AFRL-RQ-WP-TR-2019-0145

SCALABLE INFERENCE FOR RARE EVENTS
Computational Methods for Estimating Probability of Tail Events

Tuhin Sahai, Benjamin Zhang, Youssef Marzouk, and Quan Long
United Technologies Research Center

AUGUST 2019
Final Report

AIR FORCE RESEARCH LABORATORY
AEROSPACE SYSTEMS DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7542
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the USAF 88th Air Base Wing (88 ABW) Public Affairs Office (PAO) and is available to the general public, including foreign nationals.

Copies may be obtained from the Defense Technical Information Center (DTIC) (https://discover.dtic.mil/).

AFRL-RQ-WP-TR-2019-0145 has been reviewed and is approved for publication in accordance with assigned distribution statement.

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
August 2019	Final	16 March 2016 – 15 August 2019

4. TITLE AND SUBTITLE
SCALABLE INFERENCE FOR RARE EVENTS
Computational Methods for Estimating Probability of Tail Events

5a. CONTRACT NUMBER
FA8650-16-C-7646

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
61101E

6. AUTHOR(S)
Tuhin Sahai, Benjamin Zhang, Youssef Marzouk, and Quan Long

5d. PROJECT NUMBER
DARPA

5e. TASK NUMBER

5f. WORK UNIT NUMBER
Q1NG

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
United Technologies Research Center
411 Silver Lane
East Hartford, CT 06108

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Aerospace Systems Directorate
Wright-Patterson Air Force Base, OH 45433-7542
Air Force Materiel Command
United States Air Force

Defense Advanced Research Projects Agency/Tactical Technology Office (DARPA/TTO)
3701 N. Fairfax Drive
Arlington, VA 22203

10. SPONSORING/MONITORING AGENCY ACRONYM(S)
AFRL/RQVC

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)
AFRL-RQ-WP-TR-2019-0145

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

13. SUPPLEMENTARY NOTES
PA Clearance Number: 88ABW-2020-0252; Clearance Date: 27 Jan 2020.

14. ABSTRACT
This report presents key findings and results in the Scalable Inference for Rare Events (SIRE) project (FA8650-16-C-7646). Under the project, we discovered deep theoretical connections between Koopman operator theory and rare event simulation in stochastic differential equations. We then developed a generalized approach for constructing efficient importance sampling methods for linear stochastic differential equations using the Kolmogorov Backward (Ornstein-Uhlenbeck) operator. We show that this approach is a special case of the Koopman operator approach. Additionally, we constructed rotorcraft models that capture critical stall phenomena that was used for computation. We then demonstrate large deviations-based importance sampling and splitting methods on rotorcraft and electrical models.

15. SUBJECT TERMS
rare events, failure predictions, tail events, Koopman operator, importance sampling, rotorcraft stall

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
a. REPORT	b. ABSTRACT	c. THIS PAGE			Philip S. Beran
Unclassified	Unclassified	Unclassified	SAR	81	19b. TELEPHONE NUMBER (Include Area Code) N/A

Table of Contents

List of Figures

List of Tables

1 Summary

This report presents key findings and results in the Scalable Inference for Rare Events (SIRE) project (FA8650-16-C-7646). We achieved the following goals:

1. Discovered deep theoretical connections between Koopman operator theory and rare event simulation in stochastic differential equations.

2. Developed a generalized approach for efficient importance sampling methods for linear stochastic differential equations using the Kolmogorov Backward (Ornstein-Uhlenbeck) operator. This approach is a special case of the Koopman operator approach. The Koopman operator approach demonstrated a 5000X and 1000X variance reductions on the Van der Pol and Duffing oscillators respectively. Moreover, this approach is able to deal with degenerate noise cases that arise in real world systems.

3. Found all the eigenfunctions of the Ornstein-Uhlenbeck operator, which will be useful for developing sampling methods.

4. Developed heuristics for efficient importance sampling methods for linear stochastic partial differential equations.

5. Constructed rotorcraft models that capture critical stall phenomena that was used for computations.

6. Formulated and applied a new left-eigenvector based large deviations-based importance sampling and splitting methods to rotorcraft and electrical models. The approach demonstrated a 30X variance reduction on the electrical system model and 36X variance reduction on the rotorcraft system. Additionally, it demonstrated a 2000X variance reduction in non-normal dynamical systems (existing large deviations based methods did not work in this setting). The splitting approach demonstrated a 10X variance reduction on a simple rotorcraft model.

7. Developed a new formulation for accelerating large deviations-based approaches using fast stochastic optimal control solvers.

2 Introduction

Uncertainty is pervasive in engineering systems. Improperly treating or ignoring uncertainty in the modeling and simulation of engineering systems may lead to costly consequences. Uncertainty quantification for rare events is both difficult and important. Since they do not occur very often, one cannot obtain enough data to study their mechanisms fully. However, at the same time, rare events are often linked to the most catastrophic consequences.

In contemporary engineering design, one studies a system, its failure rates, and expectations sensitive to rare events via computational models. These quantities of interest are relevant in a variety of models in which the risk of failures may be tiny, but the cost of failure may be catastrophic. Common examples include the probability of insolvency of an insurance firm [1], the stalling and crashing of an airplane or rotorcraft in flight [2], or the loss of information packets in communications engineering [3]. Good engineering design makes failure rates rare, and to study them properly in computational models requires efficient rare event simulation.

Stochastic differential equations are useful for modeling uncertain dynamics or systems that are influenced by noise. Often, the evolution of the state itself is not the main object of interest; rather, quantities of interest called observables of the system are more important. These quantities may be the reliability of a system, or its average operating conditions over a fixed period of time.

Due to the stochastic nature of the dynamics, these observables are often probabilities or expectations taken with respect to the probability measure induced by the stochastic dynamical system. However, these probability measures are intractable to work with exactly, so uncertainty quantification via sampling is often considered instead. Sampling methods, known as Monte Carlo methods, involve simulating many sample trajectories of the dynamical system and then computing expectations of interest by taking sample averages. These methods rely the law of large numbers, which allows exact expectations to be approximated by sample averages.

Observables of interest that are sensitive to rare events are of utmost importance as they are often consequential for the design and optimization of engineering systems. For example, in aerospace systems, one wants to choose design parameters such that the probability that the state enters some dangerous region does not exceed some threshold. Or, one may wish to study the probability measure conditioned on a rare event, which is relevant for studying mechanisms that lead to the rare event. Simple Monte Carlo methods become prohibitively expensive when expectations of interest are sensitive to rare events. Without many samples from the event of interest, the errors of these Monte Carlo estimators can be orders of magnitudes larger than the quantity of interest, thereby making the approximation useless. For this reason, there exists a large body of literature on variance reduction which seeks to make sampling methods more efficient. These methods include importance sampling, particle splitting, subset simulation, control variates, etc [1]. To attain variance reduction using these methods successfully, one typically uses theoretic and heuristic tools to gain some insight into the nature of the rare events of interest, and then use the insight to devise more efficient sampling methods.

Rare event analysis and simulation is well-studied in chemical and industrial system. Here, the rare events involve studying the probability of transitioning from one metastable state to another, or the probability that the buffer in a queuing system overflows. What is lacking is the study of rare event systems exclusively for physical and engineering systems – which may not be adequately described by established analyses.

In this project, we focus on importance sampling and splitting for stochastic differential equations. Importance sampling involves sampling from an alternative probability measure so that

more samples that reach the relevant parts of the state space can be obtained. One hopes to reduce the variance of the Monte Carlo estimator by using an appropriately chosen alternative measure. Systematic approaches to finding an alternative measure for stochastic dynamical systems include the cross-entropy method [4, 5] and large deviations-based methods. For stochastic differential equations, the Kolmogorov equations arising from stochastic calculus provides another way to characterize certain expectations of the system. In particular, a linear partial differential equation called the Kolmogorov Backward Equation (KBE) describes the evolution of expectations of the system over time. If one could solve the KBE exactly, no sampling would be necessary. Given the solution to the KBE, however, one can devise a zero-variance importance sampling estimator. Cross-entropy based approaches attempt to approximate the optimal measure by solving an optimization problem that minimizes the Kullbeck-Leibler divergence between the optimal measure and a parametrized family of probability measures.

Sequential Monte Carlo methods have been developed as an alternative to importance sampling. In these methods, one pushes samples towards regions of interest in an iterative fashion. Subset simulation [6] is one approach that is used in the engineering reliability community. They consider estimating tail probabilities of a high-dimensional distribution by sequentially pushing samples into the tails. The method involves expressing rare event probabilities as a product of conditional probabilities. These less rare conditional probabilities are found by using Markov chain Monte Carlo. However, to apply the method for stochastic differntial equations, one needs to discretize the system to obtain a high-dimensional distribution. This method is discretization-dependent, which makes the method difficult to analyze for infinite dimensional systems.

Particle splitting is another sequential Monte Carlo method. Like subset simulation, the philosophy is to partition the state space into a sequence of nested sets, so that rare event probabilities can be computed as a product of conditional probabilities. The method creates a branching process in which samples that reach rarer sets are split into child samples. However, one would need to choose the placement of the nested sets wisely for the method to be computationally worthwhile. Systematic approaches to finding the levels are have been found to have connection with large deviations-based sampling approaches [7].

No matter what approach one uses to create efficient estimators for SDEs, they are all directly or indirectly approximating the zero-variance estimator when it exists. It is often stated in literature that solving the KBE exactly is not a wise task to pursue as the dimension of the PDE scales with the dimension of the state space. In this paper, we show that a direct spectral solution to the KBE is possible for linear systems, known as Ornstein-Uhlenbeck processes. While the Kolmogorov Forward Equation (KFE) (also known as the Fokker-Planck equations), which describes the evolution of the probability distribution of the state over time, can be solved easily for linear systems in arbitrary dimension by a system ODEs, the KBE has only been solved in one dimension or under restrictive assumptions. To find a spectral solution to the KBE, we show how one can find all the eigenfunctions of the KBE evolution operator for linear systems in arbitrary dimension. Using the spectral approximation, we devise and efficient importance sampling estimators for Ornstein-Uhlenbeck processes.

While the framework developed in this paper is most relevant to linear stochastic dynamical systems in finite-dimensional state space, intuition drawn from the finite-dimensional case can be extended to linear stochastic partial differential equations and even to nonlinear systems. We explore these in the numerical examples.

We also explore large deviations-based importance sampling uses insight from large deviations theory for dynamical systems, known as the Freidlin-Wentzell theory, to construct efficient estimators. This approach converts this optimal sampling problem into a stochastic dynamic programming problem, which one opts to solve instead [8, 9].

Using the rich set of theoretical techniques from large deviations, past research has provided a computational paradigm for constructing good importance sampling estimators for rare event simulation. For example, one of the ways the exponentially tilted biasing distribution was introduced came from the proofs in Cramér's theorem. Later on, it was found to also be an efficient importance sampling distribution for certain problems [10].

One of the key insights that allowed large deviations to provide a unifying framework for rare event simulation is through a control-theoretic approach to the classical theory. The proofs provided computational methods to constructing good biasing distributions for importance sampling. Recent work in dynamic importance sampling [11] showed a way of finding good importance sampling distributions through a differential game theoretic interpretation which amounts to finding the solution to an Isaacs partial differential equation. For rare event simulation, this equation was shown to be equivalent to solving a Hamilton-Jacobi PDE or a sequence of variational problems. While many past methods were restricted certain classes of problems, this control theoretic approach to large deviations and rare event simulation admitted a general framework that easily adapted to problems as broad as queues [12] and small noise diffusions [8]. At the same time, [7] showed that that this approach also results in provably good level sets for particle splitting algorithms, thus unifying importance sampling and particle splitting methods.

Our goal under the SIRE program was to construct novel algorithmic approaches for predicting rare events in engineering systems. Real world systems, also present degenerate noise situations that can be challenging for current approaches. Under Phase I of the program, UTRC and MIT developed an large deviations theory based methodology that demonstrated 36X variance reduction in the rotorcraft system. This approach extends current rare event sampling methods to non-normal dynamic systems by biasing the system using the *left eigenvector* of the system. Under Phase II of the program, the team constructed an electrical system model to further challenge the algorithm construction. We developed an splitting approach that exploits the importance functions of the left eigenvector biasing to demonstrate 10X variance reduction on the smaller version of the rotorcraft model. To address degenerate noise, a challenge that spans multiple areas of rare events application, we developed a novel theory of Koopman operator based importance sampling. This approach can deal with nonlinearities, non-normality, degenerate noise situations, and black box systems in one framework. In fact, the left eigenvector biasing approach can be motivated using the Koopman operator framework. The approach was demonstrated on wide variety of problems including a formulation of the electrical system.

3 Background and Preliminaries

Here we review background information on stochastic dynamical models for physical systems. We also review efficient Monte Carlo methods for stochastic differential equations.

3.1 Stochastic Differential Equations and Kolmogorov Equations

We study dynamical models whose uncertainties come from external sources such as environmental factors or thermal noise, as opposed to parametric uncertainties. For this reason, we focus on rare event simulation for stochastic differential equation (SDE) models. Here, we review some basic tools used to describe and analyze SDEs following [13–15]. An SDE is a continuous time stochastic process defined on continuous state space. The system is typically written as a dynamical system that is driven by Brownian motion

$$\begin{cases} dX_t = \mathbf{A}(X_t)\,dt + \mathbf{B}(X_t)\,dW_t \\ X_0 \sim p_0 \end{cases} \tag{1}$$

where the state, X_t, evolves on $\mathbb{R}^d$, the drift, $\mathbf{A}$, is a mapping from $\mathbb{R}^d$ to itself, and the diffusion matrix, $\mathbf{B}$, is a mapping from $\mathbb{R}^d$ to the space of real-valued $d \times r$ matrices, and W_t is a standard r-dimensional Brownian motion. The initial condition X_0 is distributed according to an initial probability density p_0 on $\mathbb{R}^d$. Often, the initial condition may be deterministic, in which the density is a delta function. The quantities of interest relating to this system are expectations and probabilities, which are taken with respect to some probability measure. The probability measure $\mathbb{P}$ *induced* by the SDE system is a path-space defined measure on the set of functions from $[0,T]$ to $\mathbb{R}^d$. Define $\mathbb{E}^x[f(X_t)] = \mathbb{E}[f(X_t)|X_0 = x]$, where the expectations is taken with respect to $\mathbb{P}$. One of the main tools for studying Markov processes is the *infinitesimal generator*, $\mathscr{A}$, defined as

$$\mathscr{A}\psi = \lim_{t\to 0} \frac{\mathbb{E}^x[\psi(X_t)] - \psi(x)}{t} \tag{2}$$

where $\psi \in \mathscr{D}(\mathscr{A})$ are functions for which the limit is defined. For SDEs, the generator can be written in a closed form in terms of the drift and diffusion terms

$$\mathscr{A}\psi = \langle \mathbf{A}(x), \nabla\psi\rangle + \frac{1}{2}\mathrm{Tr}\left[\mathbf{B}(x)\mathbf{B}(x)^*\nabla^2\psi\right]. \tag{3}$$

The infinitesimal generator is related to two partial differential equations that describe the evolution of densities and statistics of a given SDE known as the Kolmogorov equations. The Kolmogorov backward equation describes the evolution of statistics, namely expectations of functions with respect to the density of states over time. Define $\Phi(t,x) = \mathbb{E}^x[f(X_t)]$, then

$$\frac{\partial\Phi(t,x)}{\partial t} = \mathscr{A}\Phi(t,x) \tag{4}$$

$$\Phi(0,x) = f(x). \tag{5}$$

The Kolmogorov forward equation, also commonly known as the Fokker-Planck equation, describes the evolution of the probability density function of the state in $\mathbb{R}^d$. The equation is found by considering the L^2 adjoint of the infinitesimal generator:

$$\frac{\partial}{\partial t}p(t,x) = \mathscr{A}^* p(t,x) \tag{6}$$

$$p(0,x) = p_0(x) \tag{7}$$

where

$$\mathscr{A}^*\pi = \nabla \cdot \left(-\mathbf{A}(x)\pi + \frac{1}{2}\nabla \cdot (\mathbf{B}(x)\mathbf{B}^*(x)\pi) \right). \tag{8}$$

Here π must be a probability density function on $\mathbb{R}^d$, that is, $\pi(x) > 0$ for all $x \in \mathbb{R}^d$ and the integral of π over $\mathbb{R}^d$ must equal one.

In theory, rare event simulation can be solved by solving either of these equations. In practice solving either of these equations exactly is intractable when the dimension of the state space is large. Instead, researchers approximate the quantities of interest by using Monte Carlo methods.

3.2 Monte Carlo Methods

Monte Carlo methods are a broad class of algorithms that estimate integrals and expectations via sampling. The methods are remarkable robust and applicable to a wide range of applications. The methods rely on the fact that in stochastic systems sample averages tend to converge to their theoretical expectations, a property known as the law of large numbers. Here we give a general overview and basic notions of Monte Carlo methods based on [1, 16].

3.2.1 Simple Monte Carlo for Rare Event Simulation

Suppose we are given an SDE as presented in Equation 1. An event of interest may be, for example, the state of the system X_t visiting some region $E \subset \mathbb{R}^d$ at time T. This event would be rare if the state does not visit the region very often. Denote $\rho = \mathbb{P}(X_T \in E)$ given the system starts at some fixed initial condition $X_0 = x$. A Monte Carlo estimator for ρ is

$$\hat{\rho} = \frac{1}{M}\sum_{i=1}^{M} \mathbb{1}_{X_T^i \in E} \tag{9}$$

where X^i are M independent simulations of the SDE, and $\mathbb{1}_{x\in A}$ is an indicator function that returns 1 if $x \in A$, and 0 otherwise. The typical metric for evaluating the quality of a Monte Carlo estimator is the mean-squared error $\mathbb{E}\left[(\rho - \hat{\rho})^2\right]$, which can be decomposed as the sum of the squared bias and the variance. We only consider unbiased estimators, so the variance is of utmost importance. The variance of this simple estimator is

$$\mathrm{Var}[\hat{\rho}] = \frac{\rho - \rho^2}{M}. \tag{10}$$

For rare event probabilities, $\rho \ll 1$, so we may approximate $\mathrm{Var}[\hat{\rho}] \approx \frac{\rho}{M}$. While typically we say an estimator is good when the variance is small, for rare event probabilities, the absolute variance is misleading as a small variance may not necessarily reflect an efficient estimator. Instead, the relative error (or coefficient of variation), which is the ratio between the standard error and the quantity of interest, is a better metric as it compares the variance of the estimator with the magnitude of the quantity of interest:

$$\text{Relative error} = \frac{\sqrt{\mathrm{Var}[\hat{\rho}]}}{\rho} \approx \frac{1}{\sqrt{M\rho}}. \tag{11}$$

From this, we see that with this simple Monte Carlo estimator, the number of samples required to keep the relative error at a reasonable level (around one) is equal to $1/\rho$. If ρ were a rare event probability, this number may be very large, and the estimator will be prohibitively expensive for large models.

3.2.2 *Importance Sampling and the Zero-Variance Estimator*

Importance sampling is a generic variance reduction method for improving the efficiency of Monte Carlo methods. Let's consider the finite dimensional case first. Let X be an $\mathbb{R}^d$-valued random variable distributed according to probability density $\eta(x)$ and suppose we wish to estimate

$$\rho = \mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x)\eta(x)\ \mathrm{d}x. \tag{12}$$

A simple Monte Carlo estimator for this integral is

$$\hat{\rho} = \frac{1}{M}\sum_{i=1}^{M} f(X^i) \tag{13}$$

$$X^i \sim \eta \text{ i.i.d..} \tag{14}$$

With some algebra, one can show that the estimator variance is

$$\mathrm{Var}[\hat{\rho}] = \frac{\mathrm{Var}[f(X)]}{M}, \tag{15}$$

which implies, as noted earlier that the only way to reduce the variance in simple Monte Carlo estimators is to use more samples. Importance sampling involves sampling from an alternative density $\pi(x)$ such that variance of the corresponding importance sampling estimator. Since the samples are not drawn from true distribution, each sample must be weighted to reflect its relative importance with respect to the nominal distribution η. We require that the support of $f(x)\eta(x)$ be contained in the support of $f(x)\pi(x)$. Importance sampling works because of the following identity

$$\rho = \mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x)\eta(x)\ \mathrm{d}x = \int_{\mathbb{R}^d} \frac{f(x)\eta(x)}{\pi(x)}\pi(x)\ \mathrm{d}x. \tag{16}$$

An importance sampling estimator is then

$$\hat{\rho}_{IS} = \frac{1}{M}\sum_{i=1}^{M} f(\tilde{X}^i)\frac{\eta(\tilde{X}^i)}{\pi(\tilde{X}^i)} \tag{17}$$

$$\tilde{X}^i \sim \pi \text{ i.i.d..} \tag{18}$$

The variance of the importance sampling estimator is

$$\mathrm{Var}[\hat{\rho}_{\mathrm{IS}}] = \frac{1}{M}\mathrm{Var}_{\pi}\left[\frac{f(\tilde{X})\eta(\tilde{X})}{\pi(\tilde{X})}\right]. \tag{19}$$

Here, we see that the variance of the estimator can be reduced by either increasing the number of samples used, M, or by judiciously choosing an importance sampling estimator π. In fact, as long as $f(x) \geq 0$, there exists a closed form expression for the optimal importance sampling distribution. The so-called optimal importance sampling distribution

$$\pi_{\mathrm{opt}}(x) = \frac{f(x)\eta(x)}{\int f(x)\eta(x)\ \mathrm{d}x} \tag{20}$$

results in a zero-variance importance sampling estimator meaning that a single sample will produce the true value ρ exactly. However, the issue with this is that to sample from the optimal distribution, one needs to know the quantity of interest exactly, which makes the original task moot. Therefore, most efficient implementations of importance sampling involve characterizing the zero-variance estimator in other ways in an attempt to approximate sampling from it. These approaches include the cross-entropy method [4], subset simulation [6], and large deviations-based methods [10].

Importance sampling can also be applied to infinite-dimensional distributions, such as path-space defined measures induced by SDEs. Denote $\mathbb{P}$ to be the probability measure induced by the stochastic process in Equation 1. Importance sampling involves constructing a new stochastic process $\{\tilde{X}\}$ that induces a different measure $\mathbb{Q}$ such that the rare event of interest occurs more often. Since the rare event occurs more often, each sample must be reweighed according to their relative importance. That is, we construct an estimator

$$\hat{\rho}_{IS} = \frac{1}{M}\sum_{i=1}^{M} \mathbb{1}_{\tilde{X}_T^i \in E} \cdot \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}(\tilde{X}^i) \tag{21}$$

where $\tilde{X}^i$ are independently sampled from $\mathbb{Q}$. The variance of this estimator is dependent on the choice of $\mathbb{Q}$:

$$\mathrm{Var}[\hat{\rho}_{IS}] = \frac{1}{M}\mathrm{Var}\left[\mathbb{1}_{\tilde{X}_T \in E}\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right] = \frac{1}{M}\left[\mathbb{E}_{\mathbb{Q}}\left[\left(\mathbb{1}_{\tilde{X}_T \in E}\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right)^2\right] - \rho^2\right]. \tag{22}$$

From here, we see that we can construct $\mathbb{Q}$ so that the variance of the estimator is reduced and the importance weight is easy to compute. We also require that $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{Q}$. Generally, the construction of $\mathbb{Q}$ is quite difficult, however, for stochastic differential equations, the alternative stochastic process $\tilde{X}$ and the importance weights can be found by Girsanov's theorem. Suppose $\tilde{X}$ is given by the SDE

$$\begin{cases} \mathrm{d}\tilde{X}_t &= \left[\mathbf{A}(\tilde{X}_t) + \mathbf{B}(\tilde{X}_t)\mathbf{u}(t,\tilde{X}_t)\right]\,\mathrm{d}t + \mathbf{B}(\tilde{X}_t)\,\mathrm{d}W_t \\ \tilde{X}_0 &= x. \end{cases} \tag{23}$$

Then $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{Q}$ and the importance weight is

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} = \exp\left[-\int_0^T \langle \mathbf{u}(t,\tilde{X}_t),\, \mathrm{d}W_t\rangle - \frac{1}{2}\int_0^T \|\mathbf{u}(t,\tilde{X}_t)\|^2\,\mathrm{d}t\right]. \tag{24}$$

We can see that the problem of constructing a good probability measure $\mathbb{Q}$ for importance sampling can be parametrized in terms of finding the relevant function $\mathbf{u}(t,x)$. For SDEs, much of the literature for importance sampling boils down to finding a good $\mathbf{u}(t,x)$ so that the variance of the resulting estimator is small, or better yet, zero. Much of the rare event simulation literature is devoted to using other tools to construct a good estimator.

Here we discuss a choice of $\mathbf{u}(t,x)$ so that the resulting importance sampling estimator will have zero variance. The biasing function comes from the Doob transform

$$u(s,x) = \mathbf{B}(x)^*\nabla\log\Phi(t,x) = \mathbf{B}(x)^*\frac{\nabla\Phi(t,x)}{\Phi(t,x)} \tag{25}$$

where $\Phi(t,x)$ is the solution to the Kolmogorov backward equation given as follows

$$\begin{cases} \frac{\partial \Phi}{\partial t} + \mathscr{A}\Phi = 0 \\ \Phi(T,x) = f(x). \end{cases} \tag{26}$$

We provide a proof for this fact in the following proposition:

Proposition 3.1. *Let $\Phi(t,x) = \mathbb{E}[f(X_T)|X_t = x]$ be the solution to the KBE. Then the biasing function*

$$u(t,x) = \boldsymbol{B}^*(x)[\nabla \log \Phi(t,x)] \tag{27}$$

will satisfy

$$f(\tilde{X}_T)\exp\left[-\int_0^T \langle u(s,\tilde{X}_s),\, dW_s\rangle - \frac{1}{2}\int_0^T \|u(s,\tilde{X}_s)\|^2\, ds\right] = h(0,x) \tag{28}$$

with probability one.

Proof. We compute the stochastic integral

$$\int_0^T \langle u(s,\tilde{X}_s),\, \mathrm{d}W_s\rangle.$$

Let $g(t,x) = \log\Phi(t,x)$, we apply Itô's formula:

$$\begin{aligned}
\mathrm{d}g =& \frac{\partial}{\partial t}\log\Phi(t,x)\,\mathrm{d}t + \langle \nabla \log\Phi(t,x),\, \mathrm{d}\tilde{X}_t\rangle + \frac{1}{2}\mathrm{Tr}\left[\nabla^2[\log\Phi(t,x)](dX_t)(dX_t)^*\right] \\
=& \frac{1}{\Phi}\frac{\partial}{\partial t}\Phi\,\mathrm{d}t + \left\langle \frac{1}{\Phi}\nabla\Phi, A(\tilde{X}_t) + BB^*\frac{\nabla\Phi}{\Phi}\right\rangle \mathrm{d}t + \left\langle \frac{\nabla\Phi}{\Phi}, BdW_t\right\rangle \\
&+ \frac{1}{2}\mathrm{Tr}\left[BB^*\frac{\nabla^2\Phi}{\Phi}\right]\mathrm{d}t - \frac{1}{2}\mathrm{Tr}\left[BB^*\frac{(\nabla\Phi)^2}{\Phi^2}\right]\mathrm{d}t \\
=& \frac{1}{\Phi}\left(\frac{\partial}{\partial t}\Phi + \langle\nabla\Phi, A\rangle + \frac{1}{2}\mathrm{Tr}\left[BB^*\nabla^2\Phi\right]\right)\mathrm{d}t + \frac{1}{2}\left\langle \frac{B^*\nabla\Phi}{\Phi}, \frac{B^*\nabla\Phi}{\Phi}\right\rangle \mathrm{d}t \\
&+ \left\langle \frac{\nabla\Phi}{\Phi}, BdW_s\right\rangle \\
=& \frac{1}{2}\|u(t,\tilde{X}_t)\|^2\,\mathrm{d}t + \langle u(t,\tilde{X}_t), dW_t\rangle
\end{aligned}$$

This implies that

$$\int_0^T \langle u(s,\tilde{X}_s), dW_s\rangle = \log h(T,x) - \log h(0,x) - \frac{1}{2}\int_0^T \|u(s,\tilde{X}_s)\|^2\,\mathrm{d}s.$$

Plugging this into the above statement, we have

$$\begin{aligned}
f(\tilde{X}_T)\exp\left[-\int_0^T \langle u(s,\tilde{X}_s),\, \mathrm{d}W_s\rangle - \frac{1}{2}\int_0^T \|u(s,\tilde{X}_s)\|^2\,\mathrm{d}s\right] &= f(\tilde{X}_T)\exp\left[\log h(0,x) - \log h(T,x)\right] \\
&= f(\tilde{X}_T)\frac{h(0,x)}{f(\tilde{X}_T)} \\
&= h(0,x).
\end{aligned}$$

□

This result should not be surprising – the zero-variance estimator is just alternative way for writing the quantity of interest. Any good variance reduction method typically tries to approximate the zero-variance estimator in an indirect way even if it is not a direct approximation. Again, we state that typically one does not directly pursue solving the KBE even though the PDE is linear.

