$= e^{At}\,\mathbb{E}\left[(x_0-\mu_0)\,(x_0-\mu_0)^\top\right]\,\left(e^{At}\right)^\top, \tag{A.9}$$

$$= \Phi(t)\;P_0\;(\Phi(t))^\top, \tag{A.10}$$

where $\Phi(t) \triangleq e^{At}$ is the *state-transition matrix*. Equation (A.10) tells that the flow of the state covariance matrix is a congruence transformation by the state-transition matrix. Notice that the state covariance flow, unlike the state mean flow, does not depend on $u(t)$, thanks to the cancelation of the particular integral in (A.8).

One can also obtain an ODE for $P_x(t)$ as follows:

$$\frac{d}{dt}P_x(t) = \frac{d}{dt}\left[\underbrace{\Phi(t)\;P_0}_{U}\;\underbrace{(\Phi(t))^\top}_{V}\right], \qquad \text{(from (A.10))} \tag{A.11}$$

$$= U\,\frac{dV}{dt} + \frac{dU}{dt}\,V, \qquad \text{(by chain rule)} \tag{A.12}$$

$$= (\Phi(t)\;P_0)\;\frac{d}{dt}(\Phi(t))^\top + \left(\frac{d}{dt}(\Phi(t)\;P_0)\right)\;(\Phi(t))^\top, \tag{A.13}$$

$$= (\Phi(t)\;P_0)\;\underbrace{e^{A^\top t}}_{(\Phi(t))^\top}\;A^\top + \left(\Phi(t)\,\underbrace{\frac{dP_0}{dt}}_{0} + \underbrace{\frac{d\Phi}{dt}}_{Ae^{At}}\;P_0\right)\;(\Phi(t))^\top, \tag{A.14}$$

$$= \Phi(t)\;P_0\;(\Phi(t))^\top\;A^\top + A\;\underbrace{e^{At}}_{\Phi(t)}\;P_0\;(\Phi(t))^\top, \tag{A.15}$$

$$= P_x(t)\;A^\top + A\,\Phi(t)\;P_0\;(\Phi(t))^\top, \qquad \text{(using (A.10))} \tag{A.16}$$

$$= P_x(t)\;A^\top + A\,P_x(t). \tag{A.17}$$

Using (A.1b), we can compute the output covariance $P_y(t)$ from the state covariance $P_x(t)$, as follows:

$$P_y(t) = C\,P_x(t)\;C^\top. \tag{A.18}$$

A.2 Stochastic LTI Flow

Derivations similar to above can be done for the stochastic LTI system

$$dx(t) = Ax(t)\,dt + Bu(t)\,dt + G\,d\mathcal{W}, \tag{A.19}$$

resulting the state mean update equation $\dot{\mu}_x(t) = A\mu_x(t) + Bu(t)$, which is same as the corresponding equation for deterministic LTI flow case, i.e. equation (A.5). The covariance propagation equation (A.17) generalizes to

$$\frac{d}{dt}P_x\left(t\right) = AP_x\left(t\right) + P_x\left(t\right)A^\top + GQG^\top, \tag{A.20}$$

where Q is the covariance of the process noise $\mathcal{W}$, which is assumed to be a Wiener process.

A.3 Stochastic LTI Map

For general stochastic case

$$x\left(k+1\right) = Ax\left(k\right) + Bu\left(k\right) + G\mathcal{W}\left(k\right), \tag{A.21}$$

the sample path becomes

$$x\left(k\right) = A^k x_0 + \sum_{j=0}^{k-1} A^{k-j-1}\left(Bu\left(j\right) + G\mathcal{W}\left(j\right)\right). \tag{A.22}$$

The mean evolves as $\mu_x\left(k+1\right) = A\mu_x\left(k\right) + Bu\left(k\right)$, and hence

$$\mu_x\left(k\right) = A^k\mu_{x_0} + \sum_{j=0}^{k-1} A^{k-j-1}Bu\left(j\right). \tag{A.23}$$

The covariance evolves as $\Sigma_x(k+1) = A\Sigma_x(k)A^\top + GQG^\top$, and hence

$$\Sigma_x(k) = A^k\Sigma_{x_0}\left(A^k\right)^\top + \sum_{j=0}^{k-1} A^{k-j-1}GQG^\top\left(A^{k-j-1}\right)^\top. \tag{A.24}$$

