FORWARD AND BACKWARD UNCERTAINTY QUANTIFICATION: METHODS, ANALYSIS, AND APPLICATIONS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Wayne Isaac Tan Uy

August 2019

FORWARD AND BACKWARD UNCERTAINTY QUANTIFICATION:

METHODS, ANALYSIS, AND APPLICATIONS

Wayne Isaac Tan Uy, Ph.D.

Cornell University 2019

The field of uncertainty quantification (UQ) deals with physical systems described by an input-output mapping. Randomness is present either in the parameters of the input or in the quantities of interest which are functionals of the output. This dissertation studies problems that arise in both forward and backward UQ and proposes methods to address them. Forward UQ quantifies the probability distribution on the output when the uncertainty on the input is propagated through the mapping. A probabilistically accurate and computationally efficient surrogate model is necessary to avoid numerous solutions of the input-output map. In contrast, backward UQ infers the distribution on the input assuming that the law of the output is known. Interest is on the well-posedness of this stochastic inverse problem and in particular, the uniqueness of the resulting solution.

We survey existing methods developed to tackle problems in these settings and examine challenges associated with them. With respect to forward UQ, commonly used surrogate models may not possess requisite convergence properties or may be constructed without regard to the distribution on the input. We therefore develop an adaptive method based on Voronoi cells that constructs the surrogate accurately in high probability regions of the input and in regions where the input-output mapping exhibits substantial variation. For backward UQ, recently proposed approaches make assumptions on the law of the input

which do not necessarily guarantee recovery of its true distribution. As such, we investigate what additional information must be specified on the input to solve this inverse problem and outline how approaches based on optimization, the principle of maximum entropy, and Bayes' theorem can incorporate such information. We conclude by considering a different type of inverse problem where the objective is on identifying samples of the input which yield large quantities of interest. This enables one to study the law of the input conditioned on extreme events and predict which inputs cause such events. We achieve this through a general framework which leverages on multifidelity surrogate models for input-output maps and machine learning classifiers.

BIOGRAPHICAL SKETCH

Wayne Isaac Tan Uy was born on September 11, 1991 in Davao City, Philippines. He graduated from the Philippine Science High School – Southern Mindanao Campus in 2008, started his university studies in Manila, Philippines from 2008–2009, and obtained his bachelor's degree in Mathematical Sciences with first class honours from the Nanyang Technological University, Singapore in 2013. He commenced his PhD studies in Applied Mathematics at Cornell University on August 2013 and earned his master's degree in 2016. During his graduate studies, he was the recipient of the Cornelia Ye Outstanding Teaching Assistant Award in 2016 and received teaching awards from the Math and Computer Science departments in 2016 and 2017, respectively. He also held internship positions at Sandia National Laboratories, Livermore in the summer of 2016 and at the Basque Center for Applied Mathematics in Bilbao, Spain in the summer of 2017.

«Ellos vienen buscando oro, ¡id vosotros también a su país a buscar otro oro que nos hace falta! Recuerda, sin embargo, que no es oro todo lo que reluce.»

Noli me tangere, 1887

ACKNOWLEDGEMENTS

Six years ago, I was on the waiting list for admission to CAM and was only accepted at the last minute when another student gave up his or her spot. It seems as if my admission had been an accident and I am grateful to CAM for taking a chance on me. CAM, and the people who have led it through the years, ensured that I had the adequate resources to flourish during my time in graduate school.

My adviser, Prof. Mircea Grigoriu, took me under his tutelage without anticipating how many headeaches I was going to cause him later on. Yet I sincerely appreciate the time he has spent mentoring me and how he was patient enough to let me mature as an academic, constantly challenging me with questions that I initially feared but relished eventually. His wisdom, insights, and perspectives on math, applied math, and writing have greatly influenced my philosophy as a researcher. It is difficult to live up to his reputation and I can only aspire to have an outstanding career that he has had and continuous to have. Mulțumesc din inimă.

My special committee members have also been indispensable in my graduate school journey. Prof. Christopher Earls introduced me to the field of UQ and I have sought his perspectives in this field many times, including book recommendations and possible career opportunities. His guidance is the reason why I am starting to find my direction in this area. Prof. Timothy Healey is a wonderful teacher and I owe my training in mathematical analysis to him. It was obvious from his notes how much effort he devoted in preparing for the lectures in the two classes I took with him. I would like to emulate his approach in structuring math classes when it's my time to teach in the future.

Outside Cornell, I am fortunate to also have mentors who have introduced me to topics complementary to my research. Dr. Kevin Carlberg hired me as an

intern at Sandia National Laboratories in the summer of 2016 without expecting how much stress I was going to cause him that time. He really pushed me to be more open-minded and to explore different directions in UQ. Until this day, I am still benefiting from the practical skills I acquired under his mentorship. I wouldn't have also finished my project without the support of my collaborators and fellow co-interns. In the summer of 2017, Prof. Elena Akhmatskaya and Dr. Tijana Radivojević offered me an intern position at the Basque Center for Applied Mathematics in Bilbao, Spain. There, I was able to hone my computational skills and I got a glimpse of how research centers in Europe operate. My research stay was also funded in part by travel grants from the Cornell Graduate School and the Mario Einaudi Center, for which I am indebted to.

The contents of this thesis are also in part due to the professors at Cornell who have imparted their knowledge through the classes I have taken with them and also to my colleagues at CAM and in Prof. Grigoriu's group who have provided me advice and with whom I have discussed my ideas. I am grateful to Danielle and Franz for taking care of me during my graduate studies.

Finally, I owe my profound gratitude to my parents and to my family for being immensely supportive during my time in graduate school. It was difficult for them to comprehend my research and the frustrations associated with an academic job since I am the first in my family to pursue this career path. It was also not until I was mature enough when I realized that my mother gave up her career to raise my siblings and I. It is because of her that I have a career today and I will always strive to compensate for her sacrifice. Know that I am also willing to give up my career for you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Uncertainty quantification is based on physical systems expressed as an input-output relationship where the objective is on understanding how uncertainty in the inputs affect uncertainty in the outputs, and vice versa. For example, consider a simple system given by the linear elliptic PDE

$$-\nabla \cdot (a(x,\xi) \cdot \nabla u(x,\xi)) = 0, \quad x \in D \subset \mathbb{R}^d \tag{1–1}$$

subject to boundary conditions. The coefficient field $a(x,\xi)$, and consequently the solution $u(x,\xi)$, is a function of the parameter $\xi \in \mathbb{R}^{n_\xi}$ and the spatial variable x. Problems in UQ often revolve around the mapping $Q : u(x,\xi) \to \mathbb{R}^{n_q}$ which computes functionals of the solution $u(x,\xi)$ representing quantities of interest $Q(\xi)$. These problems can be broadly classified into categories which include, but are not limited to, forward and backward UQ which will be the subject of the next 3 chapters of this thesis. We describe these categories in what follows.

Let ξ be a random vector defined on $(\Omega, \mathcal{F}, P)$. Forward UQ aims to approximate the probability law of $Q(\xi)$ given the law of ξ. Since Monte Carlo simulation is computationally expensive, surrogate models $\widetilde{Q}(\xi)$ are constructed that approximate $Q(\xi)$. The commonly used surrogates are stochastic Galerkin and stochastic collocation, however, [44] underscores the need for a new class of surrogate models. This is because the stochastic Galerkin method does not guarantee that $E[\widetilde{Q}(\xi)^p]$ converges to $E[Q(\xi)^p]$ for $p > 2$ while in stochastic collocation, the interpolation nodes are chosen without regards to the distribution of ξ. In Chapter 2, we consider the SROM-based surrogate, a collocation-type surrogate that is constructed by partitioning the range of ξ using Voronoi cells and computing a first-order Taylor expansion in each partition. It is shown that this sur-

rogate possesses comparable convergence properties as stochastic collocation but has significantly lower $L^p(\Omega)$ error for the same computational budget especially if the number of collocation nodes is large. The remainder of this chapter investigates the construction of an adaptive algorithm for the SROM-based surrogate that aims to address the following questions: How can the surrogate be constructed sequentially so that it rigorously targets high probability regions of $\xi(\Omega)$ where $Q(\xi)$ has strong nonlinearity? How should the collocation points be chosen and how must the partitons be refined between succeeding iterations? Mathematical analyses and computational experiments are conducted to examine these issues.

In contrast, backward UQ is the reverse of forward UQ which aims to identify the law of ξ given the law of $Q(\xi)$ for $n_q < n_\xi$. Unlike Bayesian inverse problems, the randomness here is due to the randomness in ξ, not from observation noise in $Q(\xi)$. This is ill-posed since distinct pdfs on ξ may yield the same pdf on $Q(\xi)$ when the former is propagated through $Q(\xi)$. In Chapter 3, a survey of recently proposed approaches is presented which include the disintegration theorem for probability measures [14] and an optimization approach based on the Karhunen-Loève expansion [10]. To compensate for the ill-posedness, these methods make assumptions about the probability distribution of ξ. We therefore construct examples which demonstrate that these assumptions may not guarantee the recovery of the true law of ξ and motivate the importance of such objective. The remainder of Chaper 3 then examines: what additional information on ξ must be specified in order to address the ill-posedness of this stochastic inverse problem? We illustrate how this inverse problem can be solved using tools such as optimization, principle of maximum entropy, and Bayes' theorem which incorporate this additional information.

Due to the ill-posedness in backward UQ, we conclude this thesis in Chapter 4 with a different inverse problem in which the distribution of ξ and $Q(\xi)$ are known. We desire to identify samples of the input random field that yield extreme response which can be useful in characterizing the distribution of ξ conditional on extreme events or to predict which samples of ξ cause such events. This objective offers a different perspective to existing literature in reliability which focuses on computing failure probabilities and also utilizes concepts introduced in Chapters 2 and 3. Existing work related to our objectives developed indicators which are functionals of $Q(\xi)$ that signal extreme events [26]. This chapter investigates the question: can indicators always be found, and if so, can they be validated and calibrated in a computationally efficient manner? As an alternative to using indicators, we design a general framework that consists of a multifidelity surrogate for $Q(\xi)$ and machine learning classifiers and demonstrate how their synergy addresses our objective.

CHAPTER 2

AN ADAPTIVE METHOD FOR SOLVING STOCHASTIC EQUATIONS BASED ON INTERPOLANTS OVER VORONOI CELLS

An adaptive collocation-based surrogate model is developed for the solution of equations with random coefficients, referred to as stochastic equations. The surrogate model is defined on a Voronoi tessellation of the samples of the random parameters with centers chosen to be statistically representative of these samples. We investigate various interpolants over Voronoi cells in order to formulate surrogates and analyze their convergence properties. Unlike Monte Carlo solutions, relatively small numbers of deterministic calculations are needed to implement surrogate models. These models can be used to generate large sets of solution samples with a minimum computational effort. In this work, we propose a framework for an adaptive construction of the surrogate such that by refining the Voronoi cells, the mapping between the random parameters and the solution is incorporated. A rigorous refinement measure which is quantitatively indicative of the performance of the surrogate is used to drive adaptivity. We present numerical examples that compare this surrogate with other collocation-based surrogates and demonstrate the theoretical aspects of the adaptive method.

2.1 Introduction

We consider the mapping $Z \mapsto U(x,t,Z)$ where Z represents a random vector defined on the probability space $(\Omega, \mathcal{F}, P)$ while x and t represent the spatial and temporal variables, respectively. Such mappings arise in applications, for instance problems in mechanics [13], in which the response $U(x,t,Z)$ solves the

forward stochastic equation $\mathcal{L}(U(x,t,Z)) = 0$ where the operator $\mathcal{L}$ typically characterizes a stochastic (partial) differential equation wherein Z encodes the uncertainty in the parameters of the physical model.

In order to quantify the behavior of the stochastic response $U(x,t,Z)$, probabilistic quantities of interest such as moments of U, distribution functions, probabilities of failure, etc., are sought after in applications. The only general method to accomplish this task is Monte Carlo. The method proceeds by generating a large number of samples of Z and solving the forward equation for each sample. As such, the implementation is straightforward and the computational cost does not scale with the dimension of the random vector Z. However, the method poses a computational burden for thousands of response calculations. To ameliorate the computational demand of Monte Carlo, surrogate models of the response $U(x,t,Z)$ have been introduced in which an approximation of the response is constructed in the image space of the random vector, $\Gamma = Z(\Omega) \in \mathbb{R}^d$. More specifically, a surrogate model is a mapping $Z \mapsto \widetilde{U}(x,t,Z)$ in which $\widetilde{U}(x,t,Z)$ approximates $U(x,t,Z)$ in some sense and for which samples of the response are cheaper to procure. We will drop the spatial and temporal variables for notational convenience in what follows and assume that $U(Z)$ is real-valued for simplicity.

Two general methods for constructing surrogate models for stochastic equations have flourished over the years, namely polynomial chaos or stochastic Galerkin method [74,75] and stochastic collocation method [13,24,32,43,46,55, 61,80]. Stochastic Galerkin method projects the response $U(Z)$ onto a set of polynomial basis functions which are orthogonal with respect to a probability measure. While the resulting polynomial surrogate $\widetilde{U}(Z)$ is convenient to evaluate,

$\widetilde{U}(Z)$ only converges to $U(Z)$ in $L^2(\Omega)$ and samples of $\widetilde{U}(Z)$ are prone to oscillatory behavior, cf. [30,31]. The lack of guarantee for $L^p(\Omega)$ convergence for $p > 2$ implies that $\widetilde{U}(Z)$ is impractical for use in extreme events modeling. In addition, even though $L^2(\Omega)$ convergence guarantees convergence in distribution, the rate of convergence of the latter can be slow so that truncated levels, which needs to be kept low in applications for computational reasons, may yield unsatisfactory approximations. Stochastic collocation, on the other hand, proceeds by selecting a finite number of collocation or interpolation nodes $z_k \in \mathbb{R}^d$ in $\Gamma = Z(\Omega)$, evaluating $U(z_k)$, and constructing an interpolant $\widetilde{U}(z)$ over Γ. Smolyak-based sparse grid collocation has been proposed [61,80] to overcome the curse of dimensionality when Z is high-dimensional. The performance of stochastic collocation hinges on the choice of interpolating functions and the location of the collocation nodes, aside from the regularity of the response $U(Z)$. It has been shown in [46] and in [61, Theorem 3.7] that under some assumptions on $U(Z)$, and for appropriate choices of the interpolant and collocation nodes, $\widetilde{U}(Z)$ converges to $U(Z)$ in $L^\infty(\Omega)$. A drawback of this method is that the quality of $\widetilde{U}(Z)$ may deteriorate as the number of collocation nodes increases [44] and that, except for [1, 24], the nodes are oftentimes selected without regard to the probability law of Z.

In an effort to improve the performance and accuracy of surrogate models, adaptive methods and domain decomposition techniques have been implemented which capture the behavior of the response. An adaptive procedure for stochastic collocation was used in [46] wherein collocation nodes are sequentially incorporated into the surrogate. The hierarchical surplus, which is the difference between the response and the interpolant at the preceeding iteration, is used as a criterion to determine which set of nodes will be prioritized at each

iteration. This surplus serves as an indicator of where $U(Z)$ exhibits substantial variation in Γ. Other variations in the indicator include the hierarchical surplus weighted by the probability density function value at a node as in [24] as well adjoint-based a posteriori error estimates in place of the surplus as in [43].

Domain decomposition techniques have also been applied in conjunction with adaptive methods to obtain local surrogate models in subdomains of the image of the random vector Z, denoted by Γ, thereby addressing the deficiencies of global surrogates. For example, the works [32, 74, 75] employed rectangular meshes to partition Γ after which polynomial chaos or stochastic collocation method is performed in each partition to obtain a local surrogate $\widetilde{U}_k(Z)$. Refinement criteria were then proposed to determine which partitions to refine. In these works, the decay rate of the variance of $\widetilde{U}_k(Z)$ was used as the refinement measure while partitions chosen to be refined are evenly split along the dimensions of Z which are deemed important. Performing such procedure in high dimensions, however, is challenging as refining a rectangular partition in this manner yields a large number of offspring partitions. Furthermore, rectangular partitions are not well-suited for partitioning non-hypercube domains Γ. In contrast to this type of domain decomposition, the authors in [77, 78] pursue a partitioning of Γ via a Delaunay triangulation based on Newton-Cotes nodes in Γ. The refinement criterion for a simplex partition is then based on a very crude approximation of the local $L^2(\Omega)$ error between $U(Z)$ and $\widetilde{U}_k(Z)$. A major difficulty of this approach in high dimensions is that $2^d + 1$ evaluations of the forward model are needed to interpolate over the initial triangulation, where d is the dimension of Z. We further remark that the refinement and stopping criteria considered in the aforementioned works do not clearly quantify the performance of the surrogate model being refined.

To address some of the difficulties associated with existing surrogate models, we employ a novel surrogate model $\widetilde{U}(Z)$ introduced in [40]. As in stochastic collocation, a finite number of interpolation nodes $z_k \in \Gamma$ is selected where the set $\{z_k\}$ constitutes a Stochastic Reduced Order Model (SROM) of the random vector Z. Unlike stochastic collocation, the set $\{z_k\}$ is chosen such that it captures statistical properties of Z. A Voronoi tessellation on the samples of Z is then constructed using $\{z_k\}$ as centers and this tessellation establishes a partitioning of the image $\Gamma \in \mathbb{R}^d$. On each Voronoi cell, a local surrogate model $\widetilde{U}_k(Z)$ is formulated by performing a Taylor expansion of $U(Z)$ at $Z = z_k$. As demonstrated in [40], $\widetilde{U}(Z) \to U(Z)$ in $L^p(\Omega)$ and almost surely under mild conditions, and that the computational cost scales linearly with the dimension of Z. In addition, the surrogate response $\widetilde{U}(Z)$ can be conveniently utilized to estimate probabilistic quantities of interest related to the response. We observe, however, that because the mapping $Z \mapsto U(Z)$ is not incorporated in partitioning Γ then two mappings $Z \mapsto V(Z)$ and $Z \mapsto W(Z)$ will yield surrogates with identical Voronoi tessellations of Γ for identical Z, thereby neglecting to account for the variations unique to each mapping.

We henceforth propose a framework for an adaptive method for the SROM-based surrogate described in [40] which takes into account the mapping $Z \mapsto U(Z)$ in the domain decomposition. The method sequentially builds a SROM $\{\widetilde{z}_k\}$ for Z that prioritizes regions of Γ with high probability and regions which manifest sharp variations and sensitivity with respect to Z. While most engineering applications are interested in the convergence of moments of $\widetilde{U}(Z)$, i.e. $|E[\widetilde{U}(Z)^p] - E[U(Z)^p]| \to 0$, we focus our efforts on $L^p(\Omega)$ convergence which is stronger. Consequently, we propose refinement and stopping criterion for the adaptive method based on the L^p error and these criteria are quantita-

tively indicative of the performance and accuracy of $\widetilde{U}(Z)$ for a given computational budget. The surrogate model is constructed via various interpolants over Voronoi cells and we analyze the convergence properties of these interpolants.

In what follows, we first provide a review on collocation-based surrogates, namely stochastic collocation and the SROM-based surrogate. A comparison of the L^p error of these methods under varying stochastic dimension is presented via numerical examples. It is shown that stochastic collocation can underperform as it does not take into account the probability law of Z. We then elaborate on the proposed adaptive approach for SROM-based surrogates, address issues of implementation, and investigate its mathematical properties. The benefits of using the adaptive SROM-based surrogate, in which the probability law of Z is used to determine the location of the next SROM node, over the regular SROM-based surrogate is numerically demonstrated by comparing their L^p errors. Computationally, the two constructions differ in that the adaptive method has more evaluations of $U(Z)$, which is relatively inexpensive. The comparison was performed using examples with large stochastic dimension for test response functions and for a response derived from a stochastic PDE. Furthermore, we observe that the gap between the L^p errors of the adaptive and the regular SROM-based surrogate increases as p increases.

2.2 Collocation-based surrogates

We consider two types of collocation-based surrogates, namely stochastic collocation and the SROM-based surrogate. Both surrogates are constructed using a set of collocation nodes and interpolating functions. The convergence proper-

ties of these surrogates are examined and their construction and implementation are compared using numerical examples.

2.2.1 Stochastic collocation

Stochastic collocation method [46–48, 61, 80] constructs an interpolant $\widetilde{U}(Z)$ of $U(Z)$ over Γ. The process can be summarized in the following steps:

1. Select collocation points $\mathbf{z}_m \in \Gamma \subset \mathbb{R}^d$, $m = 1, \ldots, n$ and evaluate $U(\mathbf{z}_m)\ \forall m$.
2. Construct an interpolating function $\ell_m(Z)$ for each node $\mathbf{z}_m$ such that $\ell_m(\mathbf{z}_m) = 1$.
3. Formulate the surrogate as $\widetilde{U}_n(Z) = \sum_{m=1}^{n} U(\mathbf{z}_m)\ell_m(Z)$.

More specifically, suppose that $Z(\Omega) = \Gamma = [a_1, b_1] \times \cdots \times [a_d, b_d] \subset \mathbb{R}^d$. Denote by $Z^i = \{z_1^i, \ldots, z_{m_i}^i | z_k^i \in [a_i, b_i]\ \forall\, k\}$ the 1-dimensional set of collocation nodes in the i-th coordinate axis and let ℓ_k^i be the one-dimensional interpolating function corresponding to the node z_k^i, i.e. $\ell_k^i(z_k^i) = 1$. An interpolant for $U(Z)$ can be constructed as:

$$\widetilde{U}(Z) = \sum_{k_1=1}^{m_i} \cdots \sum_{k_d=1}^{m_d} \ell_{k_1}^1(z_{k_1}^1) \cdots \ell_{k_d}^d(z_{k_d}^d) \cdot U(z_{k_1}^1, \ldots, z_{k_d}^d) \tag{2–1}$$

which requires $m_1 \times \cdots \times m_d$ collocation nodes given by the Cartesian product $Z^1 \times \cdots \times Z^d$, i.e. the full grid.

Because (2–1) suffers from the curse of dimensionality, sparse grids offer a more practical approach as the response is only evaluated at a subset of the full grid. An example of a 3-dimensional sparse grid using Newton-Cotes equidistant nodes in each coordinate axis is shown in Figure 2.1. The cardinality of

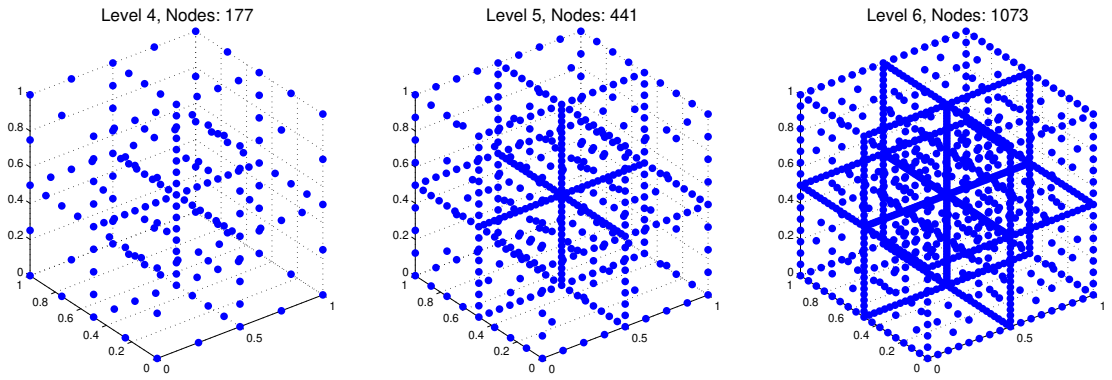


Figure 2.1: Illustration of sparse grid using Newton-Cotes equidistant nodes for $d = 3$ for different sparse grid levels. The nodes from a preceding level is a subset of the nodes in the succeeding level.

this subset can be significantly lower than that of the full grid in high dimensions. The interpolant over this sparse grid is built through Smolyak's algorithm which involves taking the tensor product of one-dimensional interpolating functions in a special manner. Unlike (2–1), the sparse grid interpolant is constructed iteratively from one sparse grid level to another but retains the same functional form. The differences are that the summation is now over the sparse grid nodes and that the weight of each basis function is the surplus between $U(Z)$ and the sparse grid interpolant in the preceding sparse grid level. An elaborate treatment of Smolyak's algorithm and sparse grid stochastic collocation is presented in [46] and the references therein.

If $\widetilde{U}$ is the sparse grid interpolant constructed using linear hat basis functions and nested Newton-Cotes equidistant nodes in each coordinate axis, the interpolation error is [46]:

$$||U - \widetilde{U}||_{L^\infty} = O(N^{-2} \cdot |\log_2 N|^{3(d-1)})$$

where N is the number sparse grid nodes assuming that $U(z) \in C^2(\Gamma)$.

In 1-dimension, i.e. $d = 1$, the commonly used interpolating functions are the

Lagrange basis functions and linear hat basis functions. For Lagrange interpolation, if $U \in C^n([a,b])$, $U^{(n+1)}(z)$ needs to be bounded on $[a,b]$ to guarantee uniform convergence. Because of this stringent regularity condition, the Lagrange interpolant can be prone to oscillatory behavior especially if the collocation points z_m are not dense at the boundary of Γ [6]. As such, collocation points such as Gauss-Legendre nodes or Clenshaw-Curtis nodes have to be employed. On the other hand, linear hat basis functions have local support and only require bounded second derivatives to ensure uniform convergence. Newton-Cotes equidistant nodes on $[a,b]$ (with endpoints included) is the most commonly used set of collocation points.

In summary, the key ingredients for stochastic collocation are the choice of interpolating functions and the choice of collocation nodes. The selection of collocation nodes is usually performed without regard for the probability law of Z. Recent works such as [1, 24] used Newton-Cotes equidistant nodes and utilized the probability density function of Z to determine where to add more collocation points in Γ. However, the resulting (sparse) grid is still structured and rectangular. To overcome these concerns, we introduce a different class of collocation-based interpolants that incorporate the probability law of Z into the construction of $\widetilde{U}(Z)$ and are well-suited for non-rectangular grids and for non-hypercube domains.

2.2.2 SROM-based surrogates

An SROM-based surrogate model $\widetilde{U}(Z)$ is a collocation-based surrogate for the stochastic response $U(Z)$. As with stochastic collocation, its construction entails

the selection of collocation nodes and the selection of the interpolating function. We now elaborate how the collocation nodes can be chosen using the concept of an SROM.

Consider a random vector $Z \in \mathbb{R}^d$ defined on the probability space $(\Omega, \mathcal{F}, P)$. A Stochastic Reduced Order Model (SROM) of size m is a discrete random vector $\widetilde{Z}$ which takes on values from the set $\{\widetilde{z}_k\}_{k=1}^m \subset \Gamma \subset \mathbb{R}^d$ with corresponding probabilities $\{p_k\}_{k=1}^m$. $\widetilde{Z}$ is defined on the same probability space as Z with $\{\widetilde{z}_k\}$ and $\{p_k\}$ chosen such that $\widetilde{Z}$ is statistically representative of Z. For instance, $\widetilde{Z}$ can be constructed such that the first two moments and the distribution of $\widetilde{Z}$ closely resemble that of Z. Different methods on the construction of SROMs are elaborated in [36, pp. 464-474] and the references therein. The non-uniqueness of the Stochastic Reduced Order Model is not an issue since we are only interested in obtaining samples which capture the statistics of the target random vector Z.

With the concept of an SROM for Z defined, an SROM-based surrogate model $\widetilde{U}(Z)$ can then be constructed using interpolants over Voronoi cells. In what follows, we consider Taylor-based interpolants and the Sibson's interpolant.

2.2.2.1 Taylor-based interpolants

The construction of an SROM-based surrogate model $\widetilde{U}(Z)$ using Taylor interpolants can be summarized in the following steps [40].

1. Generate a large number N of independent samples $\{z_k\}_{k=1}^N \subset \Gamma$ of Z.

2. For a specified m, $m \ll N$, construct an SROM $\widetilde{Z} = \{\widetilde{z}_k\}_{k=1}^m \subset \Gamma$ of size m for Z.

3. Partition the set $\{z_k\}_{k=1}^N$ into m subsets $\{\Gamma_j\}_{j=1}^m$ defined by $\Gamma_j = \{z \in \Gamma : \|z - \widetilde{z}_j\| < \|z - \widetilde{z}_l\| \text{ for } l = 1, \ldots, m, l \neq j\}$. If $\|z_k - \widetilde{z}_j\| = \|z_k - \widetilde{z}_l\|$, we can arbitrarily assign z_k to Γ_j or Γ_l. This step essentially performs a (discrete) Voronoi tessellation of Γ with the $\{\widetilde{z}_k\}_{k=1}^m$ as the centers. We therefore refer to Γ_j as Voronoi cells with $\cup_{j=1}^m \Gamma_j = \Gamma$.

4. On each Voronoi cell Γ_j, perform a zeroth-order or first-order Taylor expansion of Z at $\widetilde{z}_j$ which gives rise to

$$\widetilde{U}_m(Z) = \sum_{k=1}^{m} \mathbb{1}_{(Z \in \Gamma_k)} U(\widetilde{z}_k) \tag{2–2}$$

or

$$\widetilde{U}_m(Z) = \sum_{k=1}^{m} \mathbb{1}_{(Z \in \Gamma_k)} [U(\widetilde{z}_k) + \nabla U(\widetilde{z}_k) \cdot (Z - \widetilde{z}_k)], \tag{2–3}$$

respectively, where $\nabla U(\widetilde{z}_k)$ denotes the gradient of $U(z)$ at $z = \widetilde{z}_k$.

An illustration of the resulting surrogate model is presented in Example 1 and Figure 2.2.

Example 1. Let $Z = (Z_1, Z_2)$ where $Z_1 \sim F_1^{-1}(\Phi(Y_1)), Z_2 \sim F_2^{-1}(\Phi(Y_2))$ with Φ being the standard normal CDF, $Y_1, Y_2 \sim N(0,1), \text{cov}(Y_1, Y_2) = 0.3$ and F_1, F_2 are the CDFs of Beta$(2,6)$ and Beta$(6,6)$, respectively. Suppose that we want to construct a surrogate for $U(Z) = \sin(Z_1 + 3(Z_2 - 0.2))$. The left plot of Figure 2.2 shows an SROM $\widetilde{Z}$ of size $m = 15$ and the right plot exhibits the response function $U(Z)$ and the surrogate model. We remark that the nodes of $\widetilde{Z}$ are concentrated in regions of Z with high probability.

In sum, an SROM-based surrogate as developed in [40] uses an SROM of size m for Z as its collocation nodes while the interpolating functions are piece-

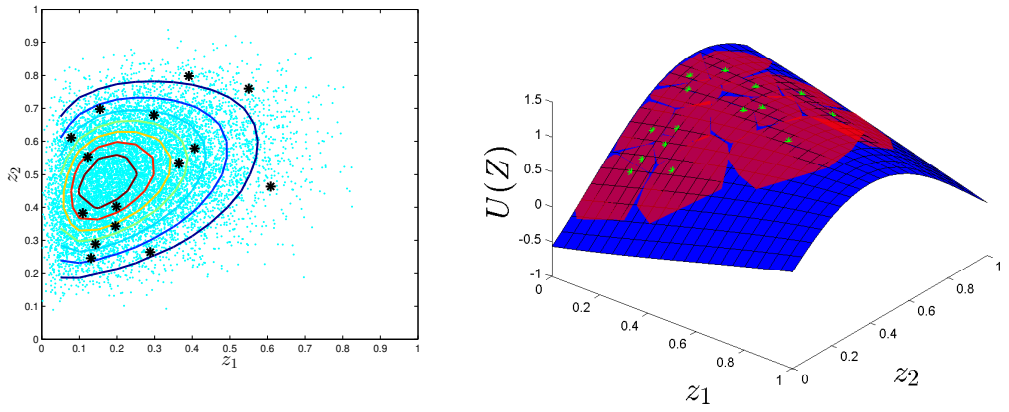


Figure 2.2: Left – SROM $\widetilde{Z}$ (black asterisks) for Z in Example 1 with samples of Z (blue dots) and contour of the joint pdf of Z. Right – Illustration of the surrogate $\widetilde{U}_m(Z)$ (red surface) for $U(Z)$ (blue surface) using first-order Taylor interpolants for $m = 15$ where Z is defined as in Example 1. The green nodes are the SROM nodes.

wise constant functions in each Voronoi cell as in (2–2) or, as shown in Figure 2.2 (right), piecewise tangent hyperplanes in each Voronoi cell as in (2–3). The interpolating functions are chosen as such in order to accommodate the fact that the collocation points no longer lie on a rectangular grid. We note that higher-order Taylor expansions can be used as interpolants over each Voronoi cell. However, this would require stringent regularity conditions of $U(Z)$ to guarantee convergence of $\widetilde{U}_m(Z)$ as we shall see below. The computational cost would also scale non-linearly to compute the higher-ordered partial derivatives, making it impractical in high dimensions.

We now briefly remark on some computational aspects involving the SROM-based surrogate and subsequently review its important properties as presented in [40]. We focus on the SROM-based surrogate using 1st-order Taylor interpolants because the surrogate in (2–2) is a coarse approximation as it does not adequately capture the variation of $U(Z)$. Hence, $\widetilde{U}_m(Z)$ in what follows refers to (2–3). Using $\widetilde{U}_m(Z)$, statistical information of $U(Z)$ can now be approximated

at a cheaper computational cost. For example, the p-th order moments of $U(Z)$ can be estimated as follows [36, pp. 470]:

$$E[U(Z)^p] \simeq E[\widetilde{U}_m(Z)^p] \simeq \sum_{k=1}^{m} \frac{n_k}{N} \sum_{z_j \in \Gamma_k} \frac{1}{n_k} (U(\widetilde{z}_k) + \nabla U(\widetilde{z}_k) \cdot (z_j - \widetilde{z}_k))^p$$

where n_k is the cardinality of the set $\{j | z_j \in \Gamma_k\}$. Evidently, it is substantially cheaper to evaluate $U(\widetilde{z}_k) + \nabla U(\widetilde{z}_k) \cdot (z_j - \widetilde{z}_k)$ compared to the true response. The fact that $E[U(Z)^p]$ can be estimated by $E[\widetilde{U}_m(Z)^p]$ can be justified through the properties below.

Theorem 1. *Let $\widetilde{U}_m(Z)$ be as in* (2–3)*. Suppose that $Z(\Omega) = \Gamma \subset \mathbb{R}^d$ is an open convex domain such that $U(Z) \in C^2(\Gamma)$. Furthermore, assume that the second-order partial derivatives of $U(Z)$ are bounded, i.e. $\left|\frac{\partial^2 U(z)}{\partial z^{(r)} \partial z^{(s)}}\right| \leq M$ for all $r, s = 1, \ldots, d, z \in \Gamma$. For $z \in \Gamma_k$, where Γ_k is a Voronoi cell of Γ with center $\widetilde{z}_k$, we then have that for some constant C_k*

$$|U(z) - \widetilde{U}_m(z)| \leq \frac{C_k}{2} \|z - \widetilde{z}_k\|^2. \tag{2–4}$$

Proof. Let $z \in \Gamma_k$. By Taylor's remainder theorem, cf. [3],

$$U(z) - \widetilde{U}_m(z) = \frac{1}{2} (z - \widetilde{z}_k)^T \; H(U(\widetilde{\xi}_k)) \; (z - \widetilde{z}_k)$$

where $\widetilde{\xi}_k$ is a point on the line segment joining z and $\widetilde{z}_k$ and $H(U(\widetilde{\xi}_k))$ is the Hessian matrix of $U(z)$ evaluated at $z = \widetilde{\xi}_k$. The symmetry of the Hessian matrix implies that $H(U(\widetilde{\xi}_k)) = Q^T \Lambda Q$ where Q is an orthogonal matrix, Λ is a diagonal matrix with entries $\lambda_i(\widetilde{\xi}_k)$, and that Q, Λ depend on $\widetilde{\xi}_k$. Hence, if $y := Q \cdot (z - \widetilde{z}_k)$,

$$\begin{aligned} |U(z) - \widetilde{U}_m(z)| &= \frac{1}{2} |y^T \Lambda y| = \frac{1}{2} \left| \sum_{i=1}^{d} \lambda_i(\widetilde{\xi}_k) \, y_i^2 \right| \\ &\leq \frac{1}{2} \left(\max_i |\lambda_i(\widetilde{\xi}_k)| \right) \|y\|^2 \leq \frac{C_k}{2} \; \|z - \widetilde{z}_k\|^2 \end{aligned} \tag{2–5}$$

where C_k is a bound on $\max_i |\lambda_i(\widetilde{\xi}_k)|$ owing to the fact that the entries of the Hessian matrix are bounded. □

Using Theorem 1, we can easily prove convergence properties of the SROM-based surrogate.

Corollary 2. *Suppose that the second-order partial derivatives of $U(Z)$ are bounded. Consider a refining sequence of SROMs for Z, that is, if $\widetilde{Z}_m = \{\widetilde{z}_k\}_{k=1}^m$ and $\widetilde{Z}_{m+1} = \{\widetilde{z}_k\}_{k=1}^{m+1}$ are SROMS for Z, $\widetilde{Z}_m \subset \widetilde{Z}_{m+1}$. We then have that $\widetilde{U}_m(Z) \to U(Z)$ almost surely as $m \to \infty$. In addition, if $\exists\, p \geq 1$ such that $U(Z) \in L^p(\Omega)$ then $\widetilde{U}_m(Z)$ also converges to $U(Z)$ in $L^p(\Omega)$ [40].*

Proof. See [40], p. 273. □

Corollary 2 illustrates how the SROM-based surrogate (2–3) possesses strong convergence properties under very mild conditions. Because a.s. and L^p convergence each imply convergence in distribution, we have that for every $u \in \mathbb{R}$, $P(\widetilde{U}_m(Z) \leq u) \to P(U(Z) \leq u)$ as $m \to \infty$. In addition, convergence in L^p implies convergence in p^{th} order moments, i.e. $E[\widetilde{U}_m(Z)^p] \to E[U(Z)^p]$, which is often of interest in engineering applications. Hence, the convergence properties of (2–3) are comparable to that of the sparse grid stochastic collocation using linear hat basis and Newton-Cotes equidistant nodes under identical regularity conditions on $U(Z)$.

2.2.2.2 Sibson-based interpolant

We extend the class of interpolants introduced in [40] for the SROM-based surrogate by considering an interpolant from geoscience. Subsequently, we will state and prove properties of this interpolant and comment on its implementation.

Suppose that we have at our disposal $\{(\widetilde{z}_k, U(\widetilde{z}_k))\}_{k=1}^m$ where $\widetilde{z}_k \in \mathbb{R}^d$ and that we are concerned with approximating the mapping $Z \mapsto U(Z)$ on $\Gamma \subset \mathbb{R}^d$ where Γ is the interior of the convex hull of the nodes $\{\widetilde{z}_k\}_{k=1}^m$. This can be accomplished using Sibson's interpolation [68] in which the value of the interpolant at a new point $z^* \in \Gamma$ is obtained as a weighted sum of $U(\widetilde{z}_k)$ for $\widetilde{z}_k$ close to z^*. The weights are computed by determining how much of the volume in each Voronoi cell of the original tessellation generated using $\{\widetilde{z}_k\}_{k=1}^m$ as the centers is lost upon retessellation with $z^* \cup \{\widetilde{z}_k\}_{k=1}^m$ as the new centers.

More formally, assume that we have a Voronoi tessellation of $\mathbb{R}^d$, not Γ, denoted by $\{\Gamma_k\}_{k=1}^m$ with $\{\widetilde{z}_k\}_{k=1}^m$ as the centers. Two Voronoi cells Γ_i and Γ_j are considered neighbors if Γ_i and Γ_j have a common boundary. For a given $z^* \in \Gamma$ the value of $U(z^*)$ can be approximated by a surrogate $\widetilde{U}_m(z^*)$ as follows:

1. If $z^* \in \{\widetilde{z}_k\}_{k=1}^m$ then $\widetilde{U}_m(z^*) = U(\widetilde{z}_k)$ for some $k \leq m$.
2. Otherwise, construct a Voronoi tessellation $\{\widetilde{\Gamma}_k\}_{k=1}^m \cup \{\widetilde{\Gamma}^*\}$ of $\mathbb{R}^d$ with $\{\widetilde{z}_k\}_{k=1}^m \cup \{z^*\}$ as the corresponding centers.
3. For each $k = 1, \ldots, m$, compute

$$w_k(z^*) := \frac{\lambda(\Gamma_k \cap \widetilde{\Gamma}^*)}{\lambda(\widetilde{\Gamma}^*)} \tag{2–6}$$

   where $\lambda(\cdot)$ is the Lebesgue measure. The numerator of (2–6) represents how much volume from Γ_k is lost and absorbed into $\widetilde{\Gamma}^*$ upon the inclusion of z^* as a Voronoi center. We remark that $\lambda(\Gamma_k \cap \widetilde{\Gamma}^*)$ and $\lambda(\widetilde{\Gamma}^*)$ are well-defined because $\widetilde{\Gamma}^*$ is a bounded region owing to the fact that z^* is in the interior of the convex hull.
4. Set $\widetilde{U}_m(z^*) = \sum_{k=1}^m w_k(z^*) U(\widetilde{z}_k)$ to be the value of the surrogate at z^*.

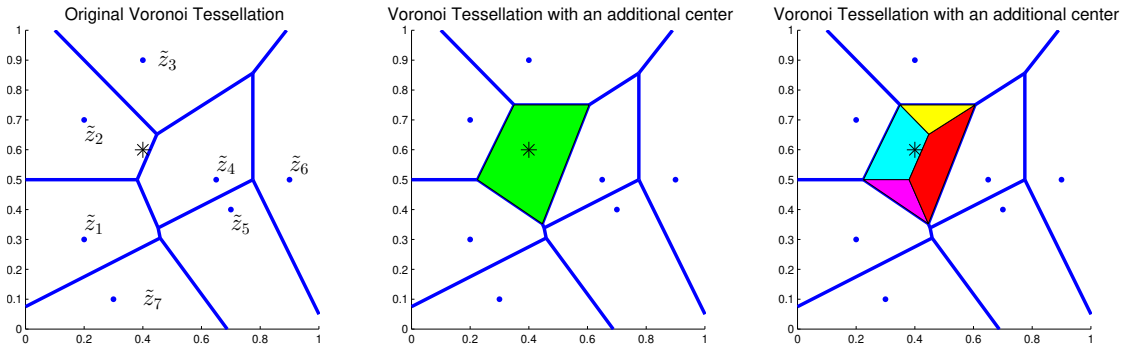


Figure 2.3: Illustration of the computation of the weights in Sibson's interpolant.