3.2.3 Splitting Methods

In splitting methods, we create a series of nested super sets of the rare events, i.e., $C_i \subset C_{i-1}, i = 1, \ldots, M$, where C_M represents the set of rare events. Particularly, the probability of event C_i is defined as follows: $P(C_i) = P(T_{c_i} < T)$, where $T_{c_i} = \inf(t|X(t) \in C_i)$, where T is given as the terminal time. Equivalently, the state space is partitioned into a sequence of differences of nested subsets $C_i - C_{i-1}, i = 0, .., M$. Intuitively, the supersets of the rare events are less and less rare and the conditional entering probabilities of the subset from its superset is not "small". The splitting methods are based on the following decomposition of the probability of the rare events:

$$\begin{aligned} P(C_M) &= P(C_M, C_{M-1}) = P(C_M|C_{M-1})P(C_{M-1}) \\ &= \ldots = P(C_0)\prod_{i=1}^{M} P(C_i|C_{i-1}), \end{aligned} \tag{29}$$

where $P(C_i|C_{i-1})$ is the conditional probability of the events where the dynamic path firstly enters set C_i before the terminal time T given that the initial state of the particle is on the threshold of C_{i-1}. Consequently, we estimate $P(C_i|C_{i-1})$ separately and multiply the estimations together to obtain the estimation of the probability of the rare event - $P(C_M)$.

The importance function, $\phi(X) : R^n \to R$, is used to define the interfaces between the sets $C_i, i = 1, \ldots, M$, n is the dimension of the state space. The fact that the number of particles which start from C_0 would decrease dramatically towards the direction of C_M motivates the splitting of the particles. Specifically, we clone $R_i - 1$ particles as the particle reaches C_i from C_{i-1}.

We can estimate the conditional entrance probability of C_i by

$$P(C_i|C_{i-1}) \approx \frac{N_i}{N_{i-1}R_{i-1}},$$

where N_{i-1} is the number of particles which entered C_{i-1}, R_{i-1} is the splitting ratio on level $i-1$, N_i is the number of particles which entered C_i. The estimation of the rare event probability is then

$$P(C_M) \approx \frac{N_M}{Nr},$$

with $r = \prod_{i=1}^{M-1} R_i$ and N is the number of initial particles, N_M is the number of particles which reach C_M in the simulations.

There are several numerical parameters which can be tuned to obtain the optimal performance of the above estimator, for instance, the number of levels (thresholds) and the importance function. The choices of these parameters affect the performance of the estimator. For example, if the R_i is too small, there might be no trajectories which can reach the rare events. If the R_i is too big, the number of particles can explode which causes numerical instability. The threshold is

commonly chosen as follows

$$\Delta_\phi = \frac{\log R}{n} \tag{30}$$

where n defines the rareness of the events, specifically, it can be the magnitude of the reciprocal of the random perturbation in the system. A heuristic argument is the following [7]:

$$\begin{aligned} P(\text{particle born at threshold } j \text{ reaches } j-1) &\approx e^{-n \inf I_\Delta(\phi)} \\ &\approx e^{-n \inf I_\Delta(r(C_j^n) - r(C_{j-1}^n))} \\ &\approx e^{-n \log R/n} = \frac{1}{R} \end{aligned}$$

Hence, only one particle will make it to the next level in a probability sense.

In our study, we apply one of the popular methods – RESTART (REpetitive Simulation Trials After Reaching Thresholds)[17–19]. In RESTART, the initial particles split when they reach any threshold C_j and the particles get terminated when they reach the threshold where they were born, so that most of the particles get terminated prematurely once they are identified as moving away from the direction of the rare events. The original particles never get prematurely terminated until the ending time of the simulation. It can be shown that the estimator is unbiased [18, 19].

In RESTART, we define two extra events: B_i: events at which the particle enters C_i from its complement $\overline{C_i}$; D_i: events at which the particle reaches $\overline{C_i}$ from C_i. If events B_i happen, $R_i - 1$ copies of the path are saved, while each copy is tagged using integer i, meaning that it is born on level i. The seeds are terminated when D_i happens or the simulation time ends.

3.3 Discretization of Stochastic Ordinary and Partial Differential Equations

Some methods for thoughtful discretization of stochastic ODEs and PDEs. Numerical methods for deterministic ordinary and stochastic differential equations cannot be applied to SODEs or SPDEs without care – one cannot simply apply the standard Euler or Runge-Kutta methods to stochastic equations. Discretizing SODEs and SPDEs is a fundamentally different problem and deriving numerical schemes that approximate their solutions must done with caution.

The main schemes we use to approximate solutions of SODEs are the Euler-Maruyama scheme, Milstein scheme [20] and the Runge-Kutta schemes as described by Rossler [21].

The main scheme we use to approximate solutions of SPDEs is the exponential Euler scheme, and its variants. Typically, one considers stochastic PDEs as an abstract stochastic evolution equation [22] of the form

$$\begin{cases} \mathrm{d}X_t &= [AX_t + F(X_t)]\ \mathrm{d}t + \mathrm{d}W_t^Q \\ X_0 &= x \end{cases} \tag{31}$$

where X_t evolves on some Hilbert space H with inner product $\langle \cdot, \cdot \rangle_H$ (e.g. L^2 with the standard inner product), A is a self-adjoint linear operator from H to H, F is a nonlinear term on H, and W_t^Q is a standard Q-Wiener process in which Q is a positive-definite self-adjoint operator on H. Let $\{e_k\}_{k \in \mathcal{N}}$ be the eigenvectors of Q with eigenvalues $q_k > 0$. A Q-Wiener process can be expressed as

$$W_t^Q = \sum_{k=1}^{\infty} \sqrt{q_k} \beta_k(t) e_k \tag{32}$$

where $\beta_k(t)$ are independent real-valued Wiener processes. Often, the solution to this SPDE will not be differentiable in the classical sense due to the rough nature of the forcing. Therefore, existance of a solution is usually treated in the mild sense in the form of an integral equation

$$X_t = e^{At}x + \int_0^t e^{A(t-s)}F(X_s)\ \mathrm{d}s + \int_0^t e^{A(t-s)}\ \mathrm{d}W_s. \tag{33}$$

The SPDEs we study in this report are *semilinear* as the drift term can be typically written as the sum of a Laplacian and a nonlinear term

$$\frac{\partial u}{\partial t} = \Delta u + F(u) + \frac{\partial^2 W}{\partial t \partial x} \tag{34}$$

$$u(0,x) = u_0(x) \tag{35}$$

$$u(t,0) = u(t,1) = 0. \tag{36}$$

For example, if $F(u) = \partial u/\partial x$, then we have the linear stochastic advection-diffusion equations. If $F(u) = -4w(1-w^2)$, then we have the stochastic Allen-Cahn equations. For our purposes, we typically assume that the systems are driven by spacetime White noise, meaning that Q is the identity operator, so that any orthonormal basis would constitute an eigenbasis of Q. For simplicity of notation, we write $W = W^I$.

The abstract stochastic evolution equations are still generally infinite-dimensional, therefore, typically one studies the evolution of the solution on a finite-dimensional subspace. Let $\{e_k\}_{k\in\mathbb{N}}$ be an orthonormal basis of H which are eigenfunctions of A with eigenvalues λ_k. Define the projection operator $P_N : H \longrightarrow H$ to be

$$P_N f = \sum_{k=1}^{N} \langle f, e_k\rangle_H e_k. \tag{37}$$

Applying the projection operator on the abstract evolution equation gives us the semidiscrete form

$$\begin{cases} \mathrm{d}X^N(t) = \left[A^N X^N(t) + F_N(X^N(t))\right]\ \mathrm{d}t + \ \mathrm{d}W^N(t) \\ X^N(0) = P_N x = x^N \end{cases} \tag{38}$$

where $X^N = P_N X$, and W^N

$$W^N(t) = \sum_{k=1}^{N} \beta_k(t) e_k(x). \tag{39}$$

The integral form of the solution is then

$$X_t^N = e^{A_N t}x^N + \int_0^t e^{A_N(t-s)}F_N(X_s^N)\ \mathrm{d}s + \int_0^t e^{A_N(t-s)}\ \mathrm{d}W_s^N. \tag{40}$$

The exponential Euler scheme uses the integral form of the semidiscrete form to devise a numerical scheme. Let $Y_k = X^N(t_k)$ and Δt to be the temporal step size. We have

$$Y_{k+1} = e^{A_N \Delta t}Y_k + A_N^{-1}(e^{A_N\Delta t} - I)F_N(Y_k) + \int_{t_k}^{t_{k+1}} e^{A_N(t_{k+1}-s)}\ \mathrm{d}W_s^N. \tag{41}$$

Let $Y_{k,i} = \langle e_i, Y_k \rangle_H$. By projecting the approximate solution at time t_k onto the orthonormal basis $\{e_k\}_{k\in\mathbb{N}}$, we have the equation

$$Y_{k+1,i} = e^{-\lambda_i \Delta t} Y_{k,i} + \frac{1 - e^{-\lambda_i \Delta t}}{\lambda_i} F_N^i(Y_k) + \sqrt{\frac{1}{2\lambda_i}(1 - e^{-2\lambda_i \Delta t})} R_k^i \tag{42}$$

where R_k^i are i.i.d. standard normal random variables.

Given certain regularity conditions on $F(u)$ (typically globally Lipschitz), the scheme can be proven to be strongly convergent with order $1/3$. However, when $F(u)$ has nonlinearities that grow superlinearly, this scheme is provably *divergent* [23]. This is an issue when trying to solve equations such as the stochastic Allen-Cahn. There are two main approaches to fixing the convergence issue. The first is to use an implicit exponential Euler scheme – in this case, for each time step, one would need to perform a nonlinear solve. This added computational cost is not appealing for practical purposes. The other approach is to somehow mollify the nonlinearity, for example by truncating it outside some ball so that the truncated nonlinearity is globally Lipschitz [24].

We consider the *tamed* exponential Euler method, in which the nonlinearity is scaled down depending on the discretization level and size of the nonlinearity [25]. We have

$$Y_{k+1,i} = e^{-\lambda_i \Delta t} Y_{k,i} + \frac{1 - e^{-\lambda_i \Delta t}}{\lambda_i (1 + \Delta t \|F_N(Y_k)\|_H)} F_N^i(Y_k) + \sqrt{\frac{1}{2\lambda_i}(1 - e^{-2\lambda_i \Delta t})} R_k^i \tag{43}$$

This method is explicit and is shown to converge for equations with cubic nonlinearities, e.g., the stochastic Allen-Cahn.

For the case where $A = \gamma\Delta$, we have

$$e_k(x) = \sqrt{2} \sin(k\pi x) \tag{44}$$

with eigenvalues $\lambda_k = \gamma(k\pi)^2$ for $k \in \mathbb{N}$.

4 Engineering Models

Before we discuss the methodological contributions of this project, we discuss the rotorcraft and electrical models that were used for some of the computational results.

In this section, we formulate the rotorcraft model. We use a rigid beam rotor analysis code called GENTRIM [26]. It is designed for fast rotor performance analysis based on blade element methods (BEM) and includes a trim control solver. Airfoil tables are used for section aerodynamics. We use the publicly available NACA0012 airfoil data for the airfoil table. For high fidelity aerodynamics, the model can be coupled with computational fluid dynamics code. In this paper, the aerodynamics is only based on airfoil table data and surrogate models. Typically, a uniform induced velocity is used and blade flapping dynamics can easily be included. With given flight (ambient) and rotor conditions (such as RPM, desired thrust and hub moments), GENTRIM trims the rotor with cyclic control inputs ($\theta_0, \theta_{1C}, \theta_{1S}$), and outputs corresponding performance metrics (such as rotor torque, propulsive force, and lift over drag ratio). A summary of its input and output variables are given in Table 1.

When the aerodynamics of the model is based on airfoil tables, the fidelity of the aerodynamic forces depend on the airfoil table resolution. Fine resolution of the Mach number and angle of attack lead to more accurate predictions of aerodynamic forces. Additionally, stall characteristics can also be included in the airfoil tables. While GENTRIM does not include aeroelastic modeling, the source code is fully accessible to the user. Therefore any necessary modifications to the states for performing online updates can be made relatively easily.

Simple vehicle dynamics can be included for angle of attack stability analysis. Typically, helicopters are unstable in response to upward gusts during forward flight. The upward gusts (V_z) cause a higher angle of attack, which increases rotor thrust. If the thrust vector is located ahead of the fuselage CG point, the increased thrust produces a nose-up pitching moment. The noise-up attitude further increases the rotor disk angle of attack, hence leading to an instability. Even if the thrust vector is located behind the fuselage CG point, the rotor flapping dynamics makes the high angle of attack lead blade to tilt backwards due to the asymmetry in dynamic pressure in the advancing and retreating sides. This effect also contributes to destabilization. To mitigate stability issues, we require the net pitching moment (M_{CG}) to be such that $\partial M_{\text{CG}}/\partial V_z < 0$. To this end, a horizontal tail with an airfoil is introduced to balance the pitching moment of the rotor. The simplified equations of motion can be expressed as:

$$\begin{aligned} \dot{\theta} &= q, \\ \dot{q} &= \frac{M_{CG}}{I_y}, \\ M_{CG} &= M_R + T ds_R - L_H ds_H. \end{aligned} \tag{45}$$

In the above equations, θ is the pitch of the rotorcraft, q is the pitch rate, M_R is the moment around the rotor hub, $T ds_R$ is the moment due to rotor thrust, and $L_H ds_H$ is the moment due to lift from the horizontal tail. Thus, depending on the size of the horizontal tail, rotor characteristics, and flight conditions, a vehicle can, under certain conditions, encounter instabilities and potentially crash. To convey the complexity of the model and the variables we are considering, we refer the reader to Table 1 and to Figure 1. Figure 1 pictorially shows how the model parameters and variables depend on each other.

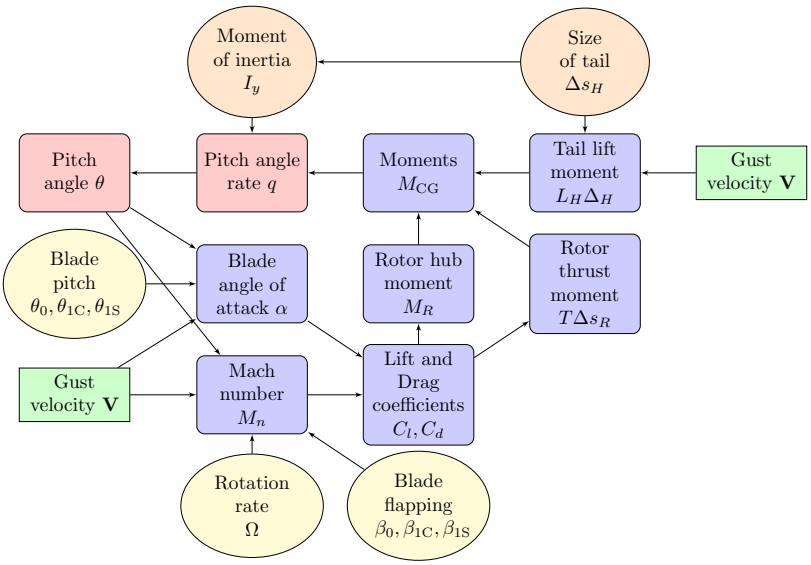

Figure 1: Schematic of rotorcraft model and their interdependencies.

Table 1: Parameters of the rotorcraft and trim model. The input and output variables of the trim control are labeled accordingly.

I/O	**Variable Type**	**Variable**	**Description**
Inputs	Geometry	Radius	Blade radius
		Twist distribution	Blade twist distribution
		Chord distribution	Blade chord length distribution
	Aerodynamics	Airfoil table $(C_l, C_d = f(M_{\text{tip}}, \alpha))$	C81 table for each airfoil, containing C_l and C_d as function of Mach number and angle of attack
	Operating Condition	M_{tip}	Tip Mach number
		C_T	Rotor thrust coefficient
		C_{M-x}	Hub roll moment (+ starboard up)
		C_{M-y}	Hub pitching moment (+ nose up)
		Pressure	Ambient pressure
		Temperature	Ambient temperature
	Flight Condition	μ	Advance ratio $\frac{V_\infty}{V_{\text{Tip}}}$
		α_s	Shaft angle of attack
Outputs	Forces/Moments	C_Q	Torque
		C_X	Drag or propulsive Force
	Blade Pitch Control	$\theta_0, \theta_{1C}, \theta_{1S}$	$\theta_{\text{bl}} = \theta_0 + \theta_{1C}\cos(\psi) + \theta_{1S}\sin(\psi)$
	Blade Flapping	$\beta_0, \beta_{1C}, \beta_{1S}$	$\beta = \beta_0 + \beta_{1C}\cos(\psi) + \beta_{1S}\sin(\psi)$
	Rotor Performance	L/D_e	$\frac{L}{D_e} = \frac{C_L}{C_H + C_Q/\mu}$

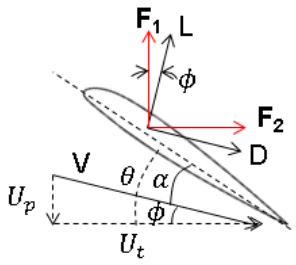

Figure 2: Airfoil section pitching.

4.1 Rotor Blade Element Method (BEM)

The GENTRIM code evaluates rotor thrust, torque, horizontal force, roll and pitch hub moments, and rotor lift over drag ratio L/D_e with given input parameters using the blade element method (BEM). It can also trim the rotor for specified thrust and hub moments, and outputs pitch control settings. The blade element method is used to compute integrated hub forces and moments. At a given blade section (Figure 2), the normal and tangential forces are computed as follows,

$$\begin{aligned}
\alpha &= \theta - \phi, \\
U_p &= V_i \cos(\beta) + r\dot{\beta} + V_x \cos(\alpha_s)\sin(\beta)\cos(\psi) - V_x \sin(\alpha_s)\cos(\beta) - V_z \cos(\alpha_s)\cos(\beta), \\
U_t &= \Omega r \cos(\beta) + V_x \cos(\alpha_s)\sin(\psi) - V_z \sin(\alpha_s)\sin(\psi), \\
C_l, C_d &= f(\alpha, M, \omega_{C_l}, \omega_{C_d}), \\
F_1 &= C_l \cos(\phi) - C_d \sin(\phi), \\
F_2 &= C_l \sin(\phi) + C_d \cos(\phi).
\end{aligned} \tag{46}$$

Here, α is the angle of attack, V_i is the induced velocity, β is the blade flapping angle, ψ is the azimuth angle, α_s is the shaft angle of attack, M_n is the Mach number, Ω is the blade rotational speed, r is the section radial location, V_x and V_z are freestream velocity components, and ω_{C_l} and ω_{C_d} are uncertainty weighting factor. In this paper, there is no uncertainty when computing the coefficients of lift and drag. This model feature may be considered for future research. The blade pitch angle is described using the following equation,

$$\theta = \theta_0 + \left(\frac{r}{R} - 0.75\right)\theta_{tw} + \theta_{1s}\sin(\psi) + \theta_{1c}\cos(\psi) + \theta_{el}, \tag{47}$$

where, θ_0 is the collective pitch angle, θ_{tw} is the twist angle, θ_{1s} and θ_{1c} are the cyclic pitch angles, and θ_{el} is the elastic deflection. Once section F_1 and F_2 forces are computed, integrated hub loadings at a given azimuth angle are computed as

$$\begin{aligned}
\text{Thrust} &= \int_{r_0}^{\text{Tip}} F_1 \cos(\beta) dr, \\
\text{Torque} &= \int_{r_0}^{\text{Tip}} F_2 r \cos(\beta) dr.
\end{aligned} \tag{48}$$

To compute the induced velocity on the rotor blades, a linear inflow model is used as shown in following expressions

$$\begin{aligned} \frac{V_i}{V_{\text{tip}}} &= \lambda_0(1 + k_x \cos(\psi) + k_y r \sin(\psi)), \\ \lambda_0 &= \frac{C_T}{2\sqrt{\mu^2 + (\mu \tan(\alpha_s) + \lambda_0)^2}}, \end{aligned} \tag{49}$$

where C_T is the thrust coefficient and μ is the advance ratio which is given as,

$$\mu = \frac{V_\infty \cos(\alpha_s)}{V_{\text{tip}}}. \tag{50}$$

The second equation for λ_0 is solved using Newton iterations for given thrust and advance ratio conditions. The lift from the horizontal tail is computed as

$$L_H = \frac{1}{2}(V_x^2 + V_z^2) A_t C_{l_\alpha} \alpha_t, \tag{51}$$

where A_t is the area of the tail airfoil. To determine the lift coefficient of the tail airfoil, we use an affine model $C_l = C_{\alpha_0} + C_{l_\alpha} \alpha_t$, where C_{α_0} is the lift coefficient when the angle of attack is zero. We model stall of the airfoil by cutting off the lift coefficient past $\pm 15^\circ$. That is, any angle of attack outside $\pm 15^\circ$ has zero lift coefficient. While in reality, the angle of attack does not go to zero immediately outside this region, we model it as described since we do not simulate the rotorcraft for angles of attack outside this range. For an approximation to the NACA0012 airfoil we have $C_{l_\alpha} = 2\pi$. For detailed equations for rotor blade element method we refer the reader to [27, 28].

4.2 Aircraft Electric Model

Collins Aerospace, a business unit of UTC, is a leading supplier of aerospace electric systems for both commercial and military platforms, and supports the aerospace industry initiative towards more electric and all-electric aircraft, optimizing performance, decreasing operating and maintenance costs, increasing dispatch reliability, and reducing gas emissions. To realize such aircraft benefits, UTRC is supporting Collins Aerospace to develop disruptive change in technologies that address reduction of volume and weight while increased reliability of the electric system, goals which align closely with the SIRE program.

A typical airplane electric power system comprises of distributed power generators and loads including energy storage on an electric network. Some of the loads operate continuously while others operate intermittently depending on the flight mission. There are multiple power generators to increase redundancy and improve availability of the system under fault condition. The power distribution system has the ability to route the electric power from various sources to the loads through contactor switches. Such a system with distributed sources and dynamic loads is complex and requires significant integration and testing effort to qualify for flightworthiness.

In emerging and future aircraft platforms the electrical power required during a flight mission will continue to grow (from the KiloWatt to MegaWatt range); while the system must be robust and should maintain power to flight critical loads. The demand for higher electric power implies lower source and distribution impedance resulting in potential for increased fault energy. Also the

loads in such an electric system are dynamic and exhibit transient constant power characteristics which translate into negative dynamic resistances. This results in dynamic coupling between power electronics loads and sources which can potentially destabilize or create resonances in the power system. The dynamic loads could also exceed the power generator capability in certain mode of operation. In addition, the electric sub-components have potential failure modes when exposed to electrical power quality non-compliance. The components will operate effectively only if its input power quality is within specification limits. Some other examples of issues related to the electric system are harmonic distortion in current and voltages resulting in stress on the system, EMI/EMC, environmental effect causing material properties break-down, corona or partial discharge.

Given the aforementioned challenges, design for higher efficiency, reliability and reduced weight system requires a strong capability in electric power system component modeling, fault analysis and design in compliance with the aircraft system requirements. Of which, electric system stability is one of the predominant issues that should be addressed at an earlier stage in the design without resorting to extensive hardware testing and hence taken for study in this effort.

We will address the *stability* concerns of a representative electrical system. The system comprises of a three phase AC source (generator) feeding an active rectifier (AC/DC converter) which provides a common DC bus voltage for a motor driven load and a battery based energy storage system. There are three main modes of operation with respect to the power flow. During normal operation, the power flows from the generator to the motor through the rectifier and the inverter. During regeneration mode, the power flows from the motor to the battery through the inverter and DC/DC converter. During high transient load condition, the power flows from battery to the motor through the DC/DC converter and inverter. The design goal is to guarantee *stable* operation during all three modes of operation. The electric system architecture chosen for analysis is fundamental to different aircraft platforms, consequently the results from this work will be broadly applicable. The first step in analyzing the system using SIRE method is to develop appropriate models.

To start the computations we have constructed the model described below (for testing).

4.2.1 Simple Model

To emulate the rare event behavioral characteristics of the electric system and its loads, a preliminary non-linear model of a Buck Converter is made. The Buck converter as shown in Figure 3 is representative of several aircraft system converters and can be used as a test case to model the rare event dynamics. It is a DC-DC converter that steps input voltage down to a desired output voltage. It comprises of aircraft DC source V_s, switch S, diode D, inductor L, output capacitor C, and variable load R. The assumption is that the DC source is fixed in amplitude and the loads are linear in nature. A switching model is made in the MATLAB Simulink environment where in the initial conditions of the state variables can be updated dynamically during a simulation run. The state variables in this model are the inductor current and capacitor voltage and the simplified model is expressed as:

$$\begin{bmatrix} \dot{i}_L \\ \dot{V}_0 \end{bmatrix} = \begin{bmatrix} -\frac{R}{L} & \frac{-1}{L} \\ \frac{1}{C} & 0 \end{bmatrix} \begin{bmatrix} i_L \\ V_0 \end{bmatrix} + S_n \begin{bmatrix} V_s \\ 0 \end{bmatrix} + \sigma \begin{bmatrix} 1 \\ 0 \end{bmatrix} \, \mathrm{d}\dot{W}_t \tag{52}$$

where the switching function S_n takes the values of 1 and 0 when the switch S is on or off respectively. The critical inductance for continuous conduction of the Buck converter is

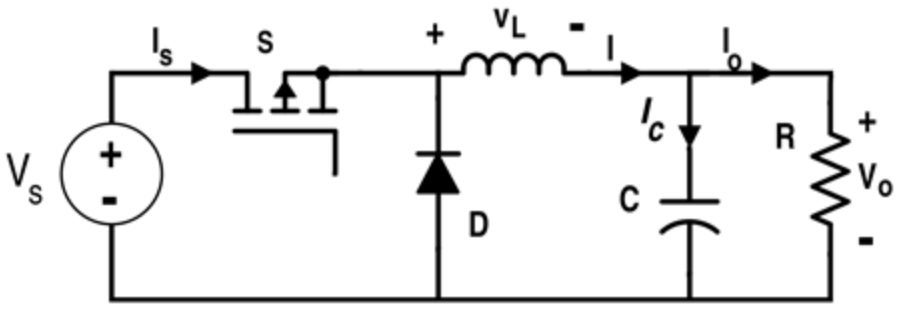

Figure 3: Buck Converter circuit.

$L_{crit} = (1-D)TR/2$, where D is the average duty cycle for with the switch S is ON, and T is the total switching period. The inductor is also modeled as a nonlinear element as it is subject to saturation of the core at higher flux densities. Thus during dynamic mode of operation, if the critical condition for continuous conduction (L_{crit}), or flux saturation of inductor occurs, the performance of buck converter is compromised and can be considered as rare event failure.