APPENDIX B

SOME RESULTS ON DISTRIBUTIONAL COMPARISON

B.1 Computing ${}_2W_2\left(\mathcal{B}\left(\alpha,\beta\right),\mathcal{B}\left(\beta,\alpha\right)\right)$

We denote $I_t^{-1}\left(\alpha,\beta\right)$ as the inverse of the beta CDF, $I_x\left(\alpha,\beta\right) \triangleq \dfrac{B\left(x;\alpha,\beta\right)}{B\left(\alpha,\beta\right)}$ as the regularized incomplete beta function, and $B\left(x;\alpha,\beta\right) \triangleq \displaystyle\int_0^x z^{\alpha-1}\left(1-z\right)^{\beta-1}\,dz$ as the incomplete beta function.

Theorem 24. ${}_2W_2\left(\mathcal{B}\left(\alpha,\beta\right),\mathcal{B}\left(\beta,\alpha\right)\right) = \sqrt{\dfrac{\alpha\left(\alpha+1\right)+\beta\left(\beta+1\right)}{\left(\alpha+\beta\right)\left(\alpha+\beta+1\right)} - 2\left(\dfrac{\beta}{\alpha+\beta}-\mathcal{J}\right)}$,
where $\mathcal{J} \triangleq \dfrac{1}{\beta+1}\displaystyle\int_0^1 \left(I_t^{-1}\left(\alpha,\beta\right)\right)^{1-\alpha}\left(1-I_t^{-1}\left(\alpha,\beta\right)\right)^{1-\beta}\left(I_t^{-1}\left(\beta,\alpha\right)\right)^{\beta+1}$
${}_2F_1\left(\beta+1,1-\alpha;\beta+2;I_t^{-1}\left(\beta,\alpha\right)\right)\,dt$.

Proof. From (4.3), we have

$$ {}_2W_2^2\left(\mathcal{B}\left(\alpha,\beta\right),\mathcal{B}\left(\beta,\alpha\right)\right) = \int_0^1 \left(I_t^{-1}\left(\alpha,\beta\right) - I_t^{-1}\left(\beta,\alpha\right)\right)^2\,dt. \tag{B.1} $$

The following identities, stated without proof, will be useful for the evaluation of (B.1).

Property 1.

$$ \int I_t^{-1}\left(a,b\right)\,dt = \frac{\left(I_t^{-1}\left(a,b\right)\right)^{a+1}}{\left(a+1\right)B\left(a,b\right)}\,{}_2F_1\left(a+1,1-b;a+2;I_t^{-1}\left(a,b\right)\right) + \text{ constant.} $$

Property 2.

$$ \int \left(I_t^{-1}\left(a,b\right)\right)^2\,dt = \frac{\left(I_t^{-1}\left(a,b\right)\right)^{a+1}}{\left(a+1\right)B\left(a,b\right)}\left({}_2F_1\left(a+1,1-b;a+2;I_t^{-1}\left(a,b\right)\right) - \right. $$

$$ {}_2F_1\left(a+1,-b;a+2;I_t^{-1}(a,b)\right)+\text{ constant.}$$

Property 3.

$I_0^{-1}(a,b)=0$, *and* $I_1^{-1}(a,b)=1$.

Property 4. *(Gauss Theorem)*

$${}_2F_1(A,B;C;1)=\frac{\Gamma(C)\,\Gamma(C-A-B)}{\Gamma(C-A)\,\Gamma(C-B)}.$$

Property 5.

$$\frac{d}{dt}I_t^{-1}(a,b)=B(a,b)\left(I_t^{-1}(a,b)\right)^{1-a}\left(1-I_t^{-1}(a,b)\right)^{1-b}.$$

Using Properties 2 and 3, we get

$$\int_0^1\left(I_t^{-1}(\alpha,\beta)\right)^2\,dt=\frac{1}{(\alpha+1)\,B(\alpha,\beta)}\left[{}_2F_1(\alpha+1,1-\beta;\alpha+2;1)-{}_2F_1(\alpha+1,-\beta;\alpha+2;1)\right]. \tag{B.2}$$