An example of how the weights in Sibson's interpolant are computed is illustrated in Figure 2.3. The nodes $\{\widetilde{z}_k\}$ are denoted by the blue dots with their corresponding Voronoi cells in the leftmost plot while the asterisk represents z^*. The updated Voronoi tessellation after inserting z^* as a Voronoi center is shown in the middle plot which corresponds to step 2 in the above procedure. The weights $w_k(z^*)$ for $\widetilde{z}_k$, $k = 1, \ldots, 4$, can be computed as the ratio of the area of the respective colored regions $(\Gamma_k \cap \widetilde{\Gamma}^*)$ in the rightmost plot with the area of the green colored region $(\widetilde{\Gamma}^*)$ in the middle plot. Furthermore, $w_k(z^*) = 0$ for $k \geq 5$ which demonstrates that Sibson's interpolant only uses the values at the neighboring nodes.

Sibson's interpolant possesses a number of properties that have been stated and proven in [28, 65, 67]. We review some of them here.

Property 3 (Sibson 1980)**.** *The functions $w_k(z^*)$ satisfy the local coordinates property in that*

$$z^* = \sum_{k=1}^{m} w_k(z^*)\widetilde{z}_k$$

for all $z^* \in \Gamma \subset \mathbb{R}^d$.

Property 4. *The interpolant $\widetilde{U}_m(z^*)$ has linear precision. In other words, if the true response U is linear, $U(z^*) = \widetilde{U}_m(z^*)$ $\forall z^* \in \Gamma$.*

The second property easily follows from the local coordinates property. Indeed, if U is linear,

$$U(z^*) = U\left(\sum_{k=1}^{m} w_k(z^*)\widetilde{z}_k\right) = \sum_{k=1}^{m} w_k(z^*)U(\widetilde{z}_k) = \widetilde{U}_m(z^*).$$

Property 5. *$\widetilde{U}_m(z^*)$ is continuous for all $z^* \in \Gamma$ and is continuously differentiable for all $z^* \in \Gamma$, $z^* \notin \{\widetilde{z}_k\}_{k=1}^m$.*

Property 6. *For $d = 1$, i.e. in the 1-dimensional case, Sibson's interpolant and collocation with linear hat basis functions coincide. For $d = 2$, it has been shown in [28] that Sibson's interpolant has gridded bilinear precision. In other words, suppose that $\{\widetilde{z}_k\}_{k=1}^m$ lie on a rectangular lattice, i.e. $\{\widetilde{z}_k\}_{k=1}^m$ are the Newton-Cotes equidistant nodes in $\Gamma \subset \mathbb{R}^2$. Sibson's interpolant then coincides with collocation using $\{\widetilde{z}_k\}_{k=1}^m$ and linear hat basis functions.*

Based on the above property, Sibson's interpolant is a generalization of collocation using linear hat basis functions for arbitrarily located collocation nodes, for any dimension. Hence, using this interpolant, the nodes can be chosen in accordance with the probability law of Z. However, the properties above are only valid in the interior of the convex hull of $\{\widetilde{z}_k\}_{k=1}^m$. Efforts to extend the validity of the interpolant beyond the interior have been undertaken as in [8, 9]. These approaches introduce a different or additional geometric construction and we do not pursue them here for simplicity.

It is now straightforward to incorporate Sibson's interpolant into the SROM-based surrogate model – we simply let the $\{\widetilde{z}_k\}_{k=1}^m$ above to be an SROM for the random vector Z. Subsequently, we analyze the convergence properties of Sibson's interpolant when employed to construct surrogate models for a stochastic response $U(Z)$. In the following, we will assume that Γ is the interior of the convex hull of $\{\widetilde{z}_k\}_{k=1}^m$.

Proposition 7. *Suppose that $U(Z)$ is continuous on Γ. Consider a refining sequence of SROMs for Z. We then have $\widetilde{U}_m(Z) \to U(Z)$ as $m \to \infty$.*

Proof. Let $\epsilon > 0$ be given and let $\{\widetilde{z}_k\}_{k=1}^m \subset \Gamma$ be the SROM nodes which partition Γ into the corresponding Voronoi cells $\{\Gamma_k\}_{k=1}^m$. For $z^* \in \Gamma$, we can always choose m to be large enough so that $|U(\widetilde{z}_k) - U(z^*)| < \epsilon, \ \forall\, k$ such that $w_k(z^*) \neq 0$ due to the continuity of $U(Z)$ at $Z = z^*$. By noting that $\sum_{k=1}^m w_k(z^*) = 1$, the discrepancy can be bounded by

$$\begin{aligned}\left|\widetilde{U}_m(z^*) - U(z^*)\right| &= \left|\sum_{k=1}^m w_k(z^*)U(\widetilde{z}_k) - U(z^*)\right| = \left|\sum_{k=1}^m w_k(z^*)(U(\widetilde{z}_k) - U(z^*))\right| \\ &\leq \sum_{k=1}^m w_k(z^*)|U(\widetilde{z}_k) - U(z^*)| < \epsilon.\end{aligned}$$

□

Proposition 8. *Suppose that $U(Z)$ is differentiable on Γ and that $\left|\frac{\partial U(z)}{\partial z^{(i)}}\right| \leq M$ for $i = 1, \ldots, d$ and $z \in \Gamma$. Consider a refining sequence of SROMs for Z. We then have that for $p \geq 1$, if $U(Z) \in L^p(\Omega)$, $\widetilde{U}_m(Z) \to U(Z)$ in $L^p(\Omega)$ as $m \to \infty$.*

Proof. By the Mean Value Theorem,

$$|U(\widetilde{z}_k) - U(z)| \leq \|\nabla U(\xi)\| \cdot |\widetilde{z}_k - z\| \leq \sqrt{d}M|\widetilde{z}_k - z\|$$

where ξ lies on the line segment joining z and $\widetilde{z}_k$. As a consequence,

$$\begin{aligned}\left|\widetilde{U}_m(z) - U(z)\right| &\leq \sum_{k=1}^m w_k(z)|U(\widetilde{z}_k) - U(z)| \leq \sqrt{d}M \max_{k,\, w_k \neq 0} |\widetilde{z}_k - z\| \\ &\leq \sqrt{d}M \max_{z \in \Gamma} \max_{k,\, w_k \neq 0} |\widetilde{z}_k - z\|\end{aligned}$$

where $\max_{z \in \Gamma} \max_{k,\, w_k \neq 0} |\widetilde{z}_k - z\| \to 0$ as $m \to \infty$, i.e. for $\epsilon > 0$, $\exists N$ such that $\max_{z \in \Gamma} \max_{k,\, w_k \neq 0} |\widetilde{z}_k - z\| < \epsilon$ for $m \geq N$. The claim now holds true if $p = \infty$. On the other hand,

$$\|\widetilde{U}_m(Z) - U(Z)\|_{L^p(\Omega)}^p = \int_\Gamma |\widetilde{U}_m(z) - U(z)|^p \, dF(z) < (\sqrt{d}M\epsilon)^p.$$

□

Remark. The proof above demonstrates that the L^p error of the SROM-based Sibson's interpolant is bounded by the size of the Voronoi cell as in the case of the SROM-based surrogate models introduced earlier. It is also straightforward to show that the conditions on differentiability and bounded derivatives of $U(Z)$ in Proposition 8 can be relaxed by assuming that $U(Z)$ is uniformly continuous on Γ.

In addition to its convergence properties, the SROM-based Sibson's interpolant guarantees that the surrogate does not attain unrealistic values:

Property 9. *If* $U_{min} = \min_{z\in\Gamma} U(z)$ *and* $U_{max} = \max_{z\in\Gamma} U(z)$, *then* $P(\widetilde{U}_m(Z) \in [U_{min}, U_{max}]) = 1$.

This property easily follows from the fact that $\sum_{k=1}^{m} w_k(z) = 1$.

This is important in applications in which it is preferable that the surrogate respects the physics of the model aside from guaranteeing convergence.

Despite the flexibility offered by Sibson's interpolant, its limitations are clear: the computation of the weights w_k poses difficulties in high stochastic dimension, a challenge that is not encountered in the 2-d or 3-d applications in computational geometry, geoscience, etc. For one, volumes of Voronoi cells need to be approximated. Since determining the exact boundaries of Voronoi cells might be impractical in high dimensions, a counting process can instead be utilized to estimate volumes through Monte Carlo integration. Secondly, for z^* close to the boundary of the convex hull of $\{\widetilde{z}_k\}_{k=1}^m$, the boundaries of the resulting Voronoi cell $\widetilde{\Gamma}^*$ can extend well beyond the interior of the convex hull Γ even though its Lebesgue measure is guaranteed to be finite. Consequently, a "bounding box" encompassing Γ is used in applications to impose bounds on the extent of $\widetilde{\Gamma}^*$,

making it more feasible to estimate its volume. The accuracy of this interpolant therefore depends on the size of the bounding box as well as the number of samples for the counting process.

2.2.3 Comparison of collocation-based surrogates

After providing a survey of collocation-based surrogates, a comparison of their key features is summarized in Table 2.1.

	SROM-based			Collocation (linear hat)	
	0th-order Taylor	1st-order Taylor	Sibson	Full grid	Sparse grid
Gradient needed?	✗	✓	✗	✗	✗
Exact for linear response?	✗	✓	✓	✓	✓
Is $\widetilde{U}_m(z)$ within bounds?	✓	✗	✓	✓	✗

Table 2.1: Comparison of collocation-based surrogates.

In this section, we are primarily interested in investigating the effect of the choice of collocation nodes through a comparison of the SROM-based surrogate with sparse grid stochastic collocation for varying stochastic dimension and for the same computational budget. The 1st-order Taylor interpolant will be used for the SROM-based surrogate while Newton-Cotes equidistant nodes and linear hat basis function will be used for sparse grid stochastic collocation. We have chosen to focus on the 1st-order Taylor interpolant because it provides a balance between accuracy and ease of implementation. The specification of the response is outlined in the following example.

Example 2. Consider the stochastic response $U(Z) = \arctan(15 \cdot \|Z - 0.7\|^2) -$

$\arctan(15 \cdot \|Z - 0.3\|^2)$, $Z \in \mathbb{R}^d$, where $Z_i \sim F^{-1}(\Phi(Y_i))$ with Φ being the standard normal CDF, $Y_i \sim N(0,1), \text{cov}(Y_i, Y_j) = 0.85$ for $i \neq j$ and F_i is the CDF of Beta(3, 3). An illustration of the probability law of Z and a plot of $U(z)$ for $z \in [0,1] \times [0,1]$ is shown in Figure 2.4 for $d = 2$.

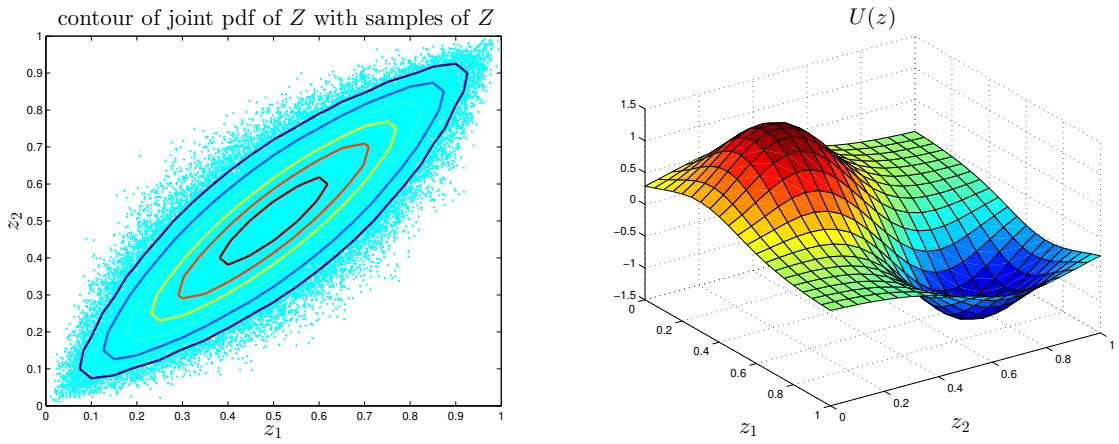


Figure 2.4: Joint pdf of Z (left) and $U(z)$ (right).

Figures 2.5, 2.6, 2.7, exhibit the L^p errors $\|U(Z) - \widetilde{U}(Z)\|_{L^p(\Omega)}$ for $d = 2, 4, 6$, respectively, obtained via Monte Carlo approximations. For each figure, the blue solid curve (denoted by SpGrid) and the red dashed curve (denoted by SROM) are the L^p errors for the sparse grid surrogate and the SROM-based surrogate, respectively. Each figure contains 4 subplots which correspond to $p = 1, 2, 3, \infty$. The x-axis of each subplot refers to the sparse grid level which is characterized by the number of evaluations of $U(z)$ required to form the interpolant. Hence, for each sparse grid level with n_l nodes, $\lfloor n_l/(d+1) \rfloor$ SROM nodes are used to ensure that the number of computational units between both surrogates is similar.

As the L^p error comparison plots demonstrate for this example, sparse grid stochastic collocation outperforms SROM-based surrogate for $d = 2$ even though the sparse grid nodes do not align with the high probability regions of Z. In Figure 2.8, we show the location of the sparse grid nodes for the 8^{th}

sparse grid level (characterized by the location of the sparse grid nodes as in Figure 2.1) and the corresponding SROM nodes for the same computational budget, together with the contour of the response. It is therefore not surprising that the sparse grid collocation perform well because the nodes are adequately distributed in the domain. However, this trend is reversed as the stochastic dimension is increased, as can be seen in Figures 2.6 and 2.7. For skewed distributions for Z in high dimensions, a large sparse grid level, which implies a large computational expense, is necessary for the sparse grid nodes to capture the most probable regions of Z. This numerical example thus underscores the importance of incorporating the probability law of Z into the location of the collocation nodes.

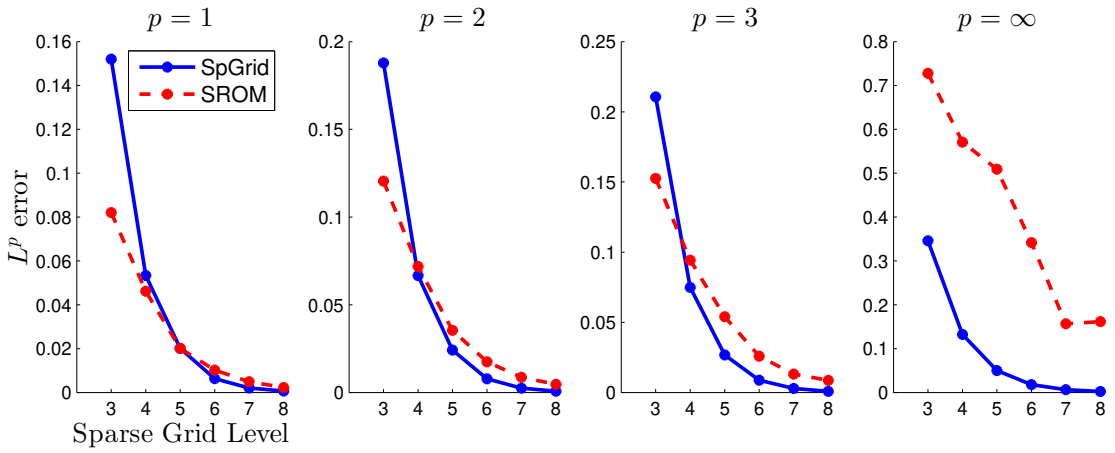


Figure 2.5: L^p error comparison of SROM-based surrogate and sparse grid stochastic collocation for $d = 2$. Sparse grid level/Number of Nodes: 4/65, 6/321, 8/1537.

2.3 Adaptive SROM-based surrogate

We now focus on the main contribution of this work. While the SROM-based surrogate is capable of capturing the probability law of Z, we can further extend

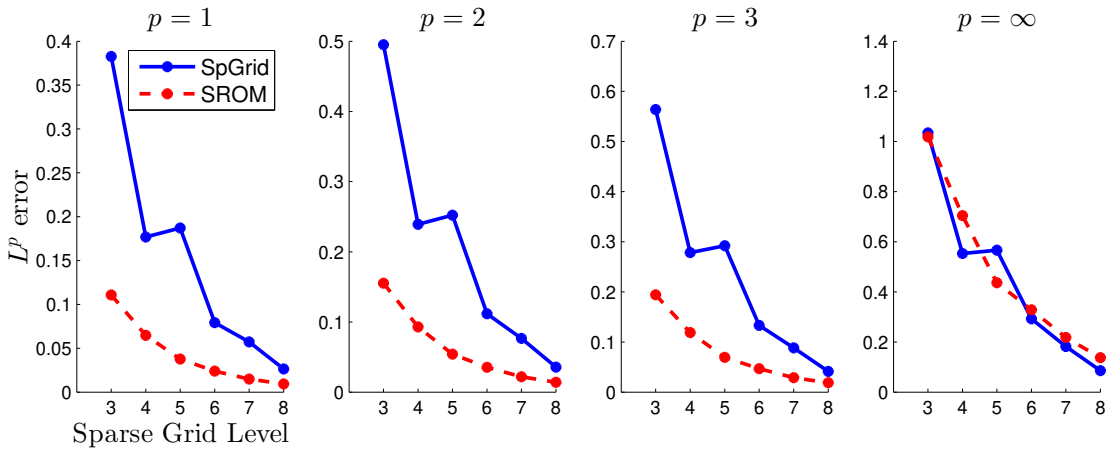


Figure 2.6: L^p error comparison of SROM-based surrogate and sparse grid stochastic collocation for $d = 4$. Sparse grid level/Number of Nodes: 4/401, 6/2929, 8/18945.

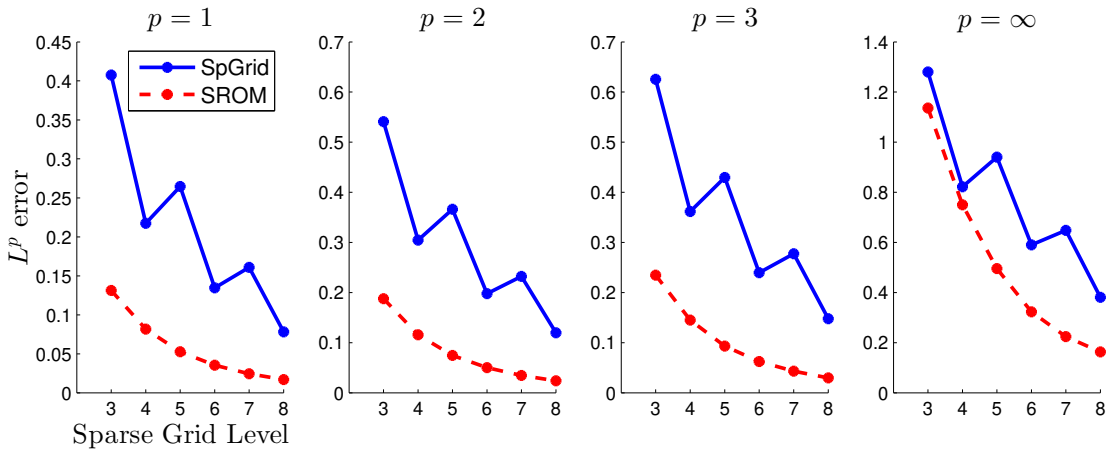


Figure 2.7: L^p error comparison of SROM-based surrogate and sparse grid stochastic collocation for $d = 6$. Sparse grid level/Number of Nodes: 4/1457, 6/15121, 8/127105.

this surrogate in order to also capture the regions of Z for which $U(Z)$ manifests substantial variation. This can be achieved by constructing the SROM-based surrogate in a sequential manner. We offer two approaches for the adaptive construction wherein one employs a global sampling strategy while the other uses a local sampling strategy. The algorithm presented below which incorporates global sampling has some similarities with an algorithm coined for the deterministic case in a terrain modeling application [7], although both algorithms

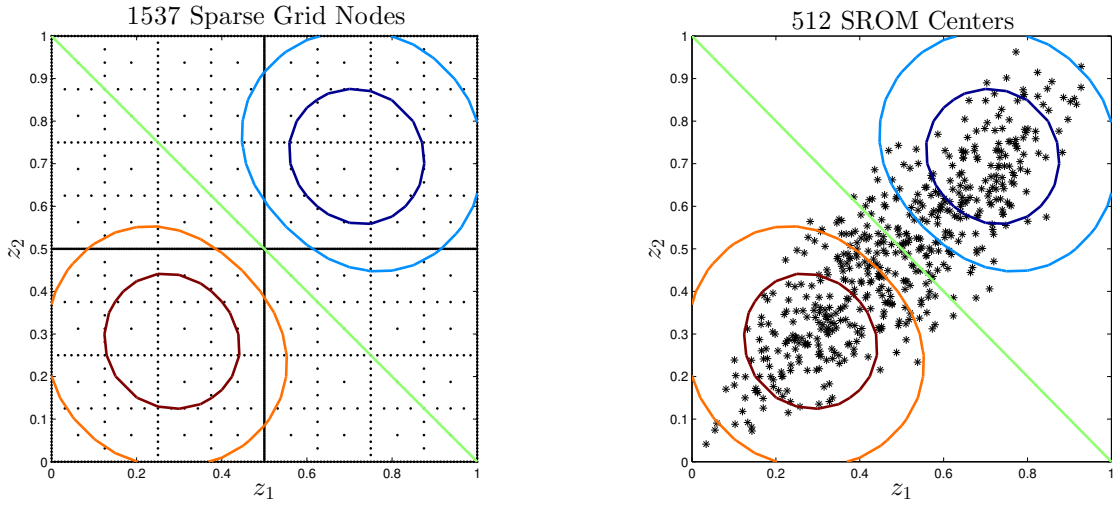


Figure 2.8: Location of sparse grid nodes for the 8th level (left) and SROM nodes (right) with contours of $U(z)$.

have been developed independently.

2.3.1 Description and illustration of the algorithm

As before, we will use the SROM-based surrogate with 1st-order Taylor interpolant due to reasons specified above. The main idea behind the adaptive algorithm is that instead of simultaneously computing the gradients at all Voronoi centers, the centers at which the gradients will be evaluated will be chosen sequentially. Therefore, we will assume that the following parameters are specified:

- p – the order of the L^p error ($p \geq 1$).
- ϵ – tolerance level for the L^p error.
- τ – the available computational units/budget.

2.3.1.1 Adaptive method with global sampling

The global sampling approach proceeds with the initialization step in which we generate an SROM $\widetilde{Z} = \{\widetilde{z}_k\}_{k=1}^{M}$ of Z and evaluate $U(\widetilde{z}_k)\ \forall k$, but not $\nabla U(\widetilde{z}_k)$, within the available computational budget τ. In order to obtain an initial surrogate model for $U(z)$, we then construct an SROM $\widetilde{Z}^*$ of $\widetilde{Z}$ of size $m \ll M$ (i.e. $m = 1$), where the SROM nodes of $\widetilde{Z}^*$ is a subset of $\{\widetilde{z}_k\}_{k=1}^{M}$. The initial surrogate model (2–3) can then be constructed using the SROM nodes $\widetilde{Z}^*$ which induces an initial partitioning of Γ through $\{\Gamma_i\}_{i=1}^{m}$. At every iteration, we select a partition where the error of the surrogate is largest and select an SROM node $\widetilde{z}_k$ from this partition under some criteria. The gradient at this SROM node is then computed, a new partition corresponding to this SROM node is then constructed, and the current surrogate is updated owing to the new partition. This procedure is repeated until the computational budget τ is exhausted or if the error is smaller than the specified tolerance ϵ.

The construction of the adaptive SROM-based surrogate with global sampling can thus be summarized in Algorithm 1.

Algorithm 1 Adaptive method for SROM-based surrogate with global sampling

1: Generate an SROM $\widetilde{Z}$ of Z given by $\mathcal{A} = \{\widetilde{z}_k\}_{k=1}^{M}$ and evaluate $U(\widetilde{z}_k)$.
2: Generate an SROM $\widetilde{Z}^*$ of $\widetilde{Z}$ given by $\{\widetilde{z}_k^*\}_{k=1}^{m} \subset \mathcal{A}$, $m \ll M$, and evaluate $\nabla U(\widetilde{z}_k^*)$.
3: Construct a Voronoi tessellation $\{\Gamma_i\}_{i=1}^{m}$ of Γ using $\{\widetilde{z}_k^*\}_{k=1}^{m}$ as the centers.
4: Compute the surrogate $\widetilde{U}_m(z)$ and update $\mathcal{A} \leftarrow \mathcal{A} \setminus \{\widetilde{z}_k^*\}_{k=1}^{m}$.
5: **while** budget τ is not exhausted **and** approximate $\|\widetilde{U}_m(z) - U(z)\|_{L^p(\Gamma)}^p > \epsilon$ **do**
6: Select *partition* $\Gamma_k \in \{\Gamma_i\}_{i=1}^{m}$ with the largest approximate $L^p(\Gamma_k)$ error.
7: *Select node* $\widetilde{z}_j \in \Gamma_k \cap \mathcal{A}$ and compute $\nabla U(\widetilde{z}_j)$.
8: *Refine the partition* Γ_k.
9: Update the surrogate $\widetilde{U}_{m+1}(z) \leftarrow \widetilde{U}_m(z)$, the set $\mathcal{A} \leftarrow \mathcal{A} \setminus \{\widetilde{z}_j\}$, and $m \leftarrow m+1$.
10: **end while**

We now address the technicalities encountered in Algorithm 1. At each iteration, the L^p error of the surrogate is used as the refinement measure to drive adaptivity. Mathematically, the L^p error of $\widetilde{U}_m(z)$ on the *partition* Γ_k is equivalent to

$$\begin{aligned} \|U(z) - \widetilde{U}_m(z)\|^p_{L^p(\Gamma_k)} &= \int_{\Gamma_k} |U(z) - \widetilde{U}_m(z)|^p \, dF(z) \\ &= E[|U(z) - \widetilde{U}_m(z)|^p \,|\, Z \in \Gamma_k] \cdot P(\Gamma_k), \end{aligned}$$

i.e., the refinement measure of Γ_k is equal to the average p-th order surplus of the surrogate weighted by the probability of that *partition*. As the refinement measure cannot be computed exactly in practice, we approximate $P(\Gamma_k)$ by determining the proportion of samples of Z such that $z \in \Gamma_k$ whereas the average p-th order surplus conditional on Γ_k can be approximated using the response values $U(\widetilde{z}_j)$ for $\widetilde{z}_j \in \Gamma_k \cap \mathcal{A}$ where $\{\widetilde{z}_j\}$ is obtained from the initialization step above. More explicitly, if we denote by $n_k := \#\{\widetilde{z}_j | \widetilde{z}_j \in \Gamma_k \cap \mathcal{A}\}$,

$$E[|U(z) - \widetilde{U}_m(z)|^p \,|\, Z \in \Gamma_k] \approx \frac{1}{n_k} \sum_{i=1}^{n_k} |U(\widetilde{z}_i) - \widetilde{U}_m(\widetilde{z}_i)|^p.$$

As the iteration of the adaptive algorithm with global sampling progresses, it is possible that $\{\widetilde{z}_j | \widetilde{z}_j \in \Gamma_j \cap \mathcal{A}\}$ may eventually be empty for some partition Γ_j and that the L^p error may not be approximated. To preempt such a scenario, we will assume that when the user specifies τ computational units in the initialization stage, this already includes n_c units of contingency response evaluations of $U(z)$, i.e. $\tau = M + g + n_c$ where g is the maximum number of gradient calculations. If this situation occurs, we construct an SROM of size $\lceil P(\Gamma_j) n_c \rceil$ for Z conditioned on Γ_j and evaluate $U(z)$ on these newly obtained SROM nodes to approximate the L^p error on this *partition*. Subsequently, we add these SROM nodes to $\mathcal{A}$ and we update n_c as $n_c \leftarrow n_c - \lceil P(\Gamma_j) n_c \rceil$.

We can also obtain an approximate confidence interval to quantify the error

in our estimation of the refinement measure of Γ_k. Let $Y_i = |U(\widetilde{z}_i) - \widetilde{U}_m(\widetilde{z}_i)|^p$, $i = 1, \ldots, n_k$. By the Central Limit Theorem, $S_{n_k} = \dfrac{1}{n_k} \displaystyle\sum_{i=1}^{n_k} Y_i$ is asymptotically $\mathcal{N}\left(\mu, \frac{\sigma^2}{n_k}\right)$ where $\mu = E[|U(z) - \widetilde{U}_m(z)|^p \,|\, Z \in \Gamma_k]$ while $\sigma^2 \approx \dfrac{1}{n_k - 1} \displaystyle\sum_{i=1}^{n_k} (Y_i - S_{n_k})^2$. Confidence bands for μ are now immediate from this asymptotic distribution.

The choice of refinement measure we have pursued implies that the L^p error over Γ will be used as a stopping criterion for the adaptive process. The same techniques described above are applicable in approximating the L^p error on the whole probability space. Thus, the adaptive process terminates either when the approximate L^p norm is less than the specified tolerance or when the budget τ is exhausted. The latter occurs precisely when the maximum number of gradient evaluations has been reached or when there are no more contingency response evaluations available.

2.3.1.2 Adaptive method with local sampling

An alternative to the adaptive algorithm with global sampling as presented above is possible through a local sampling procedure. Essentially, instead of selecting all of the samples $\widetilde{z}_k$ at which to evaluate the response at the beginning of the algorithm, some of these samples are selected as the adaptive construction progresses. The initialization step is still the same, however, rather than generating an SROM $\widetilde{Z}$ of size M as in before, we only generate an SROM $\widetilde{Z} = \{\widetilde{z}_k\}_{k=1}^{M'}$ where $M' \ll M$ and evaluate $U(\widetilde{z})\, \forall k$ but not $\nabla U(\widetilde{z}_k)$. Subsequently, at each iteration, for each partition Γ_i, we construct an SROM of size m_i for Z conditioned on Γ_i and evaluate $U(z)$ on these SROM nodes. This local sampling scheme removes the need to determine the location of all nodes $\widetilde{z}_k$ at the initialization step.

The construction of the adaptive SROM-based surrogate with local sampling is summarized below:

Algorithm 2 Adaptive method for SROM-based surrogate with local sampling

1: Perform Steps 1-4 as in Algorithm 1.
2: **while** budget τ is not exhausted **and** approximate $\|\widetilde{U}_m(z) - U(z)\|^p_{L^p(\Gamma)} > \epsilon$ **do**
3: Perform Lines 6-9 as in Algorithm 1.
4: **for** each Γ_i **do**
5: Generate an SROM $\widetilde{Z}\,|\,\Gamma_i$ of $Z\,|\,\Gamma_i$ given by $\{\widetilde{z}_k\}_{k=1}^{m_i}$ and evaluate $U(\widetilde{z}_k)$.
6: Update the set $\mathcal{A} \leftarrow \mathcal{A} \cup \{\widetilde{z}_k\}_{k=1}^{m_i}$.
7: **end for**
8: **end while**

The size m_i of the SROM $\widetilde{Z}$ conditioned on Γ_i can be obtained by solving a constrained least squares problem. More specifically, the values of m_i can be chosen such that the proportion of samples $\widetilde{z}_k$ in Γ_i for which the response has been computed is equal to $P(\Gamma_i)$, the probability of the *partition* Γ_i. Let N be the number of partitions Γ_i in the current iteration and let $n_i := \#\{\widetilde{z}_j | \widetilde{z}_j \in \Gamma_i \cap \mathcal{A}\}$. The condition can then be mathematically expressed as $\frac{n_i+m_i}{\sum_{j=1}^N n_i + \sum_{j=1}^N m_i} = P(\Gamma_i)$ for $i = 1, \ldots, N$ which is a linear system with nonnegative constraints for m_i. A cheap estimate of $P(\Gamma_i)$ can be obtained as before.

For a practical implementation of the adaptive algorithm with local sampling, we have imposed a cap on the maximum number of response and gradient calculations for a given computational budget. This is because as the adaptive construction progresses, the partitions decrease in size and it becomes counterintuitive to compute more response evaluations to approximate the L^p error. By imposing a limit on the response and gradient calculations, the local sampling scheme then encounters the same problem as in the global sampling scheme wherein the L^p error may not be approximated for a particular *partition* for some iteration. Consequently, we address the issue similarly using contin-

gency response evaluations. The stopping criterion under the local sampling approach is now similar as the global sampling approach.

Finally, using the L^p error as the refinement criterion has implications on the performance of the global and local sampling schemes we have proposed. The local sampling scheme allows one to obtain a surrogate at lesser expense since not all of the sample responses that are used to approximate the L^p error are generated during the initialization step. However, for some response functions and for some choices of the probability law for Z, the global sampling scheme can be favorable in that the approximate L^p errors of the partitions are more accurate at the early stages of the adaptive construction when the sizes of the partitions are large.

Remark. If a posteriori error estimates of $U(Z) - \widetilde{U}(Z)$ conditioned on Z are available, as in [43], the L^p error of the a posteriori error estimate can be used instead for the refinement measure. The L^p error gives us a metric over the probability space rather than at a single sample of Z only.

We now elaborate on some computational aspects and implementation issues of the adaptive algorithm.

2.3.1.3 Choice of parameters

The initialization step requires that values for p, ϵ need to be specified. Suppose that we are interested in performing a forward sensitivity analysis on the quantity of interest $q(Z) := \ell(U(x,t,Z))$ where $U(x,t,Z)$ is a solution to a PDE. If the objective is to examine the k-th order moments of $q(Z)$ for $k = 1, \ldots, n$, then it would be natural to set $p = n$ as convergence in L^p also guarantees convergence

in L^q for $q \leq p$.

On the other hand, the choice of ϵ can be inferred from the following inequality. Let $U^h(x,t,Z)$ be the discretized solution of $U(x,t,Z)$ in the time and physical space and let $\widetilde{U}^h(x,t,Z)$ be the SROM-based surrogate based on $U^h(x,t,Z)$. If $q^h(Z) \coloneqq \ell(U^h(x,t,Z))$ and $\widetilde{q}^h(Z) \coloneqq \ell(\widetilde{U}^h(x,t,Z))$,

$$\|q(Z) - \widetilde{q}^h(Z)\| \leq \|q(Z) - q^h(Z)\| + \|q^h(Z) - \widetilde{q}^h(Z)\|.$$

The second term in the inequality right hand side is the error obtained by approximating the discretized solution by a surrogate and is precisely the stopping criterion of the adaptive method above. The first term however is due to the error of using a discretized solution in place of the analytic solution and bounds for the physical discretization error are usually available using a posteriori error estimates. Hence, ϵ must be chosen such that it is of the same order of magnitude as the error due to the physical discretization.

2.3.1.4 Selection of new node and choice of partitioning strategy

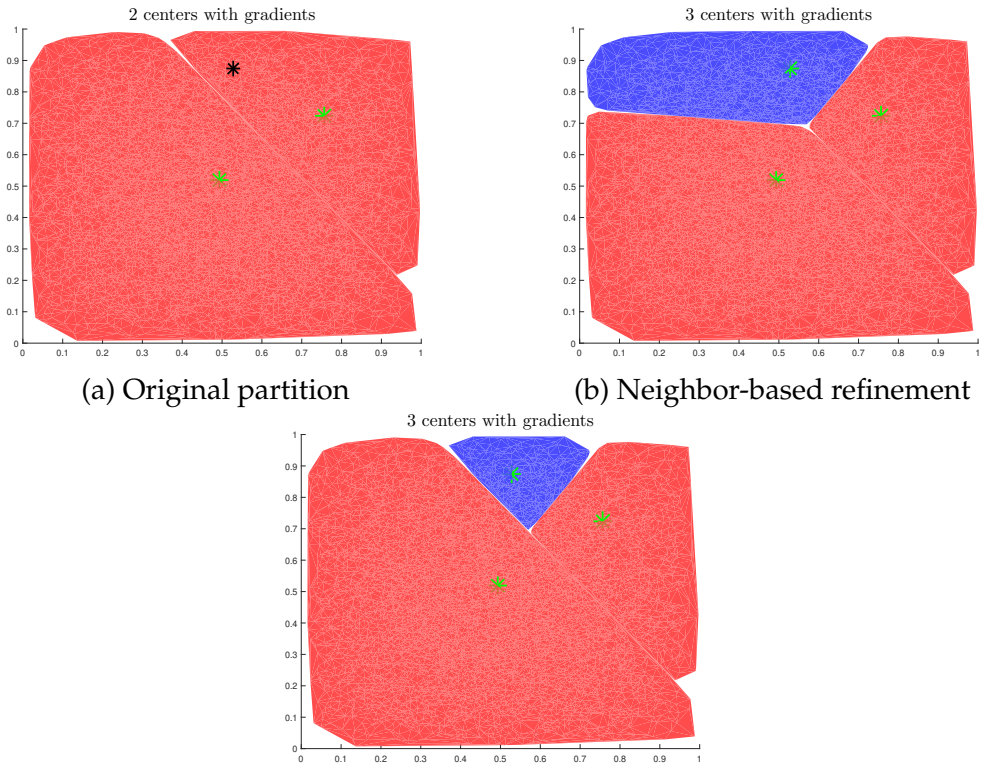
We now address how to select the new node $\widetilde{z}_j$ as outlined in line 7 of Algorithm 1 (and the corresponding line in Algorithm 2) and how to carry out the partitioning of Γ_k that is necessary for both algorithms.

Suppose that the partition with the largest approximate L^p error has been identified. Two schemes on the selection of the new node have been investigated, namely the surplus-based scheme and surplus-pdf-based scheme. Under the former scheme, we select the node $\widetilde{z}_j \in \Gamma_k \cap \mathcal{A}$ for which the surplus $|U(\widetilde{z}_j) - \widetilde{U}_m(\widetilde{z}_j)|$ is the largest. The latter scheme, however, selects the node for which $|U(\widetilde{z}_j) - \widetilde{U}_m(\widetilde{z}_j)|^p f(\widetilde{z}_j)$ is maximized and this quantity is precisely the inte-

grand of the $L^p(\Omega)$ error assuming that Z has joint pdf $f(z)$. The surplus-based scheme has the advantage in that the joint pdf of Z need not be known explicitly. Despite this convenience, it may result in computing gradients in regions of low probability especially in the initial iterations, when the partitions are relatively large. This in turn can adversely affect the L^p error of the surrogate in the subsequent iterations.

We have also investigated two types of strategies for refining the partitions Γ_k which are the neighbor-based refinement and cell-based refinement. Using the notation in Algorithms 1 and 2, Figure 2.9(a) shows the original partition where the green asterisks denote $\widetilde{z}_k$ for which $\nabla U(\widetilde{z}_k)$ has been computed and the black asterisk is $\widetilde{z}_j$ which resides in the partition Γ_k. Under neighbor-based refinement, a Voronoi tessellation of Γ with $\widetilde{z}_j$ as an additional Voronoi center is performed to obtain the top right plot. Essentially, only neighboring Voronoi cells of Γ_k are affected by this partitioning strategy as the number of partitions increases even though this is not evident in Figure 2.9(b). To visualize this, one can consider an example in 2 stochastic dimensions with a unit square domain where the Voronoi cells are squares of identical sizes. It is clear that refining the upper right most Voronoi cell will not affect the Voronoi cells in the lower left region of the domain. In contrast, for cell-based refinement, only a Voronoi tessellation conditioned on Γ_k with the black and the green asterisk as centers is performed to carry out the refinement in order to obtain the bottom plot. Hence, the resulting partitions in this case are not Voronoi cells.

From a computational perspective, it is clear that the cell-based refinement is less expensive and can be efficiently implemented using a tree data structure. As for the neighbor-based refinement, storing and updating the distance of each



(a) Original partition

(b) Neighbor-based refinement

(c) Cell-based refinement

Figure 2.9: Original partition (top left) and illustration of neighbor-based (top right) and cell-based (bottom) refinement.

sample of z_i of Z (that we have generated to construct the SROM $\widetilde{Z}$) from its corresponding Voronoi center at every iteration will be convenient to speed up the sample clustering in the refinement process. However, we shall see later that the neighbor-based refinement strategy will be advantageous especially in high stochastic dimension.

An illustration of the adaptive method with global sampling using neighbor-based refinement is shown in Figure 2.10 where the surplus-based scheme was used to select the new node in a partition. As can be seen, the adaptive method prioritizes regions of Γ with high probability and with high variation in $U(z)$. In summary, the adaptive method and the direct construction of the SROM-based surrogate as outlined in Section 2.2.2.1 are similar in that both surrogates use the

first-order Taylor interpolant in each Voronoi cell and both have the same number of gradient calculations. However, both methods differ in that the adaptive method has more evaluations of $U(z)$ in order to approximate the L^p error needed for the refinement criterion. In addition, the nodes $\widetilde{z}_j$ at which we compute $\nabla U(\widetilde{z}_j)$ in the adaptive method are located in regions of high probability in Z and large variation in $U(Z)$.