5 The Koopman Operator and Rare Event Simulation

One of the major contributions of this project makes the connection between the Koopman operator, specifically its stochastic version, and efficient rare event simulation algorithms. From an intuitive standpoint, existing rare event simulation methods are often concerned with the *dynamics of the state* in that methods try to find clever ways of pushing the state of a system into its rare regions in some direct fashion. On the other hand, the Koopman operator deliberately avoids this focus, rather it studies dynamical systems by looking at the *dynamics of the observables*. The key insight of our new approach is treating rare event probabilities as observables and seeing how they evolve through the lens of the Koopman operator. Doing so allows the Koopman operator to inform how a system evolves towards rare events.

In this section, we first review the theoretical properties of the Koopman operator and its implications for the analysis of nonlinear dynamical systems. Next, we review a class of numerical methods called dynamic mode decomposition (DMD) that allow one to practically study the Koopman operator. Afterwards, we explain the connection between the Koopman operator and the Kolmogorov backward equations, and present the new strategy for rare event simulation in general SDE systems. We present a special case of our method to linear SDE systems and describe how this can provide intuition on designing better estimator for linear stochastic PDEs.

5.1 The Koopman Operator and the Dynamics of Observables

We begin by discussing the Koopman operator of a deterministic continuous-time dynamical system. This review is based on [29]. Consider an autonomous differential equation on state space $S \subset \mathbb{R}^d$

$$\begin{cases} \dot{x}_t &= g(x_t) \\ x_0 &= x \end{cases} \tag{53}$$

where g is a function from S to itself and $x \in S$ is the initial condition. This dynamical system induces a flow map $F^t : S \longrightarrow S$ such that $x_t = F^t(x_0)$ is a solution to the differential equation. The traditional way of studying dynamical systems is to characterize the flow map, and therefore, the solution paths of the dynamical system. An alternative way of studying dynamical systems is to study the dynamics of the observables of the system. An *observable* is a function h that maps from the state space to the real numbers. This a natural perspective to study dynamical systems, for example, in computational fluid dynamics, the state x_t is a very high dimensional discretization of a solution to the Navier-Stokes equation, while an observable may be the velocity of the fluid at a particular point. While the evolution of the state is given by the flow map, the evolution of the observable is given by the *Koopman operator*.

Let $\mathscr{H}$ be the space of admissable observables, which we may assume is some Hilbert space of functions (typically L^2). The Koopman operator describes the evolution of observables on $\mathscr{H}$.

Definition 5.1. *The Koopman operator $\mathscr{K}^t : \mathscr{H} \longrightarrow \mathscr{H}$ is defined as*

$$\mathscr{K}^t h = h \circ F^t \tag{54}$$

The key feature of the Koopman operator is that it is always linear even if the underlying

dynamical system is nonlinear. Observe that if $h_1, h_2 \in \mathscr{H}$ and $c \in \mathbb{R}$, then

$$\begin{aligned}\mathscr{K}^t(h_1 + ch_2) &= (h_1 + ch_2) \circ F^t \\ &= h_1 \circ F^t + c(h_2 \circ F^t) \\ &= \mathscr{K}^t h_1 + c\mathscr{K}^t h_2.\end{aligned}$$

The tradeoff for obtaining this linear relationship is that the Koopman operator is fundamentally infinite dimensional even if the dimension of the state space of the nominal dynamical system is finite.

Note that for any given dynamical system, there is not a single Koopman operator, rather there is a so-called *one-parameter semigroup* of Koopman operator. This means that for any $t, s \in \mathbb{R}$, we have that $\mathscr{K}^{t+s} = \mathscr{K}^t \circ \mathscr{K}^s$. This fact is crucial for relating the stochastic Koopman operator with the Kolmogorov equations.

The Koopman operator enables spectral analysis for nonlinear dynamical systems. Since $\mathscr{K}^t$ is a linear operator, we may ask for its eigenvalues and eigenfunctions. Specifically, we are interested in the case where the Koopman operator has discrete spectra, so that one can express observables in terms of the eigenfunctions. Let $(\lambda_i, \phi_i(x))$ be an eigenvalue-eigenfunction pair where

$$\mathscr{K}^t \phi_i(x) = e^{\lambda_i t} \phi_i(x). \tag{55}$$

Here, $\phi_(x)$ is called the Koopman eigenfunction and λ_i is the Koopman eigenvalue. If the set of Koopman eigenfunctions form a complete set, one can write evolution of any observable in a closed expression. Express the observable h as a linear combination of the Koopman eigenfunctions

$$h(x) = \sum_{i=1}^{\infty} \mu_i^h \phi_i(x).$$

Then the evolution of $h(x)$ can be expressed as

$$\begin{aligned}h \circ F^t &= \mathscr{K}^t h \\ &= \mathscr{K}^t \sum_{i=1}^{\infty} \mu_i^h \phi_i(x) \\ &= \sum_{i=1}^{\infty} \mu_i^h \mathscr{K}^t \phi_i(x) \\ &= \sum_{i=1}^{\infty} \mu_i^h e^{\lambda_i t} \phi_i(x).\end{aligned}$$

Furthermore, this allows us to write the flow map in terms of the Koopman eigenfunctions as well. Let $x : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ be the *full state observable*. Then assuming that it can be written as a linear combination of the Koopman eigenfunctions,

$$x = \sum_{i=1}^{\infty} \mu_i \phi_i(x),$$

we can write the flow map as

$$\begin{aligned} x_t = F^t(x_0) &= \mathscr{K}^t x_0 \\ &= \mathscr{K}^t \sum_{i=1}^{\infty} \mu_i \phi_i(x_0) \\ &= \sum_{i=1}^{\infty} \mu_i e^{\lambda_i t} \phi_i(x_0). \end{aligned}$$

Having the eigenfunctions of the Koopman operator is very useful. Naturally, one may ask if stochastic dynamical systems have an analogous operator. We discuss the *stochastic Koopman operator* and show, for the SDE case, that it is closely related to the generator of the SDE. Suppose we have an autonomous stochastic differential equation as given in Equation 1.

Definition 5.2. *The stochastic Koopman operator, denoted as $\mathscr{K}^{S,t}$ is defined as*

$$\mathscr{K}^{S,t} f(x) = \mathbb{E}\left[f(X_t)|X_0 = x\right] \tag{56}$$

where the expectation is evaluated over the distribution of the process at time t.

Again, we do not have a single Koopman operator, rather we have a semigroup of Koopman operators indexed by t. In semigroup theory, it is standard to study the infinitesimal generator of the semigroup by evaluating

$$\mathscr{K}^S f = \lim_{t\to 0} \frac{\mathscr{K}^{S,t} f(x) - f(x)}{t} = \lim_{t\to 0} \frac{\mathbb{E}[f(X_t)|X_0 = x] - f(x)}{t}. \tag{57}$$

Compare this expression with Equation 2. One can see that the generator of the stochastic Koopman operator semigroup is identical to the infinitesimal generator of the stochastic differential equation. By the spectral mapping theorem [29], the elements of the Koopman operator semigroup share the same eigenfunctions and eigenvalues as their generator. Now recall that the Kolmogorov backwards equation's evolution operator is exactly the infintesimal generator of the SDE. Therefore, we can find approximate solutions of the KBE if we can find the eigenfunctions of the stochastic Koopman operator.

Eigenfunctions of the stochastic Koopman operator have the property that the evolution of their expectations can be computed easily. Let ϕ_i be an eigenfunction of the generator of the stochastic Koopman operator with eigenvalue λ_i. Then we have

$$\mathbb{E}[\phi(X_t)|X_0 = x] = e^{\lambda_i t}\phi(x). \tag{58}$$

In the next section, we review dynamic mode decomposition (DMD), which is a class of methods for finding the Koopman eigenfunctions.

However, even though a direct approach is not recommended for these equations, they can still provide insight into how one approach characterizing rare events in dynamical systems.

5.2 Dynamic Mode Decomposition Methods

Dynamic mode decomposition (DMD) methods are a class of numerical algorithms that attempt to find the Koopman eigenfunctions of a dynamical system using sample paths or even real data of the dynamical system. The method was originally developed as means of model order reduction for complex fluid flows. Time series data of snapshots of the high-fidelity fluid simulations were used study or find their lower dimensional behavior. The DMD method produced modes that described global features of the flow such as coherent structures. The corresponding eigenvalues of the modes described the growth, decay, or oscillatory behavior over time. One of the main features of the method was that it was valid even if the flow had nonlinear behavior. The original DMD method as described by Schmid [30] was presented as a purely numerical linear algebra method. It was later shown in [31] that the modes produced by DMD was related to spectral quantities of the Koopman operator.

5.2.1 Extended DMD and the Koopman operator

Extended DMD [32] is a generalization of the standard and exact DMD methods and makes the connection to the Koopman operator more clear. Consider a discrete-time dynamical system on $S \subset \mathbb{R}^d$ $z_{k+1} = F(z_k)$. This dynamical system has an associated Koopman operator $\mathscr{K}$ defined on some set of admissible observable functions $\mathscr{H}$.

Suppose we have data pairs $\{(x_1, y_1), \ldots, (x_m, y_m)\}$ defined such that $y_k = F(x_k)$. Define a basis of K observables $\mathscr{D} = \{\psi_1, \ldots, \psi_K\} \subset \mathscr{H}$. The span of the basis is a finite-dimensional subspace of $\mathscr{H}$. Extended DMD aims to directly approximate the Koopman operator on $\mathscr{H}$ by finding a matrix $\mathbf{K} \in \mathbb{R}^{K \times K}$ on the subspace $\mathscr{H}_{\mathscr{D}}$ defined by the span user-defined basis. Let ϕ be an arbitrary element of $\mathscr{H}$. Following the notation of [32], ϕ can be expressed as

$$\phi = \sum_{k=1}^{K} a_k \psi_k = \Psi \mathbf{a} \tag{59}$$

where $\Psi = [\psi_1(x) \cdots \psi_K(x)]$ and $\mathbf{a} \in \mathbb{R}^K$ is the vector of coefficients. To find the approximate finite dimensional representation of the Koopman operator, we pose a least problem. The exact action of the true Koopman operator on ϕ gives us

$$\mathscr{K}\phi = \phi \circ F = (\Psi \circ F)\mathbf{a}. \tag{60}$$

The approximate Koopman operator $\tilde{\mathscr{K}}$ would act on ϕ via a linear transformation of the coefficients

$$\tilde{\mathscr{K}}\phi = \Psi(\mathbf{K}\mathbf{a}). \tag{61}$$

To find $\mathbf{K}$, we use the data snapshot pairs $\{(x_i, y_i\}$ to setup a least squares problem in which we search for the $\mathbf{K}$ that minimizes the residual between Equation 60 and Equation 61 on the snapshot pairs. The cost function is

$$J = \frac{1}{2} \sum_{m=1}^{M} \left| ((\Psi \circ F)(x_m) - \Psi(x_m)\mathbf{K})\mathbf{a} \right|^2. \tag{62}$$

Note that since $y_m = F(x_m)$, we may write

$$J = \frac{1}{2} \sum_{m=1}^{M} \left| (\Psi(y_m) - \Psi(x_m)\mathbf{K})\mathbf{a} \right|^2. \tag{63}$$

The $\mathbf{K}$ that minimizes this cost function is $\mathbf{K} = \mathbf{G}^{+}\mathbf{A}$, where

$$\mathbf{G} = \frac{1}{M}\sum_{m=1}^{M} \Psi(x_m)^{*}\Psi(x_m) \tag{64}$$

$$\mathbf{A} = \frac{1}{M}\sum_{m=1}^{M} \Psi(x_m)^{*}\Psi(y_m) \tag{65}$$

and $^{+}$ denotes the Moore-Penrose pseudoinverse. Typically one does not need to explicitly form $\mathbf{K}$ since we are interested more in the approximations of the Koopman eigenfunctions and eigenvalues in this basis. Eigenvectors of the $\mathbf{K}$ correspond to approximations of the Koopman eigenfunctions. Indeed, if ξ_k is an eigenvector with eigenvalue μ_k, then $\chi_j = \Psi\xi_k$ is the approximate eigenfunction. If the data snapshots came from a continuous-time dynamical system with time spacing Δt, and μ_k is the approximate Koopman eigenvalue of the discretized system, then $\lambda_k = \log(\mu_k)/\Delta t$ is the approximate Koopman eigenvalue of original system.

Extended DMD makes the connection of DMD methods to the Koopman operator more clear. Convergence details about DMD is explored and proven in [33]. In this viewpoint, the standard and exact DMD methods are a special case of extended DMD in which the basis of observables are the linear monomials. In some sense, it is a linear approximation of the dynamical system via the Koopman operator.

5.2.2 Generator extended dynamic mode decomposition

Generator extended dynamic mode decomposition (gEDMD) is another method that estimates the Koopman eigenfunctions of a dynamical system. Here, one works with the infinitesimal generator of the Koopman operator. It has been established that EDMD converges to a Galerkin method with respect to some user-defined measure. In other words, gEDMD basically amounts to a Galerkin approximation of the backward evolution operator, which we can use to find approximate Doob transforms.

Given a fixed basis of observables $\{\psi_i(x)\}_{i=1}^{N}$, the action of the generator is

$$\mathscr{A}\psi_i(x) = \mathbf{A}(x)^{T}\nabla\psi_i(x) + \frac{1}{2}\mathrm{Tr}\left[\mathbf{B}\mathbf{B}^{*}\nabla^{2}\psi_i(x)\right]. \tag{66}$$

Define the matrices $\mathfrak{A}$ and $\mathfrak{G}$ to be

$$\mathfrak{A}_{ij} = \int \mathscr{A}[\psi_j(x)]\psi_i(x)\,\mathrm{d}\nu \tag{67}$$

$$\mathfrak{G}_{ij} = \int \psi_i(x)\psi_j(x)\,\mathrm{d}\nu \tag{68}$$

where ν is some user-defined measure. Then the finite dimensional approximation to the Koopman generator is $\mathbf{K} = \mathfrak{G}^{-1}\mathfrak{A}$. The Koopman eigenfunction can be found by considering the eigenvectors of $\mathbf{K}$. Let v_i be the ith eigenvector of $\mathbf{K}$ with eigenvalue λ_i. Then

$$\phi_i(x) = \sum_{j=1}^{N} v_{ij}\psi_j(x) \tag{69}$$

is an approximate Koopman eigenfunction with eigenvalue λ_i.

5.3 Nonlinear Systems and Rare Events

As the systems are interest are always SDEs, we drop the qualifier 'stochastic' when discussing the stochastic Koopman operator and its eigenfunctions. The Koopman eigenfunctions provide a way to decompose the rare event probability according to different time scales. This is done by expressing the indicator function over the rare event region in terms of the eigenfunctions. Observe that we may write

$$\begin{aligned}\mathbb{P}[X_T \in E | X_0 = x] &= \mathbb{E}[\mathbb{1}_E(X_T)|X_0 = x] \\ &= \mathbb{E}\left[\sum_{i=1}^{\infty} c_i \phi_i(X_T)\Big| X_0 = x\right] \\ &= \sum_{i=1}^{\infty} c_i \mathbb{E}[\phi_i(X_T)|X_0 = x] \\ &= \sum_{i=1}^{\infty} c_i e^{\lambda_i t}\phi_i(x).\end{aligned}$$

Recall that our main goal is to develop efficient importance sampling estimators for SDE systems and that the zero-variance estimator corresponds to applying the Doob transform to the system. Exact application of the Doob transform requires the solution of the Kolmogorov backward equation. The KBE evolution operator's eigenfunctions are exactly the Koopman eigenfunctions, which can be found by applying DMD methods. We propose directly approximating the zero-variance estimator with the Koopman eigenfunctions.

The problem involves estimating the quantity $\mathbb{E}[f(X_T)|X_0 = x]$ for the SDE

$$\begin{cases} \mathrm{d}X_t &= \mathbf{A}(X_t)\ \mathrm{d}t + \mathbf{B}(X_t)\ \mathrm{d}W_t \\ X_0 &= x \end{cases} \tag{70}$$

whose infinitesimal generator is

$$\mathscr{K}^S f(x) = \langle \mathbf{A}(x), \nabla f\rangle + \frac{1}{2}\mathrm{Tr}\left[\mathbf{B}\mathbf{B}^*\nabla^2 f\right].$$

Let $\{\phi_i\}_{i=1}^N$ be the eigenfunctions of $\mathscr{K}^S$ with eigenvalue $\{\mu_i\}_{i=1}^N$. We expand the function in terms of the eigenfunctions

$$f(x) \approx \tilde{f}(x) = \sum_{i=1}^{N} f_i \phi_i(x). \tag{71}$$

Then an approximate solution to the KBE 5 is

$$\tilde{\Phi}(t,x) = \sum_{i=1}^{N} f_i e^{\mu_i(T-t)}\phi_i(x). \tag{72}$$

Then the Doob transform gives us biasing function

$$\mathbf{u}(t,x) = \mathbf{B}(x)^*\nabla \log \tilde{\Phi}(t,x) = \mathbf{B}(x)^* \frac{\sum_{i=1}^N e^{\mu_i(T-t)}\nabla\phi_i(x)}{\sum_{i=1}^N f_i e^{\mu_i(T-t)}\phi_i(x)}. \tag{73}$$

5.3.1 *Numerical Example*

We show our approach applied on to a two-dimensional nonlinear dynamical system. The system is a noisy gradient flow with a Rosenbrock potential, which is a non-quadratic potential well. The equations are

$$\mathrm{d}X_t = -\nabla V(x_t)\ \mathrm{d}t + \sqrt{\varepsilon}\ \mathrm{d}W_t \tag{74}$$

$$V(x) = x_1^2 + 0.4(x_2 - x_1^2)^2. \tag{75}$$

The probability of interest is

$$\mathbb{P}\left(\sup_{t\in[0,T]} \|X_t\| \geq 1 \middle| X_0 = 0 \right). \tag{76}$$

To find the Koopman eigenfunctions, we applied extended DMD with 10^3 trajectories over 10 seconds. We used 25 two-dimensional Hermite polynomials to be the basis. We only use the first three eigenfunctions that decay to slowest in our biasing scheme. They are plotted in Figure 4.

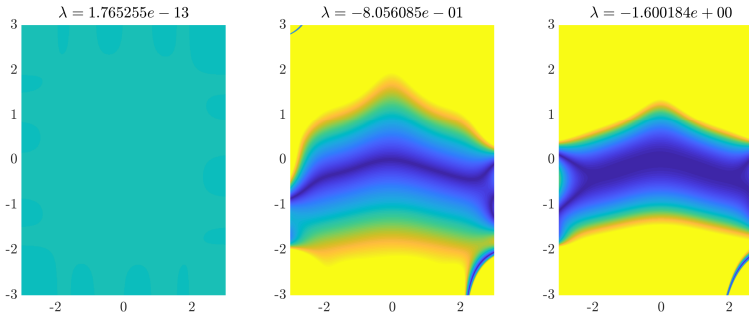

Figure 4: Koopman eigenfunctions with eigenvalues $\mu_1 = 0$, $\mu_2 = -0.81$, $\mu_3 = -1.60$.

In Figure 5, we show the exit trajectories of simple Monte Carlo and the biased samples. In Table 2, we show the variance and relative error of these methods. We see a variance reduction of a factor of 10. Since variance scales inversely with the number of samples, this implies that one can use 10 times few samples (and therefore computation time) with the Koopman biasing over standard Monte Carlo.

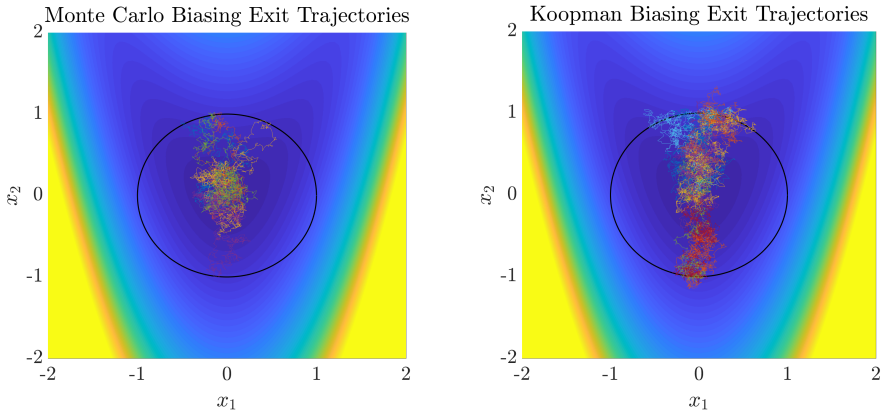

Figure 5: Exit trajectories for simple Monte Carlo and Koopman biased samples

Table 2: Results for Rosenbrock potential.

	Variance	Relative error
Monte Carlo	1.47×10^{-8}	1.55
Koopman IS	1.47×10^{-9}	0.51

$\rho_{\text{true}} = 7.6 \times 10^{-5}$, Samples per estimate: $M = 5000$

5.3.2 Van der Pol oscillator

Consider the noisy Van der Pol oscillator

$$\mathrm{d}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ \mu(1-x_1^2)x_2 - x_1 \end{bmatrix} \mathrm{d}t + \sqrt{2\varepsilon}\begin{bmatrix} \mathrm{d}W_1 \\ \mathrm{d}W_2 \end{bmatrix} \tag{77}$$

The deterministic version of this system is well-known since it exhibits a limit cycle, in which any initial condition (besides the origin) will eventually converge to the limit cycle. The remnants of the limit cycle is detectable in the stochastic case in that solutions to the stochastic Van der Pol equation cluster in a band that is centered around the limit cycle. We consider the problem of how does one 'peel' a solution of the stochastic system off of the limit cycle band. In the numerical experiments, we let $\mu = 0.3$, $\varepsilon = 0.01$ and $T = 10$, and consider estimating

$$\mathbb{P}\left(x_1(T)^2 + x_2(T)^2 > 2.7^2 | x_1(0) = -2, x_2(0) = 0\right). \tag{78}$$

We first find the stochastic Koopman eigenfunctions of the system. We apply gEDMD with basis of 2D Legendre polynomials with total order equal to 10 that is orthonormal with respect to the uniform measure on the $(x_1, x_2) \in [-6,6] \times [-6,6] \subset \mathbb{R}^2$. There are 66 elements in this basis. In Figure 6 we show the first nine Koopman eigenfunctions.

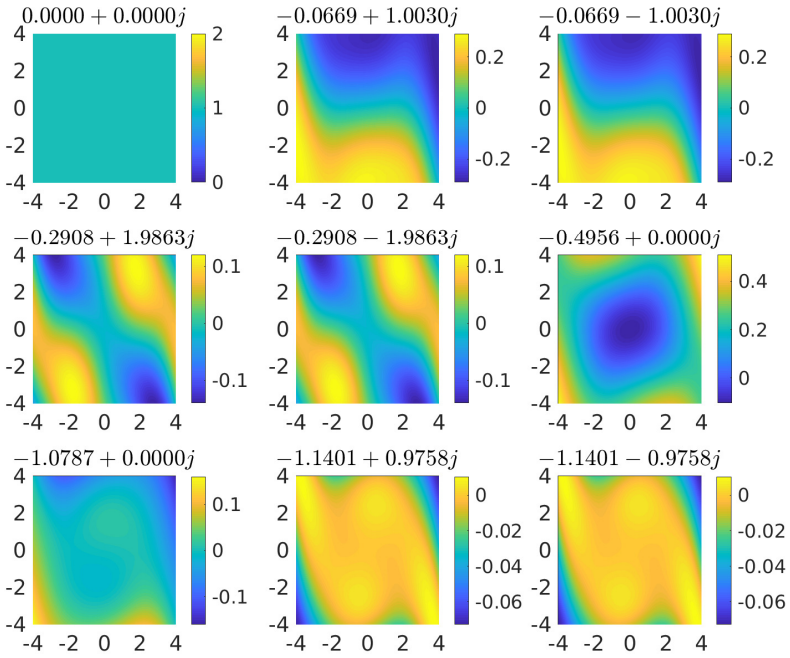

Figure 6: First nine stochastic Koopman eigenfunctions for the Van der Pol oscillator. Eigenfunctions are ordered according to the magnitude of the real part of the Koopman eigenvalues.

To find the approximate Doob transform to efficiently estimate Equation 78 we project the indicator function over the rare region of interest on to the eigenfunctions. Since the eigenfunctions are expressed in the Legendre basis, we can easily find the Koopman modes by first expanding the indicator function in terms of the Legendre basis, and then do a simple change of basis.

That is, suppose we have

$$\mathbb{1}_{\|x(T)\|>2.7}(x) \approx \sum_{i=1}^{N} c_i \psi_i(x) \tag{79}$$

where $\psi_i(x)$ are elements of the Legendre basis, and $c_i = \langle \mathbb{1}(x), \psi_i(x) \rangle_\nu$. Denote $\mathbf{c} = [c_1, \ldots, c_N]^T$. Then the Koopman modes are

$$\Xi = \mathbf{V}^{-1}\mathbf{c}. \tag{80}$$

With the modes, the approximate Doob transform is then

$$\mathbf{u}(t,x) = \sqrt{2\varepsilon}\frac{\sum_{i=1}^{N} \xi_i e^{\lambda_i (T-t)} \nabla \phi_i(x)}{\sum_{i=1}^{N} \xi_i e^{\lambda_i (T-t)} \phi_i(x)}. \tag{81}$$

In practice, we notice that applying the resulting biasing as-is does not push the samples toward the rare event region enough. Therefore, we typically scale $\mathbf{u}(t,x)$ such that approximately half of the samples the rare event of interest.

In Figure 7, we show unbiased and biased sample paths of the oscillator. In Figure 8, we show the empirically estimated probability density of the state in phase space at time T.

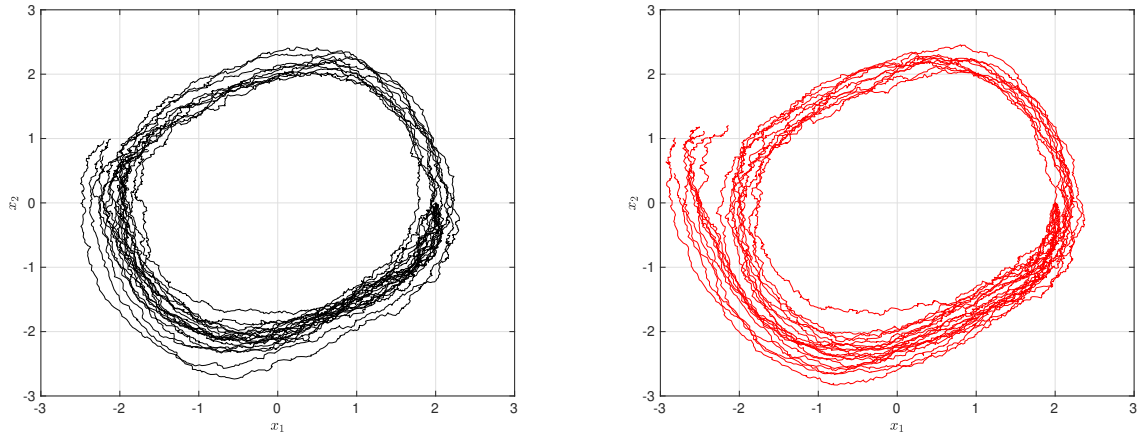

Figure 7: Left: sample paths of unbiased Van der Pol oscillator. Right: sample paths of biased Van der Pol oscillator.