Recalling that $\Gamma(k+1)=k\Gamma(k)$, Property 4 results

$${}_2F_1(\alpha+1,1-\beta;\alpha+2;1)=\frac{\Gamma(\alpha+2)\,\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}, \tag{B.3}$$

$${}_2F_1(\alpha+1,-\beta;\alpha+2;1)=\frac{\Gamma(\alpha+2)\;\beta\Gamma(\beta)}{(\alpha+\beta+1)\,\Gamma(\alpha+\beta+1)}. \tag{B.4}$$

Substituting the above expressions in (B.2), we obtain

$$\begin{aligned}\int_0^1\left(I_t^{-1}(\alpha,\beta)\right)^2\,dt &= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)},\text{ and}\\ \int_0^1\left(I_t^{-1}(\beta,\alpha)\right)^2\,dt &= \frac{\beta(\beta+1)}{(\alpha+\beta)(\alpha+\beta+1)}.\end{aligned} \tag{B.5}$$

Thus, (B.1) simplifies to

$$ {}_2W_2^2\left(\mathcal{B}\left(\alpha,\beta\right),\mathcal{B}\left(\beta,\alpha\right)\right)=\frac{\alpha\left(\alpha+1\right)+\beta\left(\beta+1\right)}{\left(\alpha+\beta\right)\left(\alpha+\beta+1\right)}-2\int_0^1 I_t^{-1}\left(\alpha,\beta\right)I_t^{-1}\left(\beta,\alpha\right)\,dt \tag{B.6}$$

To evaluate the remaining integral in (B.6), we employ integration-by-parts with $f\left(t\right):=I_t^{-1}\left(\alpha,\beta\right)$ as the first function and $g\left(t\right):=I_t^{-1}\left(\beta,\alpha\right)$ as the second. Now, we know that $\int_0^1 f\left(t\right)g\left(t\right)dt$ equals

$$ \underbrace{\left[f\left(t\right)\int g\left(t\right)dt\right]\Big|_{t=0}^{t=1}}_{\mathcal{I}}-\underbrace{\int_0^1\left(f'\left(t\right)\int g\left(t\right)\,dt\right)\,dt}_{\mathcal{J}}. \tag{B.7}$$

From Properties 1 and 3, we get

$$ \begin{aligned}\mathcal{I} &= \left[\frac{1}{\left(\beta+1\right)B\left(\alpha,\beta\right)}I_t^{-1}\left(\alpha,\beta\right)\left(I_t^{-1}\left(\beta,\alpha\right)\right)^{b+1}\,{}_2F_1\left(b+1,1-a;b+2;1\right)\right]\Big|_{t=0}^{t=1}\\ &= \frac{{}_2F_1\left(\beta+1,1-\alpha;\beta+2;1\right)}{\left(\beta+1\right)B\left(\alpha,\beta\right)}=\frac{\beta}{\alpha+\beta}.\end{aligned} \tag{B.8}$$

Further, Properties (1) and (5) yield

$$ \begin{aligned}\mathcal{J}=\frac{1}{\beta+1}\int_0^1 &\left(I_t^{-1}\left(\alpha,\beta\right)\right)^{1-\alpha}\left(1-I_t^{-1}\left(\alpha,\beta\right)\right)^{1-\beta}\\ &\left(I_t^{-1}\left(\beta,\alpha\right)\right)^{\beta+1}\,{}_2F_1\left(\beta+1,1-\alpha;\beta+2;I_t^{-1}\left(\beta,\alpha\right)\right)\,dt.\end{aligned} \tag{B.9}$$

Combining (B.6), (B.7), (B.8) and (B.9), the result follows. ■

B.2 On the Stationary Density of A Deterministic Nonlinear System with Multiple Isolated Stable Equilibria

Theorem 25. *Consider a nonlinear dynamical system $\dot{x}(t)=f\left(x(t)\right)$, having multiple isolated stable equilibria $\{x_i^\star\}_{i=1}^{n^\star}$. Let us assume that the system does not admit*

any invariant set other than these stable equilibria. Also, let $\mathcal{R}_i$ *be the region-of-attraction for the* i^{th} *equilibrium point. If the dynamics evolves from an initial PDF* ξ_0*, then its stationary PDF is given by*

$$\xi_\infty(x) = \sum_{i=1}^{n^\star} m_i^\star \delta\left(x - x_i^\star\right), \tag{B.10}$$

where $m_i^\star = \int_{\mathrm{supp}(\xi_0)\cap\mathcal{R}_i} \xi_0\left(x_0\right) dx_0$.