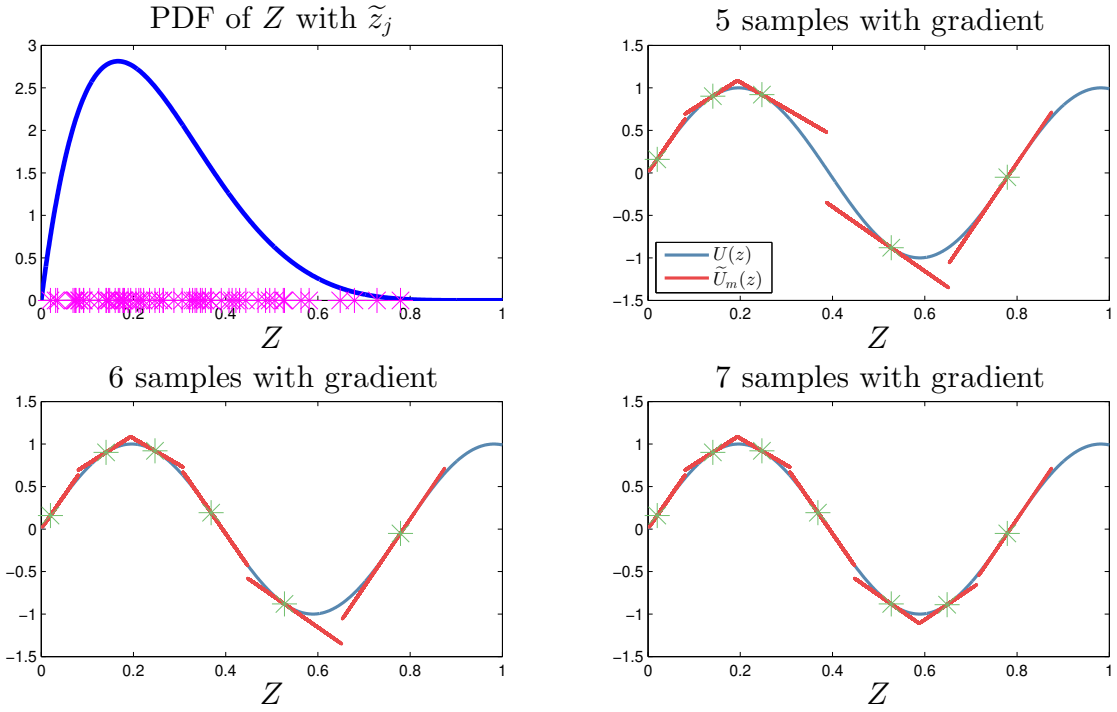


Figure 2.10: Illustration of the adaptive method with global sampling using neighbor-based refinement for $U(Z) = \sin(8Z)$, $Z \sim \text{Beta}(2, 6)$, $p = 3$. The upper left subplot shows a plot of the pdf of Z (blue) and the SROM $\widetilde{Z}$ generated during the initialization step (magenta asterisks). For the remaining subplots, the blue curve is $U(z)$, the red curve is $\widetilde{U}_m(z)$, and the green asterisks are the nodes where the gradient is available.

2.3.1.5 Post-termination of the adaptive method

Upon termination of the adaptive algorithm, it is possible that $\mathcal{A} \neq \emptyset$, i.e. there are nodes $\widetilde{z}_j$ for which $U(\widetilde{z}_j)$ is known but for which there are no computational

units left to obtain $\nabla U(\widetilde{z}_j)$. Although these nodes have served their purpose in providing validation about the SROM-based surrogate, it is natural to ask if this information can be used to improve the quality of $\widetilde{U}_m(Z)$. The challenge is that we need to use an interpolant for these nodes where the gradient information is unavailable and that this interpolant must be compatible with the 1st-order Taylor interpolant on each partition.

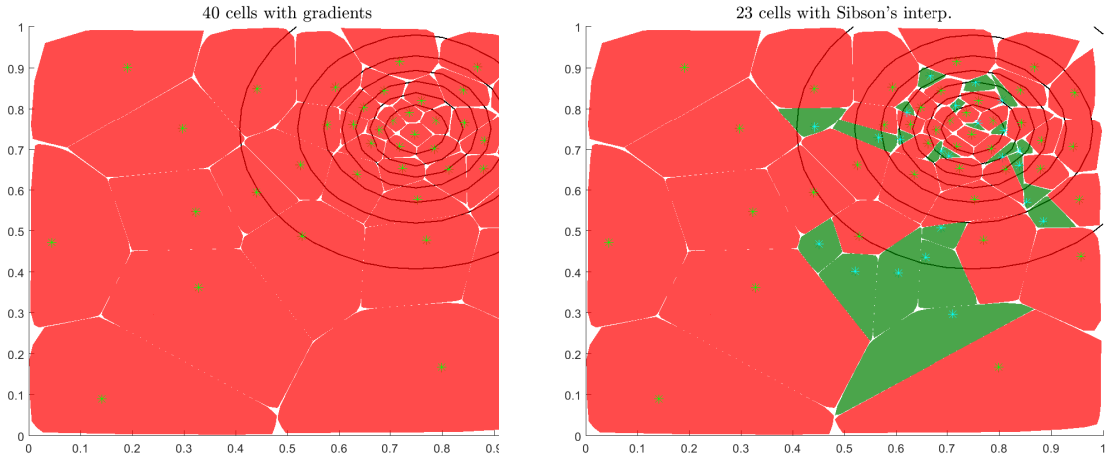
We therefore propose the construction of a hybrid surrogate model in Algorithm 3.

Algorithm 3 Hybrid SROM-based surrogate model

1: **for** each partition Γ_k **do**
2: Find $\{\widetilde{z}_j\} \in \mathcal{A} \cap \Gamma_k$ such that $|U(\widetilde{z}_j) - \widetilde{U}(\widetilde{z}_j)|^p > \epsilon$ or $|U(\widetilde{z}_j) - \widetilde{U}(\widetilde{z}_j)|^p f(\widetilde{z}_j) > \epsilon$.
3: Perform a cell-based refinement on Γ_k using $\{\widetilde{z}_j\}$ found in the previous step and the existing center of the partition Γ_k.
4: Select the nodes $\widetilde{z}_j$ from the first step such that its resulting partition within Γ_k is inside $\text{Conv}(\{\widetilde{z}_j\}_{j=1}^M)$, the set from the initialization step of the adaptive method.
5: Use Sibson's interpolant on these partitions corresponding to the nodes $\widetilde{z}_j$ found in the previous step and retain the Taylor-based interpolant on the complement of these partitions within Γ_k.
6: **end for**

This construction ensures that the hybrid surrogate is exact for linear responses and that it preserves the convergence properties of the 1st-order Taylor interpolant used in the adaptive method. An example of the hybrid surrogate is illustrated below for $U(Z) = \arctan(50 \cdot \|Z - 0.75\|^2)$, $Z \in \mathbb{R}^2$ where $Z_i \sim \text{Beta}(2, 2)$ i.i.d, $p = 2$ in which Figure 2.11 exhibits the resulting partitions of Γ while Figure 2.12 illustrates the surrogates. In Figure 2.11, the left subplot refers to the resulting partition once the adaptive method with global sampling using neighbor-based refinement terminates while the right subplot exhibits the partitioning of the hybrid surrogate with Sibson's interpolant employed in the green partitions. The interpolant in the red partitions is unchanged and is con-

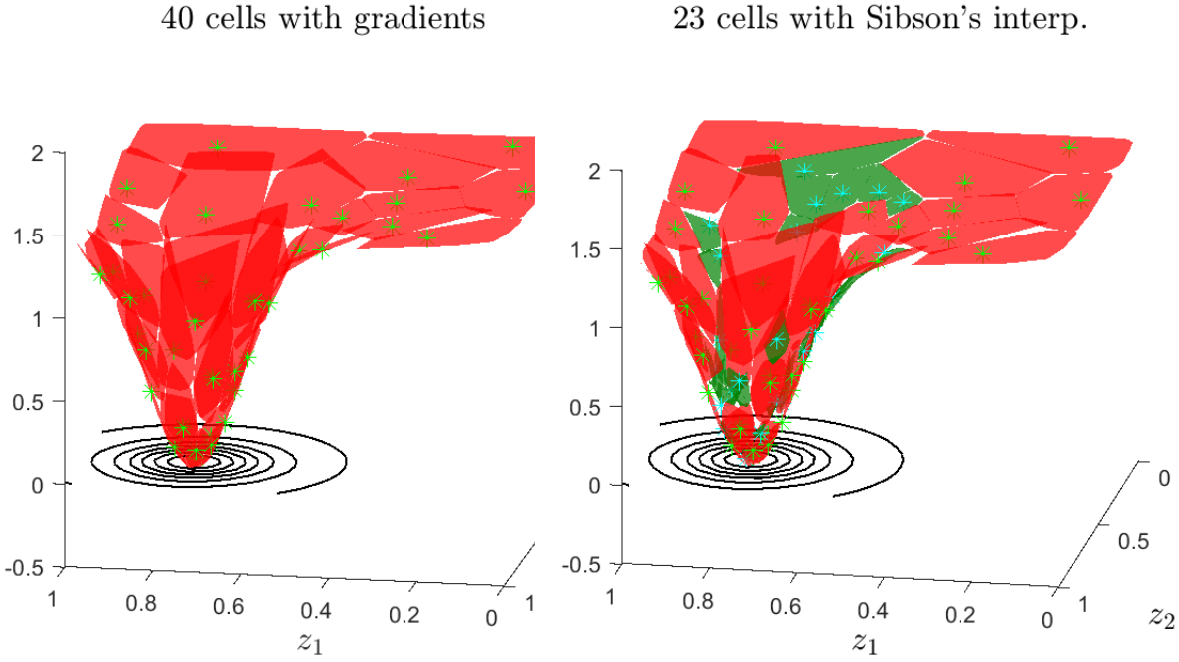
structed using the 1st-order Taylor interpolant. All the asterisks in the subplots correspond to the centers of their corresponding partitions and the contour lines of $U(z)$ are also displayed in all subplots.



(a) Original partition (b) Hybrid partition

Figure 2.11: Partition from the adaptive method (left) and for the hybrid surrogate (right).

As Figure 2.12 demonstrates, Sibson's interpolant is able to smooth out the overshoot caused by using the 1st-order Taylor interpolant as can be observed by comparing the region $(z_1, z_2) \in [0.4, 1] \times [0, 0.5]$ in both subplots. This is because for a fixed partition, the gradient of Sibson's interpolant is not constant unlike that of the 1st-order Taylor interpolant, making it more suitable for response surfaces with large curvature. This is quantitatively manifested in the L^p error plots in Figure 2.13 where the horizontal axis represents the iteration number or equivalently, the number of nodes with gradient calculated. The L^p error of the hybrid surrogate is plotted at the last iteration and the error is lower than that of the previous iteration corresponding to the termination of the adaptive algorithm. We acknowledge that we cannot prove that the hybrid surrogate reduces the L^p error of the surrogate obtained from the adaptive method. How-



(a) Original surrogate (b) Hybrid surrogate

Figure 2.12: SROM-based surrogate obtained from the adaptive method (left) and hybrid SROM-based surrogate (right).

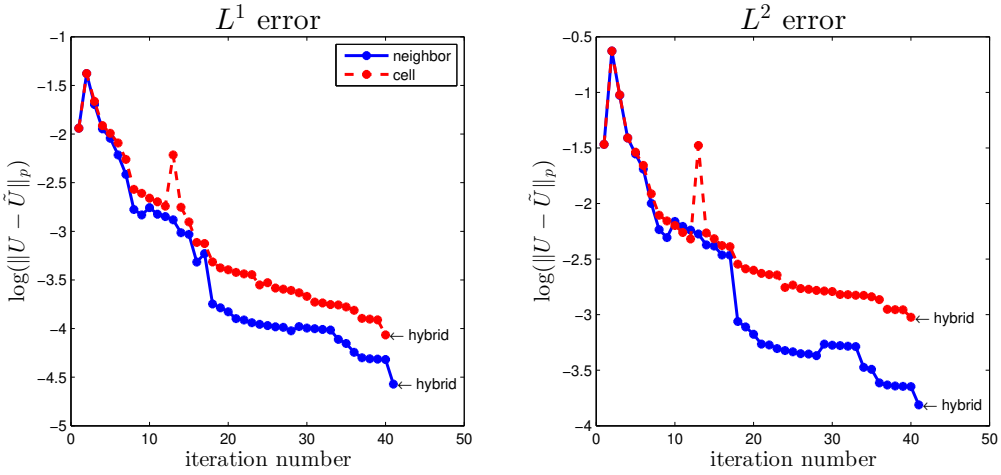


Figure 2.13: L^p error of the SROM-based surrogate and the hybrid surrogate for $p = 1$ (left) and $p = 2$ (right) as a function of the number of nodes with ∇U computed. Results for neighbor-based refinement is shown in red while that for cell-based is shown in blue.

ever, we have the consolation that Sibson's interpolant guarantees that the surrogate does not go beyond the minimum and maximum values of the response

according to Property 9.

2.3.2 Analysis of the algorithm

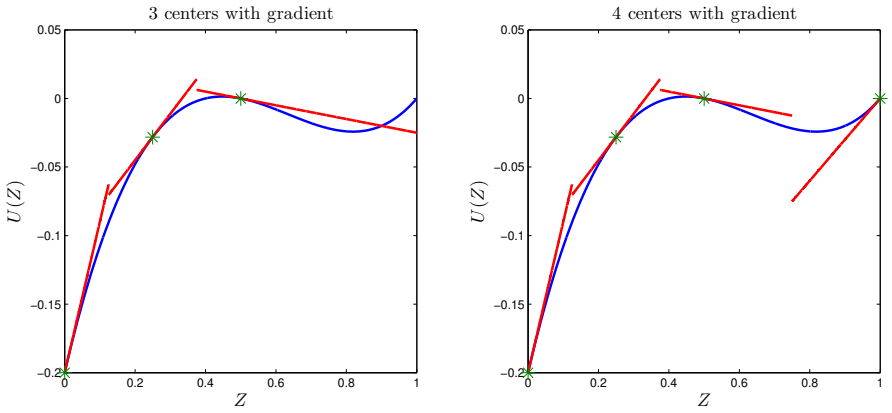


Figure 2.14: Illustration as to why the L^p is not monotonically decreasing.

As can be observed from Figure 2.13, the L^p error of the SROM-based surrogate obtained from the adaptive method is not monotonically decreasing as a function of the number of iterations. The number of iterations is the same as the number of nodes with gradient, hence, this result appears counterintuitive given that more computations have been expended. We now provide a simple analytic example to demonstrate why such situation occurs.

Example 3. Let $U(Z) = (Z-0.4)(Z-0.5)(Z-0.1)$ where $Z \sim \text{Unif}(0,1)$ and $p = 1$. In Figure 2.14, an illustration of how the adaptive method using neighbor-based refinement transitions from the 3rd iteration (left) to the 4th (right) is presented. The nodes with gradient in iteration 3 are $\{\widetilde{z}_j\}_{j=1}^3 = \{0, 0.25, 0.5\}$ while $\widetilde{z}_4 = 1$ is the additional node with gradient in iteration 4.

In order to see why the L^1 norm is not monotonically decreasing, it suffices to compare $\int_{0.75}^{1} |U(z)-\widetilde{U}_3(z)|\,dz$ and $\int_{0.75}^{1} |U(z)-\widetilde{U}_4(z)|\,dz$ because $z \in [0.75, 1]$ is the

portion of the domain for which the interpolant has changed. On the interval $[0.75, 1]$, $\widetilde{U}_3(z) = U(0.5) + U'(0.5)(z - 0.5)$ whereas $\widetilde{U}_4(z) = U(1) + U'(1)(z - 1)$. In addition, using Taylor's remainder theorem, we know that

$$U(z) - \widetilde{U}_j(\widetilde{z}_j) = \frac{1}{2}U''(\xi(z))(z - \widetilde{z}_j)^2$$

for some ξ which is a function of $z \in [0.75, 1]$. Hence, elementary calculations show that $|U(z) - \widetilde{U}_3(z)| = |(2z - 1.8)(z - 0.5)^2|$ while $|U(z) - \widetilde{U}_4(z)| = |(2z + 0.2)(z - 1)^2|$. Even though $|z - 1|^2 < |z - 0.5|^2$ for $z \in [0.75, 1]$, this is offset by $|2z - 1.8| < |2z + 0.2|$ which implies that $\|U(z) - \widetilde{U}_4(z)\|_{L^1} > \|U(z) - \widetilde{U}_3(z)\|_{L^1}$.

The example above leads to the following proposition which provides a sufficient condition to ensure that the L^p error is monotonically decreasing.

Proposition 10. *Suppose that the Hessian of $U(z)$ is constant and that all its eigenvalues are all equal. It then follows that $\|U - \widetilde{U}_{m+1}\|_{L^p} \leq \|U - \widetilde{U}_m\|_{L^p}$, i.e. the L^p error is monotonically decreasing.*

Proof. Under the stated assumptions, the discrepancy between the response and the surrogate on a partition Γ_k as shown in (2–5) can be simplified into the following. As before, $y \coloneqq Q \cdot (z - \widetilde{z}_k)$ where Q is an orthonormal matrix in the eigendecomposition of the Hessian and $\widetilde{z}_k$ is the center of the partition. We have that

$$|U(z) - \widetilde{U}_m(z)| = \frac{1}{2}\left|\sum_{i=1}^{d} \lambda_i y_i^2\right| = \frac{|\lambda|}{2}\|Q \cdot (z - \widetilde{z}_k)\|^2 = \frac{|\lambda|}{2}\|z - \widetilde{z}_k\|^2$$

where λ is the constant eigenvalue.

Hence the L^p error of the surrogate on a partition is simplified as

$$\|U(z) - \widetilde{U}_m(z)\|^p_{L^p(\Gamma_k)} = \left(\frac{|\lambda|}{2}\right)^p \int_{\Gamma_k} \|z - \widetilde{z}_k\|^{2p} \, dF(z).$$

As shown in the previous example, it only suffices to examine subsets of the partition Γ_k for which the interpolant over the subset changed from one iteration to the next. Let Γ' be such a subset such that $\Gamma' \subset \Gamma_j$ in iteration m but $\Gamma' \subset \Gamma_k$ in iteration $m+1$. Denote the center of the partition Γ_i by $\widetilde{z}_i$.

It therefore follows that

$$\begin{aligned}
\|U(z) - \widetilde{U}_{m+1}(z)\|^p_{L^p(\Gamma')} &= \left(\frac{|\lambda|}{2}\right)^p \int_{\Gamma'} \|z - \widetilde{z}_k\|^{2p} \, dF(z) \\
&\leq \left(\frac{|\lambda|}{2}\right)^p \int_{\Gamma'} \|z - \widetilde{z}_j\|^{2p} \, dF(z) \\
&= \|U(z) - \widetilde{U}_m(z)\|^p_{L^p(\Gamma')}
\end{aligned}$$

where the inequality holds by definition of Voronoi tessellation. □

In 1-dimension, the sufficient conditions are tantamount to the response having constant concavity. We also remark that the statement above holds regardless of the partitioning strategy chosen for the adaptive method. Even though the L^p error is not monotonically decreasing in general, we expect that the L^p error will have a decreasing trend due to the convergence properties of the SROM-based surrogate.

2.4 Numerical results

We now demonstrate the benefits of an adaptive construction of the SROM-based surrogate over the direct method through a variety of numerical examples in different stochastic dimensions. Since we expect that the adaptive method will be more computationally expensive, we have ensured that the number of gradient calculations for both methods is identical for the comparison to be fair.

In addition, as mentioned above, the SROM-based surrogate obtained is not unique but we do not attempt to generate multiple instances of the SROM-based surrogate in the following results.

In Example 4, a 2-d example is considered where the components of the random vector $Z = (Z_1, Z_2)$ are uniform i.i.d. in order to isolate the effect of the variation of $U(Z)$. We then compare the resulting partitioning of $\Gamma = Z(\Omega)$ obtained using direct and adaptive constructions. We continue this example with Example 5 wherein we compare the resulting partitioning of Γ for the global and local sampling scheme. Examples 6 and 7 are high-dimensional examples which investigate the performance of the direct and adaptive method with global sampling in cases where the region of high probability and the region of large variation in $U(Z)$ exactly coincide or slightly intersect. On the other hand, Examples 8 and 9 consider an application to solving a partial differential equation with random parameters in high dimensions. Lastly, a comparison between the performance of the adaptive algorithm with global and local sampling is carried out in Example 10.

In the following two examples, we visually compare the resulting partitioning of Γ for various methods of constructing the SROM-based surrogate. For instance, in Example 4, we aim to show that the adaptive method is able to track regions of high variation unlike the direct construction. In Example 5, we show how the resulting SROM nodes under local sampling can be more spread out than that produced by global sampling.

Example 4. Consider $U(Z) = \arctan(50 \cdot \|Z - 0.75\|^2)$, $Z \in \mathbb{R}^2$ where $Z_i \sim \text{Unif}(0, 1)$ i.i.d, $p = 3$. An illustration of the response is shown in Figure 2.12 above. Figure 2.15 demonstrates the resulting 40 partitions of Γ using direct

construction (left), adaptive method with neighbor-based refinement (middle), and adaptive method with cell-based refinement (right), with the contour lines of $U(z)$. Both adaptive algorithms were carried out with global sampling. With only 84 additional evaluations of $U(z)$ for both refinement strategies, the adaptive method targets regions where the variation of $U(z)$ is large (since the distribution of Z in this case is uniform).

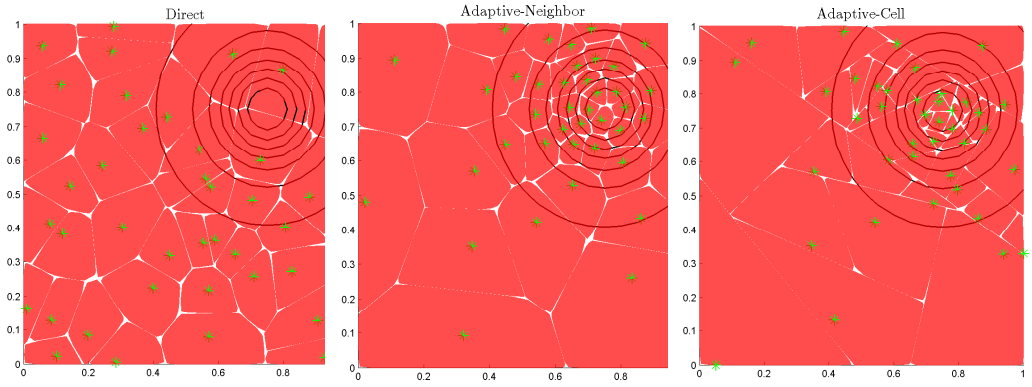


Figure 2.15: Comparison of the partitioning of Γ for 3 types of construction of SROM-based surrogates: Direct, Adaptive-Neighbor, Adaptive-Cell (left, middle, and right panels).

Example 5. Let $U(Z)$ be as in Example 4 with the same probability law for Z. Figure 2.16 demonstrates the resulting 30 partitions of Γ for the adaptive method using neighbor-based refinement with global sampling (left) and local sampling (right). All of the asterisks in both figures represent the nodes $\widetilde{z}_k$ for which $U(z)$ has been evaluated. In addition, the green asterisks represent the nodes for which $\nabla U(z)$ has been computed. A comparison of the subplots shows how the nodes acquired through local sampling are more spread out compared to that of global sampling because some of these nodes were selected as the adaptive construction progressed. Quantitatively, the marginal variance along Z_1 for the nodes obtained from global sampling is 0.0593 while the marginal variance along Z_2 is 0.0765. For comparison, the marginal variances for the nodes under local sampling are 0.0779 and 0.0829 along Z_1 and Z_2, respectively.

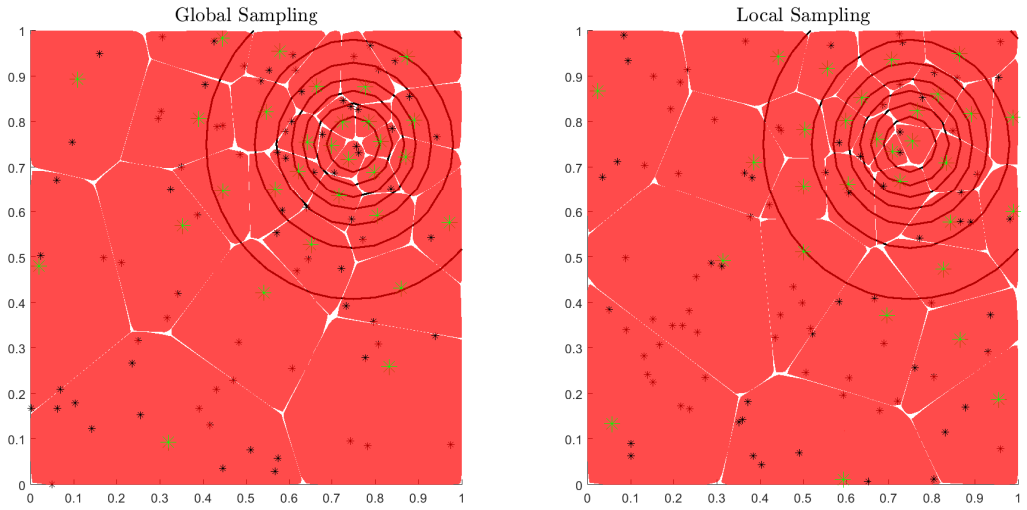
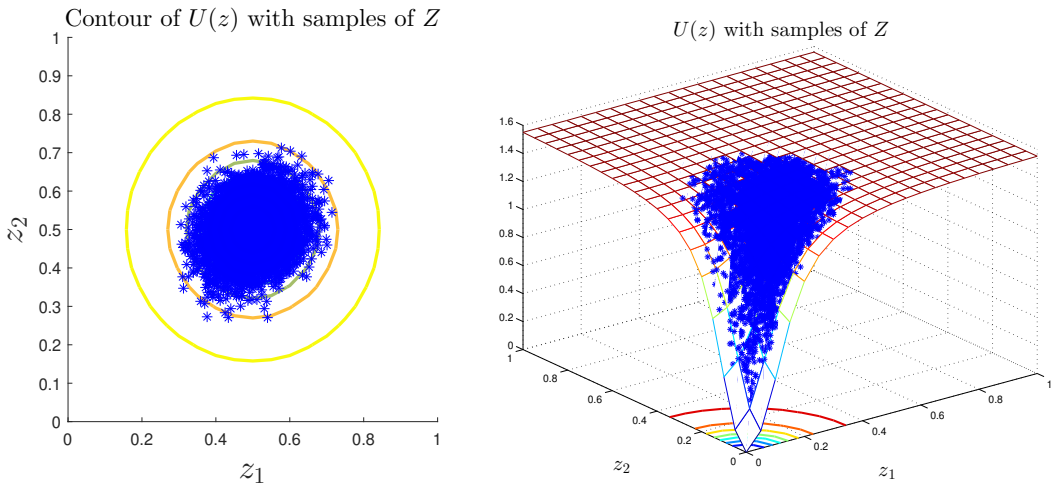


Figure 2.16: Comparison of the partitioning of Γ for 2 types of construction of SROM-based surrogates: Global sampling and Local sampling (left and right panels).

We now examine two related scenarios in high stochastic dimension. We first consider the case when the region where Z has high probability coincides with the region where $U(Z)$ has high variation as the next example illustrates. In this scenario, we expect the surrogate resulting from the direct construction to perform well as the need to locate regions of high variation is reduced. Subsequently, we will investigate the case when the region of high probability of Z is mostly concentrated on regions where $U(Z)$ has low variation. For both of these examples, global sampling was used for the adaptive algorithm.

Example 6. Let $U(Z) = \arctan(50 \cdot \|Z - 0.5\|^2)$, $Z \in \mathbb{R}^6$ where $Z_i \sim F^{-1}(\Phi(Y_i))$, $Y_i \sim N(0,1)$, $\text{cov}(Y_i, Y_j) = 0.2$ for $i \neq j$, F is the CDF of Beta$(30, 30)$, and $p = 4$. Figure 2.17 (left) shows the contour plot of the 2-d cross section of $U(z)$ for $z_i = 0.5, i > 2$ with samples of (Z_1, Z_2).

The L^p error of the SROM-based surrogate obtained under direct construction (dashed green), adaptive method using surplus-pdf-based scheme (thick



(a) Example 6

(b) Example 7

Figure 2.17: Illustration of the response and distribution of the random vector for Example 6 (left) and for Example 7 (right).

solid blue), and adaptive method using surplus-based scheme (thin solid red) is shown in Figure 2.18 for $p = 1, \ldots, 4$ as a function of the number of nodes with gradient calculated. Neighbor-based refinement was used for both adaptive methods. The L^p error under the direct construction was obtained using nodes with gradient calculations (equivalently, the number of partitions) in increments of 5. In addition, these sets of nodes do not form a refining sequence unlike in the case for the adaptive methods, i.e. the SROM nodes in the previous iteration is not a subset of the nodes in the succeeding iteration. With 235 extra evaluations of $U(z)$ for the adaptive surplus-based method and 237 extra evaluations for the surplus-pdf based method in the last iteration, we see the benefits of using an adaptive approach compared to a direct construction. In this example, there is no significant difference in the performance of the two schemes for selecting a new node.

Example 7. Set $U(Z) = \arctan(50 \cdot \|Z\|^2)$, $Z \in \mathbb{R}^{20}$ where $Z_i \sim F^{-1}(\Phi(Y_i))$, $Y_i \sim N(0,1)$, $\text{cov}(Y_i, Y_j) = 0.4$ for $i \neq j$, F is the CDF of Beta$(6, 20)$, and $p = 4$. An illustration of the 2-d cross-sectional response $U(z)$ for $z_i = 0, i > 2$ and samples of (Z_1, Z_2) is

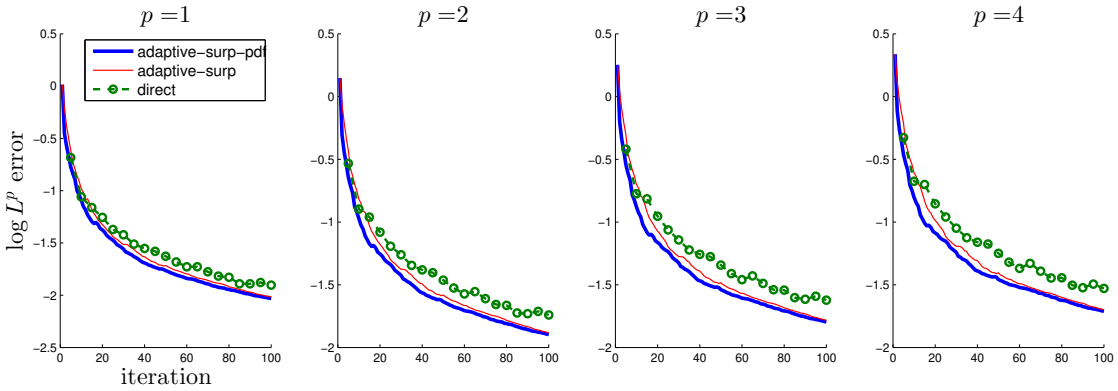


Figure 2.18: L^p error of the surrogate for the response in Example 6 obtained using 3 types of construction: Direct, Adaptive-surplus-pdf, Adaptive-surplus (Dashed, Thick Solid, Thin Solid).

shown in Figure 2.17 (right).

The L^p errors of the surrogate under the 3 different types of construction are presented in Figure 2.19 with the surplus-based scheme requiring 513 additional evaluations of $U(z)$ while the surplus-pdf-based scheme only requiring 503 more for the last iteration. Both adaptive methods outperform the direct construction in general, especially as p increases. For example, at the 80th iteration, there is a 62% decrease in the L^4 error if the adaptive surplus-pdf-based scheme is used instead of the direct construction.

The fluctuations in the higher L^p error under the direct construction result from the amplification of the error between the response and the surrogate that is not explicit in the L^1 error. Furthermore, this demonstrates the challenges associated with approximating response surfaces with tangent hyperplanes as was discussed in Section 2.3.2 for the case when the SROM nodes form a refining sequence. Even in the case where the SROM nodes of the preceeding iteration is not a subset of the succeeding iteration, as in the direct construction here, it is still possible that the L^p norm will not be monotonically decreasing. This is

because of the lack of curvature of the tangent hyperplane approximation.

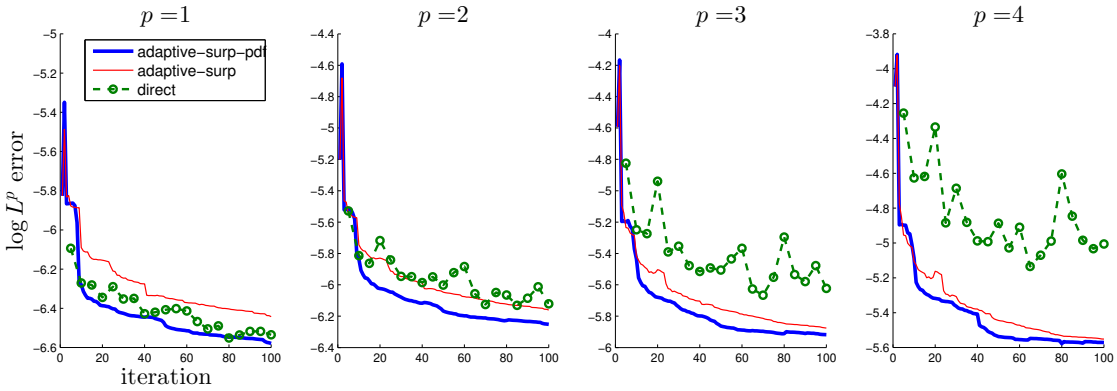
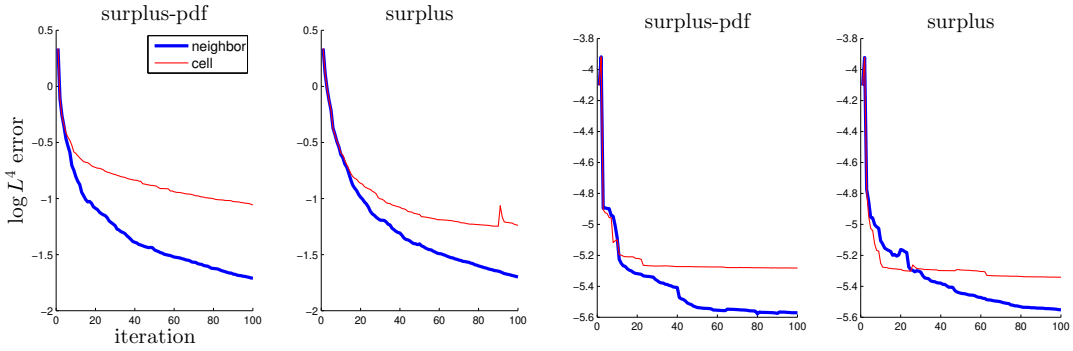


Figure 2.19: L^p error of the surrogate for the response in Example 7 obtained using 3 types of construction: Direct, Adaptive-surplus-pdf, Adaptive-surplus (Dashed, Thick Solid, Thin Solid).

In the examples above, we did not display the L^p error of the adaptive method using cell-based refinement as it has the tendency to underperform compared to the direct construction despite the latter not taking into account information about regions where $U(z)$ has high variation. This is because only 1 partition changes from one iteration to the next under cell-based refinement which makes it prone to slow convergence especially if the response is very smooth. For comparison, the number of partitions that change between consecutive iterations can be large under neighbor-based refinement especially in high dimensions. To visualize this, the simplest Voronoi cell is a hyperrectangle which has $2d$ faces in d dimensions. We quantify these observations through Figure 2.20 where a comparison is made for the L^4 error of the surrogate obtained using neighbor-based refinement (thick blue) and cell-based refinement (thin red) for Examples 6 and 7. Both schemes for selecting a new node have been considered and it is evident that the disparity can be considerable for both types of refinement. As a consequence, we will only focus on the neighbor-based refinement in the examples that follow.



(a) Example 6 (b) Example 7

Figure 2.20: Comparison of the L^4 error using the neighbor-based refinement (thick blue) and cell-based refinement (thin red) using both schemes of selecting a new node for Example 6 (left 2 subplots) and for Example 7 (right 2 subplots).

We then present examples in which the response is a solution to a stochastic partial differential equation as given in Examples 8 and 9. As before, we employ the global sampling scheme for the adaptive algorithm.

Example 8. Consider the PDE $\nabla\cdot(A(x,Z)\nabla U(x,Z)) = 0,\ x \in D$ with $D = (0,l_1)\times(0,l_2)$ and the boundary conditions $U(0,x_2) = 0, U(l_1,x_2) = 1$ for $x_2 \in (0,l_2)$ and $U_{x_2}(x_1,0) = U_{x_2}(x_1,l_2) = 0$ for $x_1 \in (0,l_1)$. $Z = (Z_1,\ldots,Z_d)$ is a random vector defined on the probability space $(\Omega,\mathcal{F},P)$. Physically, $U(x,Z)$ represents the electric potential on a specimen with $A(x,Z)$ being the conductivity field. A quantity of interest is the apparent conductivity given by

$$q(Z) = \frac{1}{l_2}\int_D A(x,Z)\frac{\partial U(x,Z)}{\partial x_1}\,dx$$

which is a random variable defined on the same probability space as Z. Our objective is to construct an SROM-based surrogate for $q(Z)$.

Suppose that $Z \in \mathbb{R}^4$ where $Z_i \sim \Phi(Y_i)$, $Y_i \sim N(0,1)$, $\mathrm{cov}(Y_i,Y_j) = 0.9$ for $i \neq j$, and $p = 4$, and that our conductivity field is given by

$$A(x,Z) = 4 + \sum_{i=1}^{2}(\sin(2\pi Z_{2i-1})\cos(2ix_1 + 2ix_2) + \cos(2\pi Z_{2i})\sin(2ix_1 + 2ix_2)).$$

The left subplot of Figure 2.21 shows samples of (Z_1, Z_2) whereas the right subplot shows the response surface of $q(Z_1, Z_2, 0.5, 0.5)$ using $l_1 = l_2 = 1$.

We remark that the construction of an SROM-based surrogate for $q(Z)$ requires calculations of $\frac{\partial U}{\partial Z_i}$ which can be obtained by differentiating both sides of the PDE with respect to Z_i and solving the resulting PDE.

Figure 2.22 shows the L^p errors of the SROM-based surrogate constructed in 3 different ways as before. The surplus-based scheme required 173 additional evaluations of $U(z)$ while the surplus-pdf-based scheme only required 150 more for the last iteration. The results show that the gap between the L^p errors of the adaptive and direct construction increases as p increases. For example, at the 55th iteration, there is a 34% decrease in the L^4 error if the adaptive surplus-pdf-based scheme is used instead of the direct construction.

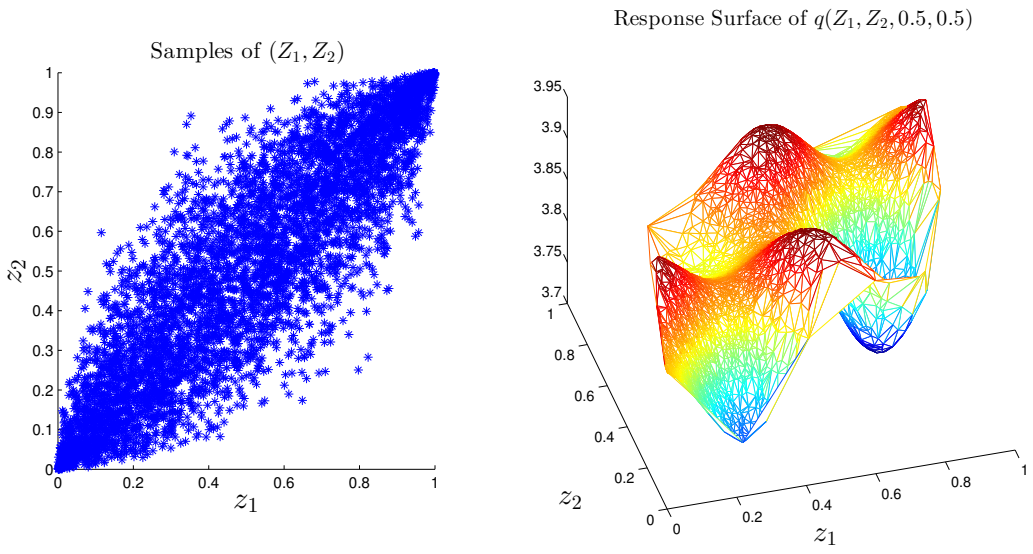


Figure 2.21: Illustration of the response and distribution of the random vector for Example 8.

Example 9. Consider the PDE in Example 8 using the same boundary conditions for $l_1 = l_2 = 0.2$. We construct the conductivity field as $A(x, Z) = F^{-1} \circ \Phi(G(x, Z))$ where F is the CDF of a Beta distribution with support $[1, 8]$ and shape pa-

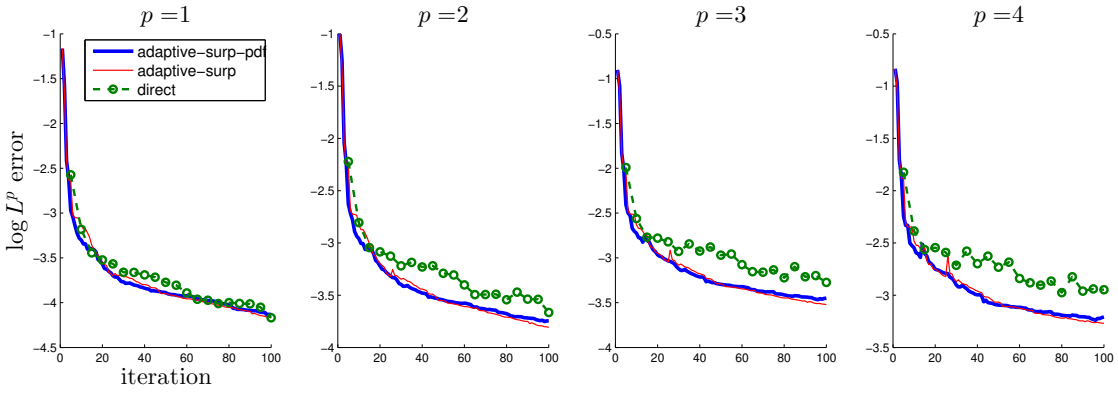


Figure 2.22: L^p error of the surrogate for the response in Example 8 obtained using 3 types of construction: Direct, Adaptive-surplus-pdf, Adaptive-surplus (Dashed, Thick Solid, Thin Solid).

rameters 2 and 6. Φ represents the CDF of the standard normal random variable with G being a homogeneous Gaussian field with mean 0, variance 1 and covariance function $c(\xi) = E[G(x,Z)G(x+\xi,Z)] = \exp(-0.5 \cdot (\xi_1^2 + 2\rho\xi_1\xi_2 + \xi_2^2))$ where $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$ with $\rho = 1000$. G is parametrically represented by a truncated Karhunen-Loève expansion taking on the form $G(x,Z) = \sum_{i=1}^{20} \sqrt{\lambda_i}\phi_i(x)Z_i$ where λ_i and ϕ_i are the eigenvalues and eigenfunctions of $c(\xi)$, respectively, and $Z_i \sim N(0,1)$, *i.i.d*. As before, we are still interested in constructing a surrogate for $q(Z)$ where $Z \in \mathbb{R}^{20}$.