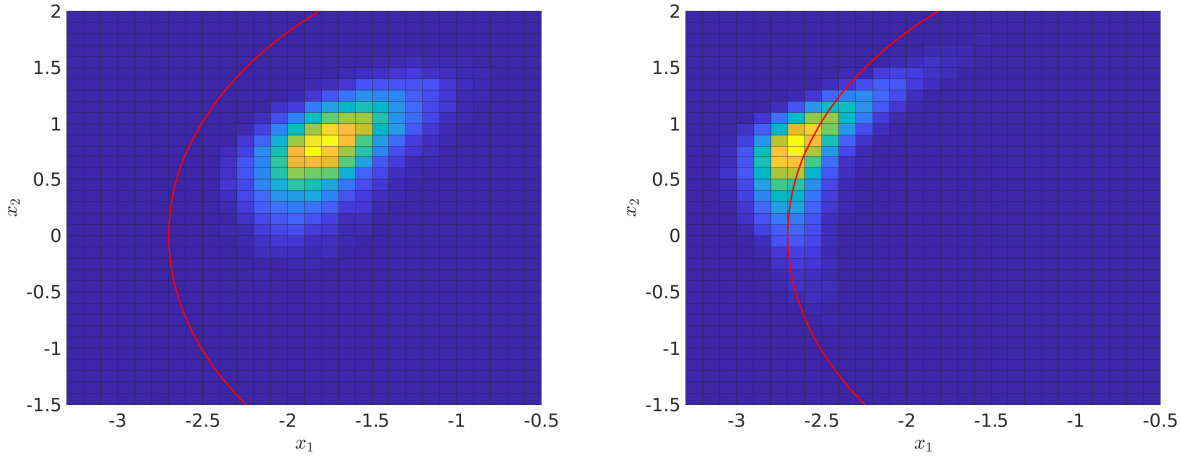

Figure 8: Left: empirical pdf of the state at time $t = T$ for the unbiased system. Right: empirical pdf of the state at time $t = T$ for the biased system. Red line denotes the boundary of the region of interest.

The true probability is estimated via brute force Monte carlo to be $\rho = 1.70 \times 10^{-5}$ using two million samples. We estimate the probability using importance sampling with only 100 samples. We obtain a variance reduction of 5000, which means that one needs a factor of 5000 fewer samples to estimate the probability of interest using importance sampling over simple Monte Carlo.

Table 3: Van der Pol oscillator results.

	Variance	Relative error
Monte Carlo	1.70×10^{-7}	24.25
Importance sampling	5.43×10^{-11}	0.43

$\rho_{\text{true}} = 1.70 \times 10^{-5}$, Samples per estimate $M = 100$

5.3.3 Duffing oscillator

Here we consider the noisy Duffing oscillator

$$\mathrm{d}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ -\delta x_2 - x_1(\beta + \alpha x_1^2) \end{bmatrix} \mathrm{d}t + \sqrt{2\varepsilon}\begin{bmatrix} \mathrm{d}W_1 \\ \mathrm{d}W_2 \end{bmatrix}. \tag{82}$$

The Duffing oscillator has the well-known feature in which there are two basins of attraction that the state oscillates in in phase space. In the deterministic version of the system has solution typically decay into one of the two fixed points while oscillating in one of the basins of attraction. In the stochastic case, small noise can perturb the state such that it can push the state from one basin of attraction to another. We estimate the probability of transitioning into the other basin of attraction given some initial condition:

$$\mathbb{P}\left[x_1(T) > 0 | x_1(0) = -1.5, x_2(0) = 0\right] \tag{83}$$

The parameters performed in the numerical experiments are $\alpha = 1$, $\beta = -1$, $\delta = 0.5$, $\varepsilon = 0.001$ and $T = 10$. Similar to the Van der Pol oscillator, we find the stochastic Koopman eigenfunctions by applying gEDMD with a basis of 2D Legendre polynomials with total order equal to 12 that is orthonormal with respect to the uniform measure on $[-3,3] \times [-3,3]$. There are 91 total elements in this basis.

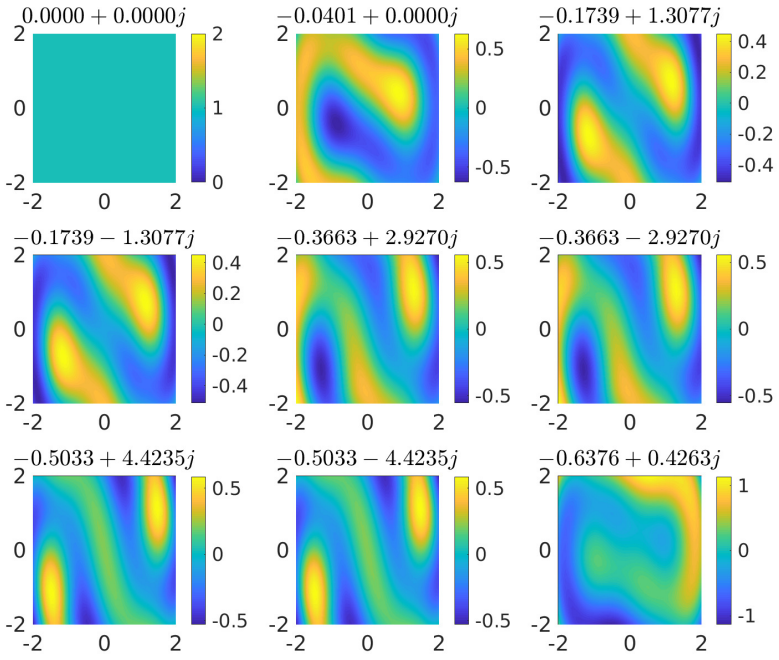

Figure 9: First nine stochastic Koopman eigenfunctions of the noisy Duffing oscillator. Eigenfunctions are ordered according to the magnitude of the real parts of the Koopman eigenvalues.

In this problem, we expand the following indicator function in terms of the Koopman eigenfunctions

$$\mathbb{1}_{x_1(T)>0}(x) = \begin{cases} 1 \text{ if } x_1(T) > 0 \\ 0 \text{ otherwise.} \end{cases} \tag{84}$$

Here, we only project the indicator function onto the first 8 Koopman eigenfunctions. Similar to the Van der Pol oscillator, we scale the biasing so that approximately half of the sample paths reach the region of interest.

In Figure 10, we show unbiased and biased sample paths of the Duffing oscillator. In Figure 11, we show the empirically estimated probability density of the state in phase space at time T.

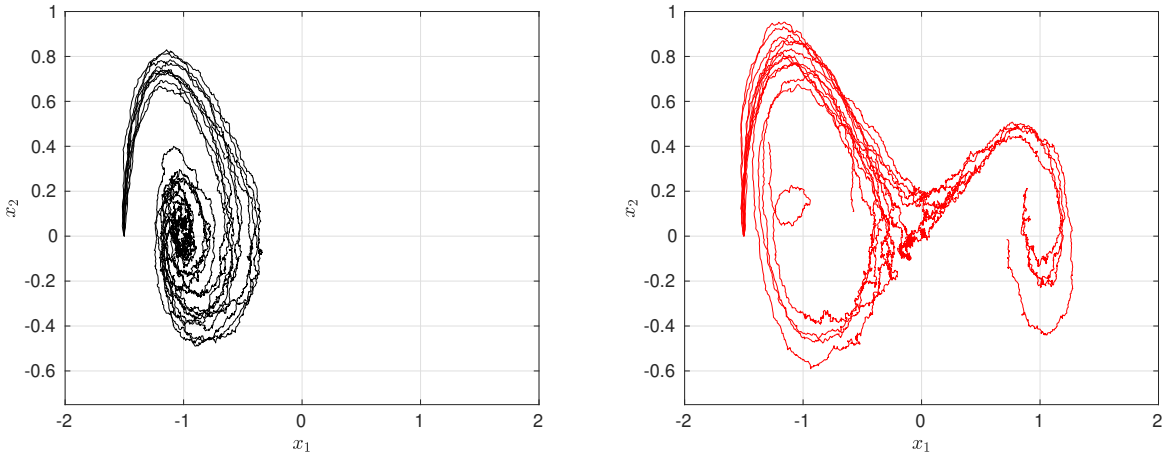

Figure 10: Left: sample paths of unbiased Duffing oscillator. Right: sample paths of Duffing oscillator.

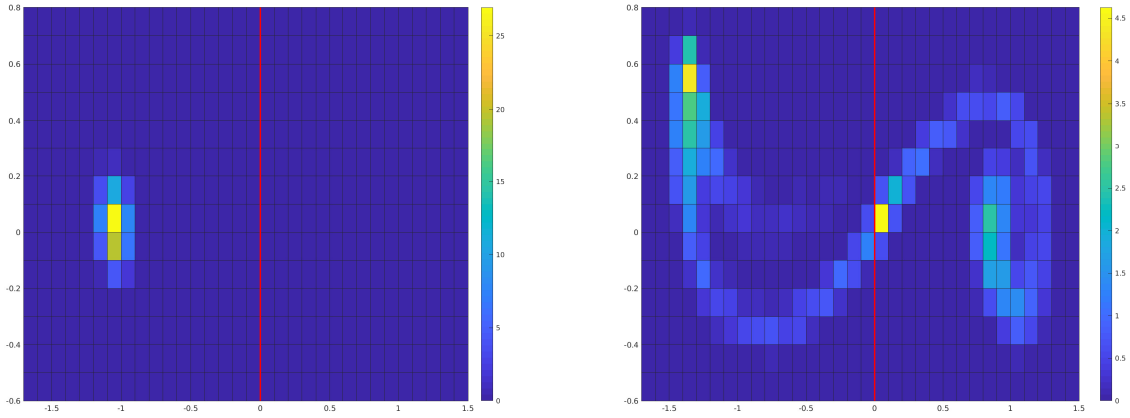

Figure 11: Left: empirical pdf of the state at time $t = T$ for the unbiased system. Right: empirical pdf of the state at time $t = T$ for the biased system. Red circle denotes the boundary of the region of interest.

The true probability is estimated via brute force Monte carlo to be $\rho = 1.70 \times 10^{-5}$ using two million samples. We estimate the probability using importance sampling with only 100 samples. We obtain a variance reduction of 5000, which means that one needs a factor of 5000 fewer samples to estimate the probability of interest using importance sampling over simple Monte Carlo.

Table 4: Duffing oscillator results.

	Variance	Relative error
Monte Carlo	4.57×10^{-8}	15.60
Importance sampling	4.38×10^{-11}	0.48

$\rho_{\text{true}} = 1.37 \times 10^{-5}$, Samples per estimate $M = 300$

5.4 Special Case: Linear Systems

In this section, we consider a special stochastic dynamical system in which the drift term is a linear function and the diffusion matrix is constant. The reason for studying these systems, commonly known as *Ornstein-Uhlenbeck processes*, is that the eigenfunctions of its infinitesimal (Koopman) generator can be found exactly. That is, no numerical approximations are needed for characterizing the eigenfunctions.

Consider the stochastic differential equation

$$\begin{cases} dX_t = \mathbf{A}X_t\,dt + \mathbf{B}\,dW_t \\ X_0 = x \end{cases} \tag{85}$$

where X_t is an element of $\mathbb{R}^d$, $\mathbf{A}$ and $\mathbf{B}$ are $d \times d$ and $d \times r$ real-valued matrices, respectively, and W_t is a standard r-dimensional Wiener process. Assume $\mathbf{A}$ is diagonalizable, and have eigenvalues with negative real parts. To be more precise, let v_i and w_i be the right and left eigenvectors of $\mathbf{A}$ with eigenvalues $-\lambda_i$, such that $\mathrm{Re}(\lambda_i) > 0$. Furthermore, assume that none of the left eigenvectors of $\mathbf{A}$ are contained within the kernel of $\mathbf{B}^*$. All these assumptions are to guarantee the existence of a unique, nondegenerate invariant measure. The infinitesimal generator of the Ornstein-Uhlenbeck process is given by

$$\mathscr{A}\psi = \langle x, \mathbf{A}^*\nabla\psi\rangle + \frac{1}{2}\mathrm{Tr}\left(\mathbf{B}\mathbf{B}^*\nabla^2\psi\right) \tag{86}$$

where f is typically a twice continuously differentiable function from $\mathbb{R}^d$ to $\mathbb{R}$. The conditions for existence and uniqueness of the eigenfunctions of the generator, as well as their properties are well-outlined in [34]. The authors manage to show that under conditions similar to our assumptions above, the OU operator has eigenfunctions when considering its actions on the space of $L^p(\nu)$ functions, where ν is the invariant measure of the OU process. Furthermore, they show that the basis of eigenfunctions is complete in $L^p(\nu)$ for $p \geq 2$, the fact that the eigenfunctions are all polynomials, and that the eigenvalues and eigenfunctions are the identical for all p. We summarize these facts in the following propositions (Theorem 3.1 and Proposition 3.1 in Metafune).

Proposition 5.3. *Let $\lambda_1, \ldots, \lambda_l$ be the distinct eigenvalues of $\mathbf{A}$. Then the spectrum of $\mathscr{A}$ is given by*

$$\left\{ \sum_{j=1}^{l} n_j\lambda_j : n_j \in \mathbb{N}_0 \right\}.$$

Moreover, the linear span of the eigenfunctions of $\mathscr{A}$ is dense in $L^p(\nu)$.

Proposition 5.4. *Suppose that u in the domain of $\mathscr{A}$ and satisfies $(\gamma - \mathscr{A})u = 0$ for some $\gamma \in \mathbb{C}$. Then u is a polynomial of degree less than or equal to $|Re(\gamma)/s(\mathbf{A})|$, where $s(\mathbf{A}) = \sup_j\{Re(\lambda_j)\}$.*

We focus on estimating the zero-variance importance sampling estimator for quantities of the form

$$\rho = \mathbb{E}[f(X_T)|X_0 = x] \tag{87}$$

for $f \in \mathscr{C}^2(\mathbb{R}^d)$. We note, however, that the methodology we present in the following sections may be used to estimate rare event probabilities. Indeed, if one were to choose f to be an indicator function over some region $E \subset \mathbb{R}^d$, then we have that

$$\rho = \mathbb{E}[\mathbb{1}_E(X_T)|X_0 = x] = \mathbb{P}(X_T \in E|X_0 = x). \tag{88}$$

Even though the indicator function is not in $\mathscr{C}^2$, we may still obtain heuristics for constructing good biasing functions for rare event simulation.

Since the KBE is a linear PDE, one can construct solutions in terms of the eigenfunctions of the infinitesimal generator. Assume that $\phi_i(x)$ and μ_i be such that $\mathscr{A}\phi_i(x) = \mu_i\phi_i(x)$ exists. Consider the Cauchy problem defined in Equation 5. Assuming that f is spanned by the eigenfunctions $\phi_i(x)$, one can check that the solution to Equation 5 is

$$\Phi(t,x) = \sum_{i=1}^{\infty} f_i e^{\mu_i(T-t)}\phi_i(x) \tag{89}$$

where

$$f(x) = \sum_{i=1}^{\infty} f_i\phi_i(x). \tag{90}$$

We describe an efficient formulation to procure all the eigenfunctions of the Ornstein-Uhlenbeck operator. While the work of [34] finds all the eigenvalues, proves the existence of, and characterizes the properties of the eigenfunctions, it does not provide the eigenfunctions. We show that the eigenfunctions are contained in the span of a suitably chosen basis. We now construct the basis and the state the matrix eigenvalue problem that will result in the eigenfunctions of the OU operator. The building blocks for the basis will come from finding eigenfunctions defined on each of the eigenspaces of $\mathbf{A}$. We first handle finding eigenfunctions of certain eigenvalue problems and then use them to define the general eigenfunctions.

5.4.1 Case I: Eigenfunctions on Eigenspaces with Real Eigenvalues

Here, we only consider eigenfunctions corresponding to real eigenvalues. We make an *ansatz* that the eigenfunctions should be of the form $\phi(x) = g(\langle x,z\rangle)$, where g is a univariate function and z is a fixed vector in $\mathbb{R}^d$. Then the eigenvalue equation is

$$\langle x, \mathbf{A}^* z\rangle\, g'(\langle x,z\rangle) + \frac{1}{2}\|\mathbf{B}^* z\|^2 g''(\langle x,z\rangle) = \mu g(\langle x,z\rangle). \tag{91}$$

Notice that if one chose z to be the left eigenvector corresponding to a real eigenvalue, then we have the following *one-dimensional* eigenvalue problem. That is, if $z = w_i$ with real eigenvalue $-\lambda_i$, then we have

$$-\lambda_i\langle x,w_i\rangle g'(\langle x,w_i\rangle) + \frac{1}{2}\|\mathbf{B}^* w_i\|^2 g''(\langle x,w_i\rangle) = \mu g(\langle x,w_i\rangle) \tag{92}$$

This differential equation is the well-known Hermite differential equation, whose solution are the generalized Hermite polynomials. From here, we claim that the solution to this eigenvalue problem is

$$\phi_j(x) = \mathrm{He}_j\left(\frac{\sqrt{2\lambda_i}}{\|\mathbf{B}^* w_i\|}\langle x,w_i\rangle\right), \tag{93}$$

with eigenvalue

$$\mu_{ij} = -j\lambda_i. \tag{94}$$

While we have written down all the eigenfunctions that lie on the eigenspaces corresponding to real eigenvalues, these are not the eigenfunctions we will be using to construct the basis. Note that

$$\phi(x) = \mathrm{He}_j(\sqrt{\lambda_i}\langle x, w_i\rangle)$$

has gradient and Hessian equal to

$$\begin{aligned}\nabla\phi(x) &= \sqrt{\lambda_i}\mathrm{He}_j'(\sqrt{\lambda_i}\langle x, w_i\rangle)w_i = \sqrt{\lambda_i}j\mathrm{He}_{j-1}(\sqrt{\lambda_i}\langle x, w_i\rangle)w_i \\ \nabla^2\phi(x) &= \lambda_i j\mathrm{He}_{j-1}'(\sqrt{\lambda_i}\langle x, w_i\rangle)w_i w_i^T = \lambda_i j(j-1)\mathrm{He}_{j-2}(\sqrt{\lambda_i}\langle x, w_i\rangle)w_i w_i^T.\end{aligned}$$

5.4.2 Case II: Eigenfunctions on Eigenspaces with Complex Eigenvalues

Consider the following two-dimensional OU process where the drift term has complex eigenvalues

$$dX_s = \begin{bmatrix} -a & b \\ -b & -a \end{bmatrix} X_s \, \mathrm{d}s + \sqrt{2}\, \mathrm{d}W_s \tag{95}$$

where $a, b > 0$, $X_s \in \mathbb{R}^2$ and W_s is a standard two-dimensional Brownian motion. The eigenvalues of the drift matrix are complex; they are $\lambda_\pm = -a \pm ib$. In [35], the authors found that the eigenfuctions of the associated OU operator

$$\left\langle x, \begin{bmatrix} -a & -b \\ b & -a \end{bmatrix} \nabla J_{m,n} \right\rangle + \Delta J_{m,n} = \mu_{m,n} J_{m,n} \tag{96}$$

are the so-called Hermite-Itô-Laguerre polynomials given by

$$J_{m,n}(z, \bar{z}; \rho) = \begin{cases} (-1)^n n! z^{m-n} L_n^{m-n}(z\bar{z}, \rho), & m \geq n \\ (-1)^m m! \bar{z}^{n-m} L_m^{n-m}(z\bar{z}, \rho), & m < n \end{cases} \tag{97}$$

where $L_n^\alpha(x, \rho)$ are the generalized Laguerre polynomials, $z = x_1 + ix_2$, and $\rho = 2/a$ defined in the appendix. The eigenvalues are

$$\mu_{m,n} = -(m+n)a - i(m-n)b.$$

Note that these polynomials are not the same as the complex-valued 2D Hermite polynomials.

Using this as a starting point, we can find eigenfunctions of $\mathscr{A}$ corresponding to complex eigenvalues. By Theorem 2.6 in [35], the OU eigenvalue problem for the 2D process can be written as

$$(-a - ib)z\frac{\partial}{\partial z} J_{m,n} + (-a + ib)\bar{z}\frac{\partial}{\partial \bar{z}} J_{m,n} + 4\frac{\partial^2}{\partial z \partial \bar{z}} J_{m,n} = \mu_{m,n} J_{m,n}. \tag{98}$$

Consider the eigenvalue problem

$$\langle x, \mathbf{A}^* \nabla\phi \rangle + \Delta\phi = \mu\phi. \tag{99}$$

Let $\lambda, \bar{\lambda}$ be the complex eigenvalue pair of $\mathbf{A}$ with left eigenvectors $w, \bar{w}$. To be clear, we define the left eigenvectors to be such that

$$w^*\mathbf{A} = \lambda w^*, \;\; \bar{w}^*\mathbf{A} = \bar{\lambda}\bar{w}^*, \tag{100}$$

which implies that

$$\mathbf{A}^* w = \bar{\lambda} w \;\; \mathbf{A}^*\bar{w} = \lambda \bar{w}. \tag{101}$$

Notice that if we let $\phi(x) = J_{m,n}(z, \bar{z})$ where $z = \langle x, w \rangle$, then

$$\nabla\phi(x) = \frac{\partial J}{\partial z} w + \frac{\partial J}{\partial \bar{z}} \bar{w} \tag{102}$$

$$\Delta\phi(x) = 2\frac{\partial^2 J}{\partial z \partial \bar{z}} \tag{103}$$

and we have

$$\langle x, \mathbf{A}^*\nabla\phi(x)\rangle + \Delta\phi(x) = \left\langle x, \mathbf{A}^* \left(\frac{\partial J}{\partial z} w + \frac{\partial J}{\partial \bar{z}} \bar{w} \right) \right\rangle + 2\frac{\partial^2 J}{\partial z \partial \bar{z}} \tag{104}$$

$$= \bar{\lambda} \langle x, w \rangle \frac{\partial J}{\partial z} + \lambda \langle x, \bar{w} \rangle \frac{\partial J}{\partial \bar{z}} + 2\frac{\partial^2 J}{\partial z \partial \bar{z}}. \tag{105}$$

This implies that $J_{m,n}(\langle x, w \rangle, \overline{\langle x, w \rangle})$ is an eigenfunction of [eigenvalue equation with identity diffusion].

5.4.3 General Eigenfunctions

We show the matrix eigenvalue problem that will lead to the eigenfunction for any eigenvalue in the spectrum. Suppose the drift matrix $\mathbf{A}$ decomposes the state space into l eigenspaces. Assume that there are l' eigenspaces corresponding to real eigenvalues and $l - l'$ eigenspaces corresponding to complex eigenvalues. We define the elements of the basis to be

$$\phi_{\mathbf{n}} = \prod_{i=1}^{l'} \mathrm{He}_{n_i}\left(\sqrt{\lambda_i}\langle x, w_i \rangle\right) \cdot \prod_{i=l'+1}^{l} J_{n_i^1, n_i^2}\left(\langle x, w_i \rangle, \overline{\langle x, w_i \rangle}\right) \tag{106}$$

where $\mathbf{n} \in \mathbb{N}_0^d$. To be clear, $n_i \in \mathbb{N}_0$ if $1 \leq i \leq l'$ and $n_i \in \mathbb{N}_0^2$ if $l' + 1 \leq i \leq l$. Note that $2l - l' = d$.

We can show that this basis is closed under the action of the OU operator.

Theorem 1. *Let ϕ_n be the eigenfunction of tensorized eigenfunctions. Let $\{\phi_n\}$ be the basis. Then the span of this basis is closed under the action of $\mathscr{A}$. That is, $\mathscr{A}V \subset V$.*

Proof. Let ϕ_i be eigenfunctions of the following operator with eigenvalue μ_i

$$\mathscr{A}_s\phi = \langle x, \mathbf{A}^*\nabla\phi \rangle + \Delta\phi.$$

Let $\phi_{\mathbf{n}} = \prod_{i=1}^{l} \phi_i(x)$, where l denotes the l eigenspaces of $\mathbb{R}^d$ defined by $\mathbf{A}$. We have that $\phi_i(x)$ are the eigenfunctions on the eigenspaces, and the basis functions are defined from them. We show

that the span of this basis is invariant under the action of $\mathscr{A}$. It suffices to show that $\mathscr{A}\phi \in V$ for some $\phi \in V$. We compute:

$$\langle x, \mathbf{A}^*\nabla\phi\rangle = \sum_{i=1}^{l} \langle x, \mathbf{A}^*\nabla\phi_i(x)\rangle \prod_{j=1, i\neq j}^{l} \phi_j(x)$$

$$\mathrm{Tr}\left[\mathbf{Q}\nabla^2\phi\right] = \sum_{i=1}^{l} \mathrm{Tr}\left[\mathbf{Q}\nabla^2\phi_i(x)\right] \prod_{j=1, i\neq j}^{l} \phi_j(x) + \sum_{i,j} \mathrm{Tr}\left[\mathbf{Q}\nabla\phi_i(x)\nabla\phi_j(x)^T\right] \prod_{k\neq i\neq j} \phi_k(x).$$

To leverage the fact that ϕ_i is an eigenfunction of $\mathscr{A}_s$, we add and subtract

$$\sum_{i=1}^{l} \mathrm{Tr}\nabla^2\phi_i(x) \prod_{j=1, i\neq j}^{l} \phi_j(x) = \sum_{i=1}^{l} \Delta\phi_i(x) \prod_{j=1, i\neq j}^{l} \phi_j(x)$$

to $\langle x, \mathbf{A}^*\nabla\phi\rangle$ to obtain

$$\langle x, \mathbf{A}^*\nabla\phi\rangle = \left(\sum_{i=1}^{l} \mu_i\right)\phi - \sum_{i=1}^{l} \Delta\phi_i \prod_{j=1, i\neq j}^{l} \phi_j(x).$$

Therefore, we have

$$\mathscr{A}\phi_{\mathbf{n}} = \left(\sum_{i=1}^{l} \mu_i\right)\phi_{\mathbf{n}} + \sum_{i=1}^{l} \mathrm{Tr}\left[(\mathbf{Q}-\mathbf{I})\nabla^2\phi_i(x)\right] \prod_{j=1, i\neq j}^{l} \phi_j(x) + \sum_{i,j,i\neq j} \mathrm{Tr}\left[\mathbf{Q}\nabla\phi_i(x)\nabla\phi_j(x)^T\right] \prod_{k\neq i\neq j} \phi_k(x).$$

By the properties of the trace operator, we may write the last term as

$$\sum_{i,j,i\neq j} \mathrm{Tr}\left[\mathbf{Q}\nabla\phi_i(x)\nabla\phi_j(x)^T\right] \prod_{k\neq i\neq j} \phi_k(x) = 2\sum_{i<j} \mathrm{Tr}\left[\mathbf{Q}\nabla\phi_i(x)\nabla\phi_j(x)^T\right] \prod_{k\neq i\neq j} \phi_k(x)$$

We just need to show that the other two terms are written as linear combinations of basis functions as well. Suppose we have that

$$\phi_{\mathbf{n}} = \prod_{i=1}^{l'} \mathrm{He}_{n_i}\left(\sqrt{\lambda_i}\langle x, w_i\rangle\right) \cdot \prod_{i=l'+1}^{l} J_{n_i^1, n_i^2}\left(\langle x, w_i\rangle, \overline{\langle x, w_i\rangle}\right) \tag{107}$$

where we assume we have l' real eigenvalues and $l - l'$ complex eigenvalue pairs. From this, we may compute [the two terms above] directly. For the Hessian term, we have two cases: ϕ_i corresponds to a real eigenvalue or a complex eigenvalue. If $\phi_i(x) = \mathrm{He}_{n_i}\left(\sqrt{\lambda_i}\langle x, w_i\rangle\right)$, then

$$\begin{aligned}\mathrm{Tr}\left[(\mathbf{Q}-\mathbf{I})\nabla^2\phi_i(x)\right] &= \mathrm{He}''_{n_i}\left(\sqrt{\lambda_i}\langle x, w_i\rangle\right)\mathrm{Tr}\left[(\mathbf{Q}-\mathbf{I})w_i w_i^T\right] \\ &= \lambda_i n_i\left(n_i - 1\right)\langle(\mathbf{Q}-\mathbf{I})w_i, w_i\rangle \mathrm{He}_{n_i-2}\left(\sqrt{\lambda_i}\langle x, w_i\rangle\right).\end{aligned}$$

If $\phi_i(x) = J_{n_i^1,n_i^2}\left(\langle x,w_i\rangle, \overline{\langle x,w_i\rangle}\right)$, then

$$\begin{aligned}
&\mathrm{Tr}\left[(\mathbf{Q}-\mathbf{I})\nabla^2\phi_i(x)\right] \\
&= \frac{\partial^2 J}{\partial z^2}\mathrm{Tr}\left[(\mathbf{Q}-\mathbf{I})w_i w_i^T\right] + \frac{\partial^2 J}{\partial \bar{z}^2}\mathrm{Tr}\left[(\mathbf{Q}-\mathbf{I})\bar{w}_i\bar{w}_i^T\right] + \frac{\partial^2 J}{\partial z\partial\bar{z}}\mathrm{Tr}\left[(\mathbf{Q}-\mathbf{I})(ww^* + \bar{w}\bar{w}^*)\right] \\
&= \langle(\mathbf{Q}-\mathbf{I})\bar{w}_i, w_i\rangle\frac{\partial^2 J}{\partial z^2} + \langle(\mathbf{Q}-\mathbf{I})w_i,\bar{w}_i\rangle\frac{\partial^2 J}{\partial \bar{z}^2} + 2\langle(\mathbf{Q}-\mathbf{I})w_i,w_i\rangle\frac{\partial^2 J}{\partial z\partial\bar{z}} \\
&= \langle(\mathbf{Q}-\mathbf{I})\bar{w}_i, w_i\rangle n_i^1(n_i^1-1)J_{n_i^1-2,n_i^2} + \langle(\mathbf{Q}-\mathbf{I})w_i,\bar{w}_i\rangle n_i^2(n_i^2-1)J_{n_i^1,n_i^2-2} \\
&\quad + 2\langle(\mathbf{Q}-\mathbf{I})w_i,w_i\rangle n_i^1 n_i^2 J_{n_i^1-1,n_i^2-1}
\end{aligned}$$

As for the third term, we have three cases: when ϕ_i and ϕ_j both correspond to real eigenvalues, both have complex eigenvalues or one corresponds to a real eigenvalue, the other a complex eigenvalue.