Proof. Since $\{x_i^\star\}_{i=1}^{n^\star}$ is the unique set of attractors, it is easy to verify that the stationary PDF is of the form (B.10); however, it remains to determine the weights $m_i^\star$. We observe that **either** $\mathrm{supp}\left(\xi_0\right) \subseteq \mathcal{R}_i$, for some $i = 1, \ldots, n^\star$, **or** $\mathrm{supp}\left(\xi_0\right)$ intersects multiple $\mathcal{R}_i$.

Now, recall that $\mathcal{R}_i \triangleq \{x_0 : \dot{x}(t) = f\left(x(t)\right), x(0) = x_0, \lim_{t\to\infty} x\left(t\right) = x_i^\star\}$. Thus, if $\mathrm{supp}\left(\xi_0\right) \subseteq \mathcal{R}_i$, then $m_i^\star = \int_{\mathrm{supp}(\xi_0)} dm_0 = \int_{\mathrm{supp}(\xi_0)} \xi\left(x_0\right) dx_0 = 1$, and consequently, $m_j^\star = 0$, $\forall j = 1, \ldots, n^\star$, $j \neq i$, since $\int \xi_\infty\left(x\right) dx = 1$. In this case, notice that $\mathrm{supp}\left(\xi_0\right) = \mathrm{supp}\left(\xi_0\right) \cap \mathcal{R}_i$.

On the other hand, if $\mathrm{supp}\left(\xi_0\right)$ intersects multiple $\mathcal{R}_i$, then only for $x_0 \in \mathrm{supp}\left(\xi_0\right) \cap \mathcal{R}_i$, the integral curves of $\dot{x}(t) = f\left(x(t)\right), x(0) = x_0$, will satisfy $\lim_{t\to\infty} x\left(t\right) = x_i^\star$. In other words, only the set $\mathrm{supp}\left(\xi_0\right)\cap\mathcal{R}_i$ contributes to $m_i^\star$, i.e. $m_i^\star = \int_{\mathrm{supp}(\xi_0)\cap\mathcal{R}_i} dm_0 = \int_{\mathrm{supp}(\xi_0)\cap\mathcal{R}_i} \xi\left(x_0\right) dx_0 < 1$.

Combining the above two cases, we conclude $m_i^\star = \int_{\mathrm{supp}(\xi_0)\cap\mathcal{R}_i} \xi_0\left(x_0\right) dx_0$. ■

B.3 Derivation of Stationary PDF (4.21)

We re-write the Itô SDE (4.20) as

$$\begin{Bmatrix} dx_1 \\ dx_2 \end{Bmatrix} = \begin{Bmatrix} x_2 \\ -\dfrac{\partial}{\partial x_1} U\left(x_1\right) - c x_2 \end{Bmatrix} dt + \begin{Bmatrix} 0 \\ 1 \end{Bmatrix} dW, \tag{B.11}$$

with $U(x_1) := \frac{1}{2}(ax_1^2 - b\cos 2x_1)$. An Itô SDE with drift nonlinearity of the form (B.11), admits [213] stationary PDF $\eta_\infty(x_1, x_2) \propto \exp\left(-\frac{c}{Q}H(x_1, x_2)\right)$, where the Hamiltonian function $H(x_1, x_2) := U(x_1) + \frac{1}{2}x_2^2$.

APPENDIX C

WASSERSTEIN CONVERGENCE OF MARGINALS

The purpose of this appendix is to prove that convergence of joint PDFs imply convergence in respective univariate marginals, but the converse is not true. Here, the convergence of PDFs is measured in Wasserstein metric ${}_2W_2$, hereafter simply denoted as W, for notational ease. We first prove the following preparatory lemma that leads to our main result in Theorem 26.