We visualize the response surface in Figure 2.23 by constructing cross-sections of $q(Z)$ along coordinate axes Z_i and Z_j for which λ_i, λ_j are of similar magnitude. The values at the remaining coordinates are set to zero, the mean of Z_i. The plots indicate that there is a larger variation in the response outside the region of high probability. On the other hand, the cross-sections of $q(Z)$ along coordinate axes where $\lambda_i \gg \lambda_j$ appear to be linear, facilitating the approximation by tangent hyperplanes in these cases.

Figure 2.24 compares the L^p error of the surrogate under the 3 types of con-

struction that we have considered in the previous examples. The direct construction of the SROM-based surrogate was performed in increments of 10 iterations where the SROM nodes do not form a refining sequence as before. The adaptive surplus-based scheme has 903 extra evaluations of $U(Z)$ compared to the direct construction for the last iteration while the adaptive surplus-pdf based scheme has 850 more than the direct construction. The advantage of the surplus-pdf-based scheme in selecting the new node is quantified in Figure 2.24. The performance of the surplus-based scheme is hampered by the fact that the regions of high variation are mostly outside the region of high probability. As a consequence, in the earlier iterations of the adaptive method when the partitions are still large, the surplus-based scheme computes gradients in regions of low probability. Because the curvature of the response in these regions tends to be larger, the resulting surpluses also tend to be larger, resulting in a surrogate that prioritizes this region in the earlier stages of the algorithm. This is because, in certain cases, a very large surplus will overshadow the impact of the factor $P(\Gamma_k)$ in the L^p error expression given by $E[|U(z) - \widetilde{U}_m(z)|^p \,|\, Z \in \Gamma_k] \cdot P(\Gamma_k)$ especially for large p which implies that partitions of smaller probability are refined in the initial stages of the adaptive construction. These factors combined affect the performance of the surrogate in the succeeding iterations, yielding a slower convergence in this example.

We conclude with an example in which we are interested in comparing the performance of the two types of sampling schemes we have proposed, namely global and local sampling.

Example 10. We revisit the response $q(Z)$ in Example 8 with the corresponding probability law for Z. The adaptive construction was carried out using the

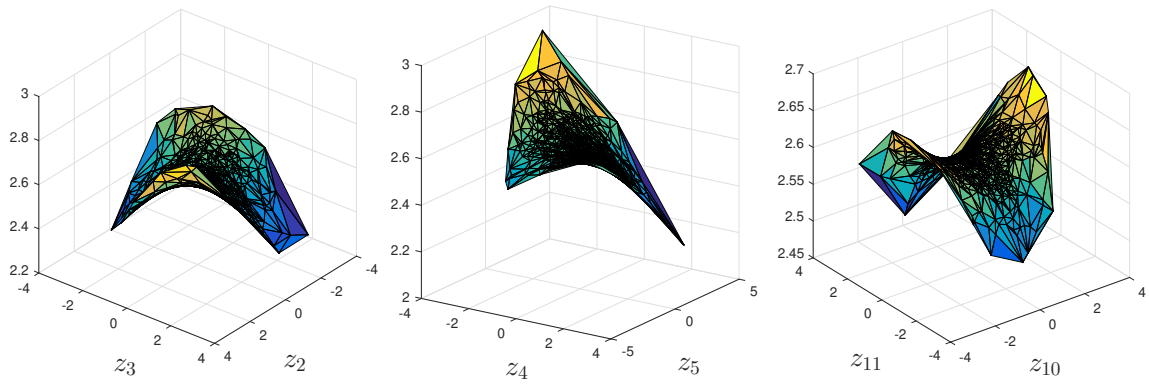


Figure 2.23: Cross-sections of the response $q(Z)$, $Z \in \mathbb{R}^{20}$, in Example 9.

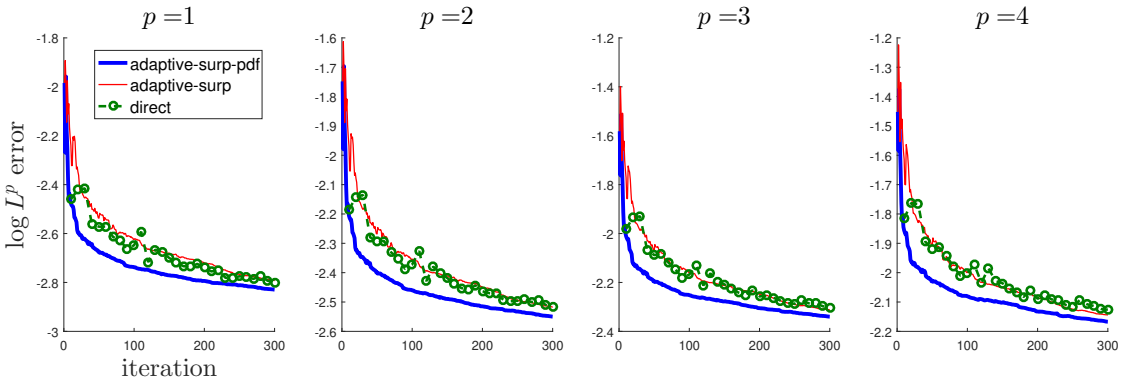


Figure 2.24: L^p error of the surrogate for the response in Example 7 obtained using 3 types of construction: Direct, Adaptive-surplus-pdf, Adaptive-surplus (Dashed, Thick Solid, Thin Solid).

surplus-pdf based scheme in selecting a new node.

Figure 2.25 shows the L^p error of the surrogates as a function of the computational budget. This is defined as the total number of response and gradient calculations for each surrogate at each iteration. We will assume that calculating a partial derivative is as costly as evaluating the response which implies that the computational budget in this example refers to the total number of times a PDE was solved in each iteration. For this example, the global sampling scheme was initialized with 100 evaluations of the response $U(z)$ while the local sampling scheme was initialized with 30 evaluations of $U(z)$ with additional samples added as the iteration progressed. As can be seen, the local sampling

scheme produces a surrogate with decent performance at a lesser expense at the beginning stages of the adaptive algorithm. However, in some cases such as this example, the global sampling scheme provides a more accurate surrogate in that the L^p errors are better approximated at the initial stages of the adaptive construction. In general, the performance between the two sampling schemes depends on the response function as well as the probability law of Z.

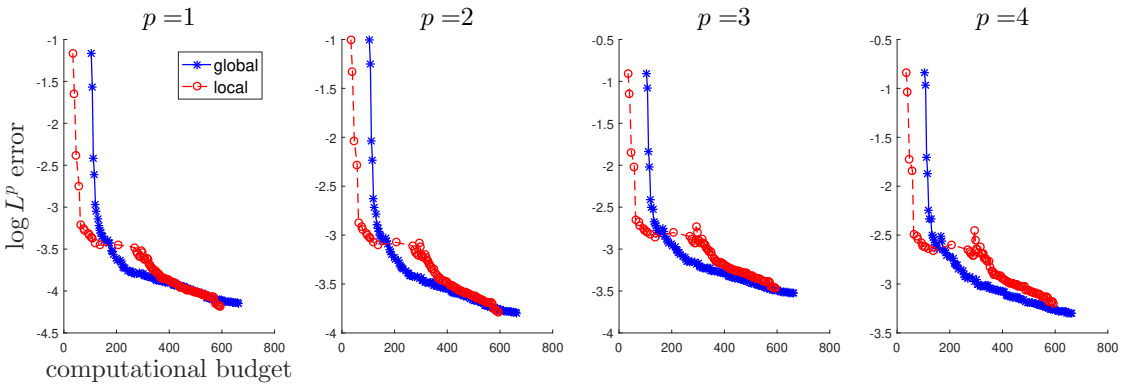


Figure 2.25: L^p error of the surrogate for the response in Example 7 obtained using 3 types of construction: Global sampling and Local sampling (Solid and Dashed).

2.5 Conclusion

We have surveyed novel and commonly used collocation-based surrogates for forward sensitivity analysis. We examined how the performance of sparse grid collocation can be adversely affected in high stochastic dimension as it does not take into account the probability law of the random input vector in the construction of the surrogate. To ameliorate this deficiency, we advocate the use of the SROM-based surrogate in which the interpolants employed are compatible with collocation nodes that are statistically representative of the input distribution. A comparison between zeroth-order Taylor, first-order Taylor, and the newly in-

troduced Sibson's interpolant, was made in terms of their implementation and convergence properties.

Since the direct construction of the SROM-based surrogate does not take into account the regions where the response exhibits large variations, we proposed an adaptive construction which simultaneously targets regions with high probability and high variation. For this adaptive method, two approaches for selecting the new node in a partition, two approaches for sequentially partitioning the probability space, and two sampling strategies were proposed and investigated, and their implementation and rate of convergence were compared. Subsequently, we analyzed how the adaptive method impacts the L^p error of the surrogate. Finally, numerical examples were furnished to demonstrate the benefits of the adaptive construction using various responses and probability distributions in different stochastic dimensions.

CHAPTER 3

SPECIFICATION OF ADDITIONAL INFORMATION FOR SOLVING STOCHASTIC INVERSE PROBLEMS

Methods have been developed to identify the probability distribution of a random vector Z from information consisting of its bounded range and the probability density function or moments of a quantity of interest, $Q(Z)$. The mapping from Z to $Q(Z)$ may arise from a stochastic differential equation whose coefficients depend on Z. This problem differs from Bayesian inverse problems as the latter is primarily driven by observation noise. We motivate this work by demonstrating that additional information on Z is required to recover its true law. Our objective is to identify what additional information on Z is needed and propose methods to recover the law of Z under such information. These methods employ tools such as Bayes' theorem, principle of maximum entropy, and forward uncertainty quantification to obtain solutions to the inverse problem that are consistent with information on Z and $Q(Z)$. The additional information on Z may include its moments or its family of distributions. We justify our objective by considering the capabilities of solutions to this inverse problem to predict the probability law of unobserved quantities of interest.

3.1 Introduction

Inverse problems emerge from applications in science and engineering when information about inputs to a system is sought given measurements of observable quantities. Suppose that the physical system is modeled by the mapping $A(x, Z) \mapsto U(x, Z)$ where $x \in \mathbb{R}^d$ is the spatial variable, $Z \in \mathbb{R}^n$ is a parameter, $A(x, Z)$ is a deterministic function, and $U(x, Z)$ is the response. One of the

most commonly investigated aspects of this mapping involves deterministic inverse problems. They deal with the estimation of the unknown parameter Z provided measurements $\{U_{x_i}\}_{i=1}^{N_{obs}}$ of U at spatial points $x_i \in \mathbb{R}^d$ which are referred to as quantities of interest. The ill-posedness of this problem is due to the non-uniqueness of the solution for Z and is commonly addressed via two well-established methods. Optimization approaches [41] solve for Z by minimizing the objective function $\sum_{i=1}^{N_{obs}} |U_{x_i} - U(x_i, Z)|^2 + \lambda^2 ||Z||^2$ where the regularization term $\lambda^2 ||Z||^2$ suppresses noisy solutions. In contrast, Bayesian approaches [45] construct the solution as a probability density function (pdf) for Z instead of a point estimate, i.e. we acquire a probabilistic solution to a deterministic problem, by specifying a prior pdf on Z based on available information on the unknown parameter. The observations $\{U_{x_i}\}_{i=1}^{N_{obs}}$ are typically assumed to be polluted by random noise whose law coupled with the prior on Z yield the posterior pdf on Z. Despite the connections between both approaches, they emphasize that additional information on Z is required to address the ill-posedness that arises from solving the inverse problem.

In this work, we focus instead on a different class of inverse problems where the unknown quantity is inherently stochastic. Using the same mapping as above, let Z be a random vector defined on the probability space $(\Omega, \mathcal{F}, P)$ such that $A(x, Z)$ is a random field while $U(x, Z)$ is the stochastic response. The stochastic inverse problem we address is the following [12, 14, 15, 34]:

> Determine the probability law of Z given probabilistic information (such as pdf) of the quantity of interest $Q(Z) \in \mathbb{R}^m$ represented by functionals of the stochastic response $U(x, Z)$.

By definition, the quantity of interest Q is a function of the response U. Ex-

amples include $Q(Z) = \max_x |U(x, Z)|$ and $Q(Z) = (U(x_1, Z), \ldots, U(x_m, Z))$ where $x_1, \ldots, x_m \in \mathbb{R}^d$ are fixed. Also, the output $Q(Z)$ may or may not contain observation noise; regardless, it is still a random quantity as it depends on Z which distinguishes it from what is encountered in the Bayesian formulation above in which $Q(Z)$, without the additive observation noise, is deterministic. This inverse problem is the direct reverse of forward uncertainty propagation which is concerned with obtaining probabilistic information of $Q(Z)$ given the probability law of Z.

Without additional information on Z, solving this stochastic inverse problem becomes challenging since distinct probability laws on Z can produce the same law for $Q(Z)$ cf. [12, pp. 1839-1840]. This ill-posedness is then compensated by imposing assumptions on the law of Z [10, 12, 14, 15, 34, 60, 81]. We mainly scrutinize two methods proposed in literature in which the only information known about Z comprises, at most, its bounded range. The first approach [12, 14, 15] uses the disintegration theorem for probability measures to obtain a pdf for Z given the pdf of $Q(Z)$. The second approach [10] aims to approximate the unknown random field $A(x, \omega)$, $\omega \in \Omega$, by solving an optimization problem; it is shown that this is conceptually identical to the stochastic inverse problem we consider above.

These existing methodologies are examined and the inadequacy of their basic implementations in recovering the true law of Z in the absence of further information on Z is used to motivate the following contribution of this work:

Identification of additional information to recover the true law of Z.

This work offers supplemental analysis to [12, 14–16] in that while the same

inverse problem is tackled, the desired properties that we seek in designing the solution are different. Our focus here is on the reconstruction of the true law which is made possible by incorporating various types of additional information on Z not considered in [12, 14–16]. We investigate what type of information Z needs to be equipped with and the corresponding tools that can be employed to solve the inverse problem consistently provided such additional information. Furthermore, we underscore the importance of recovering the true law of Z so that the resulting distribution on Z can be used to characterize unobserved quantities of interest other than those to which it was calibrated. As such, the contributions of this work aid in improving predictions based on solutions to stochastic inverse problems. The examples analyzed here are not meant to be critiques of existing literature. Instead, they offer further insight to better understand how these methods operate and suggest clues on issues that need to be accommodated to enhance their applicability.

3.2 Absence of information on the unknown random quantity

We survey existing methods based on: the disintegration theorem (Section 3.2.1) and an optimization approach (Section 3.2.2), to tackle the inverse problem. The information available on the unknown random quantity is limited to its bounded domain. It is argued that this is insufficient to recover the true law of the unknown quantity.

3.2.1 Disintegration of probability measures on generalized contours

The disintegration theorem is summarized in Section 3.2.1.1 which serves as the basis of the works [12, 14, 15]. Two design approaches for the disintegration theorem are then elaborated in Sections 3.2.1.2 and 3.2.1.3 with each containing an example applying the method. The examples underscore that the ansatz imposed by this methodology – i.e., the pdf on the generalized contour is uniform, may be restrictive as the pdf on the contours can possess complex behavior. They also show that the true pdf of the unknown quantity is under/overestimated if this ansatz is accepted.

3.2.1.1 Review of methodology

For the random vector $Z \in \mathbb{R}^n$ defined on the probability space $(\Omega, \mathcal{F}, P)$, let Q be the mapping $Q : \Gamma \to \mathcal{D}$ where $\Gamma := Z(\Omega)$ is a compact subset, $\mathcal{D} := Q(\Gamma) \subset \mathbb{R}^m$ with $m < n$, and $Q(z) = q(U(x, z))$ for $z \in \Gamma$ and for some function $q(\cdot)$ that is locally differentiable. In the rest of this work, distinction is made between pdfs constructed with respect to Lebesgue and non-Lebesgue measures. For measurable $A \subset \Gamma$ and $B \subset \mathcal{D}$, denote by $P_Z(A) = \int_A \rho_Z(z)\, d\mu_Z$ and $P_Q(B) = \int_B \rho_Q(q)\, d\mu_Q$ the probability measures on Z and Q, respectively, where $\rho_Z(z)$ and $\rho_Q(q)$ are the corresponding pdfs with respect to the specified measures μ_Z and μ_Q. Unlike in calculus-based probability, μ_Z and μ_Q are not restricted to be Lebesgue. For notation purposes, if the pdfs are constructed with respect to the Lebesgue measure, they are denoted by f instead of ρ.

The method proposed in [12, 14, 15] then addresses the following:

Given the pdf of $Q(Z)$ and the bounded range Γ of Z, estimate the pdf of Z.

The cited approaches compute probabilities in Γ by viewing the domain from a coordinate system of contours instead of the traditional Cartesian system. A summary of [14] is outlined below.

For $d \in \mathcal{D}$, define the set of points $\{z \in \Gamma | Q(z) = d\}$ to be a generalized contour. This is a generalization of contour curves when $n = 2, m = 1$. Generalized contours are equivalence classes with the relation $a \sim b$ if and only if $Q(a) = Q(b)$. Consequently, as $\mathcal{D}$ is the range of Q, the domain Γ is a union of generalized contours. A representative element from each generalized contour can then be selected which serves as an indexing mechanism across all contours. This indexing set is a m-dimensional manifold that intersects each contour once and is called a transverse parameterization. In other words, there is a bijection between points in the transverse parameterization and the generalized contours. To clarify these concepts, an example is shown in Figure 3.1a below for $n = 2, m = 1$ and $Q(z_1, z_2) = z_1 \cdot z_2$. The red thin curves are the contours of Q while the solid blue curve marked $\mathcal{L}$ ($z_2 = z_1$) and the dashed green curve marked $\mathcal{L}'$ ($z_2 = z_1^2$) are different transverse parameterizations. Because the transverse parameterization is not unique in general, we will only consider one parameterization in what follows and denote it by $\mathcal{L}$.

Every $z \in \Gamma$ can then be expressed under this new coordinate system. Let $\pi : \Gamma \to \mathcal{L}$ be the onto mapping such that for $x_{\mathcal{L}} \in \mathcal{L}$, $\pi^{-1}(x_{\mathcal{L}})$ is the corresponding generalized contour and that for $A \subset \Gamma$, $\pi(A)$ is the portion of the transverse parameterization $\mathcal{L}$ which intersects the generalized contours contained in A. We associate every $z \in \Gamma$ with $(x_{\mathcal{L}}, x_C)$ where $x_{\mathcal{L}} \in \mathcal{L}$ represents the generalized

contour in which z resides and x_C is the coordinate along the generalized contour $\pi^{-1}(x_{\mathcal{L}})$. Figure 3.1b exhibits this change of coordinate system where for this example, $x_{\mathcal{L}}$ parametrizes the arc length of the transverse from the origin while x_C parametrizes the arc length of the contour from $z_2 = 1$. The transverse $\mathcal{L}$ is parameterized by $z_2 = z_1$. The green thick curve is the contour $\pi^{-1}(x_{\mathcal{L}})$ that is indexed by $x_{\mathcal{L}} = 0.4$, that is, $x_{\mathcal{L}} \in \mathcal{L}$ is 0.4 units from the origin. Meanwhile, the x_C-coordinate of each of the 3 magenta circles on this contour is obtained by measuring the arc length of said contour from $z_2 = 1$ up to each of the 3 marked circles.

With this new coordinate system, the randomness in Z implies that there are random vectors $X_{\mathcal{L}}$ and X_C associated with $x_{\mathcal{L}}$ and x_C, respectively. In particular, for measurable $K \subset \mathcal{L}$, denote by $P_{X_{\mathcal{L}}}(K) = \int_K \rho_{X_{\mathcal{L}}}(x_{\mathcal{L}})\, d\mu_{X_{\mathcal{L}}}$ the probability measure on $X_{\mathcal{L}}$ where $\rho_{X_{\mathcal{L}}}$ is the pdf on $\mathcal{L}$ with respect to $\mu_{X_{\mathcal{L}}}$ that is specified. To solve the inverse problem, the probability of a measurable set $A \subset \Gamma$ can therefore be computed using the disintegration theorem for probability measures as follows [14, Corollary 4.1, Theorems 4.4, 4.5].

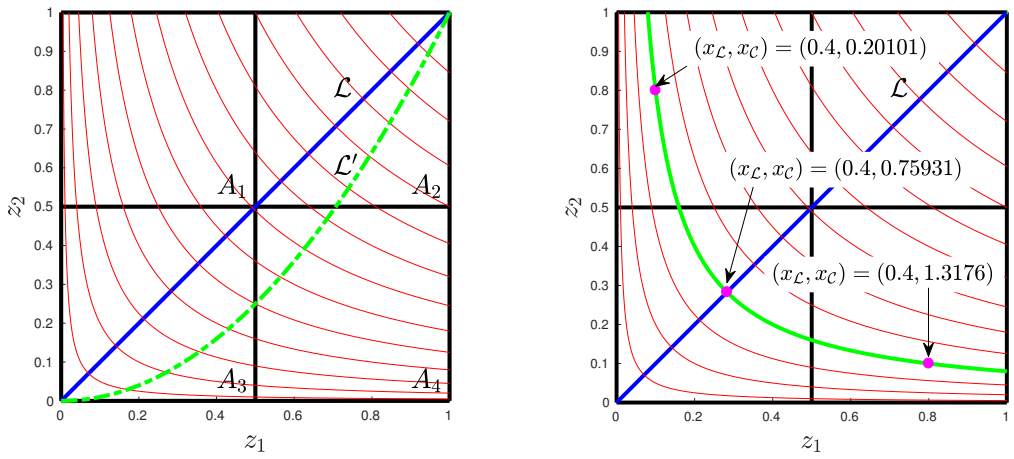
Theorem 11 (Disintegration theorem)**.** *Let P_Z be the probability measure defined by the law on Z and Q the measurable mapping between Γ and $\mathcal{D}$. For measurable $A \subset \Gamma$, P_Z admits the following disintegration*

$$P_Z(A) = \int_{\pi(A)} \int_{\pi^{-1}(x_{\mathcal{L}})\cap A} \rho_{X_C|X_{\mathcal{L}}}(x_C|x_{\mathcal{L}})\; d\mu_{X_C|X_{\mathcal{L}}}\; \rho_{X_{\mathcal{L}}}(x_{\mathcal{L}})\; d\mu_{X_{\mathcal{L}}} \tag{3–1}$$

where $\rho_{X_C|X_{\mathcal{L}}}$ is the conditional pdf on the generalized contour corresponding to $X_{\mathcal{L}}$ while the measure $\mu_{X_C|X_{\mathcal{L}}}$ satisfies

$$\mu_Z(A) = \int_{\pi(A)} \int_{\pi^{-1}(x_{\mathcal{L}})\cap A} d\mu_{X_C|X_{\mathcal{L}}}\; d\mu_{X_{\mathcal{L}}} \tag{3–2}$$

with μ_Z and $\mu_{X_{\mathcal{L}}}$ specified.



(a) (b)

Figure 3.1: Left panel: Generalized contours (thin red lines) and the two transverse parameterizations (thick blue and dashed green line) for $Q(z_1, z_2) = z_1 \cdot z_2$. Right panel: Illustration of the change in coordinate system from $z \in \Gamma$ to $(x_\mathcal{L}, x_C)$.

Figure 3.2 illustrates the region of integration in (3–1) for $A \subset \Gamma$.

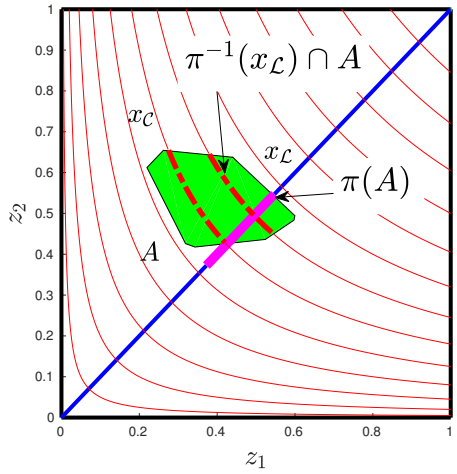


Figure 3.2: Illustration of the disintegration theorem. A is represented by the green region while the dashed red curves represent $\pi^{-1}(x_\mathcal{L}) \cap A$ and the solid magenta curve represents $\pi(A)$.

In Theorem 11, we have assumed that the disintegration of P_Z into a marginal and conditional family of measures is such that the latter two measures are absolutely continuous with respect to $\mu_{X_\mathcal{L}}$ and $\mu_{X_\mathcal{C}|X_\mathcal{L}}$ to admit the pdfs given by $\rho_{X_\mathcal{L}}(x_\mathcal{L})$ and $\rho_{X_\mathcal{C}|X_\mathcal{L}}(x_\mathcal{C}|x_\mathcal{L})$. If the probability density functions $\rho_{X_\mathcal{L}}$ and

$\rho_{X_C|X_\mathcal{L}}$ are specified, the pdf of Z can then be estimated. One such computational approach is to partition Γ into Voronoi cells and estimate (3–1) on each partition as in [15]. Due to the bijection between $\mathcal{L}$ and $\mathcal{D}$, the disintegration theorem guarantees that when the pdf of Z constructed in this manner is propagated (pushed forward) through the model, the resulting pdf on Q matches the prescribed ρ_Q. This is because $\rho_{X_\mathcal{L}}$ is fully specified given knowledge of ρ_Q from the relation $\int_K \rho_{X_\mathcal{L}}(x_\mathcal{L})\, d\mu_{X_\mathcal{L}} = \int_{Q(K)} \rho_Q(q)\, d\mu_Q$ for measurable $K \subset \mathcal{L}$. On the other hand, $\rho_{X_C|X_\mathcal{L}}$ cannot be identified solely relying on information from $\rho_Q(q)$, rendering the inverse problem ill-posed. Two distinct choices for $\rho_{X_C|X_\mathcal{L}}$ yield distinct pdfs on Z whose resulting pdf on Q when propagated through the model both match ρ_Q. It has been argued that it is reasonable to assume that $\rho_{X_C|X_\mathcal{L}}$ is uniform over the generalized contour, i.e.

$$\rho^{ansatz}_{X_C|X_\mathcal{L}}(x_C|x_\mathcal{L}) \coloneqq \rho_{X_C|X_\mathcal{L}}(x_C|x_\mathcal{L}) = \left(\int_{\pi^{-1}(x_\mathcal{L})} d\mu_{X_C|X_\mathcal{L}}\right)^{-1} \tag{3–3}$$

(see [14, equation 4.5]). The ansatz (3–3) has been employed with apparently satisfactory results in examples related to recovering the true probability law of Z in PDE models [14] as well as in inverse problems that arise in storm-surge applications [14], hydrodynamic models [15], and groundwater contamination [56].

To summarize, the method discussed constructs a *consistent* probability measure on Z, i.e. the push-forward of this measure matches the observed probability measure on $Q(Z)$. In other words, it constructs a particular pullback measure of an observed measure. It is therefore reasonable to use this method to attempt to "recover" the true probability measure/pdf on Z assuming information on $Q(Z)$ only. Of course, this can be successfully accomplished if $m = n$ and if Q is

bijective through a standard change of variables. When this is not the case, the ansatz (3–3) is likely insufficient if the objective is beyond consistency.

In order to apply the method above, the measures $\mu_Z, \mu_Q, \mu_{X_\mathcal{L}}$ have to be specified. In the next 2 sections, we examine design approaches for these measures as presented in literature [14, 15]. We also investigate the issues that may arise from application of the ansatz (3–3).

3.2.1.2 Pdfs with respect to Lebesgue measures

Following Figures 3, 4 and Equations 4.3, 4.4 of [15], we interpret $\mu_Z, \mu_Q, \mu_{X_\mathcal{L}}$ to be chosen as the Lebesgue measure so that $d\mu_Z(z) = dz$, $\mu_Q(q) = dq$, $d\mu_{X_\mathcal{L}}(x_\mathcal{L}) = dx_\mathcal{L}$ and the respective densities are $f_Z, f_Q, f_{X_\mathcal{L}}$. Using geometric arguments, it follows from (3–2) that $\mu_{X_C|X_\mathcal{L}}$ is also Lebesgue so that $d\mu_{X_C|X_\mathcal{L}} = dx_C$. As a consequence, the disintegration theorem (3–2) can be rewritten as

$$P_Z(A) = \int_{\pi(A)} \int_{\pi^{-1}(x_\mathcal{L})\cap A} f_{X_C|X_\mathcal{L}}(x_C|x_\mathcal{L})\ dx_C\ f_{X_\mathcal{L}}(x_\mathcal{L})\ dx_\mathcal{L} \tag{3–4}$$

while the ansatz (3–3) becomes

$$f^{ansatz}_{X_C|X_\mathcal{L}}(x_C|x_\mathcal{L}) = \left(\int_{\pi^{-1}(x_\mathcal{L})} dx_C\right)^{-1}. \tag{3–5}$$

Note the change in notation to emphasize that the pdfs are now with respect to the Lebesgue measure.

Demonstration of method on an example. Since all measures in the disintegration theorem have been specified, the method can now be applied to solve stochastic inverse problems. In the following, we consider a simple example to emphasize two points. First, assuming that the pdf on the generalized contours is uniform may not always enable the recovery of the true pdf of Z

for any mapping. The calculations carried out that indicate that the true pdf is under/overestimated is supported by Figure 3.4. Second, the pdf along the generalized contours can have substantial variation for different types of laws on Z. Figures 3.5 and 3.6 show how complicated the behavior can be of the pdf on the contour.

Example 11. Suppose that the inverse problem is independent of the spatial discretization x. Consider $Q(Z) = Z_1 \cdot Z_2$, $Z = (Z_1, Z_2)$ where $Z_1, Z_2 \sim U(0,1)$ and are independent. It is shown that given the pdf f_Q of Q, (1) the methodology above is unable to recover the probability law of Z and (2) the pdf along the generalized contours of Q are not uniform.

Consider the contours of Q (red thin curves) with the transverse parameterization $\mathcal{L}$ (thick blue line) where $0 \leq x_{\mathcal{L}} \leq \sqrt{2}$ in Figure 3.1a. In order to apply the proposed methodology, suppose that the support $\Gamma \coloneqq Z(\Omega) = [0,1] \times [0,1]$ is known and that f_Q is given. The pdf f_Q can be computed as follows: for $q \in (0,1]$,

$$P(Q \leq q) = \int_0^1 P\left(Z_2 \leq \frac{q}{z_1}\right) f_{Z_1}(z_1)\, dz_1 = \int_0^q dz_1 + \int_q^1 \frac{q}{z_1}\, dz_1 = q - q\log(q) \quad (3\text{–}6)$$

where f_{Z_1} is the pdf of Z_1. Thus, $f_Q(q) = -\log(q)$ for $0 < q \leq 1$. Given f_Q, it remains to determine $f_{X_{\mathcal{L}}}$ and $f_{X_C|X_{\mathcal{L}}}$ in order to compute (3–4).

As mentioned above, $f_{X_{\mathcal{L}}}$ can be uniquely obtained from f_Q. For $x_{\mathcal{L}} \in \mathcal{L}$, the generalized contour $\pi^{-1}(x_{\mathcal{L}})$ corresponding to $x_{\mathcal{L}}$ passes through the point $(z_1, z_2) = \left(\frac{x_{\mathcal{L}}}{\sqrt{2}}, \frac{x_{\mathcal{L}}}{\sqrt{2}}\right)$. Hence, the contour $\pi^{-1}(x_{\mathcal{L}})$ can be parameterized as $z_2 = \frac{x_{\mathcal{L}}^2}{2z_1}$, $x_{\mathcal{L}}^2/2 \leq z_1 \leq 1$, and we deduce the relationship $Q = \frac{X_{\mathcal{L}}^2}{2}$. We therefore have that for $x_{\mathcal{L}} \in (0, \sqrt{2}]$, $P(X_{\mathcal{L}} \leq x_{\mathcal{L}}) = P(2Q \leq x_{\mathcal{L}}^2) = F_Q(x_{\mathcal{L}}^2/2)$ which yields

$$f_{X_{\mathcal{L}}}(x_{\mathcal{L}}) = -x_{\mathcal{L}} \cdot \log(x_{\mathcal{L}}^2/2).$$

Figure 3.3 shows a plot of $f_{X_{\mathcal{L}}}$ (thick magenta line) over $\mathcal{L}$. Using (3–5) and the

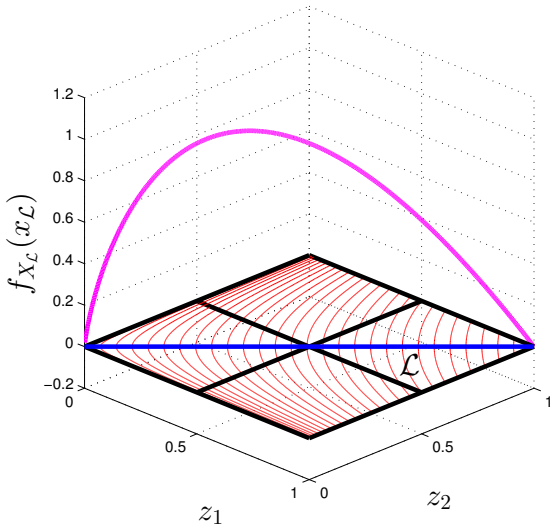


Figure 3.3: Plot of the pdf $f_{X_{\mathcal{L}}}$ (thick magenta line) over $\mathcal{L}$ (thick blue line).

arc length formula applied to the parameterization of the contour $\pi^{-1}(x_{\mathcal{L}})$ for $x_{\mathcal{L}} \in \mathcal{L}$, we obtain for any $x_C \in \pi^{-1}(x_{\mathcal{L}})$:

$$f^{ansatz}_{X_C|X_{\mathcal{L}}}(x_C|x_{\mathcal{L}}) = \left(\int_{x_{\mathcal{L}}^2/2}^{1} \sqrt{1 + \frac{x_{\mathcal{L}}^4}{4z_1^4}}\, dz_1 \right)^{-1}.$$

With the components of (3–4) specified, $P_Z(A)$ can now be approximated for any measurable $A \subset \Gamma = [0,1]^2$. We partition Γ into 4 measurable regions of equal area as shown in Figure 3.1a, namely the northwest (A_1), northeast (A_2), southwest (A_3), and southeast (A_4) regions. If the proposed methodology is able to recover the true probability law of Z given f_Q, we expect the approximations of $P_Z(A_i)$ to be close to 0.25. Numerical calculations reveal that $P_Z(A_i) \approx 0.2886, 0.2459, 0.1770, 0.2886$ for $i = 1, \ldots, 4$, respectively which shows that $P_Z(A_3)$ in particular underestimates the true probability by a significant amount.

To understand the values obtained for $P_Z(A_i)$, we plot each term in the integrand in (3–4), namely $\int_{\pi^{-1}(x_{\mathcal{L}}) \cap A_i} f^{ansatz}_{X_C|X_{\mathcal{L}}}(x_C|x_{\mathcal{L}})\, dx_C$ for $i = 1, \ldots, 4$ and $f_{X_{\mathcal{L}}}(x_{\mathcal{L}})$, both as a function of $x_{\mathcal{L}}$ in Figure 3.4. Each subplot corresponds to each quadrant. The

first term of the integrand amounts to the proportion of the generalized contour inside the quadrant and is displayed with red solid curves. The blue dashed curves meanwhile refer to $f_{x_{\mathcal{L}}}$ for values of $x_{\mathcal{L}}$ contained within the quadrant.

It is worth noting that $P_Z(A_1)$ and $P_Z(A_4)$ are the largest because of the following reasons. Firstly, the regions A_1 and A_4 are spanned by $0 \leq x_{\mathcal{L}} \leq 1$ whereas $f_{X_{\mathcal{L}}}(x_{\mathcal{L}})$ attains its maximum in the neighborhood $0.2 \leq x_{\mathcal{L}} \leq 0.8$. Secondly, these two quadrants possess more contours than regions A_2 and A_3 because every contour passing through the 2 former regions always passes through one of the latter regions. This implies that A_1 and A_4 comprise a wider range of values of $x_{\mathcal{L}}$ as evidenced by the domains of the corresponding subplots in Figure 3.4.

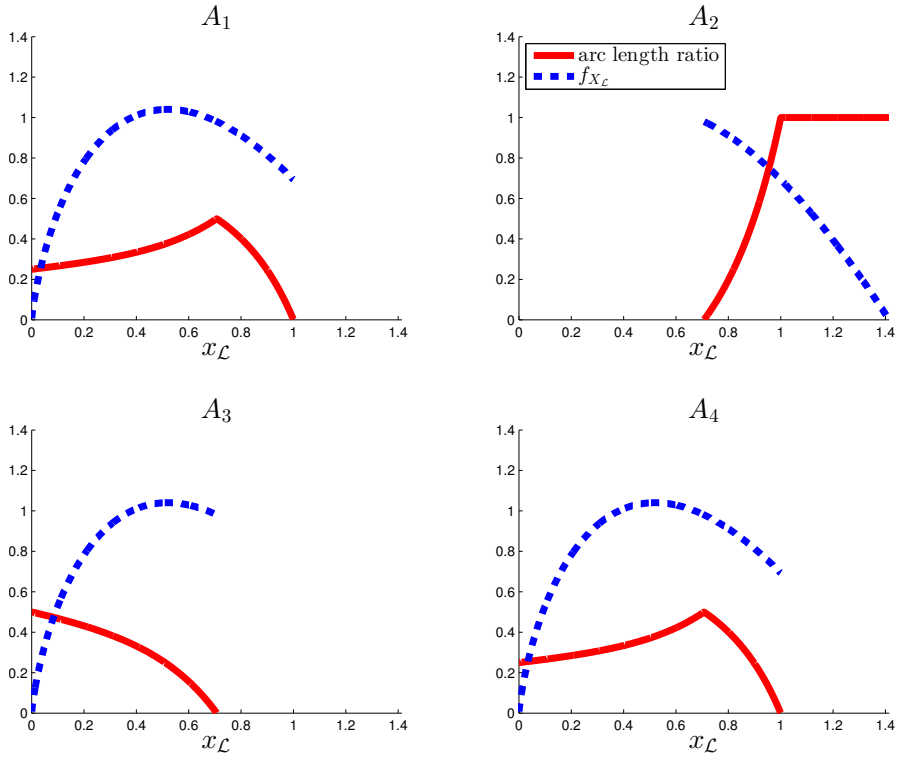


Figure 3.4: Plot of the integrand terms in (3–4) as a function of $x_{\mathcal{L}}$ for each A_i. The blue dashed curve is $f_{X_{\mathcal{L}}}$ while the red solid curve is the proportion of each contour contained in A_i.

Finally, we numerically demonstrate that for this example, the actual con-

ditional pdf along the generalized contours is not necessarily uniform (see Appendix A for the methodology). The actual conditional pdf results from disintegrating the measure P_Z associated with the uniform distribution on Z. The left panel of Figure 3.5 shows three contours of Q, the middle panel shows the corresponding actual pdfs along each contour, whereas for comparison, the right panel shows the pdf according to the ansatz (3–5). As the concavity of the contours increases, the actual pdf along the contour becomes less uniform. We see that the pdf on each contour resulting from the disintegration is not necessarily uniform even though Z is uniformly distributed on Γ. In addition, Figure 3.5 corroborates the results in Figure 3.4: the majority of the contours residing in A_3 exhibit high concavity which lead to the underestimation of $P_Z(A_3)$. If instead the $f_{X_\mathcal{C}|X_\mathcal{L}}(x_\mathcal{C}|x_\mathcal{L})$ as obtained in the middle panel of Figure 3.5 were used in computing $P_Z(A)$ in (3–4), the resulting Z would be uniformly distributed. Example 11 therefore underscores the point made at the beginning of this section that *if the objective is to recover the true pdf on Z, then the ansatz* (3–5), *which only ensures that the constructed measure/density is a pullback measure, may be insufficient to recover the structure of the true pdf in directions not informed by the data*.

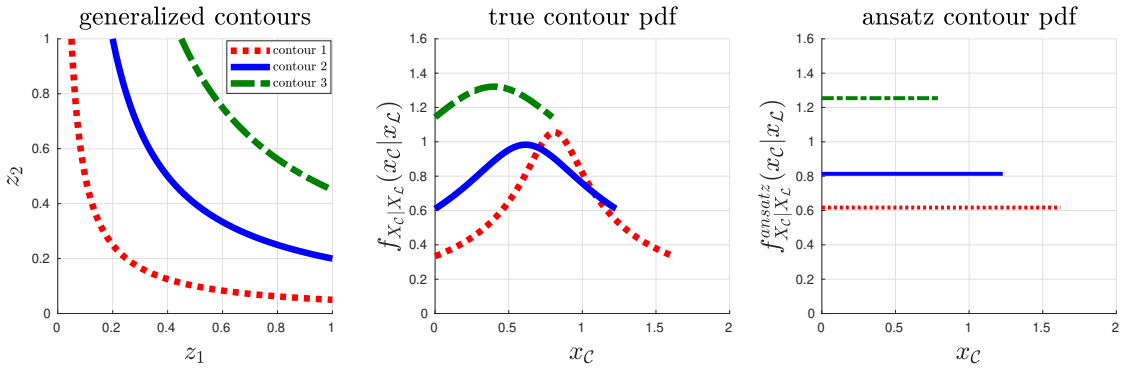


Figure 3.5: Left panel: selected contours of Q in Example 11. Middle panel: corresponding actual pdf along the contour. Right panel: corresponding pdf using the ansatz (3–5).