Case 1: both have real eigenvalues:

$$\mathrm{Tr}\left[\mathbf{Q}\nabla\phi_i(x)\nabla\phi_j(x)^T\right] = \sqrt{\lambda_i\lambda_j}\mathrm{He}'_{n_i}\left(\sqrt{\lambda_i}\langle x,w_i\rangle\right)\mathrm{He}'_{n_j}\left(\sqrt{\lambda_j}\langle x,w_j\rangle\right)\langle\mathbf{Q}w_i,w_j\rangle \tag{108}$$

$$= \sqrt{\lambda_i\lambda_j}n_i n_j\langle\mathbf{Q}w_i,w_j\rangle\mathrm{He}_{n_i-1}\left(\sqrt{\lambda_i}\langle x,w_i\rangle\right)\mathrm{He}_{n_j-1}\left(\sqrt{\lambda_j}\langle x,w_j\rangle\right) \tag{109}$$

Case 2: both have complex eigenvalues:

$$\mathrm{Tr}\left[\mathbf{Q}\nabla\phi_i(x)\nabla\phi_j(x)^T\right] = \mathrm{Tr}\left[\mathbf{Q}w_i w_j^T\right]\frac{\partial J_i}{\partial z}\frac{\partial J_j}{\partial z} + \mathrm{Tr}\left[\mathbf{Q}\bar{w}_i\bar{w}_j^T\right]\frac{\partial J_i}{\partial\bar{z}}\frac{\partial J_j}{\partial\bar{z}} \tag{110}$$

$$+ \mathrm{Tr}\left[\mathbf{Q}\bar{w}_i w_j^T\right]\frac{\partial J_i}{\partial\bar{z}}\frac{\partial J_j}{\partial z} + \mathrm{Tr}\left[\mathbf{Q}w_i\bar{w}_j^T\right]\frac{\partial J_i}{\partial z}\frac{\partial J_j}{\partial\bar{z}} \tag{111}$$

$$= n_i^1 n_j^1\langle\mathbf{Q}\bar{w}_i,w_j\rangle J_{n_i^1-1,n_i^2}J_{n_j^1-1,n_j^2} + n_i^2 n_j^2\langle\mathbf{Q}w_i,\bar{w}_j\rangle J_{n_i^1,n_i^2-1}J_{n_j^1,n_j^2-1} \tag{112}$$

$$+ n_i^2 n_j^1\langle\mathbf{Q}w_i,w_j\rangle J_{n_i^1,n_i^2-1}J_{n_j^1-1,n_j^2} + n_i^1 n_j^2\langle\mathbf{Q}\bar{w}_i,\bar{w}_j\rangle J_{n_i^1-1,n_i^2}J_{n_j^1,n_j^2-1} \tag{113}$$

Case 3: one real, one complex Since we only consider terms where $i<j$, ϕ_i corresponds to the real eigenvalues and ϕ_j corresponds to the complex eigenvalue

$$\begin{aligned}
&\mathrm{Tr}\left[\mathbf{Q}\nabla\phi_i(x)\nabla\phi_j(x)\right] \\
&= \mathrm{Tr}\left[\mathbf{Q}w_i w_j^T\right]\sqrt{\lambda_i}\mathrm{He}'_{n_i}\left(\sqrt{\lambda_i}\langle x,w_i\rangle\right)\frac{\partial J_j}{\partial z} + \mathrm{Tr}\left[\mathbf{Q}w_i\bar{w}_j^T\right]\sqrt{\lambda_i}\mathrm{He}'_{n_i}\left(\sqrt{\lambda_i}\langle x,w_i\rangle\right)\frac{\partial J_j}{\partial\bar{z}} \\
&= \sqrt{\lambda_i}n_i n_j^1\langle\mathbf{Q}w_i,w_j\rangle + \sqrt{\lambda_i}n_i n_j^2\langle\mathbf{Q}w_i,\bar{w}_j\rangle.
\end{aligned}$$

Putting everything together, this shows $\mathscr{A}$ acted on a basis function $\phi_{\mathbf{n}}$ is a linear combination of other basis functions $\phi_{\mathbf{n}'}$. Note that if $\mathbf{Q}=\mathbf{I}$, then the second term is equal to zero. In addition, if $\mathbf{A}$ is normal, that is, $\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A}$, then the third term is equal to zero as well. This implies that ϕ is exactly the eigenfunction of $\mathbf{A}$. Note that this basis are exactly the eigenfunctions if and only if $\mathbf{Q}=\mathbf{I}$ and the matrix $\mathbf{A}$ is normal (i.e. $\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A}$). □

5.5 Numerical Examples

5.5.1 Linear 2D Systems

The eigenfunctions are difficult to visualize for more than two dimensions. In these numerical examples, we show the eigenfunctions of two 2D systems – one normal and non-normal and study their eigenfunctions.

$$\mathbf{A}_{\text{norm}} = \begin{bmatrix} -1 & 0 \\ 0 & -0.3 \end{bmatrix} \tag{114}$$

$$\mathbf{A}_{\text{nonnorm}} = \begin{bmatrix} -1 & 0 \\ 2 & -0.3 \end{bmatrix} \tag{115}$$

Here we show eigenfunctions of normal and non-normal dynamics with a full rank diffusion matrix.

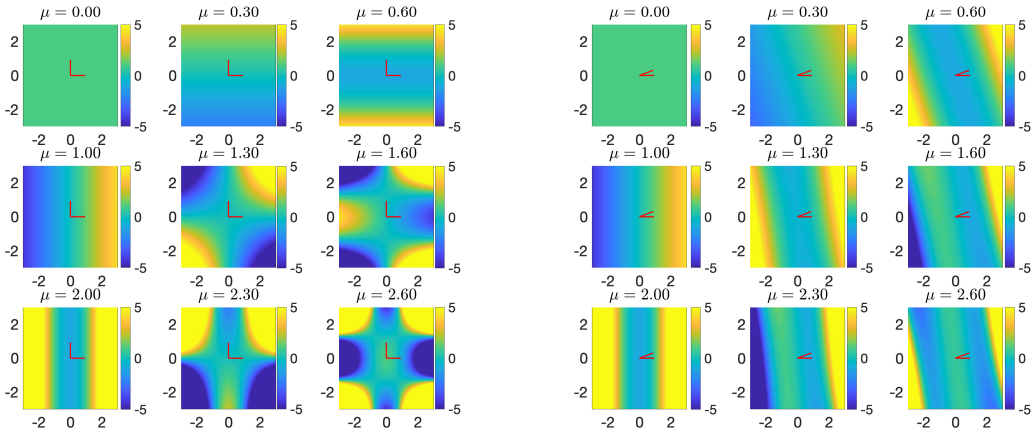

Figure 12: Eigenfunctions of normal and non-normal dynamics.

We considered the problem of estimating

$$\mathbb{P}\left(\sup_{t\in[0,10]} \|\mathbf{X}(t)\| \geq 1.2\right) \tag{116}$$

for a dynamical system with matrix $\mathbf{A}_{\text{nonnorm}}$ and diffusion matrix $\mathbf{B} = 0.1 \cdot \mathbf{I}$. To find the correct biasing, we mollified the indicator function and projected it on the first 8 even eigenfunctions using regression. The mollification is

$$f(\mathbf{x}) = \mathbb{1}\left(\|\mathbf{x}\| \geq 1.2\right) \approx \frac{1}{2} + \frac{1}{2}\tanh\left[3(\|\mathbf{x}\| - 1.2)\right].$$

We show the biasing over three time snapshots. Intuitively, this shows that one should bias in the direction of the left eigenvector unless the particles are approaching the end of the simulation time.

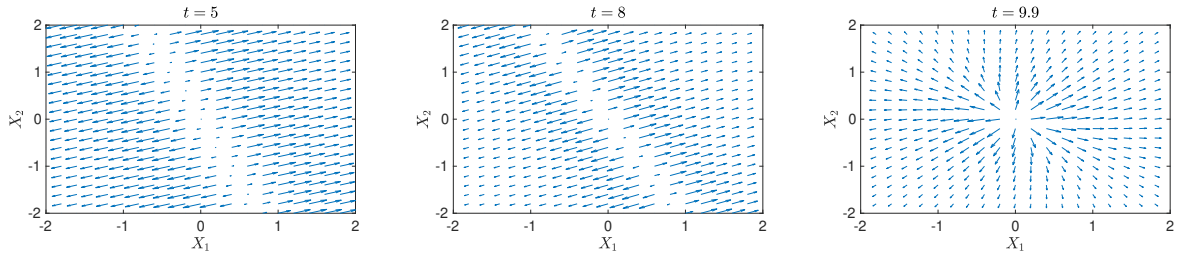

Figure 13: Three time snapshots showing biasing in the direction of the left eigenvector.

We also show the histogram of the maximum norm of the samples:

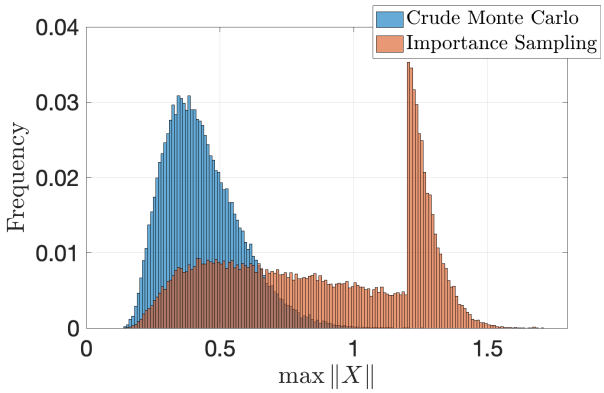

Figure 14: Importance sapmling results in more samples landing in the relevant part of the state space.

Here we summarize the simulation results. Notice that the variance is reduced by more than 2000 times.

Table 5: Results for simple 2D nonnormal dynamical system.

	Variance	Relative error
Monte Carlo	3.99×10^{-7}	15.80
Importance sampling	1.94×10^{-10}	0.31

$\rho_{\text{true}} = 4.55 \times 10^{-5}$, Samples per estimate $M = 100$

5.5.2 Stochastic Advection-Diffusion Equation

$$\frac{\partial u}{\partial t} = c\frac{\partial u}{\partial x} + \alpha\frac{\partial^2 u}{\partial x^2} + \sqrt{\varepsilon}\frac{\partial^2 W}{\partial t \partial x} \tag{117}$$

The degree of non-normality is dependent on the magnitude of the advection c compared to the diffusion term α. The ratio between the two is called the Péclet number. From the numerical examples of the simple 2D non-normal example, we saw that the biasing produced by the eigenfunctions generally pushed in the direction of the least decaying left eigenvector. Using this intuition, the biasing we apply on simulations of the advection-diffusion equation use eigenfunctions of the adjoint of the advection-diffusion operator.

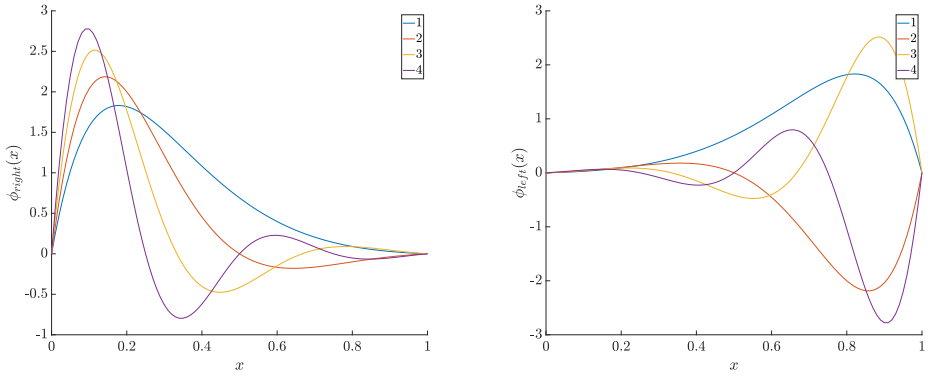

Figure 15: Eigenfunctions of the advection-diffusion operator and its adjoint.

Here we show the eigenfunctions of the advection-diffusion operator and its adjoint. Here summarize the simulation results in the following table.

Table 6: Results for stochastic advection-diffusion equation.

	Variance	Relative error
Monte Carlo	3.13×10^{-8}	1.22
Importance sampling	5.69×10^{-9}	0.52

$\rho_{\text{true}} = 1.45 \times 10^{-4}$, Samples per estimate $M = 5000$

5.5.3 Nonlinear Example: Stochastic Allen-Cahn Equation

Semilinear PDE with two metastable states. This is an example of a noisy reaction-diffusion equation. Example from [8].

$$\frac{\partial u^\varepsilon}{\partial t} = \gamma \frac{\partial^2 u^\varepsilon}{\partial x^2} - \frac{1}{\gamma} V'(u^\varepsilon) + \sqrt{\varepsilon} \frac{\partial^2 W}{\partial t \partial x} \tag{118}$$

where $V(w)$ is a double-well potential

$$V(w) = (1 - w^2)^2. \tag{119}$$

Observable of interest: magnetization of a magnetic element in a memory device

$$m(u^\varepsilon) = \int_0^1 u^\varepsilon(t,x)\, \mathrm{d}x. \tag{120}$$

Rare event of interest: rare transitions to solutions where $m(u) \geq 0$ starting from solutions where $m(u) \approx -1$.

This is a nonlinear stochastic PDE, however, as it has two metastable states, we can linearize the system about one of the two states. Define $\bar{u} := \bar{u}(x)$ to a steady state solution to the stochastic Allen-Cahn with $\varepsilon = 0$

$$\gamma \frac{\partial^2 \bar{u}}{\partial x^2} - \frac{1}{\gamma} V'(\bar{u}) = 0. \tag{121}$$

We derive the stochastic PDE governing the perturbations about this steady state solution. Let $\delta > 0$ be a small parameter, and $u^\varepsilon = \bar{u} + \delta v$. Observe that

$$\frac{\partial}{\partial t}(\bar{u} + \delta v) = \gamma \frac{\partial^2}{\partial x^2}(\bar{u} + \delta v) - \frac{1}{\gamma} V'(\bar{u} + \delta v) + \sqrt{\varepsilon} \frac{\partial^2 W}{\partial t \partial x}. \tag{122}$$

Expanding V' and taking δ to zero, we arrive at

$$\frac{\partial v}{\partial t} = \gamma \frac{\partial^2 v}{\partial x^2} - \frac{V''(\bar{u})}{\gamma} v + \sqrt{\varepsilon} \frac{\partial^2 W}{\partial t \partial x}. \tag{123}$$

5.5.4 Aircraft Electrical System: Brownian Oscillator

The aircraft electrical system as described in Section 4 can be considered as a damped harmonic oscillator forced by Brownian motion, known as a Brownian oscillator.

$$\begin{cases} \ddot{x} + 2\zeta\omega_0 \dot{x} + \omega_0^2 x = \dot{W}_t \\ x(0) = x_0, \dot{x}(0) = 0. \end{cases} \tag{124}$$

Here, x_1 can be viewed as the voltage and x_2 as the current. In our case, the rare event occurs when the voltage exceeds a certain threshold.

This second-order system can be decoupled into a first-order linear system

$$\begin{cases} \mathrm{d}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & -2\zeta\omega_0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mathrm{d}t + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mathrm{d}W_t \\ X_0 = x_0 \end{cases} \tag{125}$$

We show the eigenfunctions here.

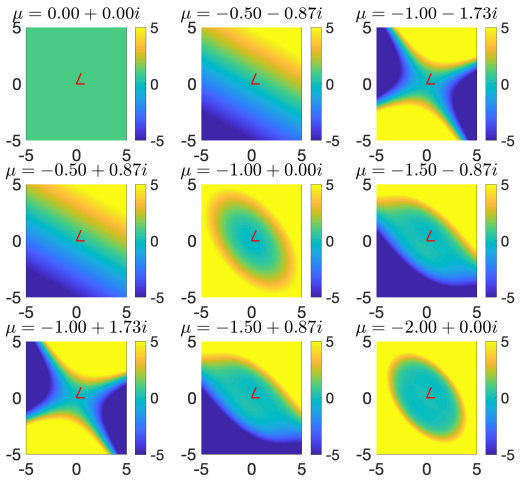

Figure 16: Brownian oscillator Koopman eigenfunctions.

We want to compute the following probability:

$$\mathbb{P}\left(\|x_{10}\| \geq 2.75 | x_0 = \dot{x}_0 = 0\right). \tag{126}$$

The samples paths and distribution of samples at time $t = 10$ for simple Monte Carlo and importance sampling are shown in the next figure:

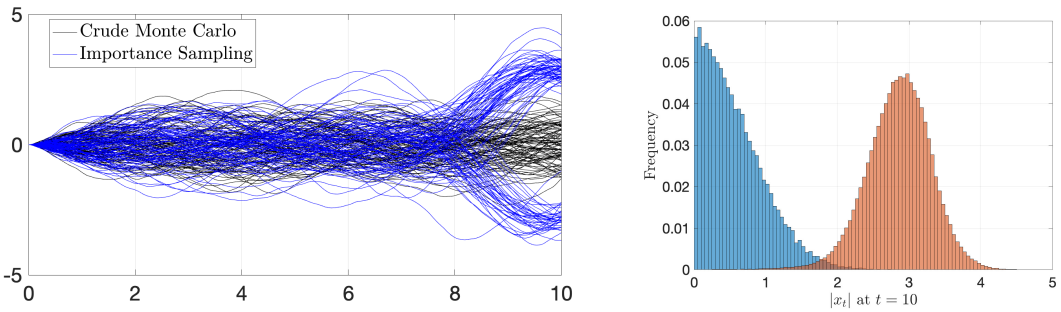

Figure 17: Sample paths and distribution of samples.

The simulation results are summarized by the following table. Notice that there is a 25 times variance reduction.

Table 7: Results for noisy Brownian oscillator.

	Variance	Relative error
Monte Carlo	1.04×10^{-7}	3.16
Importance sampling	3.87×10^{-9}	0.58

$\rho_{\text{true}} = 1.08 \times 10^{-4}$, $M = 1000$

6 Large Deviations-based Monte Carlo Methods

In this section, we show our efforts in applying large deviations-based rare event simulation methods for engineering systems. We first review the basic notions and concepts in large deviations-based simulation methods and its formulation for importance sampling and particle splitting methods. We then apply the methods on to the engineering models discussed in Section 4. Along the way, we also make connections between the large deviations-based approach and the Koopman operator approach, namely that the former is an approximation of the latter. We conclude the section with a new research direction relating to efficient implementation of these methods.

6.1 Foundational Notions

Large deviations-based approaches to rare event simulation are typically presented in the context of small noise diffusions. We consider a family of stochastic differential equations parametrized by a noise parameter ε

$$\begin{cases} \mathrm{d}X_t &= \mathbf{A}(X_t)\,\mathrm{d}t + \sqrt{\varepsilon}\mathbf{B}(X_t)\,\mathrm{d}W_t \\ X_0 &= x. \end{cases} \tag{127}$$

Define ρ_ε be a rare event probability such that $\rho_\varepsilon \to 0$ as $\varepsilon \to 0$. We first study how large deviations theory studies rare events in dynamical system.

From a pragmatic perspective, large deviations theory is a sub-branch of probability theory that allows one to compare the rarity of events without directly appealing to using their exact probabilities. This alternative description to rare events provides information on how one should construct efficient importance sampling estimators. We define some general notions from large deviations theory before restricting our attention to a special case of the theory that applies to stochastic dynamical systems.

Large deviations theory is structured around the existence of large deviations principles and rate functions. Let $\mathscr{X}$ be a complete, separable metric space.

Definition 6.1. *A rate function is a mapping I from $\mathscr{X}$ to $[0,\infty]$ such that for any $a < \infty$, the pre-image of the set $[0,a]$ under I is a compact subset of $\mathscr{X}$.*

Definition 6.2. *We say a sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ taking values in $\mathscr{X}$ has a large deviations principle with rate function I if for any open set $O \subset \mathscr{X}$,*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(X_n \in O\right) \geq -\inf_{x\in O} I(x) \tag{128}$$

and for any closed set $C \subset \mathscr{X}$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(X_n \in C\right) \leq -\inf_{x\in C} I(x). \tag{129}$$

In this abstract perspective, one can learn about the rarity a set by studying the rate function instead of the probability itself. Large deviations theory for dynamical systems is its significant research branch in itself, and is known as the Freidlin-Wentzell theory [36].

Suppose we have the SDE system

$$\begin{cases} dX_t^\varepsilon &= \mathbf{A}(X_t^\varepsilon)\,dt + \sqrt{\varepsilon}\mathbf{B}(X_t^\varepsilon)\,dW_t \\ X_0^\varepsilon &= x. \end{cases} \tag{130}$$

The large deviations principle for this system is given by the following

Theorem 2. *Let $\{X_t\}$ be the stochastic process defined above. Then the dynamical system satisfies a large deviations principle on $\mathscr{C}([0,T])$, the space of continuous functions, with rate function*

$$I(\phi(\cdot)) = \frac{1}{2}\int_0^T \left(\dot{\phi} - \boldsymbol{A}(\phi)\right)^* \left(\boldsymbol{B}(x)\boldsymbol{B}(x)^*\right)^{-1} \left(\dot{\phi} - \boldsymbol{A}(\phi)\right) ds. \tag{131}$$

Estimators for rare event probabilities are typically evaluated according to their efficiency for an entire family of SDEs rather than one at a time. Rather than desiring the estimator to have zero variance, we discuss two more notions of efficiency: strong and weak efficiency [37].

Definition 6.3. *An unbiased estimator Z^ε for ρ^ε is called strongly efficient if*

$$\lim_{\varepsilon\to 0} \frac{\mathbb{E}[(Z^\varepsilon)^2]}{(\rho^\varepsilon)^2} = 1. \tag{132}$$

Strong efficiency implies the estimator has bounded relative error for finite ε that is going to zero as ε tends towards zero. One can observe this by some simple algebra:

$$\begin{aligned} &\lim_{\varepsilon\to 0} \frac{\mathbb{E}[(Z^\varepsilon)^2]}{(\rho^\varepsilon)^2} = 1 \\ \implies &\lim_{\varepsilon\to 0} \frac{\mathbb{E}[(Z^\varepsilon)^2]}{(\rho^\varepsilon)^2} - 1 = 0 \\ \implies &\lim_{\varepsilon\to 0} \frac{\mathbb{E}[(Z^\varepsilon)^2] - (\rho^\varepsilon)^2}{(\rho^\varepsilon)^2} = 0 \\ \implies &\lim_{\varepsilon\to 0} \frac{\mathrm{Var}[(Z^\varepsilon)]}{(\rho^\varepsilon)^2} = 0. \end{aligned}$$

This implies that the square of the relative error tends towards zero as $\varepsilon \to 0$, which implies that for finite ε, the relative error is bounded.

Definition 6.4. *An unbiased estimator Z^ε for ρ^ε is called weakly efficient (also known as asymptotic or logarithmic efficiency) if*

$$\lim_{\varepsilon\to 0} \frac{\log \mathbb{E}[(Z^\varepsilon)^2]}{2\log \rho^\varepsilon} = 1. \tag{133}$$

The notion of weak efficiency takes root from the theory of large deviations, where the large deviations decay rate of the estimator matches with twice the large deviations decay of the probability of interest. That is,

$$\lim_{\varepsilon\to 0} \varepsilon \log \mathbb{E}[(Z^\varepsilon)^2] = -2\gamma \tag{134}$$

$$\lim_{\varepsilon\to 0} \varepsilon \log (\rho^\varepsilon)^2 = -2\gamma. \tag{135}$$

An estimator that is weakly efficient will have the number of samples grow subexponentially to keep the relative error constant as $\varepsilon \to 0$.

6.2 Large Deviations-Based Importance Sampling

There are many ways of arriving at large deviations-based importance sampling, however, the most direct way is to take the KBE and perform a variable transformation [8, 38]. Consider the *family* of processes parametrized by ε

$$dX_t = \mathbf{A}(X_t)\,dt + \sqrt{\varepsilon}\mathbf{B}(X_t)\,dW_t.$$

The magnitude of the noise is controlled by ε and will be important for large deviations asymptotics. The KBE for a given member of this family is

$$\begin{cases} \frac{\partial \Phi}{\partial t} + \langle \mathbf{A}(x), \nabla \Phi \rangle + \frac{\varepsilon}{2} \mathrm{Tr}\left[\mathbf{B}(x)\mathbf{B}(x)^* \nabla^2 \Phi\right] = 0 \\ \Phi(T,x) = e^{-f(x)/\varepsilon}. \end{cases} \tag{136}$$

Define $U^\varepsilon(t,x) = -\varepsilon \log \Phi(t,x)$, then we obtain a PDE given as

$$-\frac{\partial U}{\partial t} + H(x, -\mathbf{D}_x U) - \frac{\varepsilon}{2} \mathrm{Tr}\mathbf{B}\mathbf{B}^*[\mathbf{D}_x^2 U] = 0. \tag{137}$$

Since large deviations describes the how estimators and rare events behave as ε tends towards zero, the last term is typically dropped. After letting ε tend towards zero, the resulting equation is the Hamilton-Jacobi equation

$$\frac{\partial U}{\partial t} - H\left(x, -\mathbf{D}_x U\right) = 0, \tag{138}$$

where H is the Hamiltonian. This is equal to the local cumulant generating function of the SDE

$$H(x,\alpha) = \langle \mathbf{A}[x], \alpha \rangle_V + \frac{1}{2} \|\mathbf{B}^*[\alpha]\|_V^2. \tag{139}$$

The solution of the PDE are called *importance functions*, and they are important for the construction of importance sampling and splitting estimators. To obtain weakly efficient importance sampling estimators, one should choose

$$u(t,x) = -\mathbf{B}^* \nabla U(t,x), \tag{140}$$

which is consistent with the Doob transform as presented in section 3.

Furthermore, the authors in [8] prove conditions for rare event simulation for SDEs that will result in strongly efficient estimators. While solving a Hamilton-Jacobi equation is more easier than solving a KBE, it is still intractable to consider for more than a few dimensions. Instead, [8] appeals to the variational formulation of the Hamilton-Jacobi PDE.