Lemma 7. *Let ξ_1^i and ξ_2^i be the respective i^{th} univariate marginals for d-dimensional joint PDFs ξ_1 and ξ_2, supported on $\mathbb{R}_{x_1} \times \mathbb{R}_{x_2} \times \ldots \times \mathbb{R}_{x_d}$, and $\mathbb{R}_{y_1} \times \mathbb{R}_{y_2} \times \ldots \times \mathbb{R}_{y_d}$. Let $W_i \triangleq W\left(\xi_1^i, \xi_2^i\right)$, $i = 1, \ldots, d$, and $\overline{W} \triangleq W\left(\xi_1, \xi_2\right)$; then*

$$\sum_{i=1}^{d} W_i^2 \leqslant \overline{W}^2. \tag{C.1}$$

Proof. Notice that $\operatorname{supp}\left(\xi_1^i\right) = \mathbb{R}_{x_i}$, and $\operatorname{supp}\left(\xi_2^i\right) = \mathbb{R}_{y_i}$, $\forall\, i = 1, \ldots, d$. For d-dimensional vectors $x = (x_1, \ldots, x_d)^\top$, $y = (y_1, \ldots, y_d)^\top$, by definition

$$\begin{aligned} \overline{W}^2 &= \inf_{\varrho \in \mathcal{M}(\xi_1, \xi_2)} \int_{\mathbb{R}^{2d}} \| x - y \|_2^2 \;\, \varrho\left(x, y\right) \, dx \, dy \\ &= \int_{\mathbb{R}^{2d}} \| x - y \|_2^2 \;\, \varrho^\star\left(x, y\right) \, dx \, dy, \end{aligned} \tag{C.2}$$

where $\varrho^\star\left(x, y\right)$ is the optimal transport PDF supported on $\mathbb{R}^{2d}$. Clearly,

$$\xi_1^i = \int_{\mathbb{R}^{2d-1}} \varrho^\star\left(x, y\right) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_d dy_1 \ldots dy_d, \tag{C.3}$$

$$\xi_2^i \;=\; \int_{\mathbb{R}^{2d-1}} \varrho^\star(x,y)\, dx_1 \ldots dx_d dy_1 \ldots dy_{i-1} dy_{i+1} \ldots dy_d. \tag{C.4}$$

Thus, we have

$$\begin{aligned} W_i^2 &= \inf_{\varphi \in \mathcal{M}\left(\xi_1^i, \xi_2^i\right)} \int_{\mathbb{R}^2} (x_i - y_i)^2 \; \varphi(x_i, y_i) \; dx_i \, dy_i, \\ &= \int_{\mathbb{R}^2} (x_i - y_i)^2 \; \varphi^\star(x_i, y_i) \; dx_i \, dy_i, \\ &\leqslant \int_{\mathbb{R}^2} (x_i - y_i)^2 \; \widetilde{\varrho}^\star(x_i, y_i) \; dx_i \, dy_i, \end{aligned} \tag{C.5}$$

where $\widetilde{\varrho}^\star(x_i, y_i)$ is the $(i,i)^{\text{th}}$ bivariate marginal of $\varrho^\star(x,y)$. Since $\sum_{i=1}^{d} (x_i - y_i)^2 = \| x - y \|_2^2$, the result follows from (C.5), after substituting

$$\widetilde{\varrho}^\star(x_i, y_i) = \int_{\mathbb{R}^{2d-2}} \varrho^\star(x,y)\, dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_d \, dy_1 dy_{i-1} dy_{i+1} \ldots dy_d. \tag{C.6}$$

This completes the proof. ■

Theorem 26. *Convergence of Joint PDFs in Wasserstein metric, implies convergence of univariate marginals. Converse is not true.*

Proof. Using the notation of Lemma 7, when the joints ξ_1 and ξ_2 converge, then $\overline{W} = 0$. Hence from (C.1), $\sum_{i=1}^{d} W_i^2 = 0 \Rightarrow W_i = 0, \; \forall \; i = 1, \ldots, d$. However, $W_i = 0 \Rightarrow \overline{W} \geqslant 0$. Hence the result. ■