Furthermore, we evaluate how (3–5) fares as a means of regularizing across

other pdfs on $X_C|X_{\mathcal{L}}$ in the absence of information on Z. Suppose instead that Z_1, Z_2 are independent with $Z_1 \sim Beta(\nu_1, \nu_2)$ and $Z_2 \sim Beta(\tau_1, \tau_2)$. Notice that the pdf of Z in Example 11 is a special case with $\nu_1, \nu_2, \tau_1, \tau_2$ all being equal to 1. The 3 subplots of Figure 3.6 show the conditional pdf along the solid blue contour in Figure 3.5 (contour 2) for different combinations of the parameters for Z_1, Z_2. The plots reveal that the pdfs on the contours can be very complex and suggest that using the pdf along the contour as the only means of regularizing against other solutions to this inverse problem may be insufficient.

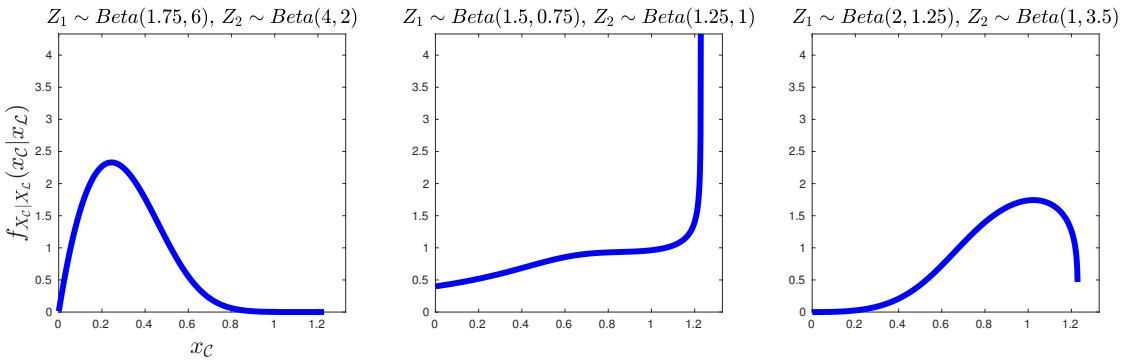


Figure 3.6: Conditional pdf $X_C|X_{\mathcal{L}}$ on contour 2 in Figure 3.5 where $Z_1 \sim Beta(\nu_1, \nu_2), Z_2 \sim Beta(\tau_1, \tau_2)$.

3.2.1.3 Pdfs with respect to non-Lebesgue measures

In contrast to Section 3.2.1.2, [14] chose the measures in (3–1) as follows: μ_Z is taken to be Lebesgue, μ_Q is defined to be the pushforward measure of μ_Z through Q (i.e. for measurable $B \subset \mathcal{D}$, $\mu_Q(B) = \mu_Z(Q^{-1}(B))$), while $\mu_{X_{\mathcal{L}}}$ is computed as $\mu_{X_{\mathcal{L}}}(K) = \mu_Q(Q(A))$ where $K = \pi(A)$. Note that μ_Q may not be Lebesgue especially if Q is nonlinear which implies that the same holds for $\mu_{X_{\mathcal{L}}}$. As before, $\mu_{X_C|X_{\mathcal{L}}}$ results from (3–2) since $\mu_Z, \mu_{X_{\mathcal{L}}}$ are now specified; however, if $\mu_{X_{\mathcal{L}}}$ is not be the Lebesgue measure, then neither is $\mu_{X_C|X_{\mathcal{L}}}$.

To further distinguish this design approach from that of Section 3.2.1.2, we emphasize that for fixed $x_{\mathcal{L}}$, $\rho^{ansatz}_{X_C|X_{\mathcal{L}}}(x_C|x_{\mathcal{L}})$ is constant with respect to $\mu_{X_C|X_{\mathcal{L}}}$. If we were to express this pdf with respect to the Lebesgue measure, i.e. we seek $f^{ansatz}_{X_C|X_{\mathcal{L}}}(x_C|x_{\mathcal{L}})$ such that $\rho^{ansatz}_{X_C|X_{\mathcal{L}}}(x_C|x_{\mathcal{L}})\, d\mu_{X_C|X_{\mathcal{L}}} = f^{ansatz}_{X_C|X_{\mathcal{L}}}(x_C|x_{\mathcal{L}})\, dx_C$, then $f^{ansatz}_{X_C|X_{\mathcal{L}}}(x_C|x_{\mathcal{L}})$ may not be constant unlike in (3–5). In addition, this approach offers a computationally efficient approximation for the case when $Z, Q(Z)$ are high-dimensional that does not require explicitly constructing $\mu_{X_C|X_{\mathcal{L}}}$ from (3–2) [14, Algorithm 1]. The publicly available code[1] is based on this design.

Demonstration of method on an example. Even with this selection of the measures in (3–1), it is shown that the ansatz (3–3) may still be unable to recover the true law of Z. We show this in the next example where the true distribution of Z in Example 11 is modified. The established results in Figures 3.5 and 3.6 of Section 3.2.1.2 will be invoked to aid in the discussion.

Example 12. We revisit the setup in Example 11 but with the true distribution of Z altered to $Z_1 \sim Beta(\nu_1, \nu_2), Z_2 \sim Beta(\tau_1, \tau_2)$, independent. Given the pdf ρ_Q of Q, it is shown that the probability law of Z cannot be recovered with the choice of ansatz (3–3).

In order to apply the disintegration theorem (3–1), $\mu_{X_C|X_{\mathcal{L}}}$ needs to be first identified from (3–2). Let P_Z^{Unif} be the probability measure of the uniform distribution on Z where Z_1, Z_2 are independent. For measurable $A \subset \Gamma$, (3–4) and the results in Section 3.2.1.2 yield the disintegration

$$P_Z^{Unif}(A) = \int_{\pi(A)} \int_{\pi^{-1}(x_{\mathcal{L}})\cap A} f^{Unif}_{X_C|X_{\mathcal{L}}}(x_C|x_{\mathcal{L}})\; f^{Unif}_{X_{\mathcal{L}}}(x_{\mathcal{L}})\; dx_C\; dx_{\mathcal{L}} \tag{3–7}$$

where $f^{Unif}_{X_C|X_{\mathcal{L}}}$ are the pdfs displayed in the middle panel of Figure 3.5. Since μ_Z is

[1] https://github.com/UT-CHG/BET

Lebesgue, $P_Z^{Unif}(A) = \mu_Z(A)$ which implies that

$$d\mu_{X_C|X_\mathcal{L}} = f^{Unif}_{X_C|X_\mathcal{L}}(x_C|x_\mathcal{L})\ dx_C \tag{3–8}$$

by comparing (3–7) with (3–2).

We now return to the setup in Example 12 where Z_1, Z_2 are actually independent beta random variables. Denote by P_Z^{Beta} the probability measure on Z corresponding to its true law. From the results in Section 3.2.1.2, its actual disintegration is

$$P_Z^{Beta}(A) = \int_{\pi(A)} \int_{\pi^{-1}(x_\mathcal{L})\cap A} f^{Beta}_{X_C|X_\mathcal{L}}(x_C|x_\mathcal{L})\ f^{Beta}_{X_\mathcal{L}}(x_\mathcal{L})\ dx_C\ dx_\mathcal{L} \tag{3–9}$$

where $f^{Beta}_{X_C|X_\mathcal{L}}$ is obtained similarly as the pdfs plotted in the panels of Figure 3.6. On the other hand, applying the ansatz (3–3) to (3–1), the inverse problem solution is

$$P_Z^{ansatz}(A) = \int_{\pi(A)} \int_{\pi^{-1}(x_\mathcal{L})\cap A} \rho^{ansatz}_{X_C|X_\mathcal{L}}(x_C|x_\mathcal{L})\ \rho_{X_\mathcal{L}}(x_\mathcal{L})\ d\mu_{X_C|X_\mathcal{L}}\ d\mu_{X_\mathcal{L}} \tag{3–10}$$

where $\rho^{ansatz}_{X_C|X_\mathcal{L}}(x_C|x_\mathcal{L})$ is a constant for fixed $x_\mathcal{L}$. If (3–10) is able to recover the true law of Z, it must be that $P_Z^{Beta}(A) = P_Z^{ansatz}(A)$ for any measurable $A \subset \Gamma$. Since the probability measure on $X_\mathcal{L}$ is the same regardless of the choice of $\mu_{X_\mathcal{L}}$, it follows that $\rho_{X_\mathcal{L}}(x_\mathcal{L})\ d\mu_{X_\mathcal{L}} = f^{Beta}_{X_\mathcal{L}}(x_\mathcal{L})\ dx_\mathcal{L}$. Hence, by comparing (3–10) with (3–9), the ansatz is able to recover the true law if and only if

$$\rho^{ansatz}_{X_C|X_\mathcal{L}}(x_C|x_\mathcal{L})\ d\mu_{X_C|X_\mathcal{L}} = f^{Beta}_{X_C|X_\mathcal{L}}(x_C|x_\mathcal{L})\ dx_C \tag{3–11}$$

which is equivalent to

$$\rho^{ansatz}_{X_C|X_\mathcal{L}}(x_C|x_\mathcal{L}) = \frac{f^{Beta}_{X_C|X_\mathcal{L}}(x_C|x_\mathcal{L})}{f^{Unif}_{X_C|X_\mathcal{L}}(x_C|x_\mathcal{L})} \tag{3–12}$$

using (3–8).

We now proceed by contradiction. Consider contour 2 (solid blue) in the left panel of Figure 3.5 and let $\nu_1, \nu_2, \tau_1, \tau_2$ be any of the values utilized in the plots of Figure 3.6. The pdf $f^{Beta}_{X_C|X_{\mathcal{L}}}(x_C|x_{\mathcal{L}})$ on this contour could then be any of these plotted pdfs. It then follows from the middle panel of Figure 3.5 and each of the panels of Figure 3.6 that $\frac{f^{Beta}_{X_C|X_{\mathcal{L}}}(x_C|x_{\mathcal{L}})}{f^{Unif}_{X_C|X_{\mathcal{L}}}(x_C|x_{\mathcal{L}})}$ is not constant for $x_{\mathcal{L}}$ fixed which contradicts the assumption on $\rho^{ansatz}_{X_C|X_{\mathcal{L}}}(x_C|x_{\mathcal{L}})$. The ansatz is therefore unable to recover the true law of Z.

Despite the examples presented in Sections 3.2.1.2 and 3.2.1.3, the methodology in [12, 14, 15] can still be useful in physical applications in that it can serve as a first model for the unknown pdf of Z. This can be later tuned or scrutinized for plausibility depending on available information on Z.

3.2.1.4 Incorporating prior information

An alternative to the method in Section 3.2.1.1 under the same specifications on the inverse problem has been developed in [16]. It assumes that some information on Z is available in the form of a prior pdf $\rho^{prior}_Z(Z)$. Analogous to Bayes' theorem, the solution to the inverse problem is a posterior pdf ρ^{post}_Z on Z that is constructed as

$$\rho^{post}_Z(Z) = \rho^{prior}_Z(Z) \cdot \frac{\rho_Q(Q(Z))}{\rho^{Q(prior)}_Q(Q(Z))}$$

for $Z \in \Gamma$ where $\rho_Q(\cdot)$ is the given pdf of Q while $\rho^{Q(prior)}_Q(\cdot)$ is the pdf of Q that is obtained by propagating ρ^{prior}_Z through the model. This solution was derived using the disintegration theorem based on conditional densities which is more general than Theorem 11 based on generalized contours. Although this method

does not explicitly deal with contours, it is related to the method in Section 3.2.1 in that $\rho_Z^{prior}(Z)$ implies a pdf $\rho_{X_C|X_\mathcal{L}}(x_C|x_\mathcal{L})$ on the generalized contours that is not necessarily uniform. We note that a sufficient condition for this method to recover the true pdf ρ_Z on Z would be if $\rho_Z^{prior} = \rho_Z$, signifying no gain in information. A more general sufficient condition only requires that the conditional pdf along the contours arising from the disintegration of ρ_Z^{prior} and the true pdf ρ_Z need to be equal [16, Sections 3, 7.3]. The methodology proposed in [16] only stresses the need for additional information to be specified on Z in order to solve the inverse problem.

3.2.2 Parametric representations of the unknown random field

This section elaborates on the second approach [10] which estimates the coefficient field of a differential equation given observations of the solution field. Section 3.2.2.1 reviews the methodology and establishes its similarity with the above inverse problem. Section 3.2.2.2 meanwhile explores two examples utilizing this method. It is shown that the probability law and the truncation level of the random variables arising from the Karhunen-Loève expansion of the solution field may be inadequate to characterize the coefficient field.

3.2.2.1 Review of methodology

Consider the stochastic equation $\mathcal{L}(U(x,\omega)) = 0$ defined on the probability space $(\Omega, \mathcal{F}, P)$ for $\omega \in \Omega, x \in D$ where the operator $\mathcal{L}$ characterizes a stochastic differential equation that depends on the random field $A(x,\omega)$. The inverse problem tackled by [10, 22, 60, 81] arising from this set-up is:

Given observations $\hat{U}$ of the solution $U(x,\omega)$, estimate the unknown field $A(x,\omega)$.

Although this problem appears different from the one posed in Section 3.1, they are in fact conceptually identical. In practice, the spatial domain is discretized so that the random fields A and U are represented as random vectors characterized by the finite-dimensional distributions of the random fields. The inverse problem is now reminiscent of the problem above. The use of the finite-dimensional distribution of $A(x,\omega)$ to characterize the field itself can be justified under the mild assumption that $A(x,\omega)$ has almost surely continuous sample paths or that it satisfies the Hölder continuity condition [36, Theorem 3.1]; see [36, 79] for more details. In what follows, it will be assumed that either assumption on A holds.

Despite this connection, proposed methodologies [10, 22, 60, 81] estimate the unknown random field by finding a finite-dimensional noise approximation, that is, $A(x,\omega) \approx A(x, Z(\omega))$ where Z is a random vector. Essentially, these methodologies construct a parametric representation of the unknown field A in which the law, the dimension of Z, and the functional form of $A(x,Z)$ are to be determined. In [22], each observed sample of U is used to acquire samples of A via optimization; these samples of A then serve to calibrate a polynomial chaos expansion (PCE) [36] for A. This optimization procedure may result in non-unique global minima yet its implications on the fitted PCE were not addressed. In [60], observed samples of U are utilized to obtain a truncated PCE of U. The unknown A is then expressed in terms of this stochastic basis. The limitations of PCE are well known [30, 31] and furthermore, as the stochastic basis for U is used as the stochastic basis for the unknown A, the truncation level of

the PCE for U can be insufficient for the PCE of A. The following illustrations tackles some of these issues.

In order to address the limitations of a PCE model for A, [10, 81] proposed to express the unknown random field as a sparse grid representation $\widetilde{A}^N(x, Y)$ of level N. This is represented as

$$\widetilde{A}^N(x, y) = \sum_{j=1}^{N} v(x, \mathbf{y}_j)\psi_j(y), \quad x \in \mathbb{R}^d, \quad y \in \Gamma \subset \mathbb{R}^k \tag{3–13}$$

for Γ bounded. Here, $Y \in \Gamma$ is a random vector whose dimension is not necessarily identical to that of Z and whose law has to be specified. In addition, $\{\mathbf{y}_j\}_{j=1}^N \subset \Gamma$ are the N sparse grid nodes, $\psi_j(y)$ are specified interpolating functions on Γ, while $v(x, \mathbf{y}_j)$ is a deterministic function of x. It only remains to address the following: How must the dimension and the law of the random vector Y be specified for the nodal values $v(x, \mathbf{y}_j)$ to be approximated through optimization?

Both [10, 81] are similar in that they employ the finite-dimensional model (3–13) for A yet they differ in how the dimension and the law of Y are prescribed. In [81], the dimension of Y is postulated to be $k = 1, 2$ in the numerical examples while the choice of law for Y is downplayed. This was demonstrated with a numerical example in which various distributions, with various support, were chosen for $Y \in \mathbb{R}^1$ to compute moments of $\widetilde{A}^N(x, Y)$. In contrast, [10] pursued an approach in selecting the dimension and law of Y based on the Karhunen-Loève (KL) expansion [36] and is detailed as follows. Consider the elliptic system

$$-\nabla \cdot (A(x, \omega)\nabla U(x, \omega)) = f(x), \quad U(x, \omega) = 0 \text{ on } \partial D \tag{3–14}$$

where $x \in D \subset \mathbb{R}^d$, $\omega \in \Omega$. Given observations $\hat{U}$ (possibly noisy) of U, the objective is to approximate the unknown field A by solving the optimization

problem

$$\min \frac{1}{2}\|U - \hat{U}\|^2 + \frac{\beta}{2}\|A\|^2 \tag{3–15}$$

for U and A under the constraint that they satisfy (3–14) and that A satisfies ellipticity constraints. The second term in (3–15) regularizes the solution A through the parameter $\beta > 0$ while the norms in (3–15) are formed using a tensor product of norms for Sobolev spaces and the $L^2(\Omega)$ norm for the probability space. A KL expansion of $\hat{U}$ is then performed to obtain for $\omega \in \Omega$

$$\hat{U}(x,\omega) = \hat{u}_0(x) + \sum_{k=1}^{\infty} \sqrt{\hat{\lambda}_k}\hat{\phi}_k(x)Y_k(\omega) \tag{3–16}$$

where $E[Y_k] = 0$, $E[Y_k^2] = 1\ \forall k$, $E[Y_kY_j] = 0$ for $k \neq j$ and $\hat{\lambda}_k$, $\hat{\phi}_k(x)$ are the eigenvalues and eigenfunctions of the covariance function of $\hat{U}(x,\omega)$. It is assumed [10, p. 10] that $\{Y_i\}_{i=1}^{\infty}$ form a basis for $L^2(\Omega)$ in the sense that every random variable with finite variance can be expressed as a linear combination of $\{Y_i\}_{i=1}^{\infty}$. As a consequence, since the optimal random field A^* for A in (3–15) satisfies $A^*(x,\cdot) \in L^2(\Omega)\ \forall x \in D$, A^* can be written as

$$A^*(x,\omega) = a_0(x) + \sum_{k=1}^{\infty} a_k(x)Y_k(\omega). \tag{3–17}$$

The deterministic functions $\{a_k(x)\}_{k=0}^{\infty}$ are determined by solving (3–15) while $\{Y_k(\omega)\}_{k=1}^{\infty}$ are obtained from (3–16). Due to the dependence of A^* on an infinite number of random variables, (3–17) is usually referred to as the solution to the infinite-dimensional problem.

For practical numerical implementation, the KL expansion in (3–16) is truncated to only consider $\{Y_k\}_{k=1}^{M}$, $M \ll \infty$ based on the decay of λ_k. The optimal solution then takes on the form $A^\dagger(x,\omega) = A^\dagger(x, Y_1(\omega), \ldots, Y_M(\omega))$, i.e. the optimal $A^\dagger$ that is sought from (3–15) is a function of $Y_1, \ldots, Y_M$ only. $A^\dagger$ is typically referred to as the solution to the finite noise problem. It was shown in [10] that

the sequence of minimizers of the finite noise problem has a subsequence that converges weakly to a minimizer of the infinite-dimensional problem under the assumption (3–17). Due to this truncation, $A^\dagger(x, Y_1(\omega), \ldots, Y_M(\omega))$ is not expected to be linear in $\{Y_k\}_{k=1}^M$ as in (3–17); as such, a sparse grid representation (3–13) for $A^\dagger(x, Y_1(\omega), \ldots, Y_M(\omega))$ using linear hat functions for ψ_j was considered in [10] to accommodate smoothness conditions on $A^\dagger$ as a function of $Y_1, \ldots, Y_M$.

In summary, [10] parameterized the unknown field $A(x, \omega)$ by a random vector Y whose dimension and distribution are based on the random variables arising from the truncated KL expansion of the observed samples $\hat{U}$. The approach pursued by [10] in specifying the random vector Y is reasonable because it is shown in [4] that if A in (3–14) depends on Y, the response U is analytical in Y. But challenges may surface from this approach which we demonstrate through examples in the next section. We clarify that the issue does not lie with the use of a sparse grid approximation for A but with the choice of the random vector Y used to construct $A^*(x, \omega)$ in (3–17).

3.2.2.2 Demonstration of method on examples

This section investigates the implications of the approach in [10]. First, unlike the PCE, the infinite set of random variables $\{Y_k(\omega)\}_{k=1}^\infty$ in (3–16) do not form a stochastic basis for $L^2(\Omega)$. For this reason, if $A \in L^2(\Omega)$, the truncated set $\{Y_k(\omega)\}_{k=1}^M$ might not even be adequate to characterize the true field A because it is unlikely that $A \in L^2(\Omega) \cap \text{span}(\{Y_k\}_{k=1}^M)$, as Example 13 will clarify. Second, even if it can be analytically shown that the random variables $\{Y_k\}_{k=1}^\infty$ in the KL expansion of U also characterize A, the truncation level employed for the practical implementation of U might not be sufficient for the implementation of A.

Example 14 lends support to this claim.

Example 13. Consider the stochastic ODE for $x \in [0,1]$, $\omega \in \Omega$:

$$-\frac{d}{dx}(A(x,\omega)\cdot\frac{d}{dx}U(x,\omega)) = 0, \quad U(0,\omega) = 0, \quad U(1,\omega) = \int_0^1 \frac{1}{A(y,\omega)}\,dy. \quad (3\text{–}18)$$

The true coefficient random field is modeled as a translation process [36] $A(x,\omega) = \alpha + (\beta-\alpha)\cdot F_{beta}^{-1}(\Phi(G(x,\omega)))$ where $\alpha = 4, \beta = 20$, F_{beta}^{-1} is the inverse cumulative distribution function (cdf) of $Beta(1,3)$, Φ is the cdf of $N(0,1)$, and $G(x,\omega)$ is a zero mean, unit variance stationary Gaussian process with Matérn covariance, i.e. $E[G(s,\cdot)G(t,\cdot)] = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}|s-t|}{\ell}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}|s-t|}{\ell}\right)$, $s,t \in [0,1]$ where K_ν is the modified Bessel function of the second kind and $\nu = \frac{5}{2}, \ell = 0.03$. To formulate the inverse problem, we approximate $A(x,\omega)$ using observed samples of $U(x,\omega)$. It is shown that the random variables arising from the KL expansion of $U(x,\omega)$ are inadequate to characterize $A(x,\omega)$.

We generate 10000 noiseless samples of $U(x,\omega)$ by solving (3–18) for each sample of $A(x,\omega)$. The analytical solution is given by $U(x,\omega) = \int_0^x \frac{1}{A(y,\omega)}\,dy = \int_0^x B(y,\omega)\,dy$ where $B(y,\omega) \coloneqq \frac{1}{A(y,\omega)}$. A KL expansion of $U(x,\omega)$ is then performed to obtain (3–16). Figure 3.7 shows samples of $A(x,\omega)$ together with their respective samples $U(x,\omega)$ while Figure 3.8 displays histograms of the samples Y_k in the KL expansion of $U(x,\omega)$ corresponding to the four largest eigenvalues. Note that the integral equations that need to be solved to obtain the eigenvalues and eigenfunctions of the covariance function of $U(x,\omega)$ have to be discretized solely for numerical implementation. Hence, we are essentially performing an eigendecomposition of the covariance matrix $\mathbf{K}$ of $\mathbf{U} = (U(x_1,\omega),\ldots,U(x_M,\omega))$ where $\mathbf{K}_{ij} = Cov(U(x_i),U(x_j))$ for $\{x_i\}_{i=1}^M \subset [0,1]$. In the following, we used an extremely fine mesh with $x_{i+1} - x_i = 0.005$ such that $M = 201$. Further decreasing the mesh size does not alter the conclusions that follow.

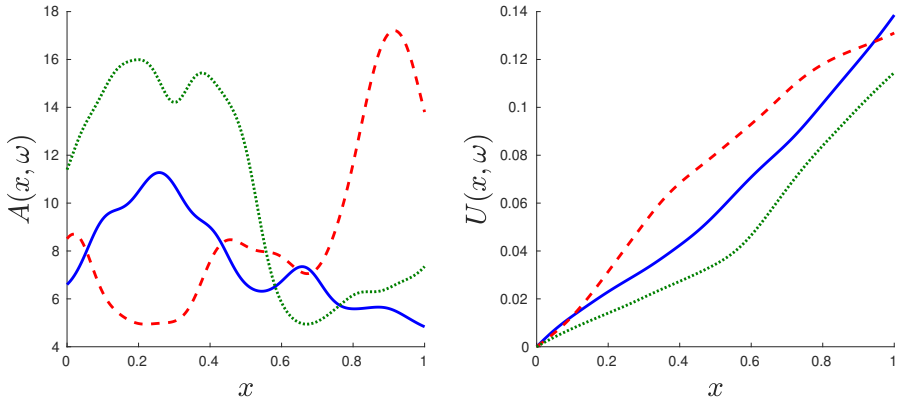


Figure 3.7: Left panel: samples of $A(x,\omega)$. Right panel: corresponding samples of $U(x,\omega)$ via (3–18).

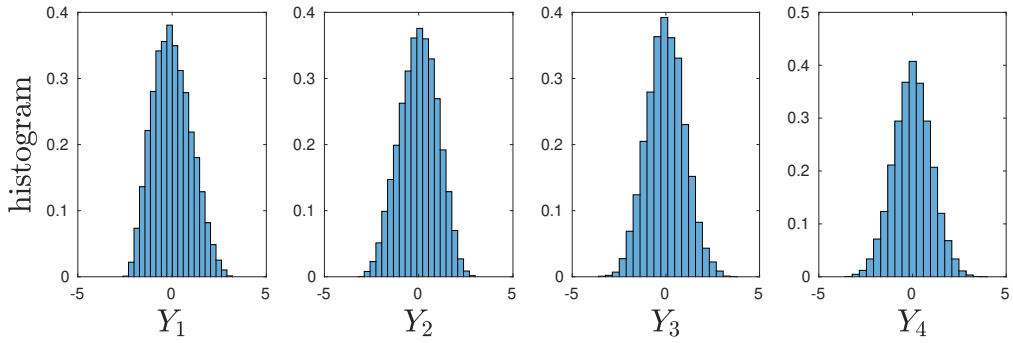


Figure 3.8: Histograms of the first 4 random variables in the KL expansion of U.

In order to approximate the unknown field $A(x,\omega)$, we solve the optimization problem by minimizing the $L^2([0,1]) \otimes L^2(\Omega)$ norm, i.e.

$$\min_{\widetilde{A}} \int_0^1 E\left[\left|U(x,\cdot) - \int_0^x \frac{1}{\widetilde{A}(y,\cdot)}\, dy\right|^2\right] dx = \min_{\widetilde{B}} \int_0^1 E\left[\left|U(x,\cdot) - \int_0^x \widetilde{B}(y,\cdot)\, dy\right|^2\right] dx. \tag{3–19}$$

We tackle the infinite-dimensional problem to obtain solutions of the form (3–17) instead of (3–13) because we do not perform truncation. As a first attempt, we characterize the random field B (instead of A) in the spirit of (3–17) to yield the expression

$$\widetilde{B}(x,\omega) = b_0(x) + \sum_{k=1}^{M} b_k(x) Y_k(\omega) \tag{3–20}$$

where $\{b_k(x)\}_{k=0}^{M}$ are to be determined. The quantity M here is not the truncation

level according to the decay of the eigenvalues of the covariance function of $U(x,\omega)$ but results from the discrete implementation of the KL expansion as discussed earlier.

It is observed from the optimization problem on the right-side of (3–19) that the minimum can be obtained if the unknown deterministic functions $b_k(x)$ satisfy

$$\sqrt{\lambda_k}\phi_k(x) = \int_0^x b_k(y)\,dy \quad \text{for} \quad k = 0,\ldots,M \tag{3–21}$$

where λ_k, $\phi_k(x)$ are eigenvalues and eigenfunctions of $Cov(U(x),U(y))$. In other words, the minimizer is achieved if $\widetilde{B}(x,\omega) = \frac{dU(x,\omega)}{dx}$ in the mean square sense [36] which results in $Cov(U(x),U(y)) = \int_0^x \int_0^y Cov(\widetilde{B}(s),\widetilde{B}(t))\,ds\,dt$ using second moment calculus. This equality is what we would have obtained from the analytic solution for U. This implies that by setting $b_k(x)$ as in (3–21), the approach pursued through (3–20) recovers the second-order statistics of B and hence that of A.

We verify this numerically by noticing that solving for $\{b_k(x)\}_{k=1}^M$ in (3–19) results in a matrix least squares problem given the 10000 realizations of $U(x,\omega)$. Figure 3.9 shows plots of the second-order statistics of the true field $A(x,\omega)$ together with the second-order statistics of the numerical solution to the inverse problem $\widetilde{A}(x,\omega) = (\widetilde{B}(x,\omega))^{-1}$ with condition (3–21) on $b_k(x)$. In particular, in the left panel of Figure 3.9, the red dotted and blue dashed curves represent $E[\widetilde{A}(x)]$ and $E[A(x)]$ respectively, while the magenta solid and green dashed curves represent $Var[\widetilde{A}(x)]$ and $Var[A(x)]$ respectively. Only 2 curves are visible because the mean and variance of $A(x,\omega)$ and $\widetilde{A}(x,\omega)$ are almost indistinguishable. On the other hand, the right panel shows the absolute value of the difference between $c(s,t) \coloneqq Cov(A(s),A(t))$ and $\widetilde{c}(s,t) \coloneqq Cov(\widetilde{A}(s),\widetilde{A}(t))$ over $s,t \in [0,1]$ which

are in good agreement.

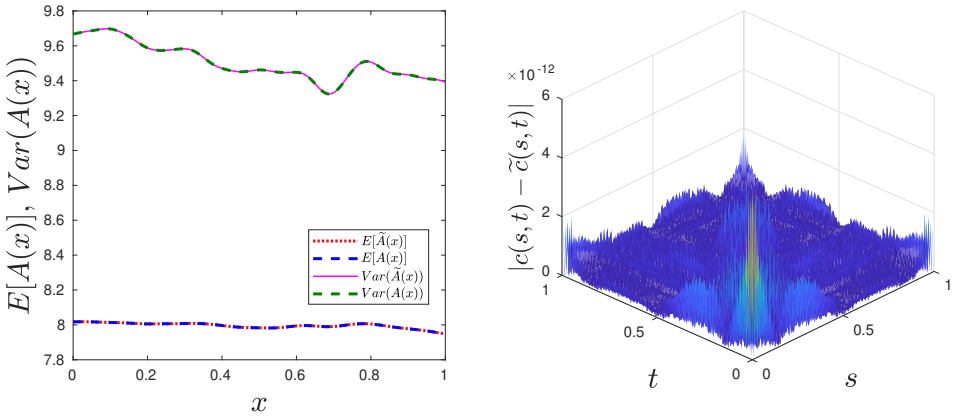


Figure 3.9: Left panel: Comparison between $E[A(x)]$ (blue dashed line) and $E[\widetilde{A}(x)]$ (red dotted line) and $Var(\widetilde{A}(x))$ (green dashed line) and $Var(\widetilde{A}(x))$ (magenta solid line). Right panel: Plot of the discrepancy between the covariance of $A(x)$ and $\widetilde{A}(x)$, i.e. $|Cov(A(s), A(t)) - Cov(\widetilde{A}(s), \widetilde{A}(t))| = |c(s,t) - \widetilde{c}(s,t)|$. $\widetilde{A}(x)$ is approximated under the first attempt.

As a second attempt, we pursue a more typical approach in which we characterize the unknown field A directly instead of its inverse, consistent with [10]. To parameterize A, we consider

$$\widetilde{A}(x,\omega) = a_0(x) + \sum_{k=1}^{M} a_k(x) Y_k(\omega) \tag{3–22}$$

where $\widetilde{A}$ solves the optimization problem (3–19) and $\{a_k(x)\}_{k=0}^{\infty}$ are to be determined. The unknown field is approximated by solving

$$\min_{\widetilde{A}} \int_0^1 E[|A(x,\cdot) - \widetilde{A}(x,\cdot)|^2]\, dx \tag{3–23}$$

for $\widetilde{A}(x)$ under the norm $L^2([0,1])\otimes L^2(\Omega)$ where $A(x,\cdot)$ refers to the true field specified in Example 13. By using the mean value theorem on $U(x,\cdot) - \int_0^x \frac{1}{\widetilde{A}(y,\cdot)}\, dy$, the boundedness of $A(x,\cdot)$, and assumptions on the boundedness of $\widetilde{A}(x,\cdot)$, minimizing (3–23) under the constraint that $\widetilde{A}(y,\cdot) > 0$ provides an upper bound for the objective function in (3–19).

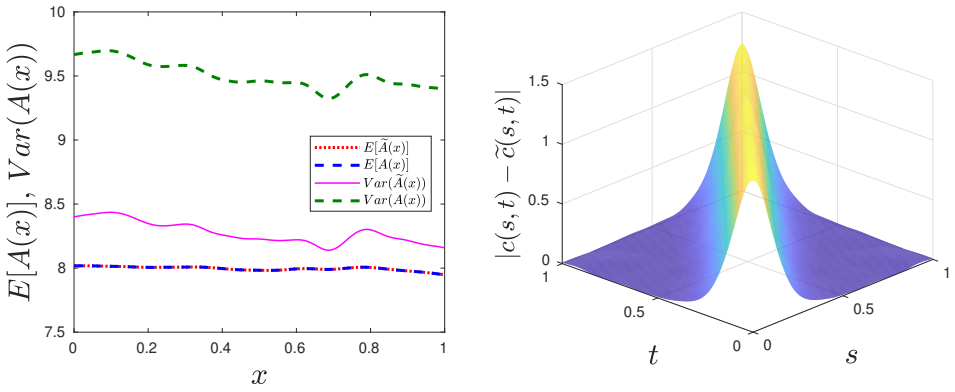


Figure 3.10: Left panel: Comparison between $E[A(x)]$ (blue dashed line) and $E[\widetilde{A}(x)]$ (red dotted line) and $Var(\widetilde{A}(x))$ (green dashed line) and $Var(\widetilde{A}(x))$ (magenta solid line). Right panel: Plot of the discrepancy between the covariance of $A(x)$ and $\widetilde{A}(x)$, i.e. $|Cov(A(s), A(t)) - Cov(\widetilde{A}(s), \widetilde{A}(t))| = |c(s,t) - \widetilde{c}(s,t)|$. $\widetilde{A}(x)$ is approximated under the second attempt.

Similar to the first attempt, a matrix least squares problem utilizing the 10000 samples of the true $A(x,\omega)$ was solved to optimize (3–23). Figure 3.10 shows, with identical legend to Figure 3.9, the second order statistics of $\widetilde{A}(x,\omega)$ under the second approach in comparison with that of $A(x,\omega)$. While $E[A(x)]$ and $E[\widetilde{A}(x)]$ are almost similar, there is a considerable difference in the variance and covariance functions between the 2 random fields. Since we minimized (3–23) instead of the optimization problem on the left-side of (3–19), whatever optimal solution we acquire from (3–19) of the form (3–22) cannot do better in matching the statistics of $A(x,\omega)$ than what we have already achieved. The derivations we have presented in the first attempt above do not hold anymore for the current attempt due to the nonlinearity of U with respect to A. In addition, increasing the number of samples of $A(x,\omega)$ to 100000 does not alter the results. This underscores that the random variables $\{Y_k\}_{k=1}^{\infty}$ arising from the KL expansion of $U(x,\omega)$ do not necessarily form a stochastic basis for $L^2(\Omega)$. This implies that $\{Y_k\}_{k=1}^{\infty}$ may be inadequate to characterize the unknown $A(x,\omega)$, even when no truncation on the basis of decay of eigenvalues is performed. However, even

when both $U(x,\omega)$ and $A(x,\omega)$ are analytically characterized by the same random variables, the KL expansion of U has to be truncated for numerical implementation. In [10], the finite set $\{Y_k\}_{k=1}^M$ resulting from this truncation is then used to characterize A. The following example shows that the truncation level for U may not be sufficient for A such that the solution to the inverse problem under this approach underestimates statistics of A.

Example 14. Consider the stochastic ODE for $x \in [0,1],\ \omega \in \Omega$:

$$\frac{d}{dx}U(x,\omega) = A(x,\omega), \quad U(0,\omega) = 0 \tag{3–24}$$

whose analytical solution is given by $U(x,\omega) = \int_0^x A(y,\omega)\,dy$. Set $A(x,\omega) = G(x,\omega)$ where $G(x,\omega)$ is a zero mean, unit variance stationary Gaussian process with spectral density [70, p. 196] $s_G(\nu) = \frac{F}{(\nu^2-\nu_0^2)^2+(2\zeta\nu\nu_0^2)^2}$ in which $\nu_0 = 20$, $\zeta = 0.1$, and F is a scaling factor such that the correlation function of A, $r_G(\tau)$, satisfies $r_G(\tau) = \int_{-\infty}^{\infty} e^{i\nu\tau} s_G(\nu)\,d\nu = 1$. To formulate the inverse problem, we approximate $A(x,\omega)$ using observed samples of $U(x,\omega)$.

As $A(x,\omega)$ is a Gaussian process, the response $U(x,\omega)$ is also a Gaussian process with mean $m_U = \int_0^x E[A(y)]\,dy = 0$ and correlation function $r_U(x,y) = \int_0^x \int_0^y r_G(s-t)\,ds\,dt$ [35]. Consequently, the KL expansion of both $U(x,\omega)$ and $A(x,\omega)$ can be expressed as

$$A(x,\omega) = \sum_{k=1}^{\infty} \sqrt{\lambda_k^G}\phi_k^G(x)Y_k(\omega), \quad U(x,\omega) = \sum_{k=1}^{\infty} \sqrt{\lambda_k^U}\phi_k^U(x)\widetilde{Y}_k(\omega)$$

where $Y_k(\omega), \widetilde{Y}_k(\omega) \sim N(0,1)\ \forall k$ and λ_k^G, $\phi_k^G(x)$ and λ_k^U, $\phi_k^U(x)$ are eigenvalues and eigenfunctions of r_G and r_U, respectively. Thus, A and U are characterized by random variables with the same law unlike in Example 13. In the solution approach in [10], the truncation level of the KL expansion of U is used to determine the number of random variables to characterize A. The truncation level M of the

KL expansion is usually deduced from the total variance of the random field which is given by $\int_0^1 Var(U(x))\,dx = \sum_{k=1}^{\infty} \lambda_k^U$. The value of M is then chosen to be the smallest integer such that $\frac{\sum_{k=1}^{M} \lambda_k^U}{\sum_{k=1}^{\infty} \lambda_k^U} \geq \alpha$ with α being close to 1.

The above principle is applied to choose M using $\alpha = 0.95$. To simulate solving the inverse problem, we do not generate samples of U through samples of A in (3–24) and estimate $r_U(x, y)$ from samples of U; rather, r_U is obtained from the relationship between r_U and r_G above. Figure 3.11 exhibits the behavior of the eigenvalues of r_U. The left panel shows the first 30 eigenvalues λ_k^U while the right panel displays $\frac{\sum_{k=1}^{M} \lambda_k^U}{\sum_{k=1}^{\infty} \lambda_k^U}$ as a function of M. It is evident that the truncation level for the KL expansion of U according to the above procedure is $M = 6$.

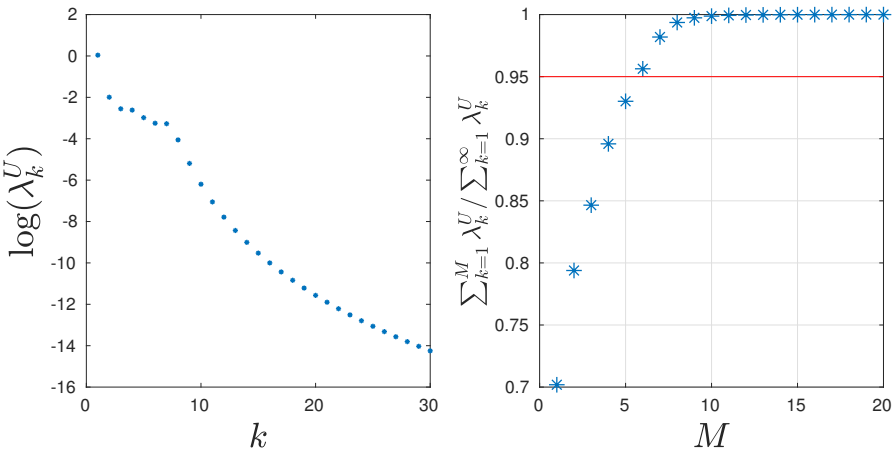


Figure 3.11: Left: First 30 eigenvalues of r_U. Right: Truncation level criterion $\frac{\sum_{k=1}^{M} \lambda_k^U}{\sum_{k=1}^{\infty} \lambda_k^U}$ vs M.

Figure 3.12 displays the behavior of the eigenvalues λ_k^G of r_G with the same legend as in Figure 3.11. In this case, the truncation level for the KL expansion of A is $M = 9$. Note that the eigenvalues of r_G decay slower than that of r_U; intuitively, this is because large values of the frequency ν are required to capture the total energy of the spectral density $2\int_0^{\infty} s_G(\nu)\,d\nu$. As $U(x, \omega)$ is obtained by integrating $A(x, \omega)$, the variation in $A(x, \omega)$ is diminished which yields a faster

decay of eigenvalues for r_U.