It is well known that this formulation has intimate connections with the large deviations rate function. The local rate function (integrand of the large deviations rate function) is the Legendre transformation of the cumulant generating function. This allows one to obtain the Hamilton-Jacobi-Bellman (HJB) equation [8, 39]

$$\frac{\partial U}{\partial t} + \inf_{\beta} \left[\langle \mathbf{D}_x U, \beta \rangle - L(x,\beta)\right] = 0 \tag{141}$$

which admits a variational form

$$U(t,x) = \inf_{\phi \in V} \left[\int_0^T L(\phi, \dot{\phi}) ds : \phi(t) = x, \phi \in E \right]. \tag{142}$$

One can then construct a biasing procedure that solves many optimization problems on the fly as one is performing Monte Carlo simulation. This would require one to solve many optimal control problems to get to the biasing of interest. While this is a more reasonable approach to the problem, it is still very computationally expensive.

To reduce computation, one can appeal to the work on subsolutions of the Hamilton-Jacobi equation [40]. They show that solutions of the HJB are not necessary to attain weak efficiency, as the use of *subsolutions* is sufficient. A subsolution is a function $\bar{U}$ that does not solve the Hamilton-Jacobi equation, rather it satisfies the following differential inequality:

$$\frac{\partial \bar{U}}{\partial t} - H\left(x, -\mathbf{D}_x \bar{U}\right) \geq 0.$$

However, weak efficiency makes no guarantees on the true performance of the estimator.A commonly used subsolution for importance sampling in SDEs is the Freidlin-Wentzell quasipotential, which is a standard tool in the study of rare transitions and reaction pathways in chemistry [36]. This quasipotential can be represented in the variational form

$$U_{FW}(x) = \inf_{\phi \in V} \left[\int_0^T L(\phi, \dot{\phi}) ds : \phi(0) = x, \phi \in E \right]. \tag{143}$$

One should compare this to Equation 142 to see that this importance function is time independent, and in fact solves the time-independent Hamilton-Jacobi equation

$$H(x, -\mathbf{D}_x U_{FW}) = 0. \tag{144}$$

However, empirical studies show that the nonasymptotic performance of the importance sampling estimator using this quasipotential is will degrade when the problem of interest contains an attractor. The theoretical explanation for this observation is well-documented in [38, 41, 42].

6.2.1 Large Deviations-Based Approach for Splitting Methods

As discussed above, using the Kolmogorov backward equation, it can be shown that the large deviations decay rate of rare event probability with initial condition $X_0 = x$ is the solution of a Hamilton-Jacobi equation

$$U_t(t,x) - H(x, -\nabla U(t,x)) = 0 \tag{145}$$

$$U(T,x) = f(x) \tag{146}$$

where $U(t,x) = -\varepsilon \log E_{t,x}\left[e^{-\frac{1}{\varepsilon}F(X^n(T))}\right]$. It can be shown that if we use $V(t,x)$ (or a subsolution of the HJB equation) as the importance function in a splitting scheme, a subexponential grows of the number of actual particles can be achieved with respect to the rareness of the events and the decay rate of the second moment is guaranteed to be better than standard Monte Carlo [19].

Consequently, constructing the splitting scheme according to the subsolution will lead to a subexponential growth of the total number of simulated particles [7, 19].

6.2.2 *New Subsolution* Ansatz

Using large deviations and a dose of stochastic calculus, [38, 41] explain the problems with the subsolution approach for biasing SDEs in the presence of an attractor and show how the second moment can become large when where $\varepsilon > 0$. The authors show that the second moment of an importance sampling estimator can be expressed as the solution to a PDE. The weak efficiency of the subsolution can be reproduced when $\varepsilon = 0$. However, via an asymptotic expansion in terms of ε, they showed that the first order term in the expansion can become arbitrarily large depending on the ratio between ε and the time horizon T.

In [42], the authors use the reasoning of [38] and provide a new approach that remedies the degrading performance of importance sampling schemes in linear self-adjoint dynamical systems. We review their approach here. Suppose we have the *linear* stochastic dynamical system

$$\begin{cases} dX^\varepsilon(t) = \mathbf{A}X^\varepsilon(t)dt + \sqrt{\varepsilon}\mathbf{B}dW_t \\ X^\varepsilon(0) = 0 \end{cases} \tag{147}$$

where $\mathbf{X}(t) \in V$ where V is a separable Hilbert space. For simplicity, assume $\mathbf{B}$ is the identity operator. The matrix $\mathbf{A}$ is self-adjoint, with eigenvectors e_k and eigenvalues $-\alpha_k$:

$$\mathbf{A}e_k = -\alpha_k e_k$$

where $\alpha_{k+1} > \alpha_k$. The system is assumed to have a spectral gap $\alpha_1 \leq 3\alpha_k$ for all $k \geq 2$. Since $\mathbf{A}$ is self-adjoint, the eigenvectors are orthogonal. The goal is to compute the probability of exit from some ball of radius L

$$\rho = \mathbb{P}\left[\sup_{s\in[0,T]} \|X^\varepsilon(s)\|_V > L\right] \tag{148}$$

via importance sampling.

Having these assumptions, the authors show that the following biasing function results in an importance sampling scheme whose second moment will be bounded for any noise level ε and time horizon T

$$u(x) = 2\alpha_1 \langle x, e_1\rangle_V e_1. \tag{149}$$

This biasing comes from a subsolution:

$$U(x) = \alpha_1\left(L^2 - \langle x, e_1\rangle_V^2\right). \tag{150}$$

In [41, 42], the authors show that given a valid subsolution, the second moment has a nonasymptotic bound of the form

$$-\varepsilon\log(Q^\varepsilon(0,x,u)) \geq \inf_{\{v:\hat{\tau}^{v,\varepsilon}\leq T\}}\left(2U(0,x) - 2\mathbb{E}U(\hat{\tau}^{v,\varepsilon}, \hat{X}^{v,\varepsilon}(\hat{\tau}^{v,\varepsilon})) + \mathbb{E}\left[\int_0^{\hat{\tau}^{v,\varepsilon}} \varepsilon\mathrm{Tr}(\mathbf{B}\mathbf{B}^*\mathbf{D}_x^2 U(\hat{X}^{v,\varepsilon}(s))ds\right]\right) \tag{151}$$

where $Q^\varepsilon(0,x,u)$ is the second moment of the importance sampling estimator corresponding with initial conditions $X^\varepsilon(0) = x$, and biasing function u. Here $\hat{X}^{v,\varepsilon}$, called the controlled process, solves

$$d\hat{X}^{v,\varepsilon}(t) = [\mathbf{A}\hat{X}^{v,\varepsilon}(t) + \mathbf{B}v(t) - \mathbf{B}u(t,\hat{X}^{v,\varepsilon}(t))]dt + \sqrt{\varepsilon}\mathbf{B}d\mathbf{W}_t \tag{152}$$

where $\hat{\tau}^{v,\varepsilon} = \inf\{t > 0 : \|\hat{X}^{v,\varepsilon}(t)\|_V \geq L\} \leq T$ almost surely. These objects come from variational representations of the second moment [41, 43].

For the subsolution described in Equation 150, the bound reduces to

$$-\varepsilon \log(Q^{\varepsilon}(0,x,u)) \geq \inf_{\{v:\hat{\tau}^{v,\varepsilon}\leq T\}} \left(2\,\mathbb{E}\langle \hat{X}^{v,\varepsilon}(\hat{\tau}^{v,\varepsilon}), e_1\rangle_V^2 + \mathbb{E}\left[\int_0^{\hat{\tau}^{v,\varepsilon}} \varepsilon\alpha_1 ds\right]\right) \geq \inf_{\{v:\hat{\tau}^{v,\varepsilon}\leq T\}} \left(2\,\mathbb{E}\langle \hat{X}^{v,\varepsilon}(\hat{\tau}^{v,\varepsilon}), e_1\rangle_V^2 \right). \tag{153}$$

This shows that the second moment using this biasing has the bound

$$Q^{\varepsilon}(0,x,u) \leq \exp\left[-\frac{2}{\varepsilon} \inf_{\{v:\hat{\tau}^{v,\varepsilon}\leq T\}} \left(\mathbb{E}\langle \hat{X}^{v,\varepsilon}(\hat{\tau}^{v,\varepsilon}), e_1\rangle_V^2 \right)\right]. \tag{154}$$

This bound shows that the second moment of the estimator only depends on the expected exit direction of the controlled process. Ultimately, [42] show a result of the following form:

Theorem 3. *Assuming* $\mathbf{A}$ *is self-adjoint, the eigenvalues have a spectral gap where* $\alpha_1 \leq \alpha_k$ *for* $k \geq 2$*, for any fixed* ε*, the second moment of the estimator is bounded above by a constant number* $C(\varepsilon)$ *for all time horizons* T*.*

The advantages of this approach are relatively clear. The computation of this biasing approach is fast as it only requires the computation of eigenfunctions and an inner product. Furthermore, having nonasymptotic bounds shows that it is provably efficient. This approach is dimension independent, meaning that the method is valid for stochastic SPDEs as well. The rationale behind this approach is that the mode that will most likely lead to an excitation away from the equilibrium point is the one that corresponds to the slowest decaying eigenvector. The intuitive idea is to get a low variance estimator, one needs the main pathway to the rare event and try to induce pathway to happen via biasing.

The disadvantages of the methods lies with its assumptions. First it requires that the system is linear and self-adjoint. Second, it requires the existence of a spectral gap. These are luxuries that are typically not present in a real engineering problem. For these reasons, we wish to extend this method to nonnormal systems, which we begin in the next section.

Left eigenvector biasing In this section we propose a new biasing for linear systems based on the left eigenvector. Our goal is to show that it is a subsolution and that its second moment is bounded. For simplicity of presentation, assume that $\mathbf{A}$ is fully diagonalizable and $\mathbf{B}$ is the identity matrix. We consider same setting as the problem above, except the operator $\mathbf{A}$ is no longer self-adjoint. Therefore, we have left eigenvectors f_k, where $f_k^*\mathbf{A} = -f_k^*\alpha_k$.

Observing the phase portrait of a nonnormal dynamical system can give insight on how one should propose a biasing scheme. In Figure 18, we show phase portraits of two two-dimensional linear dynamical systems, one with self-adjoint dynamics and the other with nonnormal dynamics. The systems share the same real and negative eigenvalues but the eigenvectors are different. To draw the paths, we set the initial conditions to be on the unit circle and run each path until they are close to the origin.

In the normal phase portrait, we see that all states immediately decay. In this case, if one wished to bias out of the basin of attraction, the least steep direction one can climb is in the direction of the least decaying eigenvector. This intuition matches the approach of [42]. However, in the nonnormal system, one can actually identify two directions for escaping from the attractor.

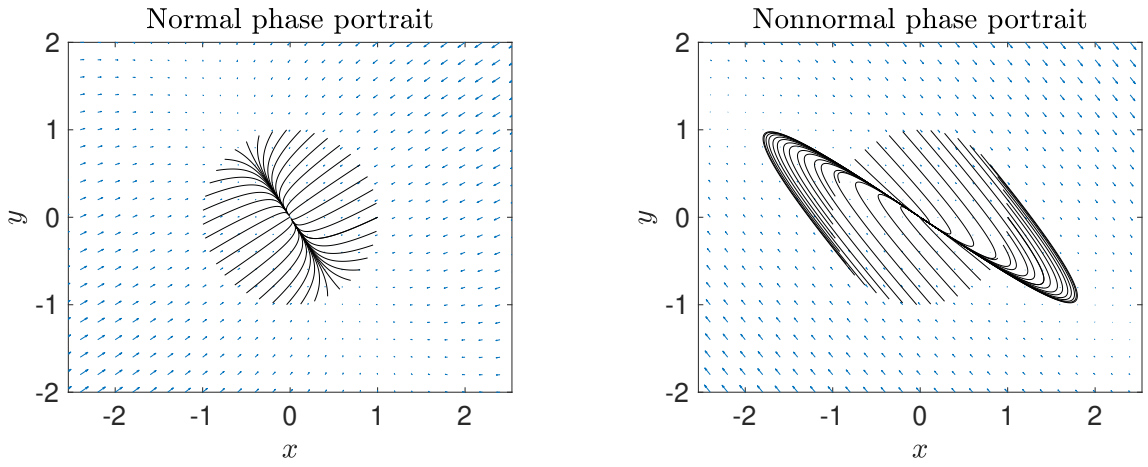

Figure 18: Phase portraits linear two dimensional dynamical systems

The first is still climbing out of the direction of the least decaying eigenvector, similar to the normal system. However, notice that states of the nonnormal system do not immediately decay – for certain initial conditions, the system actually escapes the unit ball without any biasing. In fact, there are certain initial conditions *inside the unit ball* that will escape the ball without any biasing. This is due to transient nonnormal growth, which is an extensively studied phenomenon in fluid dynamics [44, 45]. To take advantage of the nonnormal growth, one can roughly see in the phase portrait that if one biases *orthogonal* to the directions of the faster decaying eigenspace, then the transient nonnormal growth in the system will carry the state out of the unit ball. Biasing in the direction orthogonal to the faster decaying eigenmodes is equivalent to biasing in the direction of the slowest decaying left eigenvector.

From this, one may consider three biasing approaches. Following the first approach, we bias in the eigenspace that corresponds to the slowest decaying mode. This means that one has to first project the state on to the first left eigenvector and then bias in the direction of the first right eigenvector:

$$u_{\text{gad}}(x) = 2\alpha_1 \frac{\langle x, f_1 \rangle_V}{\langle e_1, f_1 \rangle_V}. \tag{155}$$

This biasing is equivalent to finding the eigendecomposition of $\mathbf{A}$, and flipping the sign of the least negative eigenvalue. This approach appears in theoretical and computational chemistry literature as the gentlest ascent dynamics (GAD) [46, 47]. This method is used to find saddle points in potential energy fields as a way of exploring multiple minima of an energy landscape. However, this method does not correspond to a subsolution, so it cannot even recover the baseline results from large deviations theory.

We can also take the approach of [42], which is to project and bias in the direction of the first eigenvector. The advantage is that this will correspond to a subsolution, and biases in the same direction as the gentlest ascent approach. Here we have the biasing function

$$u_{\text{right}}(x) = 2\alpha_1 \langle x, e_1 \rangle_V e_1. \tag{156}$$

If we take the second approach, which is biasing in the direction of the first left eigenvector,

Table 8: Summary of biasing choices.

Name	**Biasing $\mathbf{u}(\mathbf{x})$**	**Subsolution $U(\mathbf{x})$**
Gentlest Ascent	$\frac{2\alpha_1}{\lambda_1}\frac{\langle \mathbf{x}, f_1\rangle_V}{\langle e_1, f_1\rangle_V} e_1$	None
Right Eigenvector	$\frac{2\alpha_1}{\lambda_1}\langle \mathbf{x}, e_1\rangle_V e_1$	$\frac{\alpha_1}{\lambda_1^2}L^2 - \frac{\alpha_1}{\lambda_1^2}\langle \mathbf{x}, e_1\rangle_V^2$
Left Eigenvector	$\frac{2\alpha_1}{\lambda_1}\langle \mathbf{x}, f_1\rangle_V f_1$	$\frac{\alpha_1}{\lambda_1^2}L^2 - \frac{\alpha_1}{\lambda_1^2}\langle \mathbf{x}, f_1\rangle_V^2$

we have the biasing function

$$u(x) = 2\alpha_1 \langle x, f_1\rangle_V f_1. \tag{157}$$

This does correspond to a subsolution. Choose a function

$$U(x) = \alpha_1 L^2 - \alpha_1 \langle x, f_1\rangle^2, \tag{158}$$

observe that

$$\begin{aligned}\frac{\partial U}{\partial t} + \langle Ax, \mathbf{D}_x U\rangle - \frac{1}{2}|u(x)|^2 &= -2\alpha_1\langle x, f_1\rangle\langle Ax, f_1\rangle - \frac{1}{2}|2\alpha_1\langle x, f_1\rangle f_1|^2 \\ &= 2\alpha_1^2\langle x, f_1\rangle^2 - 2\alpha_1^2\langle x, f_1\rangle^2 \\ &= 0.\end{aligned}$$

Being a subsolution to the Hamilton-Jacobi equation is a first step to finding good biasing functions. The more involved next step is to bound the second moment of the corresponding estimator related to the subsolution. However, recall that using the Koopman operator, the biasing it produces for linear equations are the same as the subsolution based approach. This new *ansatz* is simply an approximation to the Koopman operator approach. This leads to more interesting questions into how one can analyze approximations of the Koopman approach by using large deviations results.

A summary of the three possible choices of biasing is presented in Table 8.

6.3 Numerical Examples

6.3.1 Linear-Quadratic Control

We briefly review a special case of the continuous time, finite horizon linear quadratic regulator problem and its solution. Most of these results are from [48, 49]. This will be important for solving the variational problem in dynamic importance sampling. Suppose we have a dynamical system where

$$\begin{cases}\dot{x}(t) = Ax(t) + Bu(t) \\ x(t_0) = x_0\end{cases}$$

for $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times p}$, $u \in \mathbb{R}^p$. Suppose we wish to minimize the following operating cost function

$$J(t_0) = \frac{1}{2}x(t_f)^T F x(t_f) + \int_{t_0}^{t_f} \frac{1}{2}u(t)^T R u(t) dt$$

where $F \in \mathbb{R}^{n\times n}, R \in \mathbb{R}^{p\times p}$ are semi-definite positive matrices. Then the control u that is optimal in operating this system is linear with the state of the system, and the gain matrix can be found by solving a Riccati equation backwards in time. We find $S \in \mathbb{R}^{n\times n}$ such that

$$\begin{aligned} \dot{S} + A^T S + SA - SBR^{-1}B^T S &= 0 \\ S(t_f) &= F. \end{aligned}$$

Then the gain matrix is $K(t) = R^{-1}B^T S(t)$, and the optimal control will be $u(t) = -K(t)x(t)$.

When we want an endpoint constraint on the control problem, we can approximate its solution through an augmented LQR problem. Suppose we want to have a terminal constraint $x(t_f) = x_T$. We do this by definiting the terminal cost function differently. Suppose we have the dynamical system as defined above, and we wish to minimize the operating cost function

$$J(t_0) = \frac{1}{2}(x(t_f) - x_T)^T F(x(t_f) - x_T) + \int_{t_0}^{t_f} \frac{1}{2}u(t)^T Ru(t)dt.$$

Define $y(t) = x(t) - x_T$. Then we have the affine dynamics

$$\begin{cases} \dot{y}(t) = Ay(t) + Bu(t) + Ax_T \\ y(t_0) = x_0 - x_T \end{cases}$$

with cost function

$$J(t_0) = \frac{1}{2}y(t)^T Fy(t) + \int_{t_0}^{t_f} \frac{1}{2}u(t)^T Ru(t)dt.$$

Then the optimal control to this affine system is equivalent to solving the LQR control problem for the following augmented system

$$\begin{aligned} \underline{\dot{y}}(t) &= \underline{A}\underline{y}(t) + \underline{B}u(t) \\ \underline{y}(t_0) &= \begin{bmatrix} x_0 - x_T \\ 1 \end{bmatrix} \end{aligned}$$

where

$$\underline{A} = \begin{bmatrix} A & Ax_T \\ 0 & 0 \end{bmatrix}, \underline{y} = \begin{bmatrix} y \\ 1 \end{bmatrix}, \text{ and } \underline{B} = \begin{bmatrix} B \\ 0 \end{bmatrix}.$$

The cost function of this LQR problem is in the form of the above equation, except with the terminal condition matrix F is augmented: $\underline{F} = \begin{bmatrix} F & 0 \\ 0 & 0 \end{bmatrix}$

$$\underline{J}(t_0) = \frac{1}{2}\underline{y}(t)^T \underline{F}\underline{y}(t) + \int_{t_0}^{t_f} \frac{1}{2}u(t)^T Ru(t)dt.$$

6.3.2 *UTRC Helicopter Model*

Exploring the rare events In this section, we simulate a few model runs to qualitatively explore conditions that may cause a helicopter in cruise to experience stall in its horizontal tail airfoil and therefore cause the helicopter to become unstable. In following simulations, the parameters of the rotorcraft are chosen according to the specifications of an MBB Bo 105 helicopter [50]. We first only consider rare events caused by random gusts, or noisy wind. The noise in the wind is modeled as an Ornstein-Uhlenbeck process, which is a mean reverting diffusion process. This is a continuous model of wind speeds where it is a constant value on average, and noise is driven by a Brownian motion. See [51] for a more thorough discussion on wind models. The wind is expressed as the following stochastic differential equation

$$d\mathbf{V} = (\mathbf{V}_0 - \mathbf{V})\,\mathrm{d}t + \sigma\,\mathrm{d}\mathbf{W}_t \tag{159}$$

where $\mathbf{V} = [V_x, V_z]^T$, the x and z velocities of the helicopter relative to the ground, $\mathbf{V}_0$ is the mean velocity, σ is the diffusion matrix, and $\mathbf{W}_t$ is a standard two dimensional Brownian motion. In the following simulations, the diffusion matrix is of the form

$$\sigma = \begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_z \end{bmatrix}.$$

Physically this models that the vertical gusts are independent of the horizontal gusts.

For all the simulations, we trim the system so that the helicopter generating enough lift to keep it in the air while cruising horizontally at a speed of approximately 44.2 meters per second (99 mi/hr), with zero vertical speed. That is, we desire the hub roll and pitch moments, C_{M-x} and C_{M-y} respectively, to be zero, while having the mean velocity to be equal to $\mathbf{V}_0 = [44.2, 0]$ m/sec. By trimming, we mean finding the blade pitch controls $\theta_0, \theta_{1C}, \theta_{1S}$ that achieve these desired flight conditions. The pitch of the aircraft is tilted forward at 0 degrees from the horizontal, so it is level with the ground. We show a few plots of the evolution the pitch over time under noisy wind conditions. We numerically integrate the system forward in time using a simple Euler-Maruyama solver [20].

In the following figures we show plots of the time evolution of the helicopter pitch under varying noisiness in the wind and sizes of the horizontal tail. Recall that in our model, the horizontal tail has an airfoil which stabilizes the rotorcraft, and that airfoil angle may exhibit stall when the vehicle pitch angle is outside the range $-15^\circ \leq \theta \leq 15^\circ$. Therefore, we are looking for environmental and design conditions such that the pitch will fall outside that value. In Figure 19, we show the time evolution of the pitch of the helicopter system while being perturbed by the noise from the wind. We plot θ and the velocity profiles as functions of time with no further trimming or other control mechanism over the course of the simulation. Here we have $\sigma_x = 0.55$ m/sec (1.23 mi/hr) and $\sigma_z = 0.55$ m/sec (1.23 mi/hr).

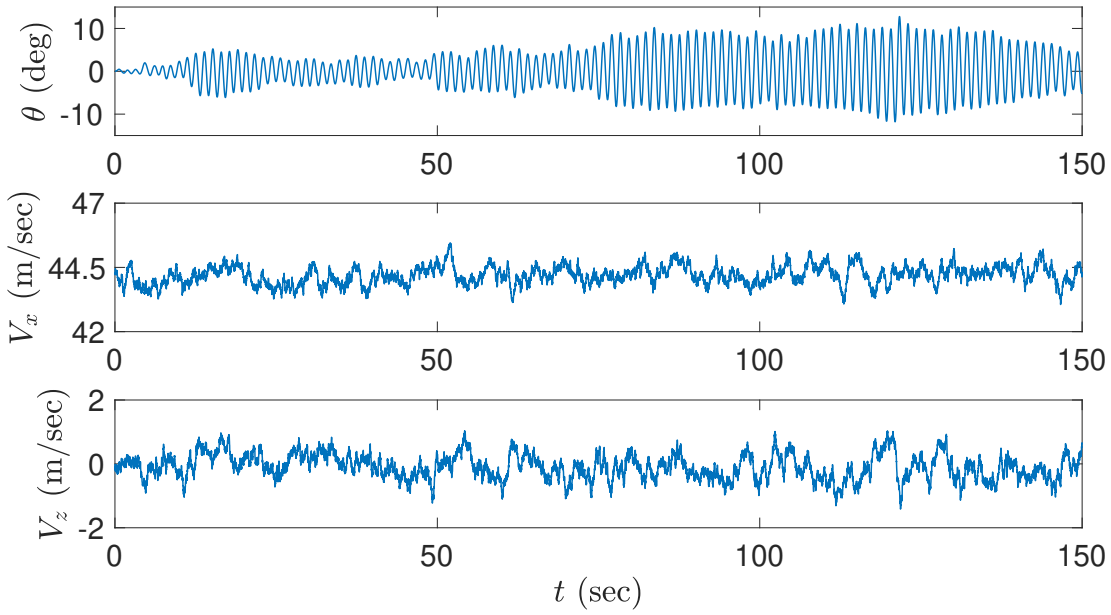

Figure 19: Evolution of the pitch angle θ over time with noise in both the horizontal and vertical velocities. We have $\sigma_x = 0.55$ m/sec, $\sigma_z = 0.55$ m/sec and $\Delta s_H = 4.7$ meters.

Figure 20 shows that with a larger entries in the diffusion matrix of the noise in the wind, the helicopter becomes less stable. We see that the pitch angle exceed 15° and becomes unstable. Here we have $\sigma_x = 4.42$ m/sec (10 mi/hr) and $\sigma_z = 2.76$ m/sec (6.2 mi/hr).

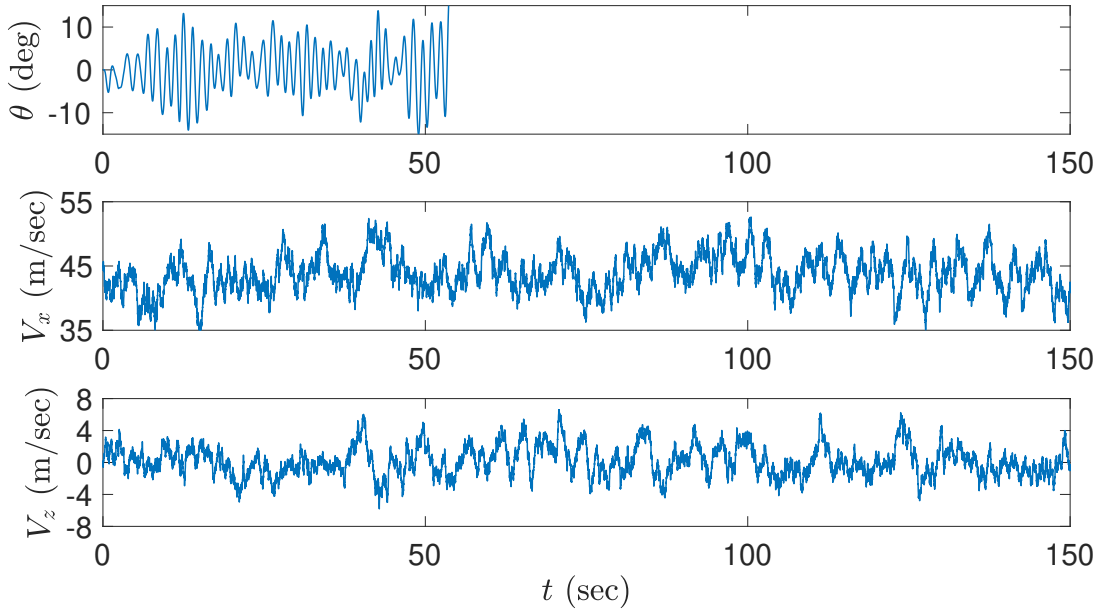

Figure 20: Under noisier, larger variance, the helicopter has larger oscillations. We have $\sigma_x = 4.42$ m/sec, $\sigma_z = 2.76$ m/sec and $\Delta s_H = 4.7$ meters.

Figure 21 shows how larger oscillations form when the tail is half the size of the usual tail

length. Eventually the sample paths leave the region of stability. Here we have $\sigma_x = 0.55$ m/sec and $\sigma_z = 0.55$ m/sec.