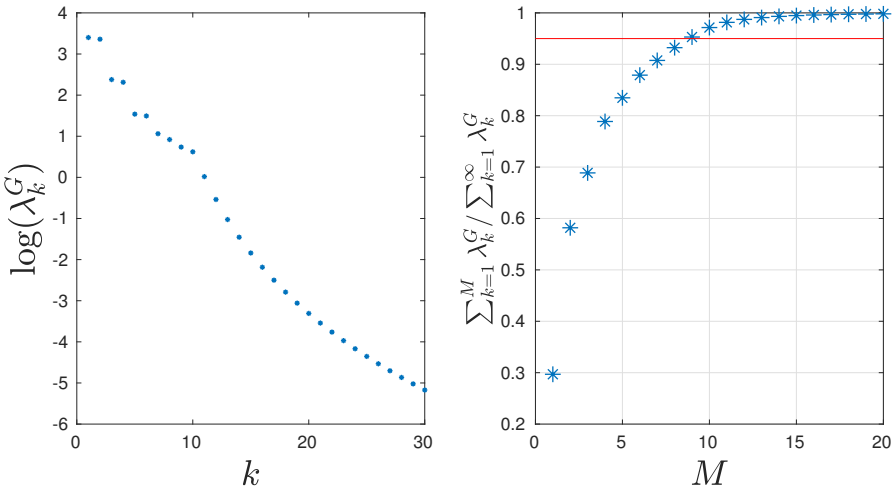


Figure 3.12: Left: First 30 eigenvalues of r_G. Right: Truncation level criterion $\frac{\sum_{k=1}^{M} \lambda_k^G}{\sum_{k=1}^{\infty} \lambda_k^G}$ vs M.

Hence, if the truncation level of the KL expansion of U is used to characterize A, the inverse problem solution would be $\widetilde{A}(x,\omega) = \sum_{k=1}^{6} \sqrt{\lambda_k^G}\phi_k^G(x)Y_k(\omega)$. Based on the discussion above, $\widetilde{A}$ may not be sufficient to capture the statistics of A and we confirm this in Figure 3.13. The four subplots in this figure represent the first 4 moments of $\sum_{k=1}^{M} \sqrt{\lambda_k^G}\phi_k^G(x)Y_k(\omega)$ for $M = 6, 9, 101$ with $M = 101$ being a sufficient approximation for $M = \infty$. We notice that for $M = 6$, $\widetilde{A}(x,\omega)$ underestimates the statistics of $A(x,\omega)$, especially for moments of even order. The underestimation can be avoided provided appropriate selection of the truncation level. Perhaps deducing the truncation level for A using other information than the observations U can resolve this issue.

To summarize, Examples 13 and 14 demonstrate that in parameterizing an unknown random field $A(x,\omega)$ by random variables $\{Y_k\}_{k=1}^{M}$ obtained from the response $U(x,\omega)$, the value chosen for M and the probability law chosen for Y_k can significantly affect the accuracy of the approximation for $A(x,\omega)$. A strategy to tackle the ill-posedness could instead require additional information being

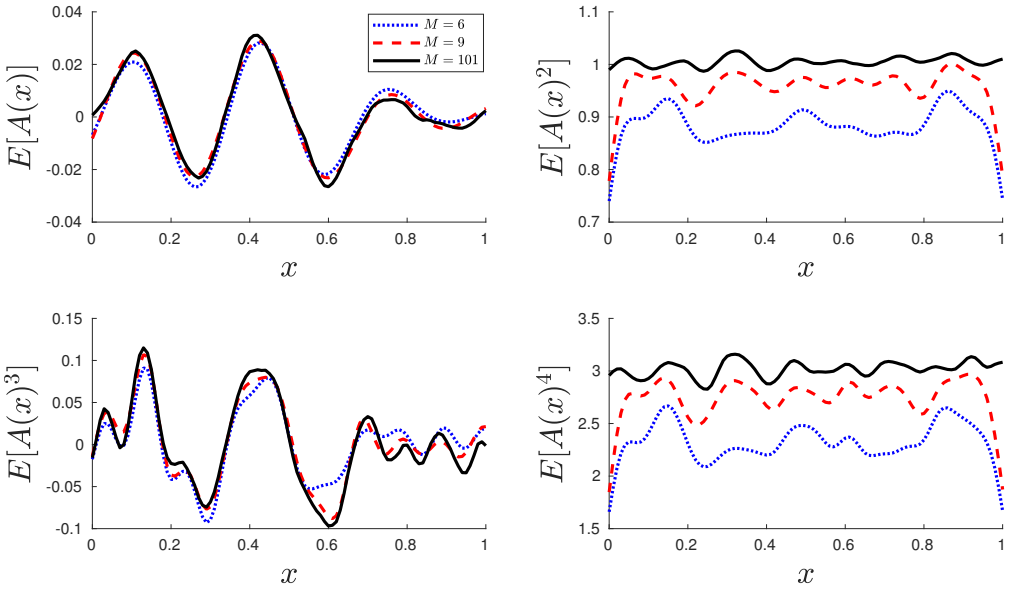


Figure 3.13: p-th order moments of $\sum_{k=1}^{M} \sqrt{\lambda_k^G}\phi_k^G(x)Y_k(\omega)$ for $p = 1, \ldots, 4$ and $M = 6$, (blue dotted line), 9, (red dashed line), and 101 (black solid line).

specified on the dimension of the random variables parameterizing A and that they belong to a family of distributions subject to unknown parameters. This will be elaborated next.

3.3 Required additional information on the unknown random quantity

Section 3.2 showed that the existing methods in general cannot recover the true probability law of a random vector Z if the information is limited to its bounded domain. We therefore devote this section to addressing the objective of this work: identify realistic additional information on Z that is required to characterize its law. The mapping in Example 11, $Q(Z) = Z_1 \cdot Z_2$, is revisited in which the true law is set as $Z_1, Z_2 \sim U(0,1)$, independent. All pdfs from this section

onwards are constructed with respect to the Lebesgue measure and are denoted by f.

It is impossible to devise a general method to determine the minimum amount of additional information that is required on Z. Different applications possess forward models with unique properties and specific information on the unknown. Hence, we outline a few scenarios with their corresponding solution methodologies. These are categorized based on what is known about Z: moment information (Section 3.3.1) or the family of distribution to which it belongs (Section 3.3.2). In each category, further subcategories are considered depending on the given information on $Q(Z)$: pdf (Sections 3.3.1.1 and 3.3.2.1) or samples (Sections 3.3.1.2 and 3.3.2.2).

3.3.1 Information on moments of Z

If information about moments of the random vector Z is available, the principle of maximum entropy [19] can be employed to determine the pdf of Z, assuming that this philosophy is accepted and that it is believed that the solution to the inverse problem resides in the subspace of maximum entropy pdfs. The principle of maximum entropy constructs the pdf $f_Z(z)$ of Z by solving

$$\begin{aligned} \underset{f_Z}{\text{minimize}} \quad & \int_\Gamma f_Z(z)\, \log(f_Z(z))\, dz \\ \text{subject to} \quad & \int_\Gamma g_k(z) f_Z(z)\, dz = \mu_k,\ \ k = 1, \ldots, N, \\ & \int_\Gamma f_Z(z)\, dz = 1. \end{aligned} \tag{3–25}$$

for some functions g_k, whose solution is derived as

$$f_Z(z) = \frac{1}{\int_\Gamma \exp[\lambda_1 g_1(z) + \cdots + \lambda_N g_N(z)]\, dz} \exp[\lambda_1 g_1(z) + \cdots + \lambda_N g_N(z)], \quad z \in \Gamma \quad (3\text{–}26)$$

where the Lagrange multipliers λ_k satisfy the relationship $\frac{\partial}{\partial \lambda_k} \int_\Gamma \exp[\lambda_1 g_1(z) + \cdots + \lambda_N g_N(z)]\, dz = \mu_k$ for $k = 1, \ldots, N$. If these multipliers exist, it can be shown that (3–26) is the unique minimizer satisfying the above constraints [19]. The succeeding sections detail how (3–26) can be used to solve the inverse problem given the pdf of $Q(Z)$ (Section 3.3.1.1) or samples of $Q(Z)$ (Section 3.3.1.2).

3.3.1.1 Pdf of $Q(Z)$

Assume that the pdf $f_Q(q) = -\log(q)$, $q \in (0, 1]$ of Q were known. The example below illustrates the above construction.

Example 15. Suppose that the only information known about Z aside from $Z \in \Gamma$ are that 1) Z_1 and Z_2 are independent and that 2) the first-order moments of Z_1 and Z_2 are within a certain range, i.e. $E[Z_1] = \mu_1 \in [0, 0.75]$, $E[Z_2] = \mu_2 \in [0.4, 1]$. It is demonstrated how the inverse problem can be solved using an entropy-based pdf.

For a fixed value of (μ_1, μ_2), the principle of maximum entropy yields the conditional pdf of Z as

$$f_Z(z_1, z_2|\mu_1, \mu_2) = \frac{\lambda_1}{(e^{\lambda_1} - 1)} \frac{\lambda_2}{(e^{\lambda_2} - 1)} \exp[\lambda_1 z_1 + \lambda_2 z_2] \quad (3\text{–}27)$$

where there is a bijective relationship between μ_i and λ_i for $i = 1, 2$ via $\mu_i = \frac{1}{1-e^{-\lambda_i}} - \frac{1}{\lambda_i}$, $\mu_i \in (0, 1)$. To estimate (μ_1, μ_2), an optimization problem can be solved which measures the discrepancy between the given pdf f_Q of Q and the one

obtained by propagating the pdf in (3–27) through the forward model Q which we denote by $\widetilde{f_Q}(\cdot|\mu_1, \mu_2)$. Mathematically, this is expressed as

$$(\mu_1, \mu_2) = \underset{(\mu_1,\mu_2)}{\operatorname{argmin}} \; d(f_Q(q), \widetilde{f_Q}(q|\mu_1, \mu_2)) \tag{3–28}$$

for some distance function d such as the L^p error, Kullback-Leibler divergence, etc. Let $F_{Z_i}(\cdot|\mu_1, \mu_2)$ and $f_{Z_i}(\cdot|\mu_1, \mu_2)$ represent the marginal cdf and pdf, respectively, of Z_i for $i = 1, 2$ based on (3–27). Elementary calculations similar to (3–6) yield

$$\widetilde{f_Q}(q|\mu_1, \mu_2) = \frac{d}{dq} \int_0^1 F_{Z_2}\left(\frac{q}{z_1}\middle|\mu_1, \mu_2\right) f_{Z_1}(z_1|\mu_1, \mu_2)\, dz_1 = \int_q^1 f_{Z_2}\left(\frac{q}{z_1}\middle|\mu_1, \mu_2\right) f_{Z_1}(z_1|\mu_1, \mu_2)\, dz_1 \tag{3–29}$$

for $q \in [0, 1]$. Figure 3.14a displays the logarithm of the L^1 error $\|f_Q(q) - \widetilde{f_Q}(q|\mu_1, \mu_2)\|_{L^1}$ for (μ_1, μ_2) in the specified ranges above. This discrepancy is minimized at $(\mu_1, \mu_2) = (0.5, 0.5)$ which corresponds to $(\lambda_1, \lambda_2) = (0, 0)$, thereby recovering the true pdf of Z: $f_Z(z_1, z_2|\mu_1, \mu_2) = \mathbb{1}_{(z_1,z_2)\in\Gamma}$ with $\mathbb{1}$ being the indicator function.

For this simple example, the pdf of Q given the pdf of Z can be computed analytically. In general, however, if $f_Z(\cdot|\theta)$ for $\theta \in \Theta$ represents the pdf of Z, propagating this through more complicated forward models Q implies evaluation of Q multiple times. An approach to ameliorate this computational burden is to use a surrogate model [36] for Q as a function of Z. This can be supplemented by the following procedure if Θ is low-dimensional, as done in [25]:

- Select M points $\{\theta_i\}_{i=1}^M \subset \Theta$.
- For each θ_i, $i = 1, \ldots, M$, propagate $f_Z(\cdot|\theta_i)$ through Q to approximate the pdf $\widetilde{f_Q}(\cdot|\theta_i)$ of Q.
- Using $\{\widetilde{f_Q}(\cdot|\theta_i)\}_{i=1}^M$, construct an interpolant for $\widetilde{f_Q}(\cdot|\theta)$ over Θ.

3.3.1.2 Samples of $Q(Z)$

In contrast to the previous section, consider that the available information on the quantity of interest Q is its N_s samples represented by $\{q^i\}_{i=1}^{N_s}$ instead of the pdf of Q. Such information is what is typically encountered in practical applications of stochastic inverse problems [25, 60]. This naturally leads to employing the Bayesian framework [45] to solve the inverse problem. If information is available on moments of Z in the form of a prior pdf, the principle of maximum entropy can be utilized to construct the likelihood function in Bayes' theorem as the next example elaborates.

Example 16. We postulate that only the following information is known about Z: 1) $Z \in \Gamma$, 2) Z_1, Z_2 are independent, and that 3) $(\mu_1, \mu_2) \coloneqq (E[Z_1], E[Z_2])$ is equally likely to take any value in the range $[0.25, 0.75]^2$. Bayes' theorem coupled with the principle of maximum entropy provide an approach to address the inverse problem.

Knowledge that (μ_1, μ_2) is equally likely in $[0.25, 0.75]^2$ translates to a prior pdf on (μ_1, μ_2) denoted by $f_\mu^{prior}(\mu_1, \mu_2) = \mathbb{1}_{(\mu_1,\mu_2)\in[0.25,0.75]^2} \frac{1}{0.5^2}$. By Bayes' theorem, the posterior pdf on (μ_1, μ_2) is

$$f_\mu^{post}(\mu_1, \mu_2|\{q^i\}_{i=1}^{N_s}) = \frac{\ell(\mu_1, \mu_2|\{q^i\}_{i=1}^{N_s}) f_\mu^{prior}(\mu_1, \mu_2)}{\int_{[0.25,0.75]^2} \ell(\mu_1, \mu_2|\{q^i\}_{i=1}^{N_s}) f_\mu^{prior}(\mu_1, \mu_2)\, d\mu_1 d\mu_2} \tag{3–30}$$

in which $\ell(\mu_1, \mu_2|\{q^i\}_{i=1}^{N_s})$ symbolizes the likelihood function. Given (μ_1, μ_2), let $f_Z(\cdot|\mu_1, \mu_2)$ as in (3–27) be the entropy-based pdf of Z while $\widetilde{f_Q}(\cdot|\mu_1, \mu_2)$ as in (3–29) be the resulting pdf when the entropy-based pdf is propagated through Q. The likelihood function is then established as

$$\ell(\mu_1, \mu_2|\{q^i\}_{i=1}^{N_s}) = \prod_{i=1}^{N_s} \widetilde{f_Q}(q^i|\mu_1, \mu_2) \tag{3–31}$$

by the independence of the samples $\{q^i\}_{i=1}^{N_s}$.

Figure 3.15a displays the posterior pdf $f_\mu^{post}(\cdot|\{q^i\}_{i=1}^{N_s})$ on (μ_1, μ_2) using $N_s = 100$ samples of Q. The maximum a posteriori (MAP) estimate is around $(\mu_1, \mu_2) = (0.5, 0.5)$ which recovers the true pdf of Z. We remark that the accuracy of this approach is in part influenced by the observed number of samples N_s of Q.

Does the entropy-based pdf for Z guarantee that the true pdf of Z can be recovered, or that the optimization problem formulated or the posterior density have a unique global minimum? The answer depends on the specific application. If not, additional information of the moments of the unknown Z need to be specified to obtain a solution to the inverse problem that can be used for prediction as elaborated in Section 3.4.2. Sufficient conditions exist which impose criteria that the moments of Z have to satisfy to uniquely determine its distribution; see [23, Theorem 3.3.11] for an example.

3.3.2 Parametric family of distributions of Z

Another strategy to combat ill-posedness is to require that the practitioner has information about the family of distributions in which the law of Z resides, subject to unknown parameters θ. This information is represented as $f_Z(\cdot|\theta)$ in which the functional form of the pdf of Z is known. We illustrate how this can be employed to solve the inverse problem if the pdf of $Q(Z)$ (Section 3.3.2.1) or samples of $Q(Z)$ (Section 3.3.2.2) is given.

3.3.2.1 Pdf of $Q(Z)$

The next example expounds on the above idea assuming that the pdf $f_Q(q) = -\log(q), q \in (0, 1]$ of Q were known.

Example 17. Suppose that it is known that Z_1, Z_2 are independent with $Z_1 \sim Beta(\nu_1, \nu_1)$ and $Z_2 \sim Beta(\nu_2, \nu_2)$ whose joint pdf is denoted by $f_Z(\cdot|\nu_1, \nu_2)$. The objective is to estimate (ν_1, ν_2) such that the pdf f_Q of Q matches the pdf obtained by propagating $f_Z(\cdot|\nu_1, \nu_2)$ through Q.

The solution methodology for this approach is identical to the optimization procedure in (3–28) in which we write the propagated pdf of $f_Z(\cdot|\nu_1, \nu_2)$ through Q as $\widetilde{f_Q}(\cdot|\nu_1, \nu_2)$. By nature of the beta distribution, $(\nu_1, \nu_2) \in (0, \infty)^2$. Figure 3.14b displays the L^1 error $\|f_Q(q) - \widetilde{f_Q}(q|\nu_1, \nu_2)\|_{L^1}$ for $(\nu_1, \nu_2) \in (0, 10]^2$ where the global minimum in this domain is attained at $(\nu_1, \nu_2) = (1, 1)$, thereby recovering the true pdf of Z. Heuristic arguments can be made to deduce that no other values for (ν_1, ν_2) outside $(0, 10]^2$ yield the global minimum. If ν_1, ν_2 are simultaneously large, $Z_1, Z_2 \to \frac{1}{2}$ a.s. which implies that $Q \to \frac{1}{4}$ a.s. Likewise, if only ν_1 is large then $Q \to \frac{1}{2}Z_2$ a.s. whose pdf does not match $f_Q(q) = -\log(q)$, and vice versa.

3.3.2.2 Samples of $Q(Z)$

In contrast to the previous example, suppose instead that the available information on Q pertains to its N_s samples $\{q^i\}_{i=1}^{N_s}$. Information on the family of distributions enables the construction of the likelihood function that is required to find the posterior pdf on the unknown parameters using Bayes' theorem. The next example highlights this idea of standard Bayesian inversion.

Example 18. Consider a model in which the following information is at the

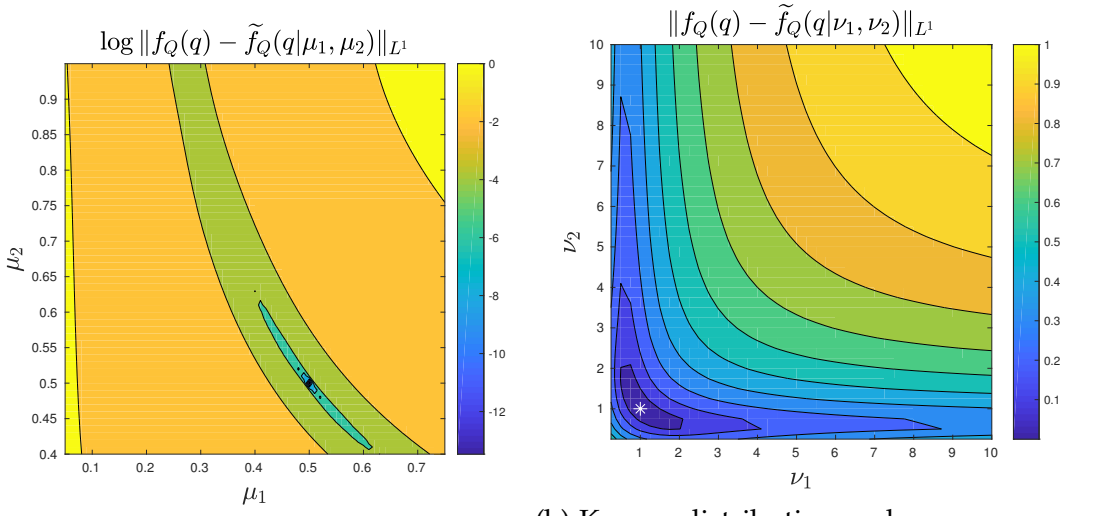
practitioner's disposal: 1) Z_1, Z_2 are independent, 2) $Z_1 \sim Beta(1, \nu_1), Z_2 \sim Beta(1, \nu_2)$, and 3) the prior pdf on (ν_1, ν_2) is characterized by $f_\nu^{prior}(\nu_1, \nu_2) = \mathbb{1}_{(\nu_1,\nu_2)\in[\frac{1}{3},3]^2} \frac{1}{0.5^2} \frac{1}{(\nu_1+1)^2} \frac{1}{(\nu_2+1)^2}$. The posterior pdf on (ν_1, ν_2) results directly from Bayes' theorem.

Since $E[Z_i] = \frac{1}{1+\nu_i}$ for $i = 1, 2$, the prior pdf f_ν^{prior} translates to a uniform prior pdf on $(E[Z_1], E[Z_2])$ with values in the range $[0.25, 0.75]^2$. The construction of the likelihood function and the posterior pdf $f_\nu^{post}(\nu_1, \nu_2|\{q^i\}_{i=1}^{N_s})$ is identical to that in (3–31) and (3–30), respectively, wherein the specified pdf on Z conditioned on (ν_1, ν_2) is propagated through Q to obtain $\widetilde{f_Q}(\cdot|\nu_1, \nu_2)$. Figure 3.15b exhibits $f_\nu^{post}(\nu_1, \nu_2|\{q^i\}_{i=1}^{N_s})$ using the same $N_s = 100$ samples of Q generated in Section 3.3.1.2. The MAP estimate hovers close to $(\nu_1, \nu_2) = (1, 1)$; with more samples of Q, the contours of $f_\nu^{post}(\nu_1, \nu_2|\{q^i\}_{i=1}^{N_s})$ center more at this point.

The above approach can be extended if practitioner believes that the true pdf of Z belongs to multiple families of distributions, each with its own set of parameters, i.e. $f_Z^1(z|\theta^1), \ldots, f_Z^{N_m}(z|\theta^{N_m})$. Bayesian model selection [39] offers a strategy to solve the inverse problem.

Since the practitioner possesses information about the family of distributions in which Z resides, this approach guarantees that the true pdf of Z can be recovered. Does this approach ensure that the optimization problem or the posterior density have a unique global minimum? As before, the answer is case dependent. For example, [50] considers a specific model in stochastic homogenization where it was analytically shown that their specified distribution on Z and the optimization problem they formulated accommodate a unique solution on the parameters of the pdf of Z.

Otherwise, additional information such as moments of Z need to be supplied to regularize against other plausible parameter values in the pdf of Z. To clarify this, assume instead that the ranges of Z_1, Z_2 are unknown yet $f_Q(q) = -\log(q)$, $Z_1 = \lambda Z_1'$, $Z_2 = \frac{1}{\lambda} Z_2'$ where $\lambda > 0, Z_1' \sim Beta(\nu_1, \nu_1), Z_2' \sim Beta(\nu_2, \nu_2)$ are specified information. In this new model, λ, ν_1, ν_2 are parameters to be approximated. Without additional information, the inverse problem possesses infinitely many solutions of the form $(\nu_1, \nu_2, \lambda) = (1, 1, \lambda)$ for any $\lambda > 0$. Specifying additional moment information on Z such as $E[Z_1]$ and $E[Z_2]$ would resolve such ill-posedness.



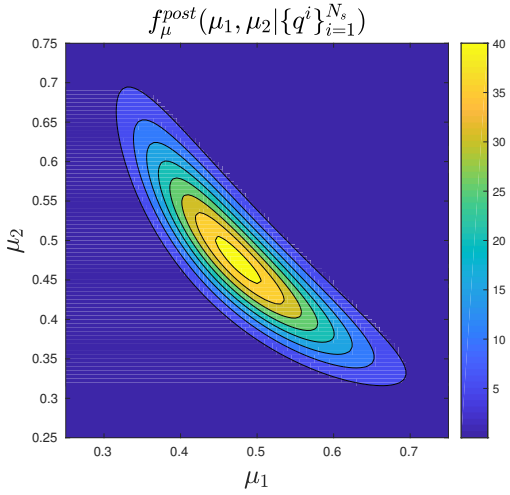
(a) Entropy-based pdf

(b) Known distribution, unknown parameters

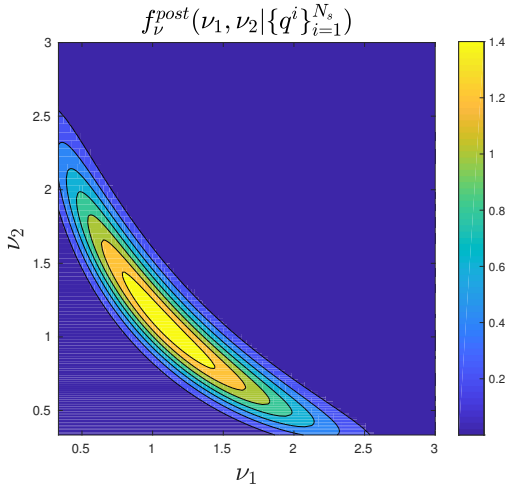
Figure 3.14: Discrepancy between the given pdf f_Q of Q and the pdf obtained by propagating the pdf $f_Z(\cdot|\theta)$ of Z through Q. Left panel: $f_Z(\cdot|\theta)$ is obtained through the principle of maximum entropy. Right panel: $f_Z(\cdot|\theta)$ is a specified distribution subject to unknown parameters. The white asterisk denotes the location of the global minimum.

3.4 Remarks

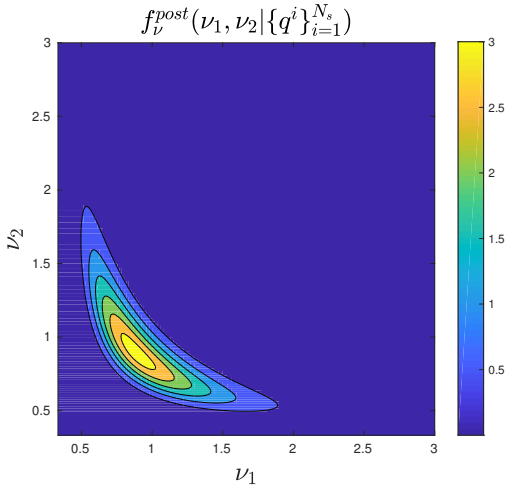
Section 3.3 dealt with the objective of this work motivated by the discussion in Section 3.2. As we have seen, existing methods may not succeed in recovering



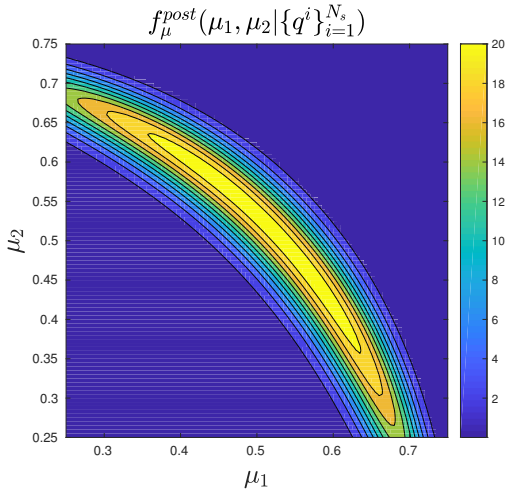
(a) Entropy-based likelihood



(b) Known distribution, unknown parameters



(c) Entropy-based likelihood



(d) Known distribution, unknown parameters

Figure 3.15: Posterior distribution of the parameters in Examples 16 and 18. Plots (a) and (c): posterior density in the (μ_1, μ_2) parameter space and in the $(\nu_1, \nu_2) = (\frac{1}{\mu_1} - 1, \frac{1}{\mu_2} - 1)$ parameter space, respectively, in which the likelihood is constructed using the principle of maximum entropy as in Section 3.3.1.2. Plots (b) and (d): posterior density in the (ν_1, ν_2) parameter space and in the $(\mu_1, \mu_2) = (\frac{1}{1+\nu_1}, \frac{1}{1+\nu_2})$ parameter space, respectively, in which the likelihood is constructed using the known family of distributions as in Section 3.3.2.2.

the true pdf on the random quantity in the absence of further information. As such, we argued on how the inverse problem should be formulated and suggested appropriate solution methods. Here, we support and clarify aspects on the model construction (Section 3.4.1) and on our objective (Section 3.4.2). In particular, the latter section highlights that in the absence of information on the unknown random quantity, the solution to the inverse problem may not be suitable to predict the law of quantities of interest other than the one it was calibrated to.

3.4.1 Posing the stochastic inverse problem

In the following, three remarks are made about the various formulations of the inverse problem considered in Section 3.3. The same notation above is used.

- *Different models on the pdf of Z result in different solutions to the inverse problem.* To clarify this point, we revisit the examples presented in Sections 3.3.1.2 and 3.3.2.2. The parameters between both models are related through $\mu_i = \frac{1}{1+\nu_i}$ and the same prior pdf characterized both sets of parameters. With the same set of $N_s = 100$ samples, we plot the posterior distributions under each model in the (μ_1, μ_2) space (Figures 3.15a and 3.15d) and in the (ν_1, ν_2) space (Figures 3.15b and 3.15c). It is clear that the contours manifest distinct behavior.
- *Difference in information content between knowing the pdf of $Q(Z)$ vs only having samples of $Q(Z)$.* Compared to having the pdf of Q as given information, only possessing samples of Q provides less information about the quantity of interest. A possible consequence of this lower information content

includes posterior densities (3–30) that are not sharp about the MAP estimate. This is remedied by requiring a large number of samples of $Q(Z)$.

- *Distinction in applying the principle of maximum entropy when moment information on Z or $Q(Z)$ is supplied.* Finally, we remark that the principle of maximum entropy has also been invoked to solve a stochastic inverse problem formulated differently from that of Section 3.3. Consider the mapping $Q : D \times \Gamma \to \mathcal{D}$ with $\Gamma \coloneqq Z(\Omega)$, D being the physical domain, and $Q(x, z) = q(U(x, z))$ for $x \in D, z \in \Gamma$ and some function q. The methodology developed in [34] seeks to address the inverse problem described as follows:

  > Determine the pdf f_Z of Z given the bounded range Γ of Z and the observed moments of Q up to order N for $x \in D$ denoted by $\hat{\mu}^p(x)$, $p = 1, \ldots, N$.

  Since the p-th order moments of Q can be expressed as integrals on Z, i.e. $\int_\Gamma Q(x, z)^p f_Z(z)\, dz$ for $x \in D$, this naturally leads to a solution based on the principle of maximum entropy in which f_Z is estimated via the optimization problem (3–25) subject to the constraint
  $\int_\Gamma Q(x, z)^p f_Z(z)\, dz = \hat{\mu}^p(x) \;\; \forall x \in D, \;\; 1 \leq p \leq N.$
  Although the principle of maximum entropy has been used as a regularizer to infer the pdf of a random vector provided information about its moments, for problems involving forward models, the absence of information on Z raises issues mentioned earlier. There is no guarantee that the proposed method is able to recover the true pdf of Z, in contrast to the authors' comment on p. B761. To see this, let f_Z^1, f_Z^2 be two pdfs of Z such that when propagated through the model, the pdf of Q is f_Q. The p-th or-

der moments of Q under both pdfs of Z are identical even though f_Z^1 has larger entropy while f_Z^2 is the true pdf of Z or vice versa. Specifying all moments of Q neither resolves the issue.

3.4.2 Validation

In Section 3.3, we considered types of information required on the unknown random quantity in order to solve the stochastic inverse problem. While some of this required information may be exigent, we argue that they are necessary to obtain solutions such that the resulting law of Z can be used to characterize other quantities of interest $\widetilde{Q}$. We remark that in relation to the methods tackled in Section 3.2, such additional information may not be necessary if the structure of the contours of Q and $\widetilde{Q}$ are similar. If this is not the case, without additional information, methods such as in Section 3.2.1.4 may result in a posterior pdf for Z whose predicted probability measure on the new quantity of interest $\widetilde{Q}$ is similar to the predicted measure on $\widetilde{Q}$ produced by the prior. The field of optimal experimental design for prediction addresses these concerns.

Here, we revisit the solutions obtained from the methods described in Sections 3.2.1.2, 3.3.1.2, and 3.3.2.2. It is demonstrated that in the absence of information on Z, the resulting solution may be inadequate to characterize the law of quantities of interest to which it was not calibrated, thereby limiting its use in practical applications.

Example 19. We revisit the forward mapping $Q(Z_1, Z_2) = Z_1 \cdot Z_2$ where $Z_1, Z_2 \sim U(0,1)$, independent, characterizes the true law on Z. The following methodologies are employed to solve the inverse problem on approximating the pdf of Z

depending on available information on Z and Q. The resulting pdf on Z is then used to predict the pdf on an unobserved quantity of interest $\widetilde{Q}(Z_1, Z_2) = Z_1 + Z_2$. The domain $Z \in [0,1]^2$ is assumed for all methods.

- Method based on the disintegration theorem using an ansatz as in Section 3.2.1.2 in which $f_Q(q) = -\log(q)$ is given and no other information on Z is required.
- Bayes' theorem with entropy-based pdf for the likelihood as in Section 3.3.1.2 in which $N_s = 100$ samples $\{q^i\}_{i=1}^{N_s}$ of Q are available and the following is known about Z: Z_1, Z_2 are independent and $f_\mu^{prior}(\mu_1, \mu_2) = \mathbb{1}_{(\mu_1,\mu_2)\in[0.25,0.75]^2} \frac{1}{0.5^2}$ is the prior pdf on $(\mu_1, \mu_2) = (E[Z_1], E[Z_2])$.
- Bayes' theorem with known family of distributions for the likelihood as in Section 3.3.2.2 in which the same $N_s = 100$ samples $\{q^i\}_{i=1}^{N_s}$ of Q are available as above and the following is known about Z: Z_1, Z_2 are independent, $Z_1 \sim Beta(1, \nu_1), Z_2 \sim Beta(1, \nu_2)$, and $f_\nu^{prior}(\nu_1, \nu_2) = \mathbb{1}_{(\nu_1,\nu_2)\in[0.75,1.25]^2} \frac{1}{0.5^2}$ is the prior pdf on (ν_1, ν_2).

The solution approach for the latter 2 methods has already been discussed. The pdf on the unobserved quantity of interest $\widetilde{Q}$ then results by computing the posterior predictive distribution. Denote by $f_\Theta^{post}(\theta|\{q^i\}_{i=1}^{N_s})$ the obtained posterior pdf on the corresponding parameter space which qualifies as the solution to the inverse problem upon application of either of the latter 2 methods. The pdf on $\widetilde{Q}(Z_1, Z_2) = Z_1 + Z_2$ is obtained through

$$f_{\widetilde{Q}}(\overline{q}|\{q^i\}_{i=1}^{N_s}) = \int_\Theta f_{\widetilde{Q}}(\overline{q}|\theta) \cdot f_\Theta^{post}(\theta|\{q^i\}_{i=1}^{N_s})\, d\theta \tag{3–32}$$

where $f_{\widetilde{Q}}(\overline{q}|\theta)$ is the pdf on $\widetilde{Q}$ obtained by propagating the conditional pdf $f_Z(\cdot|\theta)$

on Z through $\widetilde{Q}$. Elementary calculations show that

$$f_{\widetilde{Q}}(\widetilde{q}|\theta) = \begin{cases} \int_0^{\widetilde{q}} f_{Z_2}(\widetilde{q} - z_1|\theta)\, f_{Z_1}(z_1|\theta)\, dz_1 & 0 \le \widetilde{q} \le 1 \\ \int_{\widetilde{q}-1}^{1} f_{Z_2}(\widetilde{q} - z_1|\theta)\, f_{Z_1}(z_1|\theta)\, dz_1 & 1 \le \widetilde{q} \le 2 \end{cases} \tag{3–33}$$

with $f_{Z_i}(\cdot|\theta)$ being the marginal pdf of Z_i, $i = 1, 2$. In particular, under the true law on Z, $\widetilde{Q}$ has a triangular distribution whose pdf is $f_{\widetilde{Q}}(\widetilde{q}) = \widetilde{q}$ for $\widetilde{q} \in [0, 1]$ and $f_{\widetilde{Q}}(\widetilde{q}) = 2 - \widetilde{q}$ for $\widetilde{q} \in [1, 2]$.

On the other hand, the pdf f_Z^{ansatz} arising from the method based on the disintegration theorem results through this procedure:

1. Partition $\Gamma = [0, 1]^2$ into $N_{sq} = 10000$ squares $\{A_i\}_{i=1}^{N_{sq}}$ with area $(0.01)^2$.
2. Compute $P_Z(A_i)$ using (3–4) and (3–5) for $i = 1, \ldots, N_{sq}$.
3. For each A_i, let $(z_{1,i}^*, z_{2,i}^*)$ be its center. The approximate pdf of Z is then calculated as $f_Z^{ansatz}(z_{1,i}^*, z_{2,i}^*) \simeq \frac{P_Z(A_i)}{(0.01)^2}$.

If $(z_{1,i}^{SW}, z_{2,i}^{SW})$ and $(z_{1,i}^{NE}, z_{2,i}^{NE})$ represent the lower left and the upper right vertices, respectively, of A_i, the parameter along the transverse curve $x_{\mathcal{L}}$ is bounded by $\sqrt{2z_{1,i}^{SW} z_{2,i}^{SW}} \le x_{\mathcal{L}} \le \sqrt{2z_{1,i}^{NE} z_{2,i}^{NE}}$.

The left panel of Figure 3.16 displays the approximate pdf of Z produced by the method in Section 3.2.1.2 whereas the right panel shows 25000 samples of (Z_1, Z_2) drawn from this pdf through rejection sampling and interpolation. We confirm that when f_Z^{ansatz} is propagated through the forward model Q, we recover the specified pdf $f_Q(q) = -\log(q)$ as guaranteed by Theorem 11. Figure 3.17 exhibits the histogram of $Z_1 \cdot Z_2$ using 52154 samples of Z obtained from f_Z^{ansatz} together with f_Q.

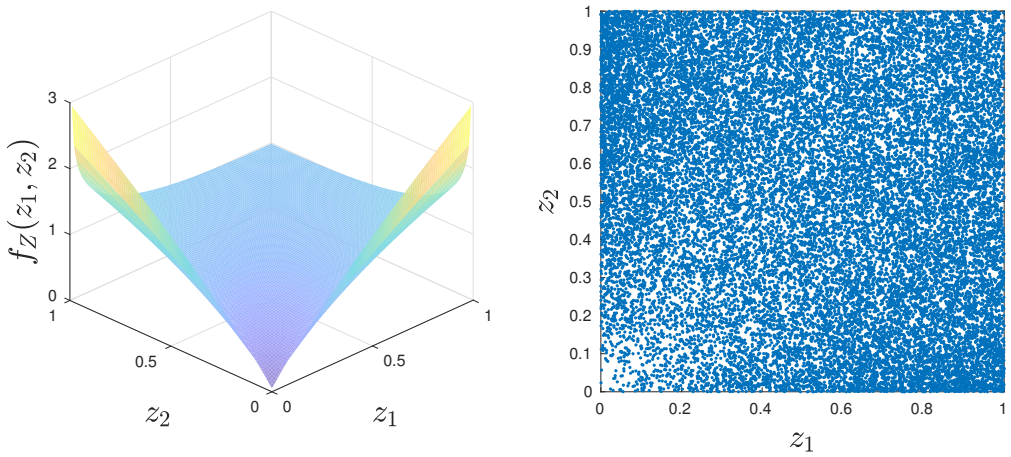


Figure 3.16: Left panel: approximate pdf of Z produced by the method in Section 3.2.1.2. Right panel: 25000 samples of (Z_1, Z_2) simulated from the pdf on the left panel.

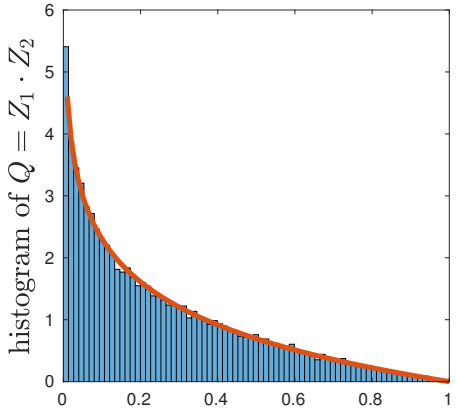


Figure 3.17: Histogram of 52154 samples of $Q = Z_1 \cdot Z_2$ where the samples of Z are drawn from f_Z^{ansatz} together with $f_Q(q) = -\log(q)$.

With the pdf f_Z^{ansatz} on Z at hand, one approach to propagate this pdf through the unobserved quantity of interest $\widetilde{Q}$ is to construct a discrete pdf approximation to $\widetilde{Q}$ using the centers of each square $(z_{1,i}^*, z_{2,i}^*)$ and their corresponding probabilities $P_Z(A_i)$. This yields distinct outcomes $\widetilde{q}_j$ of $\widetilde{Q}$ with weights $P(\widetilde{Q} = \widetilde{q}_j) = \sum_{i:z_{1,i}^* + z_{2,i}^* = \widetilde{q}_j} P_Z(A_i)$ that need to normalized to obtain the approximate pdf $f_{\widetilde{Q}}^{ansatz}$ of $\widetilde{Q}$.

We now evaluate the performance of the solution from each method to

predict the probability law of the unobserved $\widetilde{Q}$. Figure 3.18 contains 3 subplots, one for each method, plotting $f_{\widetilde{Q}}^{ansatz}$ or $f_{\widetilde{Q}}(\cdot|\{q^i\}_{i=1}^{N_s})$ together with the true pdf of $\widetilde{Q}$. A stem plot was used for the leftmost subplot to emphasize that the pdf of a discrete random variable was used to accurately approximate $f_{\widetilde{Q}}^{ansatz}$. Quantitatively, the discrepancy between the true pdf and the simulated pdf's is: $\max|f_{\widetilde{Q}}^{ansatz} - f_{\widetilde{Q}}| \approx 0.2141$ whereas $\max|f_{\widetilde{Q}}(\cdot|\{q^i\}_{i=1}^{N_s}) - f_{\widetilde{Q}}| \approx 0.0786$ for Method 2 (Bayes'+entropy) while $\max|f_{\widetilde{Q}}(\cdot|\{q^i\}_{i=1}^{N_s}) - f_{\widetilde{Q}}| \approx 0.0415$ for Method 3 (Bayes'+known family). It was also observed that increasing the number of available samples $\{q^i\}_{i=1}^{N_s}$ of the observed quantity of interest Q decreased the discrepancy for the latter 2 methods. The objective of this example was not to conclude which method is better since the available information for each was not identical. Rather, this example justifies the need to specify additional information on Z as was carried out in Sections 3.3.