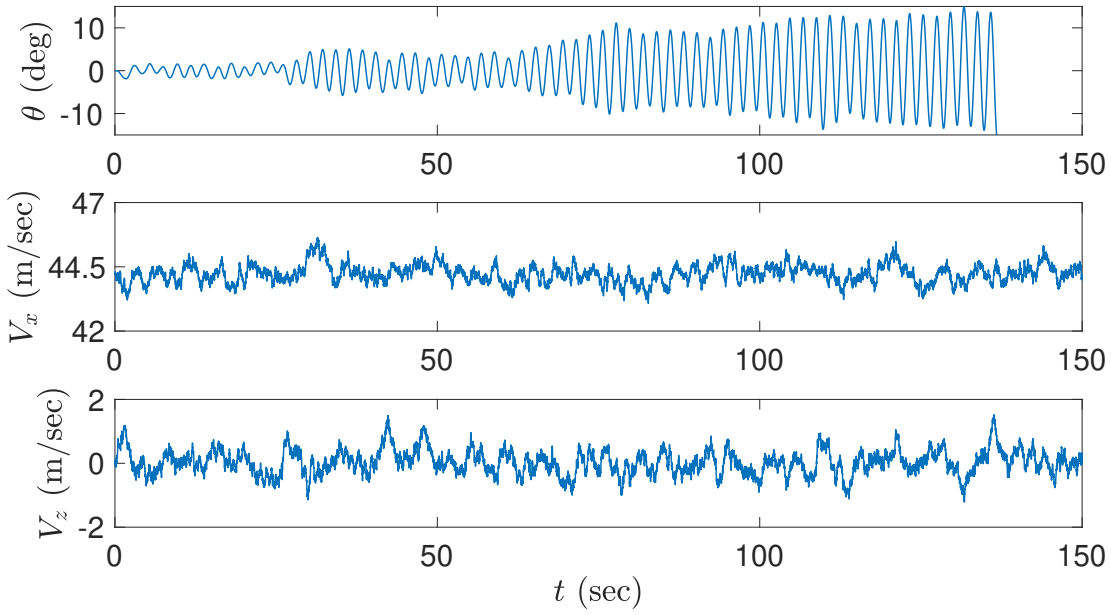

Figure 21: The helicopter with a shorter tail has larger oscillations, so it is less stable. We have $\sigma_x = 0.55$ m/sec, $\sigma_z = 0.55$ m/sec and $\Delta s_H = 2.35$ meters.

Given these simulations, a dangerous event can be defined as the helicopter pitch angle exceeding $\pm 15^\circ$. The reasoning is that since the tail airfoil is used to stabilize the rotorcraft in flight, it will fail produce lift beyond this range as the airfoil will stall. This will cause the helicopter to become unstable. Realistically, a pilot cannot apply control by trimming the helicopter continuously in time and reacting to large unexpected events. What is more likely is that the pilot can only apply trim control to the rotorcraft once every few seconds. Therefore, the rare event we define is that the helicopter pitch angle exceeds $\pm 15^\circ$ within some fixed time period. What makes this event rare is that usually one designs the tail such that the wind that causes this instability does not happen often.

Applying LDT Based on the explorations of the model and its instabilities in the previous section, we wish to estimate the probability that the vehicle pitch angle exceeds $\pm 15^\circ$ within a time window of 10 seconds, while the pilot has initially trimmed the rotorcraft so that it is level with the ground $\theta(0) = 0^\circ$. This roughly corresponds to the probability that the tail fin stalls in flight and causes instabilities within 10 seconds before the pilot can react and trim the helicopter for stability. Note that there is no additional control applied by any autonomous system or by the pilot in the 10 second period. The system is evolving on its own according to its dynamics under stochastic forcing from the noisy wind. Define E to be

$$E = \{\theta(t) | \exists t^* \in [0,10], |\theta(t^*)| \geq 15^\circ\} \subset \mathscr{AC}([0,10];\mathbb{R}).$$

We model the noise in the wind as an Ornstein-Uhlenbeck process with $\mathbf{V}_0 = [44.2, 0]^T$ m/sec, $\sigma_x = 0.89$ m/sec and, $\sigma_z = 1.56$ m/sec. We first find the large deviations rate function of the

system. Since the stochastic part of the model is only in the wind, the large deviations rate function of the system is captured by the Ornstein-Uhlenbeck process alone. Therefore we have that

$$\mathscr{L}(\mathbf{V},\dot{\mathbf{V}}) = \frac{1}{2}((\mathbf{V}-\mathbf{V}_0)+\dot{\mathbf{V}})^T(\sigma\sigma^T)^{-1}((\mathbf{V}-\mathbf{V}_0)+\dot{\mathbf{V}}) \tag{160}$$

if the deterministic model is satisfied and is equal to infinity otherwise. To find the optimal biasing for the wind at any location and time (x,t), we need to solve the following family of variational problems

$$U(x,t) = \inf_{\mathbf{V}\in\mathscr{AC}([t,t_f];\mathbb{R})} \left\{ \int_t^{t_f} \frac{1}{2}((\mathbf{V}-\mathbf{V}_0)+\dot{\mathbf{V}})^T(\sigma\sigma^T)^{-1}((\mathbf{V}-\mathbf{V}_0)+\dot{\mathbf{V}})\,\mathrm{d}s : \begin{bmatrix}\dot{\theta}\\ \dot{q}\end{bmatrix} = \begin{bmatrix} q \\ \frac{1}{I_y}M_{\mathrm{CG}}(\theta,\mathbf{V}(t))\end{bmatrix}, \theta(t)\in E \right\}. \tag{161}$$

Via the contraction principle from large deviations theory [52], we can re-write the optimization problem as an optimal control problem

$$U(x,t) = \inf_{\mathbf{u}\in\mathscr{AC}[0,t]} \left\{ \int_t^{t_f} \frac{1}{2}\|\mathbf{u}(s)\|^2 ds : \begin{bmatrix}\dot{\theta}\\ \dot{q}\\ \dot{\mathbf{V}}\end{bmatrix} = \begin{bmatrix} q \\ \frac{1}{I_y}M(\theta,\mathbf{V}(t)) \\ (\dot{\mathbf{V}}-\mathbf{V}_0)+\sigma\mathbf{u}\end{bmatrix}, \theta(t)\in E \right\}. \tag{162}$$

Intuitively, one can view this algorithm as taking control of the noise term in the model and finding a biasing path that is the "least rare" way that will cause the dynamics to go towards the rare event of interest. With a biasing path, we can find the importance weight between paths with Girsanov's theorem. Note that in this problem, we have incorporated ε into σ. Let the importance weight computed from Girsanov's theorem be

$$Z_E = \exp\left(-\int_0^{t_f}\langle \mathbf{u}(t), \mathrm{d}\mathbf{W}(t)\rangle - \frac{1}{2}\int_0^{t_f}\|\mathbf{u}(t)\|^2\,\mathrm{d}t\right). \tag{163}$$

Then the importance sampling estimator we obtain is

$$\hat{\rho}_E = \frac{1}{K}\sum_{i=1}^{K}\mathbb{1}_E(\theta^i(\cdot))Z_E^i \tag{164}$$

where K the number of independent sample paths.

We can solve the optimization problem numerically in a brute force fashion, or we can turn this problem into a linear quadratic regulator (LQR) problem by considering a linearization. Since the length of the tail is usually chosen to stabilize the rotorcraft, the net pitching moment is dominated by the moment from the tail lift. Furthermore, since the lift generated by the tail is linear with the angle of attack, we find that a linearized model is sufficient for the purposes of solving the optimization problem. That is, we have a simple harmonic oscillator that can model the dynamics of the helicopter that is linearized around a stable equilibrium point. While using this surrogate model may not exactly give us the optimal biasing, the resulting estimator will still be unbiased. Similar approaches have been considered in [41, 53] where the authors find near-optimal biasing for nonlinear dynamical systems by first considering a linearization. In Figure 22, we plot the exact net pitching moment as a function of the pitch angle θ with constant

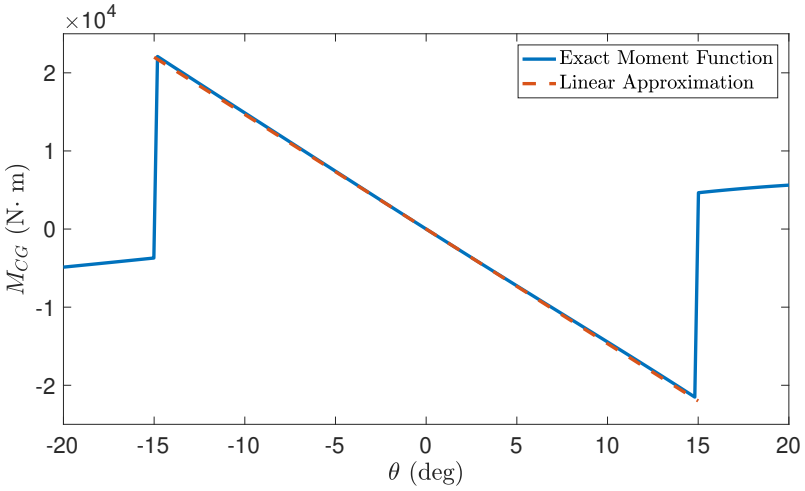

Figure 22: We plot the net moment as a function of the pitch angle θ with the zero noise in the wind and assuming the system is cruising at a velocity of $\mathbf{V}_0 = [44.2, 0]^T$ m/sec.

freestream velocity $\mathbf{V} = [44.2, 0]^T$ m/sec. We also plot a linear approximation to the model about the equilibrium point.

Note that the resulting problem is different from the standard LQR form. Usually, one wishes to find a control that drives the system towards zero. In our scenario, we wish to drive the system to a particular point. To solve our optimal control problem, we use a modified LQR algorithm presented in [49] to find the optimal biasing $\mathbf{u} = [u_x, u_z]^T$ at each time step. Note that we can replace $\mathbf{V}$ with the variable $\widetilde{\mathbf{V}} = \mathbf{V} - \mathbf{V}_0$. Let $\mathbf{u} = [u_x, u_z]^T$. For the model at hand, after linearization, we can obtain an LQR problem of the following form

$$\tilde{U}(x,t) = \inf_{\mathbf{u}\in\mathscr{AC}[t,t_f]} \left\{ \int_t^{t_f} \frac{1}{2}\|\mathbf{u}(s)\|^2 ds : \begin{bmatrix} \dot{\theta} \\ \dot{q} \\ \dot{\widetilde{V}}_x \\ \dot{\widetilde{V}}_z \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ a & 0 & b & c \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \theta \\ q \\ \widetilde{V}_x \\ \widetilde{V}_z \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \sigma_x & 0 \\ 0 & \sigma_z \end{bmatrix} \begin{bmatrix} u_x \\ u_z \end{bmatrix}, \theta(t) = x, \theta \in E \right\}. \tag{165}$$

where $a, b, c < 0$.

Since the exact time t^* that a least rare path enters the rare event may be change as a simulation is happening, and because the rare event region is split into two parts, there is an outer optimization loop to the problem in Equation 165. The outer optimization loop finds t^* and finds which direction the method should bias towards $+15^\circ$ or -15°. Finally, to make our problem more similar to the standard LQR algorithm, the endpoint constraint is replaced by a quadratic terminal penalty with a large constant. A more precise formulation of Equation 165 is given as follows:

$$\tilde{U}(x,t) = \inf_{\substack{t^*\in[t,t_f] \\ \theta_f\in\{-15^\circ,15^\circ\} \\ \theta(t)=x \\ \mathbf{u}\in\mathscr{AC}[t,t^*]}} \left\{ \frac{1}{2}C_K\left(\theta(t^*) - \theta_f\right)^2 + \int_t^{t^*} \frac{1}{2}\|\mathbf{u}(s)\|^2 ds : \begin{bmatrix} \dot{\theta} \\ \dot{q} \\ \dot{\widetilde{V}}_x \\ \dot{\widetilde{V}}_z \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ a & 0 & b & c \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \theta \\ q \\ \widetilde{V}_x \\ \widetilde{V}_z \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \sigma_x & 0 \\ 0 & \sigma_z \end{bmatrix} \begin{bmatrix} u_x \\ u_z \end{bmatrix} \right\}, \tag{166}$$

where $C_K > 0$ is a large constant. We state the algorithm we implement in Algorithm 1.

Data: Linearized model in Equation 165, exact SDE model in Equation 162, Number of desired samples K, Rare event of interest E

Result: Probability estimate $\hat{\rho}_E$

for $i = 1$ *to* K **do**

  while $\theta \notin E$ **do**

    Solve optimization problem in Equation 166 via an LQR algorithm, find $\mathbf{u}(t)$, evaluate at t_n.

    Evolve exact SDE model one time step with biasing $\mathbf{u}(t_n)$

  end

  Compute importance weight $Z_E^i = \exp\left(-\int_0^{t^*} \langle \mathbf{u}(t), d\mathbf{W}_t \rangle - \frac{1}{2}\int_0^{t^*} \|u(t)\|^2 dt\right)$.

end

return $\hat{\rho}_E = \frac{1}{K}\sum_{i=1}^{K} \mathbb{1}_E(\theta(t)) Z_E^i$

Algorithm 1: Obtain rare event probability estimate

In Figure 23, we show seven plots relating to a biased sample of the rotorcraft model. The first plot shows the evolution of θ over time. We see that the path enters the rare event around $t = 8.6s$. In the second and third plots, we show the corresponding noisy wind that causes the rare event. In the fourth and fifth plots, we show the optimal control of the noise of the wind that leads to the rare event. We see that the control is oscillatory, which makes sense. One can think of the analogy of a playground swing, where one pushes the system at the correct times to force the θ to grow over time. In the sixth plot, we show the target final time, which changes over the course of the simulation. In the seventh plot, we show which direction the biasing pushes towards, which changes over the course of the simulation. As expected, this plot is roughly periodic as well since the path approaches the two sides of the rare event periodically.

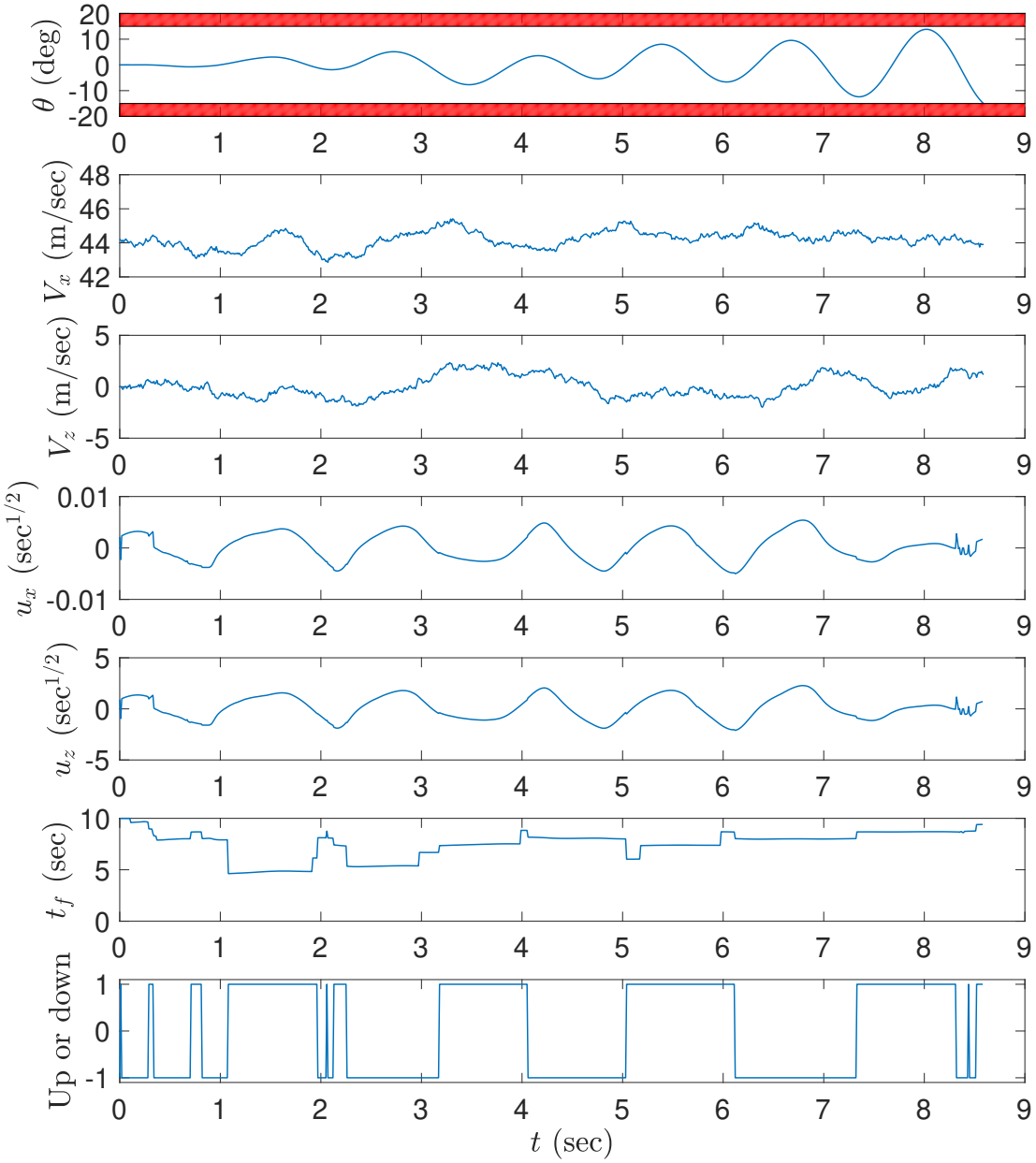

Figure 23: An example of a biased sample of the rotorcraft model that reaches the rare event by time $t = 8.6s$. The top three figures show the pitch angle and the velocity profile of the aircraft with noisy wind. The fourth and fifth figures show the applied biasing through time. The sixth figure shows the estimated end time in the rare event as the simulation is occurring. The seventh figure shows whether the algorithm is biasing towards $+15°$ or $-15°$. The simulation ends when the pitch angle reaches the rare event.

In Figure 24, we show 50 simple Monte Carlo samples of the rotorcraft system – i.e. *without* biasing – including the corresponding realizations of the wind velocities. Note that no samples reach the rare event.

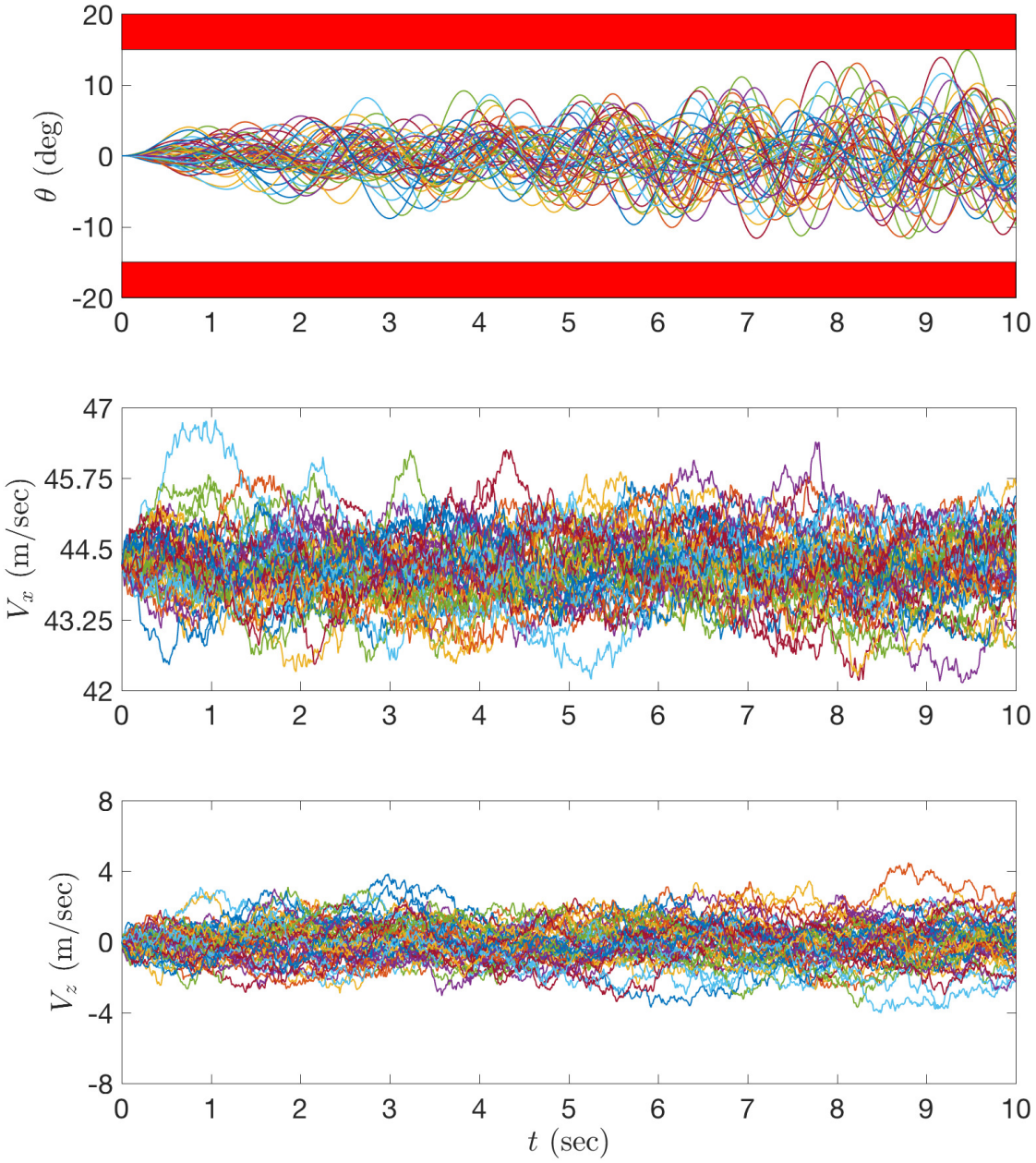

Figure 24: Simple or unbiased Monte Carlo samples of the rotorcraft model.

In Figure 25, we show 50 biased samples. It seems that all paths lead to the rare event, but not all paths to the rare event are equally important. In Figure 26 we shade each path according to its importance weight relative to the least rare path of the batch.

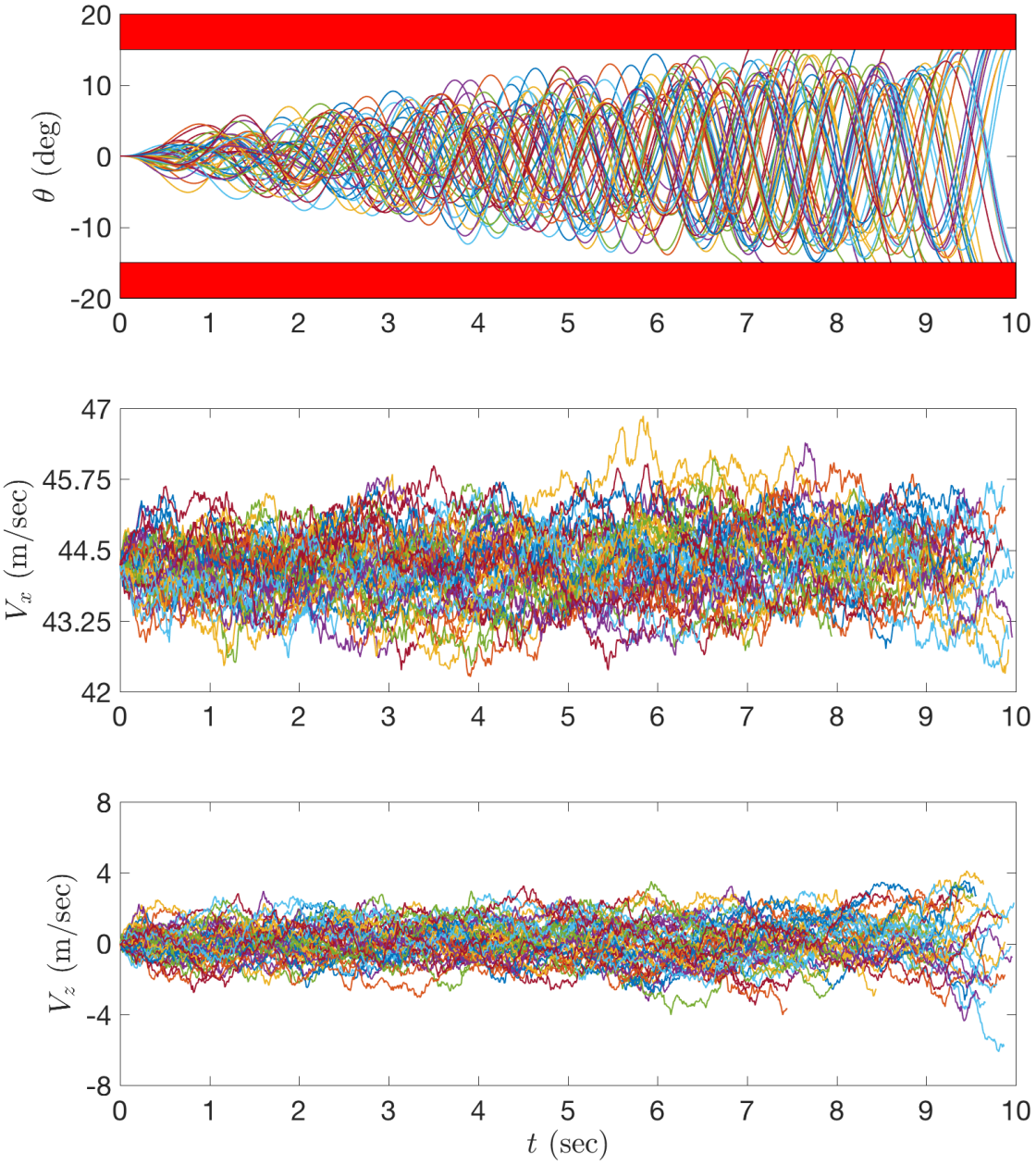

Figure 25: Biased samples of the model.

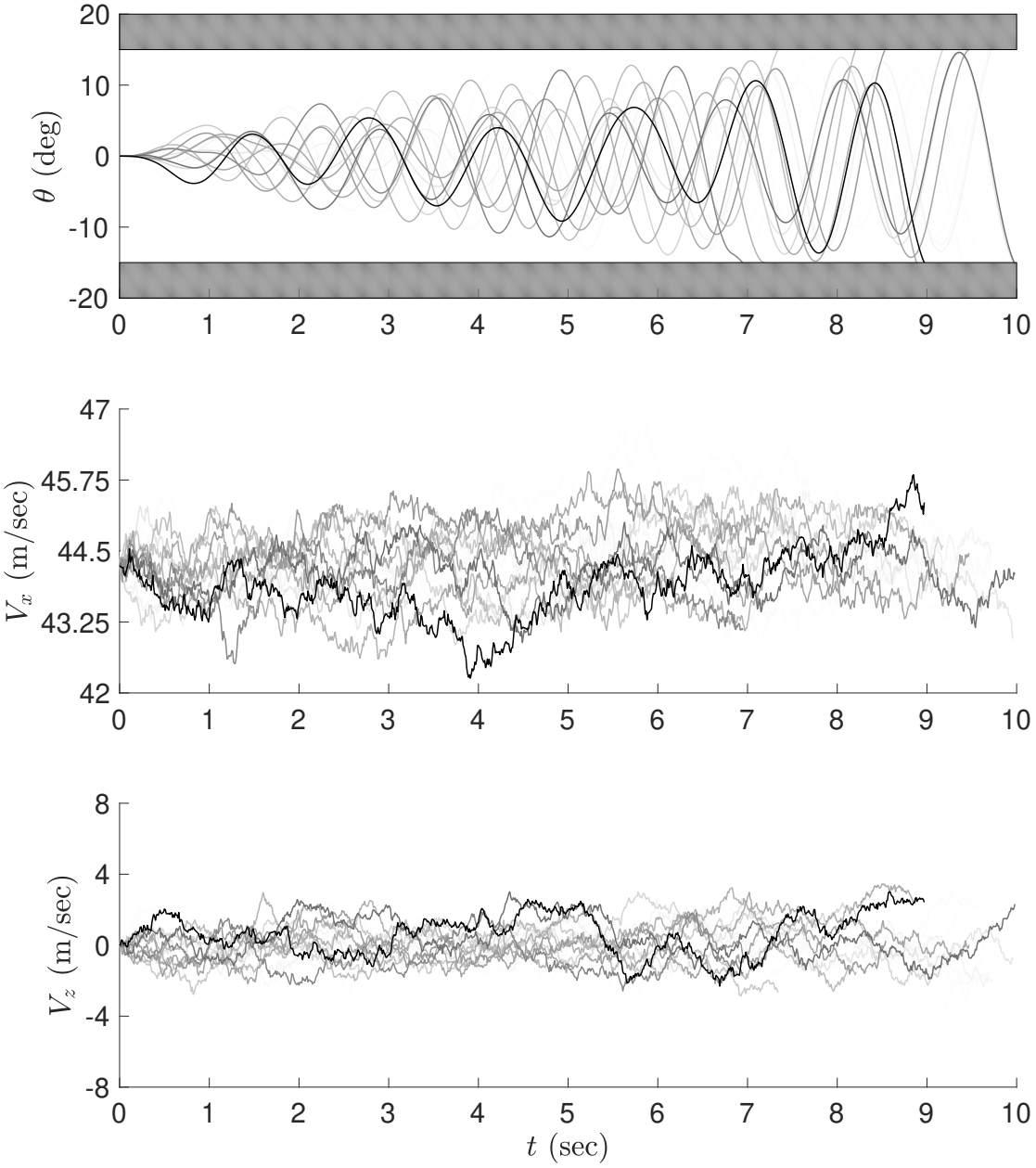

Figure 26: Weighted biased samples. Each path from Figure 25 is shaded according to its importance weight relative to the path with the largest weight in the batch.

We now state the computed probabilities based on simple Monte Carlo and dynamic importance sampling in Table 9. We take the true probability to be estimated by simple Monte Carlo with 1.5 million sample paths, $\rho_{true} = 7.83 \times 10^{-3}$. The probabilities in Table 9 are computed with 100 batches of 1000 samples for each method.

Table 9: Rotorcraft example results with $\sigma_x = 0.89$ m/sec, $\sigma_z = 1.56$ m/sec.

	Estimate $\hat{\rho}_n$	Std. Error	95% confidence	Relative error	Rel. error per sample
MC	7.80×10^{-3}	8.23×10^{-3}	$[0, 24.3] \times 10^{-3}$	1.06	33.5
DIS	7.77×10^{-3}	1.33×10^{-3}	$[5.11, 10.43] \times 10^{-3}$	0.17	5.40

Table 10: Rotorcraft example results with $\sigma_x = 0.45$ m/sec, $\sigma_z = 0.78$ m/sec.

	Estimate $\hat{\rho}_n$	Std. Error	95% confidence	Relative error	Rel. error per sample
DIS	2.65×10^{-10}	2.60×10^{-10}	$[0.05, 5.25] \times 10^{-10}$	0.98	43.8

We observe significant variance reduction, as evidenced by a roughly seven-fold reduction in relative error per sample and a narrower 95% confidence interval that places a meaningful lower bound on the failure probability. That said, the amount of variance reduction we observe for this problem is not as great as has been reported in other literature on dynamical importance sampling [8, 11]. One issue that diminishes the performance of the method is identified in [41]: the efficiency of dynamic importance sampling for small noise diffusions can deteriorate in the presence of an attractor. Our problem has an attractor, since the system is oscillating around an equilibrium point and because of the mean reverting nature of the wind. The authors in [41] note an improvement of the algorithm by considering the noise level ε into account when finding the optimal biasing; however, their approach does not extend to higher dimensions. In [42], the authors present an approach that applies to higher-dimensional problems; their algorithm, however, requires the system dynamics to be linear and self-adjoint. The latter condition, in particular, is not satisfied by the helicopter model considered here. Therefore, these fixes do not yet seem theoretically viable for real engineering problems.

Finally, we perform the same experiment to the system where σ_x and σ_z are smaller than the previous example. That is, $\sigma_x = 0.45$ m/sec and $\sigma_z = 0.78$ m/sec. The probability is computed with 100 batches of 2000 samples. Using dynamic importance sampling, we compute that the probability is 2.65×10^{-10}. It is intractable to perform crude Monte Carlo in this scenario as one would expect to see only one instance of the rare event after $\sim 10^{10}$ samples. The results are stated in Table 10. While the relative error for dynamic importance sampling is not ideal, as it has the same order of magnitude of the estimate, it still provides a useful estimate.

Applying Splitting Methods to the Rotorcraft Models We consider a simple rotorcraft model similar to the one in [54], which is defined by the following dynamic equations:

$$d\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t)dt + \delta\mathbf{f} + \sqrt{\varepsilon}\Sigma(\mathbf{x})d\mathbf{W}(t)\,, \tag{167}$$

with

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -0.275 & -0.01 \\ 9.8 & -1.4 & -0.02 \end{bmatrix}, \mathbf{f} = \begin{bmatrix} 0 \\ 6.3 \\ 9.8 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

Table 11: Comparison of RESTART estimator and standard Monte Carlo estimator.

	MC	$R=2$	$R=3$	$R=4$	$R=10$	$R=14$	$R=20$
$\bar{p}(\times 10^{-3})$	2.2	2.1	2.2	2.1	2.1	2.1	2.1
$\sqrt{V}(\times 10^{-3})$	1.5	0.66	0.4	0.32	0.21	0.194	0.196
relative error	0.75	0.33	0.2	0.16	0.11	0.092	0.098
C	0.054	0.25	0.26	0.28	0.49	0.72	1.24
$CV(\times 10^{-7})$	1.22	1.09	0.42	0.29	0.22	0.27	0.48

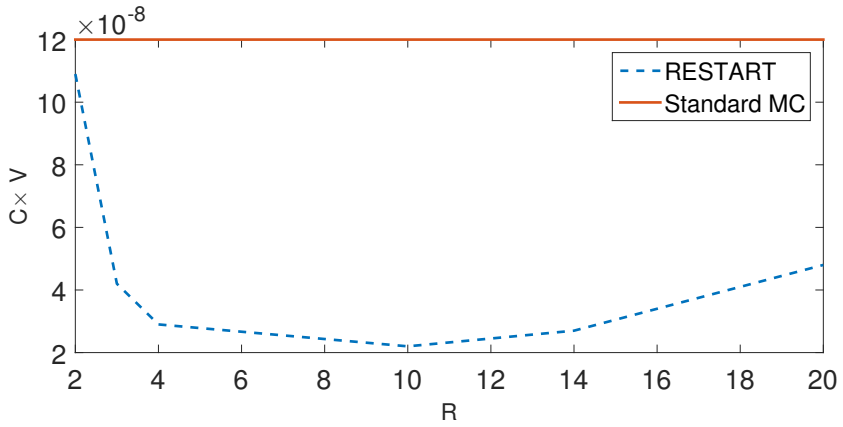

Figure 27: The computational costs of MC and RESTART to reach the same statistical tolerance.

where x_1 is the pitch angle of the helicopter, x_2 is the rate of change of the pitch angle, x_3 is the horizontal velocity, f is the thrust provided by the rotor, $\delta = 0.1$, $\mathbf{W}$ is a vector of standard Brownian motion, Σ is the standard diffusion matrix.

We used a near optimal importance function in this example. Table 11 shows the mean value ($\bar{p}$), standard deviation ($\sqrt{V}$) of the 1000-particle estimator, its relative error (standard deviation / true value), cost of a single sample (C), and the cost of the estimator under a prescribed tolerance (CV). Figure 27 shows the computational cost of the RESTART estimator using different splitting ratios R. It is observed that when $R = 10$ we obtain the optimal complexity. The complexity is not sensitive to the change of the splitting ratio when $4 \leq R \leq 14$. In addition, we empirically checked that the probability of reaching the next level is 10% (which matches the value of the optimal splitting ratio) as we pre-defined 3 thresholds.

6.4 UTRC Rotorcraft

The dynamics of the fuselage can be described as

$$\dot{\theta}_v = q \tag{168}$$

$$\dot{q} = \frac{M(\theta, v(t))}{I_y} \tag{169}$$

$$M(\theta_v, v(t)) = M_R + T s_R - L_H s_H\,, \tag{170}$$

where θ_v is the pitch angle of the fuselage, q is the rate of change of the pitch angle, M is the net pitch moment applied to the rotorcraft, where T is the thrust, s_R is the distance from the thrust to

the center of mass. The airfoil provide lift L_H. The model is implemented in software GENTRIM which solve for the rotor thrust, torque, horizontal force, roll, pitch, and hub moments, and rotor lift over drag ratio given desired input parameters. More descriptions on the model can be found in [54]. We define the absolute value of the pitch angle larger than 15^o as the rare events in which the rotor crafts become unstable and out of control.

Table 12 shows some initial results of RESTART on the UTRC rotor craft example. We noticed significant reduction of variance even when the thresholds are chosen arbitrarily.

Table 12: Comparison of RESTART estimator and standard Monte Carlo estimator.

	MC	$R = 14$
$\bar{p}(\times 10^{-3})$	7.8	8.0
$\sqrt{V}(\times 10^{-3})$	8.2	2.6
relative error	1.06	0.33

6.5 Future work: Alternate Solutions to the Stochastic HJB Equation

From the theory of subsolutions for rare event simulation, if we can solve exactly or satisfy the following set of inequalities, we may obtain an asymptotically optimal importance sampling or splitting estimator. The typical subsolutions approach to this problem is analytic as it attempts to devise cleverly constructed *ansatz* that are shown to perform well using theory.

$$\bar{U}_t(t,x) + \mathbb{H}(x, \nabla \bar{U}(t,x)) \geq 0 \text{ for } (t,x) \in [0,T) \times D \tag{171}$$

$$\bar{U}(t,x) \leq h(x) \text{ for } t \leq T, x \in \partial D \tag{172}$$

$$\bar{U}(T,x) \leq \infty \text{ for } x \in D. \tag{173}$$

Here we consider an alternative approach to solve the Hamilton-Jacobi-Bellman equation or find their subsolutions. To this end, we consider using the *Compressed Continuous Computation for Stochastic Optimal Control Package (C3SC)* presented in [55, 56]. This package can solve the stochastic HJB equation by solving a series of stochastic optimal control problems across the entire phase space. While this may seem inefficient, this method assumes that the value function (solution to the HJB equation) has low tensor rank in hopes of accelerating computation of the optimal control problem. This approach may be feasible for up to ten dimensions.

Recall that in the analysis for finding good estimators for importance sampling or splitting, a zero variance choice of biasing or importance function is possible if one can exactly solve the Kolmogorov Backward Equation, which is also known as the stochastic Hamilton-Jacobi-Bellman equation in the optimal control community.

$$\mathscr{G}^{\varepsilon}[U](t,x) = \frac{\partial U}{\partial t}(t,x) + \mathbb{H}(x, D_x U(x)) + \frac{\varepsilon}{2}\mathrm{Tr}(BB^* D_x^2 U(t,x)) = 0$$

The C3SC package solves a variant of this PDE. By finding fast solutions to the infinite-horizon stochastic optimal control problem, we procure solutions to the following PDE instead:

$$\mathbb{H}(x, D_x U(x)) + \frac{\varepsilon}{2}\mathrm{Tr}(BB^* D_x^2 U(t,x)) = 0.$$

Now, there are two tasks at hand. One is to produce solutions to this PDE via the C3SC package and test it for importance sampling and splitting. The second is to see what sort of efficiency guarantees that are preserved by using a solution to this PDE.

We briefly explain how the C3SC produces fast solution to stochastic optimal control problems. Typically a stochastic optimal control problem where the system is governed by a stochastic differential equation is solved by the Markov chain approximation method. One construct a discrete state space, discrete-time Markov chain approximation of the system, then formulate the optimal control problem as a Markov decision process. This problem is then solved by methods such as value iteration or policy iteration [57]. The difficulty in applying this approach lies with the fact that discretizing the state space for more than three or four dimensions becomes computationally prohibitive.

The novelty of the C3SC package is that it attempts to accelerate computation of the solution to the Markov decision process via compressed continuous computation. The basic idea is that the algorithm expresses high dimensional functions in a low dimensional manner by constructing a continuous version of the tensor train decomposition. The author describes the function approximation method as the *functional tensor train* approximation in which rather than having numerical entries in the tensors, one has univariate functions instead. The general functional tensor train form is in the following [56]. Suppose we have a multivariable function $f : X \longrightarrow \mathbb{R}$, where $X \subset \mathbb{R}^n$. Then

$$f(x_1, \cdots, x_n) = \sum_{i_0=1}^{r_0} \cdots \sum_{i_n=1}^{r_n} f_{i_0, i_1^{(1)}}(x_1) \cdots f_{i_n - 1, i_n}(x_n)$$

where f_{jk} are univariate scalar valued functions and r_i are called the function-train ranks. Increasing the rank increases the depth of representation the function train has with trade offs in higher dimensional computation. This product is usually written in terms of a product of matrices of the form

$$F_i(x_i) = \begin{bmatrix} f_{1,1}^{(i)}(x_i) & \cdots & f_{1,r_i}^{(i)}(x_i) \\ \vdots & \ddots & \vdots \\ f_{r_{i-1},1}^{(i)}(x_i) & \cdots & f_{r_{i-1},r_i}^{(i)}(x_i). \end{bmatrix}$$

Then by fixing $r_0 = r_n = 1$, then the product of matrices will equal a scalar function

$$f(x_1, \cdots, x_n) = F_1(x_1) \cdots F_n(x_n)$$

The package can solve high dimensional stochastic optimal control problems that do not encounter the curse of dimensionality as quickly. The caveat to this method is that while it is more scalable than traditional stochastic optimal control solvers, the method is still limit to dimensions less than 10. Anything more, the curse of dimensionality may still appear.

7 Conclusions and Future Work

The goal under the SIRE project was to develop novel rare event estimation methods for engineering systems. Under the program, we have developed a range of approaches (importance sampling and splitting) that extend the state-of-the-art. We applied these methods to rotorcraft and electrical systems that we developed under the project.

Our approaches for studying and simulating rare events exploit tools from dynamical systems theory. Our approach for dynamic systems ensures that the attracting dynamics of system attractors do not degrade performance of the rare events scheme. We used existing rare event simulation methodologies that leverage large deviations and showed their applicability to restricted classes of engineering systems. While large deviations-based importance sampling and splitting methods worked for rotorcraft and electrical systems, we reproduced difficulties of subsolution-based methods in the presence of attracting sets. We investigated ways to circumvent the issue by use of the left eigenvector in linear systems. This left eigenvector associated with slowest decaying mode was shown to lead to efficient importance sampling estimators (as summarized in the list below). We also demonstrated the splitting approach on the rotorcraft problem and formulated a novel methodology that exploits tensor decompositions for stochastic control problems.

The left eigenvector biasing work enabled us to make new connections to the study of the Koopman operator. We made a clear and novel connection between the stochastic Koopman operator and efficient importance sampling methods for stochastic differential equations. In particular, we found that the Koopman operator can motivate the left eigenvector biasing method and significantly improve the performance of the importance sampling methods in dynamic, non-normal, nonlinear, degenerate noise, and black box settings.

For future work, we propose using efficient stochastic optimal control solvers for faster and more automatic implementation of large deviations-based methods. Another route for future work involves rigorously leveraging the theoretical results of large deviations-based methods to analyze approximate the Koopman operator approaches. Moreover, we plan to extend the Koopman framework to create methods that can sample from high dimensional distributions. In essence, we can create dynamical systems that sample from arbitrary distributions. This can potentially also be used to create new Bayesian filtering methods.

Furthermore, our developed approach can be used for estimating failure in complex engineering systems, designing new systems to mitigate failures (catastrophes), and performing design of experiments for information gathering. The Koopman operator method can also be used for real time prediction by leveraging existing data driven methods for estimating the Koopman eigenfunctions. These eigenfunctions can then be used to computationally bias existing models for system evolution (akin to standard filtering methods). We plan to explore the above areas in future programs.

In summary, the novel contributions of the SIRE program and associated (demonstrated) improvements are listed below.

- Discovered deep theoretical connections between Koopman operator theory and rare event simulation in stochastic differential equations.
- Developed a generalized approach for efficient importance sampling methods for linear stochastic differential equations using the Kolmogorov Backward (Ornstein-Uhlenbeck)

operator. This approach is a special case of the Koopman operator approach. The Koopman operator approach demonstrated a 5000X and 1000X variance reductions on the Van der Pol and Duffing oscillators respectively. Moreover, this approach is able to deal with degenerate noise cases that arise in real world systems.

- Found all the eigenfunctions of the Ornstein-Uhlenbeck operator, which will be useful for developing sampling methods.
- Developed heuristics for efficient importance sampling methods for linear stochastic partial differential equations.
- Constructed rotorcraft models that capture critical stall phenomena that was used for computations.
- Formulated and applied a left-eigenvector based large deviations-based importance sampling and splitting methods to rotorcraft and electrical models. The approach demonstrated a 30X variance reduction on the electrical system model and 36X variance reduction on the rotorcraft system. Additionally, it demonstrated a 2000X variance reduction in non-normal dynamical systems (existing large deviations based methods did not work in this setting). The splitting approach demonstrated a 10X variance reduction on a simple rotorcraft model.
- Developed a new formulation for accelerating large deviations-based approaches using fast stochastic optimal control solvers.

8 References

[1] Søren Asmussen and Peter W Glynn. *Stochastic simulation: algorithms and analysis*, volume 57. Springer Science & Business Media, 2007.

[2] Benjamin Zhang, Youssef Marzouk, Byung-Young Min, and Tuhin Sahai. Rare event simulation of a rotorcraft system. In *2018 AIAA Non-Deterministic Approaches Conference*, page 1181, 2018.

[3] James Antonio Bucklew. *Introduction to rare event simulation*. Springer series in statistics. Springer, New York, 2004.

[4] Reuven Y Rubinstein and Dirk P Kroese. *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning*. Springer Science & Business Media, 2013.

[5] Wei Zhang, Han Wang, Carsten Hartmann, Marcus Weber, and Christof Schutte. Applications of the cross-entropy method to importance sampling and optimal control of diffusions. *SIAM Journal on Scientific Computing*, 36(6):A2654–A2672, 2014.

[6] Siu-Kui Au and James L Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic engineering mechanics*, 16(4):263–277, 2001.

[7] Thomas Dean and Paul Dupuis. Splitting for rare event simulation: A large deviation approach to design and analysis. *Stochastic processes and their applications*, 119(2):562–587, 2009.

[8] Eric Vanden-Eijnden and Jonathan Weare. Rare event simulation of small noise diffusions. *Communications on Pure and Applied Mathematics*, 65(12):1770–1803, 2012.

[9] Omar Kebiri, Lara Neureither, and Carsten Hartmann. Adaptive importance sampling with forward-backward stochastic differential equations. In *International workshop on Stochastic Dynamics out of Equilibrium*, pages 265–281. Springer, 2017.

[10] David Siegmund et al. Importance sampling in the monte carlo study of sequential tests. *The Annals of Statistics*, 4(4):673–684, 1976.

[11] Paul Dupuis and Hui Wang. Importance sampling, large deviations, and differential games. *Stochastics: An International Journal of Probability and Stochastic Processes*, 76(6):481–508, 2004.

[12] Paul Dupuis, Ali Devin Sezer, Hui Wang, et al. Dynamic importance sampling for queueing networks. *The Annals of Applied Probability*, 17(4):1306–1346, 2007.

[13] Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media, 2012.

[14] Bernt Øksendal. Stochastic differential equations. In *Stochastic differential equations*, pages 65–84. Springer, 2003.

[15] Grigorios A Pavliotis. *Stochastic processes and applications*. Springer, 2011.

[16] Art B. Owen. *Monte Carlo theory, methods and examples*. 2013.

[17] Gerardo Rubino and Bruno Tuffin. *Rare event simulation using Monte Carlo methods*. John Wiley & Sons, 2009.

[18] Manuel Villén-Altamirano and José Villén-Altamirano. Restart: a straightforward method for fast simulation of rare events. In *Proceedings of Winter Simulation Conference*, pages 282–289. IEEE, 1994.

[19] Thomas Dean and Paul Dupuis. The design and analysis of a generalized restart/dpr algorithm for rare event simulation. *Annals of Operations Research*, 189(1):63–102, 2011.

[20] Peter E Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 23. Springer Science & Business Media, 2013.

[21] Andreas Rößler. Runge–kutta methods for the strong approximation of solutions of stochastic differential equations. *SIAM Journal on Numerical Analysis*, 48(3):922–952, 2010.

[22] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic equations in infinite dimensions*. Cambridge university press, 2014.

[23] Matteo Beccari, Martin Hutzenthaler, Arnulf Jentzen, Ryan Kurniawan, Felix Lindner, and Diyora Salimova. Strong and weak divergence of exponential and linear-implicit euler approximations for stochastic partial differential equations with superlinearly growing nonlinearities. *arXiv preprint arXiv:1903.06066*, 2019.

[24] Sebastian Becker, Benjamin Gess, Arnulf Jentzen, and Peter E Kloeden. Strong convergence rates for explicit space-time discrete numerical approximations of stochastic allen-cahn equations. *arXiv preprint arXiv:1711.02423*, 2017.

[25] Xiaojie Wang. An efficient explicit full discrete scheme for strong approximation of stochastic allen-cahn equation. *arXiv preprint arXiv:1802.09413*, 2018.

[26] Byung-Young Min. *A physics based investigation of Gurney flaps for enhancement of rotorcraft flight characteristics*. PhD thesis, Georgia Institute of Technology, 2010.

[27] Wayne Johnson. *Helicopter Theory*. Courier Corporation, 2012.

[28] Gordon J Leishman. *Principles of Helicopter Aerodynamics*. Cambridge university press, 2006.

[29] Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1-3):309–325, 2005.

[30] Peter J Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics*, 656:5–28, 2010.

[31] Clarence W Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S Henningson. Spectral analysis of nonlinear flows. *Journal of fluid mechanics*, 641:115–127, 2009.

[32] Matthew O Williams, Ioannis G Kevrekidis, and Clarence W Rowley. A data–driven approximation of the koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, 2015.

[33] Milan Korda and Igor Mezić. On convergence of extended dynamic mode decomposition to the koopman operator. *Journal of Nonlinear Science*, 28(2):687–710, 2018.

[34] Giorgio Metafune, Diego Pallara, and Enrico Priola. Spectrum of Ornstein-Uhlenbeck operators in Lp spaces with respect to invariant measures. *Journal of Functional Analysis*, 196(1):40–60, 2002.

[35] Yong Chen, Yong Liu, et al. On the eigenfunctions of the complex ornstein–uhlenbeck operators. *Kyoto Journal of Mathematics*, 54(3):577–596, 2014.

[36] Mark Iosifovich Freidlin and Alexander D Wentzell. Random perturbations. In *Random Perturbations of Dynamical Systems*, pages 15–43. Springer, 1998.

[37] Jose Blanchet, Peter Glynn, and Kevin Leder. On lyapunov inequalities and subsolutions for efficient importance sampling. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 22(3):13, 2012.

[38] Konstantinos Spiliopoulos. Nonasymptotic performance analysis of importance sampling schemes for small noise diffusions. *Journal of Applied Probability*, 52(03):797–810, 2015.

[39] L.C. Evans. *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society, 2010.

[40] Paul Dupuis and Hui Wang. Subsolutions of an Isaacs equation and efficient schemes for importance sampling. *Mathematics of Operations Research*, 32(3):723–757, 2007.

[41] Paul Dupuis, Konstantinos Spiliopoulos, and Xiang Zhou. Escaping from an attractor: importance sampling and rest points I. *The Annals of Applied Probability*, 25(5):2909–2958, 2015.

[42] Michael Salins and Konstantinos Spiliopoulos. Rare event simulation via importance sampling for linear SPDEs. *Stochastics and Partial Differential Equations: Analysis and Computations*, pages 1–39, 2017.

[43] Amarjit Budhiraja, Paul Dupuis, Vasileios Maroulas, et al. Large deviations for infinite dimensional stochastic dynamical systems. *The Annals of Probability*, 36(4):1390–1420, 2008.

[44] Lloyd N Trefethen, Anne E Trefethen, Satish C Reddy, and Tobin A Driscoll. Hydrodynamic stability without eigenvalues. *Science*, 261(5121):578–584, 1993.

[45] Lloyd N Trefethen and Mark Embree. *Spectra and pseudospectra: the behavior of nonnormal matrices and operators*. Princeton University Press, 2005.

[46] E Weinan and Xiang Zhou. The gentlest ascent dynamics. *Nonlinearity*, 24(6):1831, 2011.

[47] Shuting Gu and Xiang Zhou. Simplified gentlest ascent dynamics for saddle points in non-gradient systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(12):123106, 2018.

[48] Arthur Earl Bryson. *Applied optimal control: optimization, estimation and control*. Routledge, 2018.

[49] Gustavo Goretkin, Alejandro Perez, Robert Platt, and George Konidaris. Optimal sampling-based planning for linear-quadratic kinodynamic systems. In *2013 IEEE International Conference on Robotics and Automation*, pages 2429–2436. IEEE, 2013.

[50] Paul Jackson, LT Peacock, and K Munson. Jane's all the world's aircraft 2003–2004. *Coulsdon, UK: Jane's Information Group*, 2003.

[51] Rafael Zárate-Miñano, Marian Anghel, and Federico Milano. Continuous wind speed models based on stochastic differential equations. *Applied Energy*, 104:42–49, 2013.

[52] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Applications of mathematics. Springer, 1998.

[53] Carsten Hartmann, Christof Schütte, and Wei Zhang. Model reduction algorithms for optimal control and importance sampling of diffusions. *Nonlinearity*, 29(8):2298, 2016.

[54] Benjamin J Zhang. A coupling approach to rare event simulation via dynamic importance sampling. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, 2017.

[55] Alex Gorodetsky, Sertac Karaman, and Youssef Marzouk. High-dimensional stochastic optimal control using continuous tensor decompositions. *The International Journal of Robotics Research*, 37(2-3):340–377, 2018.

[56] Ezra Tal, Alex Gorodetsky, and Sertac Karaman. Continuous tensor train-based dynamic programming for high-dimensional zero-sum differential games. In *American Control Conference (ACC), Milwaukee, WI, USA*, 2018.

[57] Dimitri P Bertsekas, Dimitri P Bertsekas, Dimitri P Bertsekas, and Dimitri P Bertsekas. *Dynamic programming and optimal control*, volume 1. Athena scientific Belmont, MA, 1995.

[58] Stefan Klus, Feliks Nüske, Sebastian Peitz, Jan-Hendrik Niemann, Cecilia Clementi, and Christof Schütte. Data-driven approximation of the koopman generator: Model reduction, system identification, and control. *arXiv preprint arXiv:1909.10638*, 2019.

[59] Marko Budišić, Ryan Mohr, and Igor Mezić. Applied koopmanism. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(4):047510, 2012.

[60] Zlatko Drmac, Igor Mezic, and Ryan Mohr. Data driven modal decompositions: analysis and enhancements. *SIAM Journal on Scientific Computing*, 40(4):A2253–A2285, 2018.

[61] Arnulf Jentzen and Peter E Kloeden. The numerical approximation of stochastic partial differential equations. *Milan Journal of Mathematics*, 77(1):205–244, 2009.

[62] Bernard O Koopman. Hamiltonian systems and transformation in hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 1931.

[63] J Nathan Kutz, Steven L Brunton, Bingni W Brunton, and Joshua L Proctor. *Dynamic mode decomposition: data-driven modeling of complex systems*, volume 149. SIAM, 2016.

[64] Jonathan H Tu, Clarence W Rowley, Dirk M Luchtenburg, Steven L Brunton, and J Nathan Kutz. On dynamic mode decomposition: theory and applications. *arXiv preprint arXiv:1312.0041*, 2013.

[65] Nelida Črnjarić-Žic, Senka Maćešić, and Igor Mezić. Koopman operator spectrum for random dynamical systems. *arXiv preprint arXiv:1711.03146*, 2017.

[66] M Villen-Altamirano. Restart: A method for accelerating rare event simulations. *Analysis*, 3:3.