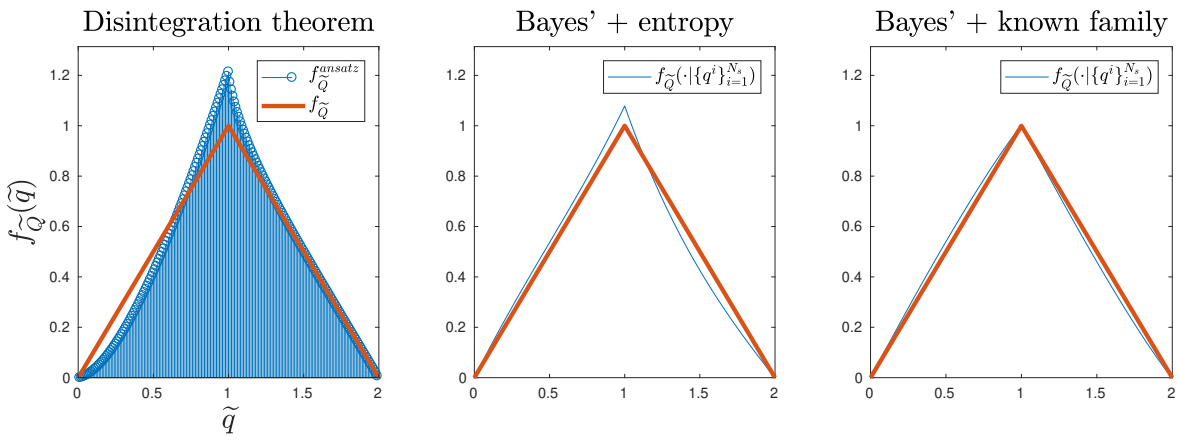


Figure 3.18: Comparison between the true pdf $f_{\widetilde{Q}}$ of $\widetilde{Q}$ and the pdf obtained by propagating through $\widetilde{Q}$ the pdf of Z stemming from methods for the inverse problem described in Example 19. A stem plot is used for $f_{\widetilde{Q}}^{ansatz}$ in the left-most plot to emphasize that an accurate discrete random variable approximation was used.

We conclude this section by showing the repercussions that may arise if the pdf f_Z^{ansatz} on Z is used to predict the pdf of more complicated quantities of inter-

est $\widetilde{Q}$. If we consider $\widetilde{Q}(Z_1, Z_2) = Z_1^2 + Z_2^2$ where $Z_1, Z_2 \sim U(0,1)$ and independent, it was proven in [76] that

$$f_{\widetilde{Q}}(\widetilde{q}) = \begin{cases} \frac{\pi}{4} & 0 \le \widetilde{q} \le 1 \\ \arcsin \frac{1}{\sqrt{\widetilde{q}}} - \frac{\pi}{4} & 1 \le \widetilde{q} \le 2 \end{cases}.$$

Additionally, we also consider $\widetilde{Q}(Z_1, Z_2) = \exp(-(Z_1^2 + Z_2^2))$. Figure 3.19 contains subplots comparing the true pdf $f_{\widetilde{Q}}$ and histogram of $f_{\widetilde{Q}}^{ansatz}$ computed via samples of Z drawn from f_Z^{ansatz} for the two aforementioned unobserved $\widetilde{Q}$'s. We see that $f_{\widetilde{Q}}^{ansatz}$ under/overestimates probabilistic properties of $\widetilde{Q}$ such as $P(\widetilde{Q} \le 0.5)$ for the left subplot and $P(\widetilde{Q} \ge 0.8)$ for the right subplot, among other properties such as moments of $\widetilde{Q}$.

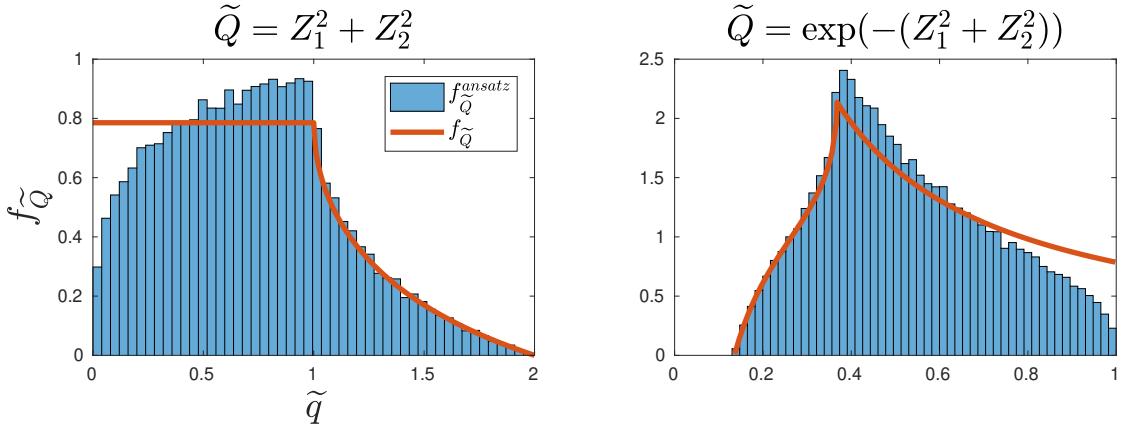


Figure 3.19: Comparison between the true pdf $f_{\widetilde{Q}}$ of $\widetilde{Q}$ and $f_{\widetilde{Q}}^{ansatz}$ for more complicated quantities of interest.

The above examples are not intended to discredit existing methods for solving stochastic inverse problems. Instead, they underscore why additional information should be specified on Z to yield solutions that can be useful in applications. In fact, existing works such as [25,50,62] impose probability distributions on the unknown random quantities after which standard mathematical tools were applied to infer the parameters that characterize these distributions.

3.5 Conclusion

This work dealt with the stochastic inverse problem of identifying the distribution of Z given probabilistic information of the quantity of interest $Q(Z)$. We surveyed general methods that have been developed to tackle the problem in which no information other than the bounded range of Z is assumed. These methods coped with the ill-posedness of this inverse problem in the following ways: [12, 14, 15] suggested an ansatz that the pdf on the generalized contours of Q is uniform while [10] obtained the probability law of Z from the Karhunen-Loève expansion of the solution to the stochastic PDE. We motivated this work by showing that this lack of additional information entails that the true pdf of Z may not be recovered.

Consequently, we argued that it is necessary for this inverse problem to be posed such that further information on Z is specified to attain solutions that are of practical use. We demonstrated that this specified information can take the form of moments of Z or the family of distributions in which Z resides subject to unknown parameters, among others. Using these information, a conjunction of tools such as Bayes' theorem, the principle of maximum entropy, and forward uncertainty propagation were utilized to solve the inverse problem in a manner that is consistent with the present information on the quantity of interest and on Z. Issues arising from this framework were also highlighted. Finally, we emphasized the need for this specified information by assessing how well the resulting solutions from the discussed methods can predict the probability law of unobserved quantities of interest. It is observed that solving the inverse problem without additional information on Z can lead to solutions that may be unreliable for prediction.

The intention of this work was not to discredit existing contributions in this area but to stress on how we believe the inverse problem must be posed for use in practical applications.

CHAPTER 4

BEYOND FAILURE PROBABILITIES: COMBINING PHYSICS-BASED SURROGATE MODELS AND MACHINE LEARNING CLASSIFIERS TO IDENTIFY INPUT RANDOM FIELDS THAT YIELD EXTREME RESPONSE

Consider a physical system modeled by a differential equation that depends on a coefficient random field. The objective of this work is to identify samples of this random field which yield extreme response as a means to: study the law of the input conditioned on rare events and predict if a random field sample causes such an event. This presents a shift from reliability engineering which focuses on computation of failure probabilities. We investigate two classification schemes that identify these samples of interest: physics-based indicators which are functionals of the input random field and surrogate models which approximate the response. As an alternative to these approaches, we propose a general framework consisting of two stages that combines the use of a physics-based surrogate model and a machine learning classifier. In the first stage, a multifidelity surrogate that requires infrequent evaluations of the full model is designed. This surrogate is then used to generate a sufficient number of samples of random fields that yield extreme events to train a machine learning classifier in the second stage. We study the analytical properties required of the surrogate model and demonstrate through numerical examples the synergy of the proposed approach.

4.1 Introduction

We motivate this work with a simple physical example involving a unit length elastic rod in 1-dimension where one end is pulled by a time-dependent force.

Suppose that the elasticity coefficient can be modeled by a random field $A(x,\omega)$ defined on the probability space $(\Omega,\mathcal{F},P)$ where $\omega\in\Omega$ and $x\in[0,1]$ refers to a location on the rod. For $t>0$, denote by $U(t,x,\omega)\in\mathbb{R}$ the displacement field whose dynamics are described by the Euler-Lagrange equation

$$\frac{\partial^2 U(t,x,\omega)}{\partial t^2}=\frac{\partial}{\partial x}\left(A(x,\omega)\frac{\partial}{\partial x}U(t,x,\omega)\right),\qquad x\in[0,1],\ t>0,\ \omega\in\Omega \tag{4–1}$$

subject to initial and boundary conditions:

$$U(t,0,\omega)=0,\qquad A(1,\omega)\left(\int_0^1\frac{dy}{A(y,\omega)}\right)^{-1}\frac{\partial}{\partial x}U(t,1,\omega)=at,\qquad t>0,\ a>0 \tag{4–2}$$

$$U(0,x,\omega)=0,\quad \frac{\partial}{\partial t}U(0,x,\omega)=\left(a\int_0^x\frac{dy}{A(y,\omega)}\right)\left(\int_0^1\frac{dy}{A(y,\omega)}\right)^{-1},\qquad x\in[0,1].$$

Oftentimes, we are interested in quantities of interest (QoI) $Q(\omega)\in\mathbb{R}$ which are functionals of the random field $U(t,x,\omega)$.

Literature in reliability engineering and rare event simulation is concerned with the computation of the (failure) probability $P(Q>\tau)$ where for large τ, $\{Q>\tau\}$ is a rare event of low probability. Commonly employed methods [59] to compute such quantities express Q as a function of the random vector $Z(\omega)\in\mathbb{R}^d$, i.e. $Q(Z(\omega))$, and include the following. Asymptotic approaches like the first-order reliability method and methods based on the large deviations principle [21] locate the point z^* on the limit state $\{z\in\mathbb{R}^d\,|\,Q(z)=\tau\}$ with the largest likelihood and approximate the failure probability at z^*. Importance sampling makes use of a change in measure so that the failure probability is approximated by weighted samples in the vicinity of the failure domain $F=\{z\in\mathbb{R}^d\,|\,Q(z)>\tau\}$. Subset simulation iteratively considers a decreasing sequence of sets, each containing F. Analytical techniques also exist for specific types of systems subject to log-normal random field coefficients [52, 53]. Finally, surrogate models replace Q by an inexpensive approximation $\widetilde{Q}$ afterwhich Monte Carlo simulation

is invoked to compute probabilities. Examples of surrogate models used for this purpose include stochastic Galerkin [51], sparse grid stochastic collocation [17], Gaussian process regression [57], artificial neural networks (ANNs) [18], and support vector machines (SVM) [2, 5, 49, 63, 66, 69].

While failure probabilities are useful in assessing risks associated with structures, materials, etc., they provide limited if any information for controlling and mitigating these risks. They do not offer insight on how to design these objects better to avoid undesired events of low probability. In this work, we therefore deviate from the existing objectives of reliability engineering and rare event simulation. Instead, we focus on identifying samples of $A(x, \omega)$ which cause $Q(\omega) > \tau$ for large τ by examining various classification schemes which determine whether or not a sample $A(x, \omega)$ yields rare events. Through a classifier, a sufficient number of samples of $A(x, \omega) \,|\, Q(\omega) > \tau$ can be obtained that enables one to: study the conditional statistics and distribution of $A(x, \omega) \,|\, Q(\omega) > \tau$, understand why rare events occur in such systems, and predict the occurrence of (unobserved) rare events. The relevance of this problem extends beyond the field of reliability engineering and includes applications such as detecting rare defects in additive manufacturing (agile production) of materials [11].

Existing studies pertinent to our objectives include works such as [26, 27]. An extensive survey of how system dynamics causes rare bursts in the values of the state variables is detailed in [27]. In [26], the authors investigated a data-driven approach to identify what induces bursts in the energy dissipation rate of the Kolmogorov flow, modeled by an inrompressible Navier-Stokes equation exhibiting chaotic behavior. An indicator that is a functional of the velocity field was deduced from solving an optimization problem that relies on the defining

equations of the system, i.e. it is physics-based. It was later validated using a large number of data based on statistics such as the rate of successful predictions and rejections. In this work, we also explore a classifying scheme through the use of physics-based indicators – quantities computable from $A(x, \omega)$ which signal $Q(\omega) > \tau$. However, we demonstrate that even in the simplest of systems, indicators may be difficult to deduce analytically and are computationally expensive to calibrate due to the infrequent occurrence of rare events. This motivates the use of a surrogate model $\widetilde{Q}$ that approximates Q; it then acts as a classifier by checking if $\widetilde{Q} > \tau$ or otherwise.

Despite the demonstrated efficiency of the above cited surrogate models for reliability engineering purposes, only a few are suitable for our objectives since many of these approaches are specifically geared to obtain convergence only in failure probability estimates. In particular, it is possible for $P(Q > \tau) \approx P(\widetilde{Q} > \tau)$ even though the measurable sets $\{\omega \,|\, Q(\omega) > \tau\}$ and $\{\omega \,|\, \widetilde{Q}(\omega) > \tau\}$, important for our setting, are substantially different. In addition, using surrogate models alone still presents challenges in rare event simulation. Machine learning classifiers, for instance, require a sufficient number of samples in the low probability regions (i.e. failure domain) for training. Also, [51] underscores that surrogate models may over or underestimate probabilities of failure.

In line with our objectives, we build on the above concerns by proposing a two-stage approach which is computationally efficient and probabilistically accurate and that leverages on the strengths of different types of classifiers. In the first stage, we formulate a multifidelity physics-based surrogate model in the spirit of [17, 51]. The surrogate model is physics-based in the sense that it is constructed using the defining equations of the physical system. The mul-

tifidelity surrogate comprises the computationally feasible surrogate model $\widetilde{Q}$ and infrequent solves of the expensive full model Q which guarantee that $Q > \tau$ whenever $\widetilde{Q} > \tau$. In the second stage, samples of $A(x, \omega)$ are then generated using the multifidelity surrogate which ensures that a large number of samples of the rare event $A(x, \omega) \,|\, Q(\omega) > \tau$ is present to accurately train a machine learning classifier. Unlike the surrogate, this classifier only considers the data and does not incorporate the system equations anymore. The classifier then serves as a mechanism to predict whether $Q > \tau$ from data.

The plan of this work is as follows. Section 4.2 attempts to discover physics-based indicators that signal rare events using analytical arguments and later supported by data. Due to the challenges of validating and calibrating indicators, Section 4.3 proposes a multifidelity surrogate approximation for the QoI as an alternative classification scheme. Finally, the two-stage approach that combines the multifidelity surrogate and machine learning classifiers is tackled in Section 4.4.

4.2 Physics-based indicators for rare events

In this section, we attempt to deduce indicators from $A(x, \omega)$ which signal $Q(\omega) > \tau$ for large τ. We motivate our proposed approach in Section 4.4 for generating samples of $A(x, \omega) \,|\, Q(\omega) > \tau$ by first investigating existing alternatives for prediction of rare events. The case when the random field $A(x, \omega)$ is represented by an infinite-dimensional noise model and a finite-dimensional noise model are addressed in Section 4.2.1 and 4.2.2, respectively. The following discussion outlines the difficulty of obtaining indicators for simple systems and

that a large number of samples of the rare event are needed to validate them.

4.2.1 Infinite-dimensional noise model

The physical example of the elastic rod in (4–1), (4–2) is revisited where the elasticity coefficient $A(x,\omega)$ is modeled by a translation Gaussian process.

Example 20. Suppose that $A(x,\omega) = \alpha+(\beta-\alpha)\cdot F^{-1}(\Phi(G(x,\omega)))$, $x \in [0,1]$, $\beta \geq \alpha > 0$ where F is the cumulative distribution function (cdf) of a $Beta(p,q)$ random variable, Φ is the cdf of the standard normal distribution, and $G(x,\omega)$ is a homogeneous zero-mean unit-variance Gaussian process with $E[G(x,\omega)\cdot G(y,\omega)] = e^{-\lambda|x-y|}$, $x,y \in [0,1]$. It can be shown that the analytical solution to (4–1), (4–2) is given by

$$U(t,x,\omega) = \left(at\int_0^x \frac{dy}{A(y,\omega)}\right)\left(\int_0^1 \frac{dy}{A(y,\omega)}\right)^{-1}, \quad t>0, \quad x\in[0,1]. \tag{4–3}$$

The quantity of interest we study is the random variable

$$Q_{max}(\omega) = \max_{x\in[0,1]} |U(1,x,\omega) - u(1,x)| \tag{4–4}$$

where $u(t,x) = atx$ is the deterministic continuum mechanics solution obtained from (4–3) if $A(x,\omega)$ were constant. We seek indicators from samples of $A(x,\omega)$ which signal $Q_{max} > \tau$ for large τ.

The simulations carried out below are based on $\alpha = 1, \beta = 10, p = 1, q = 3, \lambda = 2, a = 1$, and $\tau = 0.32$ with a spatial discretization of $\Delta x = 0.005$ to evaluate (4–3). The value of τ is chosen so that $P(Q_{max} > \tau) \approx 0.0014$. Shown in the left panel of Figure 4.1 are 10000 samples of Q_{max} while the middle and the right panels display 5 samples of $A(x,\omega)$ and 5 samples of $A(x,\omega)\,|\,Q_{max}(\omega) > \tau$, respectively. Observe that samples of $A(x,\omega)\,|\,Q_{max}(\omega) > \tau$ manifest a common trend: they

are generally increasing or decreasing. This behavior is consistent with plots of Figure 4.2 which present histograms of $A(x,\omega)|Q_{max}(\omega) > \tau$ using 10000 samples at various spatial locations $x \in [0,1]$ with the range of $A(x,\omega)$ rescaled from $[\alpha,\beta]$ to $[0,1]$. For comparison, the probability density function (pdf) of $A(x,\omega)$ for all $x \in [0,1]$ which is $Beta(p,q)$ is also included.

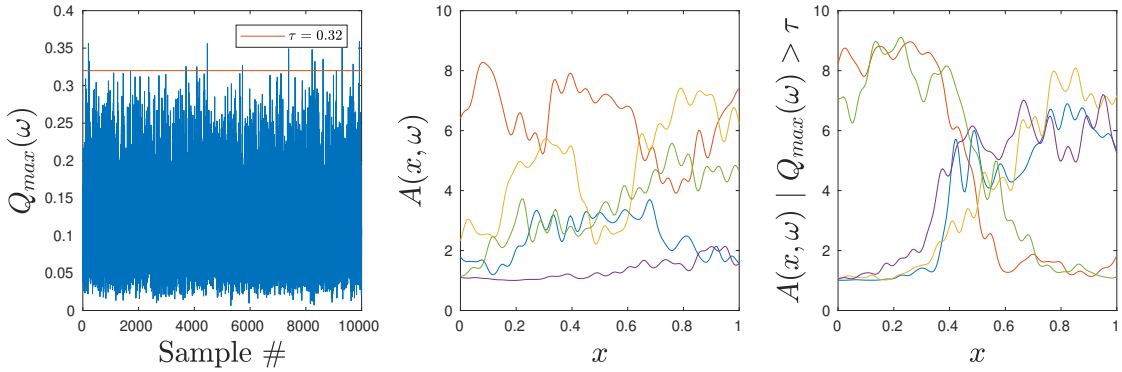


Figure 4.1: Samples of $Q_{max}(\omega)$ (left panel), $A(x,\omega)$ (middle panel), $A(x,\omega) \mid Q_{max}(\omega) > \tau$ for $\tau = 0.32$ (right panel).

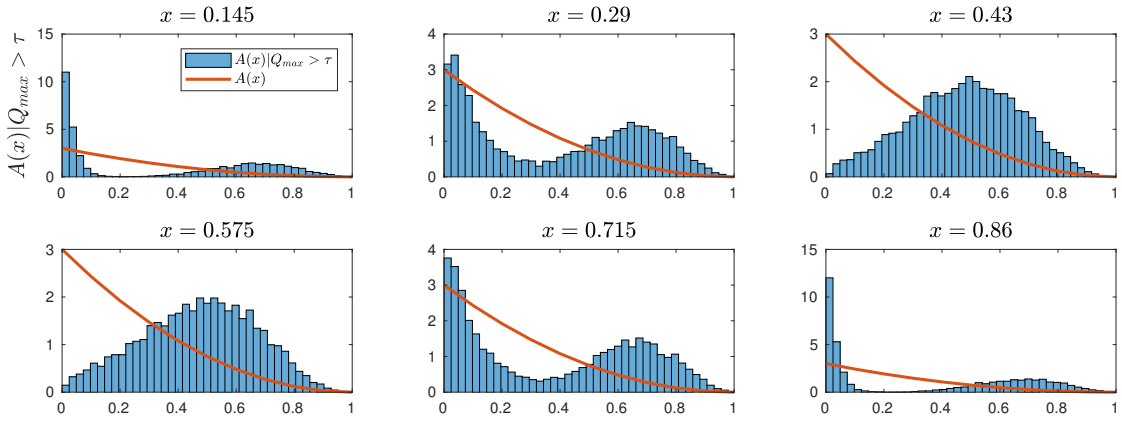


Figure 4.2: Histogram of $A(x,\omega)|Q_{max}(\omega) > \tau$ for $\tau = 0.32$ using 10000 samples compared to the pdf of $A(x,\omega)$ for a few values of $x \in [0,1]$. The range of $A(x,\omega)$ is scaled down to $[0,1]$.

We now seek indicators that aid in predicting if $Q_{max} > \tau$. Ideally, they should not resemble (4–3) since analytical solutions are unavailable in practice. But to facilitate our search, we work backwards and assume that (4–3) is known. From (4–4), Q_{max} can be computed by first finding the critical point $x = x_c(\omega) \in [0,1]$

which maximizes or minimizes $d(x,\omega) := U(1,x,\omega) - u(1,x)$ for each $\omega \in \Omega$. Since $d(0,\omega) = d(1,\omega) = 0$, the critical point $x_c(\omega)$ satisfies $d'(x_c(\omega),\omega) = 0$. Thus,

$$\frac{1}{A(x_c(\omega),\omega)} = \int_0^1 \frac{1}{A(y,\omega)}\,dy, \quad \omega \in \Omega, \tag{4–5}$$

is a necessary condition for $x_c(\omega)$, i.e., $x = x_c(\omega)$ is such that $(A(x,\omega))^{-1}$ achieves its spatial average (over $x \in [0,1]$). Substituting (4–5) to (4–4) yields an alternative expression for $Q_{max}(\omega)$:

$$\begin{aligned} Q_{max}(\omega) = |d(x_c(\omega),\omega)| &= \left| \left(\int_0^{x_c(\omega)} \frac{1}{A(y,\omega)}\,dy \right) \Big/ \left(\frac{1}{A(x_c(\omega),\omega)} \right) - x_c(\omega) \right| \\ &= \left| \int_0^{x_c(\omega)} \left(\frac{A(x_c(\omega),\omega)}{A(y,\omega)} - 1 \right) dy \right|. \end{aligned} \tag{4–6}$$

Qualitatively, for $Q_{max}(\omega)$ to be large, $1/A(y,\omega)$ must be substantially different in magnitude compared to $1/A(x_c(\omega),\omega)$ over $y \in [0, x_c(\omega)]$. This is confirmed in Figure 4.3 which displays 3 samples of $1/A(y,\omega)$ with their corresponding spatial averages $1/A(x_c(\omega),\omega)$ and $Q_{max}(\omega)$ values.

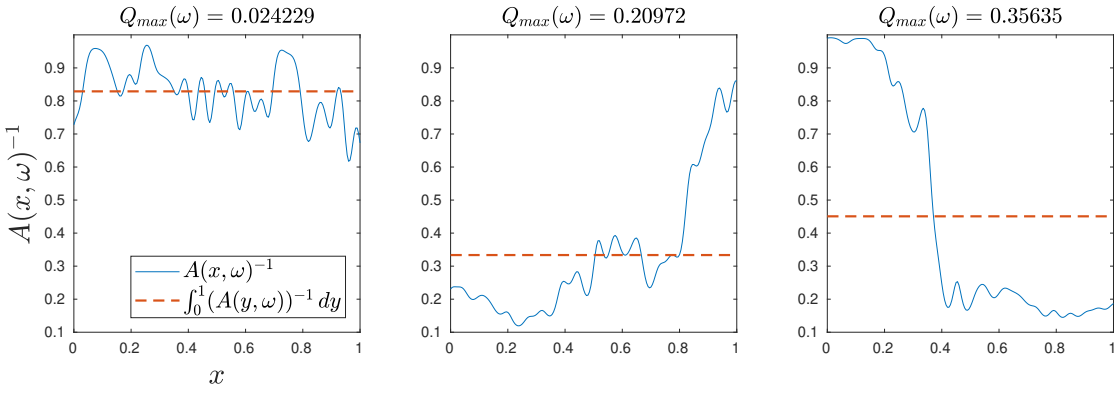


Figure 4.3: Samples of the inverse random field $A(x,\omega)^{-1}$ with their corresponding spatial averages and values of Q_{max}.

From Figure 4.3, it is speculated that two possible indicators for large Q_{max} are: the value of $1/A(x_c(\omega),\omega)$ and the frequency that $1/A(x,\omega)$ intersects its spatial average. We validate these indicators using data on $A(x,\omega)$ and $Q_{max}(\omega)$, momentarily neglecting the computational cost of obtaining such data.

Regarding the first indicator, it is observed that $1/A(x_c(\omega),\omega)$ cannot assume values close to 0 or 1 for $Q_{max}(\omega)$ to be large. Otherwise, if $1/A(x_c(\omega),\omega)$ were close to 0 or 1, the boundedness of $A(x,\omega)$ and (4–5) imply that $A(x,\omega)$ fluctuates near its spatial average. Hence, $\frac{A(x_c(\omega),\omega)}{A(y,\omega)} \approx 1$ resulting to a low $Q_{max}(\omega)$ according to (4–6). Probabilistic arguments also offer conclusions consistent with the above claims. Assume that $A(x,\omega)^{-1}$ is a homogeneous Gaussian process (despite its impossibility in our setting) on $x \in [0,1]$ such that $E[A(x,\omega)^{-1}] = \mu$ and $\mathrm{Cov}(A(x,\omega)^{-1}, A(y,\omega)^{-1}) = c(x,y)$ with $c(x,x) = \sigma^2$. From [35], $\int_0^1 A(y,\omega)^{-1}\,dy$ is a Gaussian random variable with mean μ, variance $\int_0^1\int_0^1 c(u,v)\,du\,dv$ and that the random vector $(A(x,\omega)^{-1}, \int_0^1 A(y,\omega)^{-1}\,dy)^T$ is jointly Gaussian with $Cov(A(x,\omega)^{-1}, \int_0^1 A(y,\omega)^{-1}\,dy) = \int_0^1 c(x,y)\,dy$ for all $x \in [0,1]$. In summary,

$$\begin{bmatrix} A(x,\omega)^{-1} \\ \int_0^1 A(y,\omega)^{-1}\,dy \end{bmatrix} \sim N\left(\begin{bmatrix}\mu\\ \mu\end{bmatrix}, \begin{bmatrix} \sigma^2 & \int_0^1 c(x,y)\,dy \\ \int_0^1 c(x,y)\,dy & \int_0^1\int_0^1 c(u,v)\,du\,dv \end{bmatrix}\right)$$

from which we infer

$$A(x,\omega)^{-1}\Big|\int_0^1 A(y,\omega)^{-1}\,dy = \xi \sim N\left(\mu + \frac{\int_0^1 c(x,y)\,dy}{\int_0^1\int_0^1 c(u,v)\,du\,dv}(\xi-\mu), \sigma^2 - \frac{[\int_0^1 c(x,y)\,dy]^2}{\int_0^1\int_0^1 c(u,v)\,du\,dv}\right) \tag{4–7}$$

using properties of conditioning on Gaussian distributions. Three conclusions arise from (4–7): for a fixed $x \in [0,1]$,

1. $E[A(x,\omega)^{-1}|\int_0^1 A(y,\omega)^{-1}\,dy = \xi] > \mu$ if $\xi > \mu$,
2. $E[A(x,\omega)^{-1}|\int_0^1 A(y,\omega)^{-1}\,dy = \xi] < \mu$ if $\xi < \mu$,
3. $Var[A(x,\omega)^{-1}|\int_0^1 A(y,\omega)^{-1}\,dy = \xi] < \sigma^2$.

The first and second state that for all $x \in [0,1]$, if the spatial average of the random field is large (or small), the values of the random field also tend to be

large (or small). Note that for each x, $A(x,\omega)^{-1}$ and $\int_0^1 A(y,\omega)^{-1}\,dy$ are positively correlated since in our example $\text{Cov}(A(x,\omega)^{-1}, \int_0^1 A(y,\omega)^{-1}\,dy) = \int_0^1 c(x,y)\,dy > 0$. Meanwhile, the third suggests that the variance of the random field decreases if it is known that its spatial average assumes some value. These conclusions are consistent with our discussion earlier that for large (or small) $\int_0^1 A(y,\omega)^{-1}\,dy$, values of $A(x,\omega)^{-1}$ fluctuate more closely near $\int_0^1 A(y,\omega)^{-1}\,dy$ on average leading to low $Q_{max}(\omega)$. Hence, values of the indicator $1/A(x_c(\omega),\omega)$ cannot be close to 0 or 1 for large $Q_{max}(\omega)$. Data confirms this as shown in the left panel of Figure 4.4 which displays 10000 samples of $(Q_{max}(\omega), \int_0^1 A(y,\omega)^{-1}\,dy)$ and 5000 samples of $(Q_{max}(\omega), \int_0^1 A(y,\omega)^{-1}\,dy \,|\, Q_{max}(\omega) > \tau)$. We deduce that a necessary condition for $Q_{max}(\omega) > \tau$ is that $1/A(x_c(\omega),\omega) \in [0.15, 0.6]$ approximately.

Under a similar reasoning, we intuit that a second indicator for $Q_{max}(\omega) > \tau$ is that the number of times $n_{x_c}(\omega)$ that $A(x,\omega)^{-1}$ crosses its spatial average cannot be substantial, i.e. the cardinality of the set $\{x_c(\omega) \,|\, \frac{1}{A(x_c(\omega),\omega)} = \int_0^1 \frac{1}{A(y,\omega)}\,dy\}$ is small. Notice that the first indicator does not necessarily imply the latter and vice versa. Data presented in the middle panel of Figure 4.4 supports this indicator in which 10000 samples of $(Q_{max}(\omega), n_{x_c}(\omega))$ and 5000 samples of $(Q_{max}(\omega), n_{x_c}(\omega) \,|\, Q_{max}(\omega) > \tau)$ are plotted. The plot suggests that a necessary condition for $Q_{max}(\omega) > \tau$ is that $n_{x_c}(\omega) \leq 5$.

The above discussion is summarized in the right panel of Figure 4.4 which exhibits 10000 samples of $(\int_0^1 A(y,\omega)^{-1}\,dy, n_{x_c}(\omega), Q_{max}(\omega))$ and 5000 samples of $(\int_0^1 A(y,\omega)^{-1}\,dy \,|\, Q_{max}(\omega) > \tau, n_{x_c}(\omega) \,|\, Q_{max}(\omega) > \tau, Q_{max}(\omega))$. We have therefore identified 2 indicators which serve as necessary conditions that aid in predicting if Q_{max} is large. However, these indicators were deduced assuming that the analytical solution (4–3) were known which might not be the case in practice.

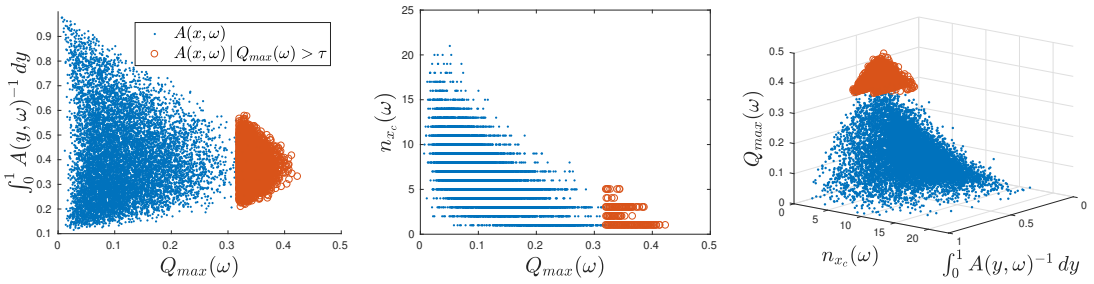


Figure 4.4: Samples of $(Q_{max}(\omega), \int_0^1 A(y,\omega)^{-1}\,dy)$ (left panel), $(Q_{max}(\omega), n_{x_c}(\omega))$ (middle panel), and $(\int_0^1 A(y,\omega)^{-1}\,dy, n_{x_c}(\omega), Q_{max}(\omega))$ (right panel). Conditional samples corresponding to $Q_{max}(\omega) > \tau$ are indicated by unfilled orange circles in each panel.

We show in the next section that the search for indicators, assuming they exist, can be facilitated in the case when $A(x,\omega)$ is represented as a finite-dimensional noise model.

4.2.2 Finite-dimensional noise model

We revisit Example 20 but examine an alternative expression for the random field $A(x,\omega)$.

Example 21. Consider the identical setup outlined in Example 20. Recast the random field $A(x,\omega)$ into a parametric form $A(x,\omega) = A(x,Z(\omega)) = a_0 + \sum_{k=1}^{\infty} Z_k(\omega)a_k(x)$ where $Z(\omega) = (Z_1(\omega), Z_2(\omega), \dots)$, $\{Z_k(\omega)\}_{k=1}^{\infty}$ are possibly non-Gaussian correlated random variables such that $E[Z_k] = 0\ \forall k$, and $\{a_k(x)\}_{k=1}^{\infty}$ are deterministic basis functions. The quantity of interest (4–4) can be rewritten as $Q_{max}(\omega) = Q_{max}(Z(\omega)) = \max_{x\in[0,1]} |U(1,x,Z(\omega)) - u(1,x)|$. The objective is to seek indicators based on components of $Z(\omega)$ that signal large $Q_{max}(Z(\omega))$.

An approach to construct such a parametric form $A(x,Z(\omega))$ for the random field $A(x,\omega)$ is through the Karhunen-Loève (KL) expansion [36]. The method

in [38] for performing the KL expansion via the singular-value decomposition (SVD) is adopted in view of numerical stability. For purposes of numerical implementation, M samples $\omega = \omega_1, \ldots, \omega_M \in \Omega$ of $A(x, \omega)$ are discretized at spatial locations $x_k = k\Delta x, k = 0, \ldots, N-1, \Delta x = \frac{1}{N-1}$, to form the $M \times N$ matrix $\mathbf{A}$ whose (j,k)-entry is $A(x_k, \omega_j)$, the value of the jth simulated sample of $A(x, \omega)$ at $x = x_k$. If $I_{M\times N}$ is the $M \times N$ identity matrix, the SVD is applied to $\mathbf{A} - a_0 I_{M\times N}$ to obtain M samples of $Z_k(\omega)$ and values of $\{a_k(x_k)\}_{k=1}^N$. Due to this approximation, the parametric form of $A(x, \omega)$ is written as a finite sum

$$A(x,\omega) = A(x, Z(\omega)) = a_0 + \sum_{k=1}^{N} Z_k(\omega) a_k(x), \quad Z(\omega) = (Z_1(\omega), \ldots, Z_N(\omega)), \quad x \in [0,1]. \tag{4–8}$$

Since $A(x, \omega) \in [\alpha, \beta]$ in Example 20, $\{Z_k\}_{k=1}^N$ are bounded random variables and hence the computer implementation of $A(x, Z(\omega))$ in (4–8) satisfies $A(x, Z(\omega)) \in [\gamma_1, \gamma_2]$ where γ_1, γ_2 approach α, β as $N \to \infty$. In the simulations below, we adhere to the representation (4–8) with $\Delta x = 0.005$.

Using the same parameter values in Example 20, the top 4 panels of Figure 4.5 plot the first four interpolated basis functions $a_k(x)$ while the bottom 4 panels contain histograms of $\{Z_k(\omega)\}_{k=1}^4$ which correspond to the random variables with the largest variance. These plots are obtained from 10000 samples of $A(x, \omega)$. In addition, each of the bottom panels show 14 asterisks which represent samples of $Z_k | Q_{max} \geq \tau$ for $k = 1, \ldots, 4$. Without making conclusions from this low sample size, it is observed that samples of Z_k for which Q_{max} is large are clustered. This behavior is not surprising as the calculations below illustrate.

Property 12. *Let $\hat{Z}, \widetilde{Z} \in \mathbb{R}^N$ be samples of $Z(\omega)$. It follows that $|Q_{max}(\hat{Z}) - Q_{max}(\widetilde{Z})| \leq \max_{x\in[0,1]} |U(1, x, \hat{Z}) - U(1, x, \widetilde{Z})|$.*

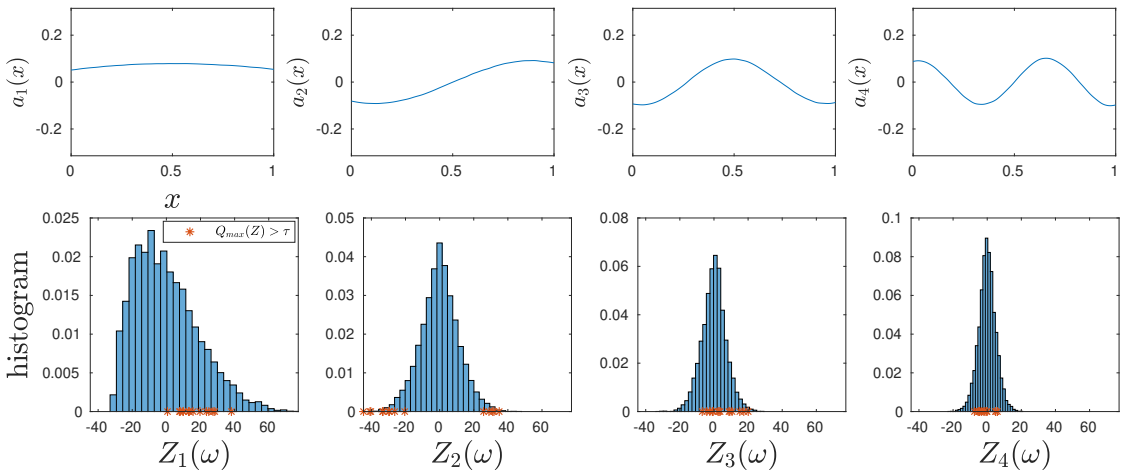


Figure 4.5: Top panels: Interpolated values of the basis functions $\{a_k(x)\}_{k=1}^4$. Bottom panels: Histogram of 10000 samples of $\{Z_k\}_{k=1}^4$ with 14 samples of $Z_k \,|\, Q_{max} > \tau$ for $\tau = 0.32$ in each subplot marked by asterisks.

This can be shown using the fact that $|\max_{x\in[0,1]} f(x) - \max_{x\in[0,1]} g(x)| \le \max_{x\in[0,1]} |f(x) - g(x)|$ for any function f, g. Hence,

$$\begin{aligned} |Q_{max}(\hat{Z}) - Q_{max}(\widetilde{Z})| &\le \max_{x\in[0,1]} \Big| |U(1,x,\hat{Z}) - u(1,x)| - |U(1,x,\widetilde{Z}) - u(1,x)| \Big| \\ &\le \max_{x\in[0,1]} \Big| (U(1,x,\hat{Z}) - u(1,x)) - (U(1,x,\widetilde{Z}) - u(1,x)) \Big|. \end{aligned}$$

Property 13. *Recall that $A(x,\omega) \in [\gamma_1, \gamma_2]$ a.s. so that $1/A(x,\omega) \in [1/\gamma_2, 1/\gamma_1]$ a.s. For $\hat{Z}, \widetilde{Z} \in \mathbb{R}^N$, there exists $C > 0$ such that $|Q_{max}(\hat{Z}) - Q_{max}(\widetilde{Z})| \le C\|A(x,\hat{Z}) - A(x,\widetilde{Z})\|_{L^2[0,1]}$.*

Since $U(1,x,Z) = \left(\int_0^x \frac{dy}{A(y,Z)}\right)\left(\int_0^1 \frac{dy}{A(y,Z)}\right)^{-1}$, let $\hat{I}(x) = \int_0^x \frac{dy}{A(y,\hat{Z})}$ and $\widetilde{I}(x) = \int_0^x \frac{dy}{A(y,\widetilde{Z})}$ so that $U(1,x,\hat{Z}) = \hat{I}(x)/\hat{I}(1)$ and $U(1,x,\widetilde{Z}) = \widetilde{I}(x)/\widetilde{I}(1)$. From [37, p. 194],

$$|U(1,x,\hat{Z}) - U(1,x,\widetilde{Z})| \le \frac{\widetilde{I}(1)|\hat{I}(x) - \widetilde{I}(x)| + \widetilde{I}(x)|\widetilde{I}(1) - \hat{I}(1)|}{\hat{I}(1)\widetilde{I}(1)}. \tag{4–9}$$

We note that $1/\gamma_2 \le \hat{I}(x), \widetilde{I}(x) \le 1/\gamma_1$ and that

$$\begin{aligned} |\hat{I}(x) - \widetilde{I}(x)| \le \frac{1}{\gamma_1^2}\int_0^x |A(y,\widetilde{Z}) - A(y,\hat{Z})|\,dy &\le \frac{1}{\gamma_1^2}\int_0^1 |A(y,\widetilde{Z}) - A(y,\hat{Z})|\,dy \\ &\le \frac{1}{\gamma_1^2}\|A(x,\hat{Z}) - A(x,\widetilde{Z})\|_{L^2[0,1]} \end{aligned} \tag{4–10}$$

for all $x \in [0,1]$ by the Cauchy-Schwarz inequality. Property 13 then follows from Property 12 and (4–9), (4–10).

Property 14. *Suppose that the parametric form in* (4–8) *is constructed with* $a_k(x)$ *appropriately scaled so that* $\int_0^1 a_k^2(x)\,dx = 1$ *for all* k. *For* $\hat{Z}, \widetilde{Z} \in \mathbb{R}^N$, *there exists* $C > 0$ *such that* $|Q_{max}(\hat{Z}) - Q_{max}(\widetilde{Z})| \leq C\left[\sum_{k=1}^{N} |\hat{Z}_k - \widetilde{Z}_k|^2\right]^{1/2}$.

This is established using Property 13 and the fact that $\int_0^1 a_j(x)a_k(x)\,dx = \delta_{jk}$ for $j,k = 1,\ldots,N$ where δ_{jk} is the Kronecker delta function. Similar results can be derived for other differential equations for which results of the form $\|U(t,x,\hat{Z}) - U(t,x,\widetilde{Z})\| \leq C\|A(x,\hat{Z}) - A(x,\widetilde{Z})\|$ hold under suitable norms.

Property 14 explains the clustering behavior of the samples of $Z_k \mid Q_{max} > \tau$ in the bottom panel of Figure 4.5. In particular, the range of $Z_2 \mid Q_{max} > \tau$ is a union of 2 disjoint sets. This is supported by Figure 4.6 which illustrates the pdf of Q_{max} conditioned on Z_2 obtained using kernel density estimation based on 250000 samples of (Z_2, Q_{max}). A consequence of this is that the failure domain $F = \{z \in \mathbb{R}^N \mid Q_{max}(z) > \tau\}$ is also a union of two disjoint sets (i.e. connected components). Determining whether $Z \in F$ requires characterizing the geometry of F which is challenging in high dimensions. This is why indicators based on components of Z are devised to aid in predicting if $Q_{max} > \tau$. From Figures 4.5, 4.6, the simplest example of an indicator could be the value that Z_2 takes, i.e. $|Z_2| > \rho$ where ρ is calibrated, which can be thought of as intimately related to the second indicator n_{x_c} in Section 4.2.1.

We also remark that if the parametric form (4–8) were constructed for $A(x,\omega)^{-1}$ instead of $A(x,\omega)$, then the random variable Z_1 associated with the basis function $a_1(x)$ in the resulting expansion can be another indicator that is

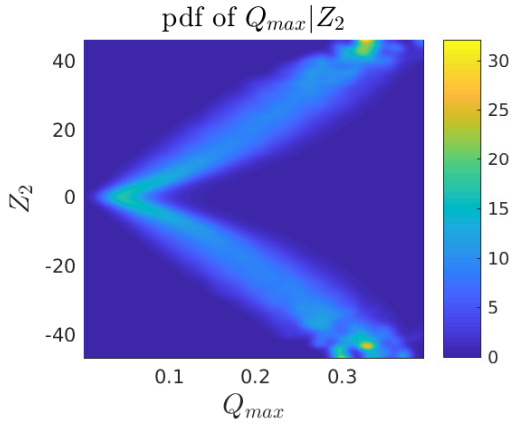


Figure 4.6: Pdf of Q_{max} conditioned on Z_2 constructed using kernel density estimation based on 250000 samples of (Z_2, Q_{max}).

closely related to the first indicator $\int_0^1 A(y,\omega)^{-1}\,dy$ introduced in Section 4.2.1. In summary, Section 4.2 highlights that indicators in certain settings may be found that act as classifiers to signal large Q_{max}. However, the discussion above also presents challenges associated with this approach. Without any physical intuition or access to any analytical formulas, identifying indicators can be difficult. Even if potential indicators can be identified, one needs to 1) verify that they correlate with large quantities of interest and 2) quantify their range of values for prediction. Doing so requires a large number of data on rare events for which simulation can be costly. In the following section, we address these issues via a multifidelity approximation to Q_{max}.

4.3 Multifidelity physics-based surrogate models

Due to the computational cost of validating and calibrating indicators, we construct a multifidelity approximation to the QoI which offers another scheme for classification. Section 4.3.1 describes the surrogate model we use while the multifidelity approach is elaborated in Section 4.3.2. The capacity of each method

as a classifier is assessed and their advantages and disadvantages are outlined.

4.3.1 SROM-based surrogate model

In the remainder of this work, we adopt the parametric form $A(x, Z)$ as in (4–8) but with Z truncated to avoid difficulties of computation in high stochastic dimension. The most straightforward approach to identify the samples of interest is to evaluate the full model $Q_{max}(Z)$. Since $Q_{max}(Z) > \tau$ is a rare event, obtaining n samples of $A(x, Z) \,|\, Q_{max}(Z) > \tau$ to understand why rare events occur requires at least $n/P(Q_{max}(Z) > \tau)$ full model solves which can be prohibitively costly. It is therefore reasonable to simulate samples of $A(x, \omega) \,|\, \widetilde{Q}_{max}(Z) > \tau$ instead where $\widetilde{Q}_{max}(Z)$ is a surrogate model that converges almost surely to $Q_{max}(Z)$. Otherwise, a surrogate model which converges to $Q_{max}(Z)$ in $L^p(\Omega)$ for p large can be a suitable approximation since if $f \in L^\infty(\Omega)$, $\lim_{p\to\infty} \|f\|_{L^p(\Omega)} = \|f\|_{L^\infty(\Omega)}$. Once the surrogate is constructed, it functions as a classifier that not only indicates if $Q_{max} > \tau$ or otherwise for a sample $A(x, Z)$ but also provides an approximate value of Q_{max}.

We remark that the mode of convergence of the surrogate is crucial for this objective. For instance, it is possible for $P(\widetilde{Q}_{max} > \tau) \approx P(Q_{max} > \tau)$ yet the samples of $A(x, Z) \,|\, \widetilde{Q}_{max} > \tau$ and $A(x, Z) \,|\, Q_{max} > \tau$ may greatly differ. We demonstrate this for a 1-dimensional input with simple QoI in Example 22 in which the polynomial chaos (PC) expansion [30, 31], which converges in $L^2(\Omega)$ but not almost surely, is used as a surrogate.

Example 22. Suppose that the response is given by $g(W) = \Phi(W)$, $W \sim N(0, 1)$. Since $g(W) \sim U(0, 1)$, $E[g(W)^2] < \infty$ and $g(W)$ admits the PC representation $g^{PCE}(W) = \beta_0 + \sum_{k=0}^{n} \beta_{2k+1} h_{2k+1}(W)$ where $h_j(W)$ are Hermite polynomials orthog-

onal with respect to the standard normal pdf while $\beta_0 = \frac{1}{2}, \beta_{2k} = 0, k = 1, \ldots, n$ and $\beta_{2k+1} = (-1)^k \frac{(2k)!}{2^{2k+1}\sqrt{\pi}(2k+1)!k!}, k = 0, 1, \ldots, n$. The objective is to compare the sets of samples $W \,|\, g(W) > \tau$ and $W \,|\, g^{PCE}(W) > \tau$ for specified τ.

We choose $\tau = 0.9825$ so that $P(g(W) > \tau) = 0.0175$. In practice, the truncation level for the PC expansion must be kept low to prevent an explosion in the number of terms if W is high-dimensional. In our simulations, we select $n = 5$ so that $P(g^{PCE}(W) > \tau) \approx 0.0190$ which has relative error 8.57%. Figure 4.7 shows histograms of 10252 samples of $W \,|\, g(W) > \tau$ and 10094 samples of $W \,|\, g^{PCE}(W) > \tau$, respectively. It is seen that the histogram due to the PCE surrogate is biased and has a thinner tail compared to the true histogram. For example, $P(W > 2.5 \,|\, g(W) > \tau) \approx 0.3641$ whereas $P(W > 2.5 \,|\, g^{PCE}(W) > \tau) \approx 0.0148$.

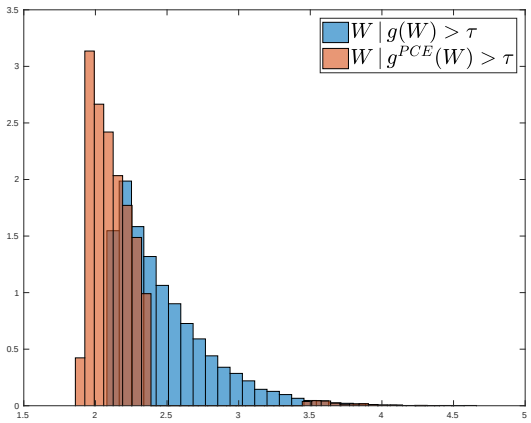


Figure 4.7: Histograms of 10210 samples of $W \,|\, g(W) > \tau$ and 10317 samples of $W \,|\, g^{PCE}(W) > \tau$ for Example 22.

In view of these issues, an SROM-based surrogate model [40] $\widetilde{U}(t, x, Z)$ for the full model $U(t, x, Z)$ is employed from which the approximation for $Q_{max}(Z)$ arises. The surrogate is constructed by partitioning the range $Z(\Omega) = \Gamma \subset \mathbb{R}^N$ of Z using Voronoi cells $\{\Gamma_k\}_{k=1}^M$ whose centers $\{z_k\}_{k=1}^M \subset \mathbb{R}^N$ are samples of Z chosen to approximate its distribution. A first-order Taylor expansion at $Z = z_k$ for every

partition is computed to obtain

$$\widetilde{U}(t,x,Z) = \sum_{k=1}^{M} (U(t,x,z_k) + \nabla_Z U(t,x,z_k)\cdot(Z-z_k))\mathbb{1}_{Z\in\Gamma_k} \tag{4–11}$$

from which the surrogate approximation $\widetilde{Q}_{max}(Z) = \max_{x\in[0,1]} |\widetilde{U}(1,x,Z) - u(1,x)|$ is then derived. Under mild conditions, it can be shown that $\widetilde{U}(t,x,Z) \to U(t,x,Z)$ almost surely and in $L^p(\Omega)$ as $M \to \infty$. From Property 12, these convergence results also hold for $\widetilde{Q}_{max}(Z)$.

To improve the performance of (4–11), in previous work [73], we proposed an adaptive construction of (4–11) that sequentially refines the surrogate $\widetilde{U}(t,x,Z)$ to enhance its accuracy in high-probability regions of Γ and regions for which the discrepancy between the QoI $Q(Z)$ and the surrogate QoI $\widetilde{Q}(Z)$ is large. This was achieved by setting the refinement criterion as $\|Q(Z) - \widetilde{Q}(Z)\|_{L^p(\Omega)}$ for specified p which also served as the stopping criterion. The L^p error was crudely approximated via $\frac{1}{n_1+\cdots+n_M}\sum_{k=1}^{M}\sum_{\ell=1}^{n_k} |Q(z_k^\ell) - \widetilde{Q}(z_k^\ell)|^p$ where $\{z_k^\ell\}_{\ell=1}^{n_k} \subset \Gamma_k$ are a few samples of Z at which we evaluate the response and from which we select the next point at which the gradient in (4–11) is calculated. Evaluating the response is less expensive than computing multiple partial derivatives and furthermore, this approximation ensures that the L^p error is estimated via contributions from each partition. As the adaptive algorithm progresses, Γ_k decreases in size and the L^p error in this partition can be approximated using a few samples only.

We now investigate the use of the SROM-based surrogate for $Q_{max}(Z)$ to produce samples that resemble $A(x,Z)\,|\,Q_{max}(Z) > \tau$. Two surrogate models are examined: denote by $\widetilde{Q}_{max}^{direct}$ the direct construction (4–11) and by $\widetilde{Q}_{max}^{adapt}$ the adaptive construction described above.

Example 23. We revisit the physical setup in Example 20 with the random field

$A(x,\omega)$ expressed in its parametric form $A(x,Z)$ (4–8). Let the true random field be

$$A(x,Z(\omega)) = a_0 + \sum_{k=1}^{10} Z_k(\omega)a_k(x), \quad Z(\omega) = (Z_1(\omega),\ldots,Z_{10}(\omega)), \quad x \in [0,1], \qquad (4\text{–}12)$$

which is (4–8) truncated to 10 basis functions only. The performance of $\widetilde{Q}_{max}^{direct}$ and $\widetilde{Q}_{max}^{adapt}$ in generating conditional samples that approximate $A(x,Z) \,|\, Q_{max}(Z) > \tau$ is compared with the computational cost of both surrogates being similar. Their use in identifying samples of $A(x,Z)$ that yield large Q_{max} is also examined.

We verified via 500000 Monte Carlo samples that the truncation level in (4–12) is sufficient for $A(x,Z) > 0$. In the calculations that follow, we adjust $\tau = 0.3172$ so that $P(Q_{max}(Z) > \tau) \approx 0.00224$ because we are now working with the truncated random field (4–12) instead of (4–8). It is assumed that calculating a partial derivative in the gradient term in (4–11) is as costly as evaluating the response $U(t,x,Z)$. We computed the gradients from (4–3) but they can also be numerically estimated by differentiating (4–1) and solving the resulting differential equations satisfied by $\nabla_Z U(t,x,Z)$ as in [73]. The surrogate $\widetilde{Q}_{max}^{direct}$ obtained comprises 370 Voronoi cells and constitutes 370 response evaluations and 370×10 gradient evaluations for a total of 4070 computational units. In contrast, $\widetilde{Q}_{max}^{adapt}$ is composed of 325 Voronoi cells which implies 3575 gradient and response calculations and an additional 465 response evaluations that are used to estimate the refinement criterion at each iteration, resulting in a total of 4040 computational units. For the adaptive construction, we used $p = \infty$ for the refinement criterion since we are targeting rare events and employed the global sampling and neighbor-based refinement configuration, following terminology in [73].

Figure 4.8 compares the histograms of Q_{max}, $\widetilde{Q}_{max}^{adapt}$ and $\widetilde{Q}_{max}^{direct}$ constructed using 10000 samples. The probability of failure estimates for each surrogate are $P(\widetilde{Q}_{max}^{adapt} > \tau) \approx 0.001834$ and $P(\widetilde{Q}_{max}^{direct} > \tau) \approx 0.001754$ obtained using 500000 Monte Carlo samples of Z. The discrepancy between the 2 surrogate models is not obvious unless we examine samples of $Z \,|\, Q_{max}(Z) > \tau$ which is relevant to our objectives. Figures 4.9 and 4.10 contrast the histograms of $Z_i \,|\, Q_{max}(Z)$ with $Z_i \,|\, \widetilde{Q}_{max}^{adapt}(Z)$ and $Z_i \,|\, \widetilde{Q}_{max}^{direct}$, respectively, for $i = 1, \ldots, 6$. These histograms result from 10000 samples of $Z \,|\, Q_{max}(Z)$, 10001 samples of $Z \,|\, \widetilde{Q}_{max}^{adapt}(Z)$ and 10022 samples of $Z \,|\, \widetilde{Q}_{max}^{direct}(Z)$. Based on these figures, the advantages of using an adaptive construction is now evident since the histograms of $Z_1 \,|\, \widetilde{Q}_{max}^{direct}$ and $Z_3 \,|\, \widetilde{Q}_{max}^{direct}$ appear biased while the modes of $Z_2 \,|\, \widetilde{Q}_{max}^{direct}$ do not appear to have the same frequency.

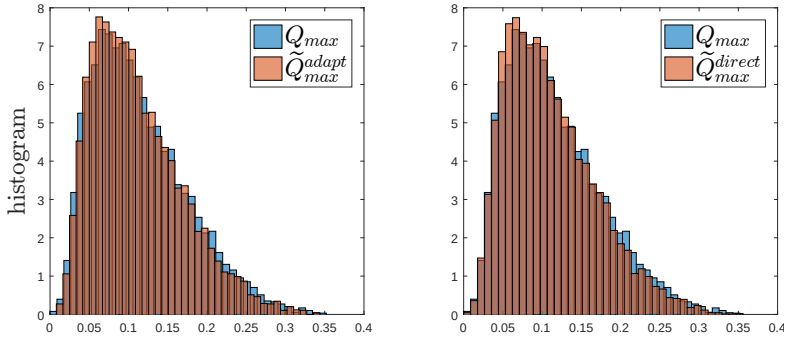


Figure 4.8: Comparison of the histogram of Q_{max} vs $\widetilde{Q}_{max}^{adapt}$ (left panel) and Q_{max} vs $\widetilde{Q}_{max}^{direct}$ (right panel) using 10000 samples.

Because of the performance of $\widetilde{Q}_{max}^{adapt}(Z)$, we further study its suitability in achieving our objectives. How well does $\widetilde{Q}_{max}^{adapt}(Z)$ classify whether $Q_{max}(Z) > \tau$ or $Q_{max}(Z) \leq \tau$ for a given Z? To accomplish this, tools and metrics commonly used in machine learning and pattern recognition [29] are adopted. Some of these metrics have also been invoked in [26, 27] to validate and calibrate indicators. These metrics can be extracted from a confusion matrix which tabulates in each entry the frequency that a classifier coincides or not with the truth; see

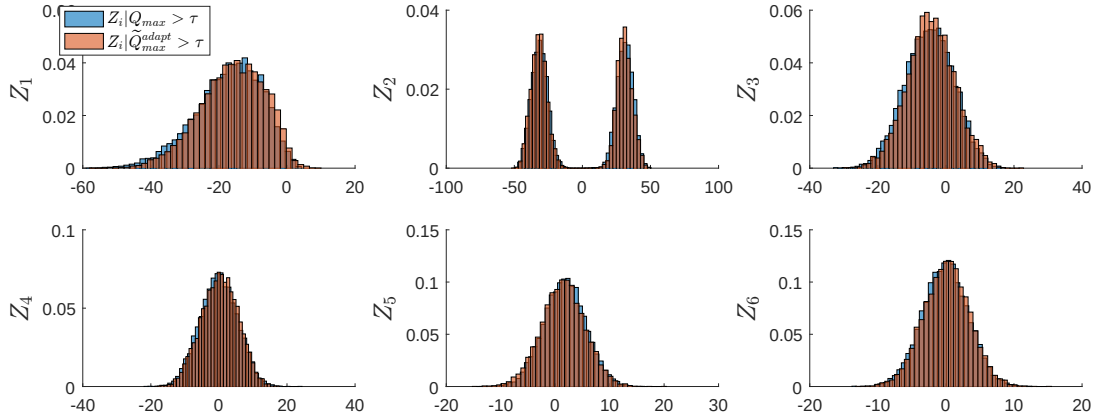


Figure 4.9: Comparison of histograms of $Z_i \,|\, Q_{max}(Z) > \tau$ and $Z_i \,|\, Q^{adapt}_{max}(Z) > \tau$ for $i = 1, \ldots, 6$.

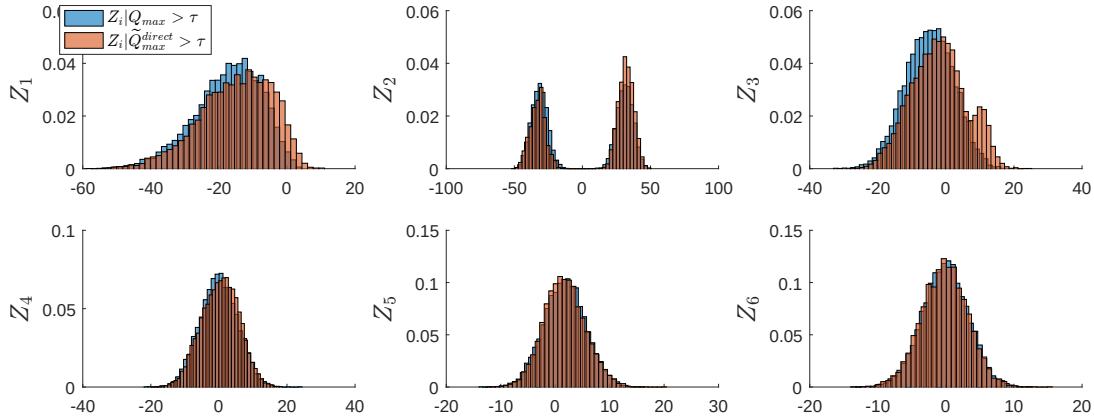


Figure 4.10: Comparison of histograms of $Z_i \,|\, Q_{max}(Z) > \tau$ and $Z_i \,|\, Q^{direct}_{max}(Z) > \tau$ for $i = 1, \ldots, 6$.

Table 4.1 for a visual definition where N_s is the total number of samples. Tables 4.2 and 4.3 chart the confusion matrices of $\widetilde{Q}^{adapt}_{max}(Z)$ generated using 10000 and 50000 samples of Z, respectively, with the latter set of samples including the former. These sets of samples are referred to as the test data.

		$Q_{max}(Z)$	
		Positive	Negative
$\widetilde{Q}_{max}^{adapt}(Z)$	Positive	$N_s P(Q_{max} > \tau, \widetilde{Q}_{max}^{adapt} > \tau)$	$N_s P(Q_{max} \leq \tau, \widetilde{Q}_{max}^{adapt} > \tau)$
	Negative	$N_s P(Q_{max} > \tau, \widetilde{Q}_{max}^{adapt} \leq \tau)$	$N_s P(Q_{max} \leq \tau, \widetilde{Q}_{max}^{adapt} \leq \tau)$

Table 4.1: Definition of the confusion matrix.

		$Q_{max}(Z)$		
		Positive	Negative	Total
$\widetilde{Q}_{max}^{adapt}(Z)$	Positive	11	1	12
	Negative	7	9981	9988
	Total	18	9982	10000

Table 4.2: Confusion matrix for $\widetilde{Q}_{max}^{adapt}(Z)$
based on 10000 samples of Z.

		$Q_{max}(Z)$		
		Positive	Negative	Total
$\widetilde{Q}_{max}^{adapt}(Z)$	Positive	84	17	101
	Negative	42	49857	49899
	Total	126	49874	50000

Table 4.3: Confusion matrix for $\widetilde{Q}_{max}^{adapt}(Z)$
based on 50000 samples of Z.

Confusion matrix	Precision	Recall	Failure rate
Table 4.2	0.9167	0.6111	0.0840
Table 4.3	0.8317	0.6667	0.1691

Table 4.4: Precision, recall, and failure rate metrics computed from Tables 4.2, 4.3.

Precision and recall metrics are calculated to assess the quality of $\widetilde{Q}_{max}^{adapt}(Z)$ as a classifier. Precision is defined as $P(Q_{max}(Z) > \tau \,|\, \widetilde{Q}_{max}^{adapt}(Z) > \tau)$, i.e. the fraction of true positives with respect to the total number of true positives and false positives while recall is defined as $P(\widetilde{Q}_{max}^{adapt}(Z) > \tau \,|\, Q_{max}(Z) > \tau)$, i.e. the fraction of true positives with respect to the total number of true positives and false negatives. In our setting, perfect precision implies that the failure domain of $\widetilde{Q}_{max}^{adapt}(Z)$ is a subset of the failure domain of $Q_{max}(Z)$ while the reverse is true for perfect recall. We also consider the failure rate which served as a basis for validation and calibration of indicators for extreme events in [27]. This is defined as the sum of the false omission rate and the false discovery rate or mathematically, $P(Q_{max}(Z) > \tau | \widetilde{Q}_{max}^{adapt}(Z) < \tau) + P(Q_{max}(Z) < \tau | \widetilde{Q}_{max}^{adapt}(Z) > \tau)$. These metrics are summarized in Table 4.4 from the confusion matrices in Tables 4.2, 4.3. It is desirable for both precision and recall to be close to 1 and for the failure rate to be near 0. Consequently, the performance of $\widetilde{Q}_{max}^{adapt}(Z)$ as a classifier can still be improved. We also remark that using a small number of samples to construct the confusion matrix can result in unstable estimates of the frequency values. Table 4.5 shows the confusion matrix for $\widetilde{Q}_{max}^{adapt}(Z)$ using another set of 10000 samples of Z. The metrics based on Table 4.5 appear inflated compared to those from Table 4.3.

$\widetilde{Q}_{max}^{adapt}(Z)$ \ $Q_{max}(Z)$	Positive	Negative	Total
Positive	20	1	21
Negative	6	9973	9979
Total	26	9974	10000

Table 4.5: Confusion matrix for $\widetilde{Q}_{max}^{adapt}(Z)$ based on a different set of 10000 samples of Z.

More computational effort can certainly be expended to make $\widetilde{Q}_{max}^{adapt}(Z)$ more accurate, however, for the SROM-based surrogate, an additional Voronoi cell requires at least $N + 1$ calculations for $Z \in \mathbb{R}^N$. Other surrogate models such as the sparse grid stochastic collocation also incur substantial cost when transitioning from one sparse grid level to another [73]. We deal with these issues by investigating the use of a multifidelity surrogate model as described in Section 4.3.2.

4.3.2 Multifidelity surrogate approach

A multifidelity surrogate leverages on various surrogate models with differing computational costs and levels of accuracy to accomplish an objective. The multifidelity approach we pursue consists of a single surrogate model and infrequent calls to the full model. Such approaches have already been applied in previous work [17, 51] in which the goal is computation of failure probabilities. Here, we adopt some ideas from these works to tackle our objectives. We first survey how these multifidelity surrogates were tailored to approximate the probability of failure $P(Q(Z) > \tau)$.

Let $\widetilde{Q}(Z)$ be the surrogate model to the QoI $Q(Z)$. Instead of the Monte Carlo estimate $\frac{1}{M}\sum_{i=1}^{M}\mathbb{1}_{\{\widetilde{Q}(z_i)>\tau\}}$ where $\{z_i\}_{i=1}^{M}$ are samples of Z, [51] advocates using estimate $\widetilde{P}_f^M = \frac{1}{M}\sum_{i=1}^{M}\mathbb{1}_{\{z_i\in\widetilde{F}\}}$ where

$$\widetilde{F} = \{z \,|\, \widetilde{Q}(z) > \tau + \gamma\} \cup \{z \,||\widetilde{Q}(z) - \tau| < \gamma \text{ and } Q(z) > \tau\}$$

for specified $\gamma > 0$. In other words, if the value of the surrogate exceeds the threshold τ by a margin γ, it is believed that value of the full model is also above τ, while if the surrogate value is close to τ, the full model is invoked to ascertain whether $Q(z) > \tau$ or $Q(z) \leq \tau$. Moreover, let $\widetilde{P}_f = P(Z \in \widetilde{F})$ be such that $\widetilde{P}_f^M$ is an estimator for $\widetilde{P}_f$ and let $p \geq 1$ such that $Q(Z), \widetilde{Q}(Z) \in L^p(\Omega)$. It was proven in [51] that for $\epsilon > 0$, $|P(Q(Z) > \tau) - \widetilde{P}_f| < \epsilon$ if $\gamma > \frac{1}{\epsilon^{1/p}}\|Q(Z) - \widetilde{Q}(Z)\|_{L^p(\Omega)}$.

In addition to [51], [17] tackled the same problem using a similar multifidelity approach. They introduced the concept of reliability which is based on an approximation to the error between $Q(Z)$ and $\widetilde{Q}(Z)$ and a constant called safety factor. Knowledge of this constant coupled with infrequent solves of the full model guarantees that their approach is able to exactly classify whether $Q(z) > \tau$ or $Q(z) \leq \tau$ for each sample z. The failure probability estimate from this method is hence identical to the Monte Carlo estimate obtained from the full model, i.e. $\frac{1}{M}\sum_{i=1}^{M}\mathbb{1}_{\{Q(z_i)>\tau\}}$.

Despite the demonstrated effectiveness of the abovementioned approaches in approximating failure probabilities, they may not be readily applied to our problem since convergence in probabilities may or may not be indicative of the samples. While the guarantee of [17] to exactly classify whether $Q(z) > \tau$ or $Q(z) \leq \tau$ at a reduced computational expense avoids this concern and accomplishes our objectives, the method is contingent on constants that need to be estimated. Likewise, the convergence bounds in [51] rely on unknown constants

whereas the more practical approach the authors proposed is geared towards convergence in estimate of the failure probability.

We therefore modify the multifidelity surrogate in [51] to obtain $\widetilde{Q}^*(Z)$ defined in Algorithm 4. Succinctly, the full model is only invoked if $\widetilde{Q}(Z) > \tau$ to ascertain whether $Q(Z) > \tau$ or $Q(Z) \leq \tau$. Unlike [51], if $\widetilde{Q}(Z) > \tau + \gamma$ for $\gamma > 0$, Algorithm 4 does not assume that $Q(Z) > \tau$ nor does it invoke $Q(Z)$ if $\widetilde{Q}(Z) > \tau - \gamma$. We note the following analytical and practical properties of $\widetilde{Q}^*(Z)$:

1. It does not rely on constants to be specified except for the threshold value τ.
2. The frequency that the full model is invoked is $P(\widetilde{Q}(Z) > \tau)$ which can be inexpensively approximated using Monte Carlo simulation beforehand.
3. The proportion of false positives of $\widetilde{Q}^*(Z)$ acting as a classifier, i.e. $P(Q(Z) \leq \tau, \widetilde{Q}^*(Z) > \tau)$, is 0. This is because from Algorithm 4, $\{\widetilde{Q}^*(Z) > \tau\} = \{\widetilde{Q}(Z) > \tau\} \cap \{Q(Z) > \tau\}$ so that $\{\widetilde{Q}^*(Z) > \tau\} \cap \{Q(Z) \leq \tau\} = \emptyset$.
4. If Monte Carlo samples of $\widetilde{Q}^*(Z)$ are generated, the quality of the surrogate $\widetilde{Q}(Z)$ can be monitored during the sampling procedure by calculating $P(Q(Z) > \tau \,|\, \widetilde{Q}(Z) > \tau)$ using the samples produced thus far. If this estimate is zero, the surrogate needs to be refined further.

Bounds on the misclassification probability $P(Q(Z) > \tau, \widetilde{Q}^*(Z) \leq \tau) + P(Q(Z) < \tau, \widetilde{Q}^*(Z) \geq \tau)$ can also be constructed but are of little practical value. Since $\{\widetilde{Q}^*(Z) \leq \tau\} = \{\widetilde{Q}(Z) \leq \tau\} \cup \{\widetilde{Q}(Z) > \tau, Q(Z) \leq \tau\}$ and that $\{\widetilde{Q}(Z) > \tau\} = \{\widetilde{Q}(Z) > \tau, Q(Z) > \tau\}$,

$$\begin{aligned} &P(Q(Z) > \tau, \widetilde{Q}^*(Z) \leq \tau) + P(Q(Z) < \tau, \widetilde{Q}^*(Z) \geq \tau) \\ &\leq P(Q(Z) > \tau, \widetilde{Q}(Z) \leq \tau) \leq P(|Q(Z) - \widetilde{Q}(Z)| \geq 0), \end{aligned}$$

however, $P(|Q(Z) - \widetilde{Q}(Z)| \geq 0)$ is usually not available in practice. We now investigate the performance of the multifidelity surrogate model applied to the setup in Example 23.

Algorithm 4 Multifidelity surrogate model $\widetilde{Q}^*(Z)$

1: Let $Q(Z)$ be the QoI and $\widetilde{Q}(Z)$ be a surrogate model.
2: For a sample z of Z:
3: **if** $\widetilde{Q}(z) > \tau$ **then**
4: Evaluate $Q(z)$.
5: Set $\widetilde{Q}^*(z) = Q(z)$.
6: **else**
7: Set $\widetilde{Q}^*(z) = \widetilde{Q}(z)$.
8: **end if**

Example 24. Consider the setup in Example 23 and let $\widetilde{Q}^{adapt}_{max}(Z)$ be the surrogate model approximation to $Q_{max}(Z)$ resulting from the adaptive construction outlined in Section 4.3.1. Denote by $\widetilde{Q}^*_{max}$ the multifidelity surrogate model according to Algorithm 4. We examine the performance of $\widetilde{Q}^*_{max}(Z)$ in classifying whether $Q_{max}(Z) > \tau$ or $Q_{max}(Z) \leq \tau$.

By generating Monte Carlo samples of $\widetilde{Q}^*_{max}(Z)$ according to Algorithm 4, we can readily investigate the convergence of the Monte Carlo estimate for the precision, $P(Q_{max}(Z) > \tau \,|\, \widetilde{Q}^{adapt}_{max} > \tau)$, at no additional computational expense. This is displayed in the left panel of Figure 4.11 which plots this quantity as a function of the number of samples of $\widetilde{Q}^{adapt}_{max}(Z) > \tau$. For completeness, the Monte Carlo estimate of the recall, $P(\widetilde{Q}^{adapt}_{max} > \tau \,|\, Q_{max}(Z) > \tau)$, as a function of the number of samples of $Q_{max}(Z) > \tau$ is plotted in the right panel, although this is acquired through extra evaluations of the full model. Both plots show that the precision and recall estimates are close to convergence after 100 samples of the conditioned event.

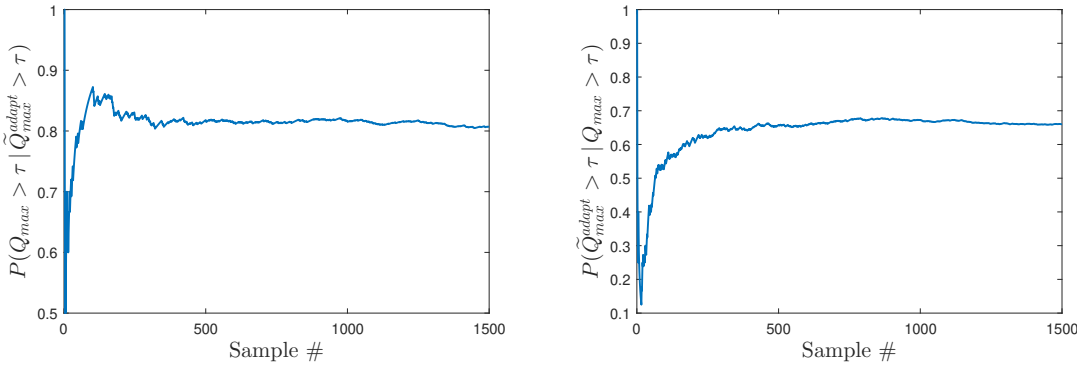


Figure 4.11: Monte Carlo estimates of precision (left panel) and recall (right panel) as a function of the number of samples of $\widetilde{Q}^{adapt}_{max}(Z) > \tau$ and $Q_{max}(Z) > \tau$, respectively.

As in Example 23, we tabulate the confusion matrix for $\widetilde{Q}^{*}_{max}(Z)$ as a classifier which we display in Tables 4.6 and 4.7 using the same test set of 10000 and 50000 samples of Z. The only difference in these confusion matrices is that the entries in the second column are updated; in particular, the false positives ($\{\widetilde{Q}^{*}_{max}(Z) > \tau\} \cap \{Q_{max}(Z) \leq \tau\}$) are now 0. The calculated metrics from Table 4.6 are 1 for precision, 0.6111 for recall, and 0.0007 for the failure rate. For Table 4.7, they are 1 for precision, 0.6667 for recall, and 0.0008 for the failure rate. Despite the perfect precision and low failure rate, the recall is still low for $\widetilde{Q}^{*}_{max}(Z)$ to be qualified as a desirable classifier that can aid in predicting if $Q_{max}(Z)$ is large. Even the recall rate were high, certain features of the multifidelity surrogate may be undesirable in certain applications. While $\widetilde{Q}^{*}_{max}(Z)$ can be utilized to generate samples of $A(x, Z) \mid Q_{max}(Z) > \tau$ to study its conditional law, a classification scheme that does not invoke the full model anymore may be preferred for purposes of prediction. We therefore investigate the use of a machine learning classifier in Section 4.4 to complement the multifidelity surrogate.

4.4 Machine learning classifiers

In the previous sections, we investigated various classification schemes to accomplish our objectives. Section 4.2 underscored that physics-based indicators from $A(x, \omega)$, if they can be found, require numerous full model solves for validation and calibration. In Section 4.3, we saw that the multifidelity surrogate model still necessitated infrequent full model evaluations for prediction. In this section, we leverage on the strengths of various classification schemes through a two-stage approach that builds on Section 4.3 and commonly used machine learning classifiers. We first assess in Section 4.4.1 the performance of only using a machine learning classifer for rare event simulations. Section 4.4.2 examines the general framework we propose.

		$Q_{max}(Z)$		
		Positive	Negative	Total
$\widetilde{Q}^*_{max}(Z)$	Positive	11	0	11
	Negative	7	9982	9989
	Total	18	9982	10000

Table 4.6: Confusion matrix for $\widetilde{Q}^*_{max}(Z)$ based on 10000 samples of Z.

		$Q_{max}(Z)$		
		Positive	Negative	Total
$\widetilde{Q}^*_{max}(Z)$	Positive	84	0	84
	Negative	42	49874	49916
	Total	126	49874	50000

Table 4.7: Confusion matrix for $\widetilde{Q}^*_{max}(Z)$ based on 50000 samples of Z.

4.4.1 Support vector machines (SVMs) for rare event simulations

In this subsection, we attempt to address our objectives of identifying input random field samples that yield extreme response through machine learning classifiers. While the classification schemes studied previously incorporated knowledge about the system equations at varying levels, classifiers traditionally used in machine learning are constructed solely based on training data. Suppose that the training set is represented by samples $\{(z_i, y_i)\}_{i=1}^{M}$ of the random vector (Z, Y) with $Z \in \mathbb{R}^d$ and $Y = \pm 1$ where the $Y = +1$ label refers to $Q_{max}(Z) > \tau$ with $Y = -1$ signifying otherwise. A machine learning classifier aims to partition these samples into two classes according to their label y_i in a manner that holds (generalizes well) for samples not observed in the data set. Once constructed, it can be used to determine the label of an arbitrary sample of Z and conclude if it resides in the failure region. Unlike surrogate models which approximate the QoI, the output of a machine learning classifier depends on the type used: logistic regression returns the probability that a sample belongs to a class, SVMs evaluate the value of a function whose sign (+ or −) indicates class membership, etc.

To support the classification process, two other sets of samples are also utilized, namely validation and test data, aside from the training data. Most machine learning classifiers depend on hyperparameters such as parameters of the optimization algorithm used to train the classifier. The parameters of the classifier then depend on these hyperparameter values. The latter needs to be specified beforehand so that the former can be found via optimization. Changing the hyperparameter values may alter the optimal parameters obtained for the

classifier, and hence the classifier itself. The validation data therefore serve to tune the values of the hyperparameters – they are selected such that the resulting optimal classifier is able to partition the validation set sufficiently. Test data, on the other hand, are invoked to report statistics such as the confusion matrices in Section 4.3.

Various algorithms exists for classification [42] yet we only concentrate on support vector machines due to its ease of implementation and geometric interpretation; the discussion in Section 4.4.2 is not catered to a specific classification algorithm. A review of the important concepts of SVMs [20,42,58] that we allude to, including its hyperparameters, is presented in Appendix B. We only implement SVMs with Gaussian kernels relying on the scale parameter γ. Despite the relative popularity of SVMs in classification problems, its performance may be hindered in rare event applications due to the scarcity of rare event samples in the training data. This implies multiple solves of the full model to guarantee a sufficient number of training samples in the failure region. To resolve this, existing work [5] suggests variants of latin hypercube sampling (LHS) [71] to obtain training data for the SVM that are adequately spread throughout the bounded range of Z. We therefore assess the performance of using an SVM trained on latin hypercube samples to classify whether or not a sample of Z corresponds to a rare event. Two synthetic examples are examined below in which the failure region is comprised of 2 disjoint sets to mimic the setting in Example 23. In these examples, latin hypercube sampling can be applied directly because the range of Z is a hypercube, unlike that in Example 23.

Example 25. Suppose that $Z = (Z_1, \ldots, Z_{10})$ where $Z_i \sim U(0,1)$ are iid random variables. Let $F = \{z \in [0,1]^{10} \,|\, ||z|| \leq 0.94\} \cup \{z \in [0,1]^{10} \,|\, ||z-1|| \leq 0.94\}$ be the failure region such that $Y = 1$ if $Z \in F$ and $Y = -1$, otherwise. For a fixed computational

budget of 10000 samples for the training data, we investigate the performance of an SVM classifier trained on latin hypercube samples in predicting if $z \in F$ or $z \notin F$ for any sample z of Z.

The probability of failure for this example is $P(Z \in F) = \frac{1}{2^9}\frac{\pi^5(0.94)^10}{\Gamma(6)} \approx 0.00268$ calculated using the formula for the volume of a ball in n-dimensions where $\Gamma(\cdot)$ is the gamma function. This is comparable to that of Example 23. Since samples generated through LHS are not fixed, 1000 sets of 10000 latin hypercube samples were produced and we selected the set with the most number of samples contained in F which was 48 samples shared between the 2 connected components of F. The SVM was trained using the Python package `scikit-learn` [64]. We varied the regularization parameter C in (B.0.2) and the kernel parameter γ using the suggested ranges $1e^{-3} \leq C \leq 1e^{6}$ and $1e^{-6} \leq \gamma \leq 1e^{3}$ with 10 evenly spaced values for each on a logarithmic scale. For each pair of values for C and γ, the trained SVM classifier was applied to 100000 test samples of Z distinct from the training data and simulated according to the true distribution of Z to compute precision and recall metrics. These are summarized in Figure 4.12. A validation data set was not utilized since we desired to examine metrics across a full spectrum of hyperparameters. The white squares for precision in the figure signify that the value was `nan`. To understand why this occurs, let $f(z)$ be the target classifier such that $f(z) > 0$ whenever $z \in F$ and $f(z) \leq 0$ otherwise and let $f^{SVM}(z)$ be the SVM classifier which predicts that $z \in F$ if $f^{SVM}(z) > 0$. Precision in this case is expressed as $P(f(Z) > 0 \,|\, f^{SVM}(Z) > 0)$ and is undefined whenever $f^{SVM}(z) \leq 0$ almost surely, ie. the trained SVM classifier always predicts that $z \notin F$ which is due to the imbalance between the classes. From Figure 4.12, a good combination of attained metric values is 0.83 for precision and 0.82 for recall achieved at $\gamma = 1$ and $C = 10$.

This example illustrates the situation when there is a large imbalance in the number of training samples in and outside the failure region. In the next example, we demonstrate that the SVM can still underperform even if this issue does not arise.

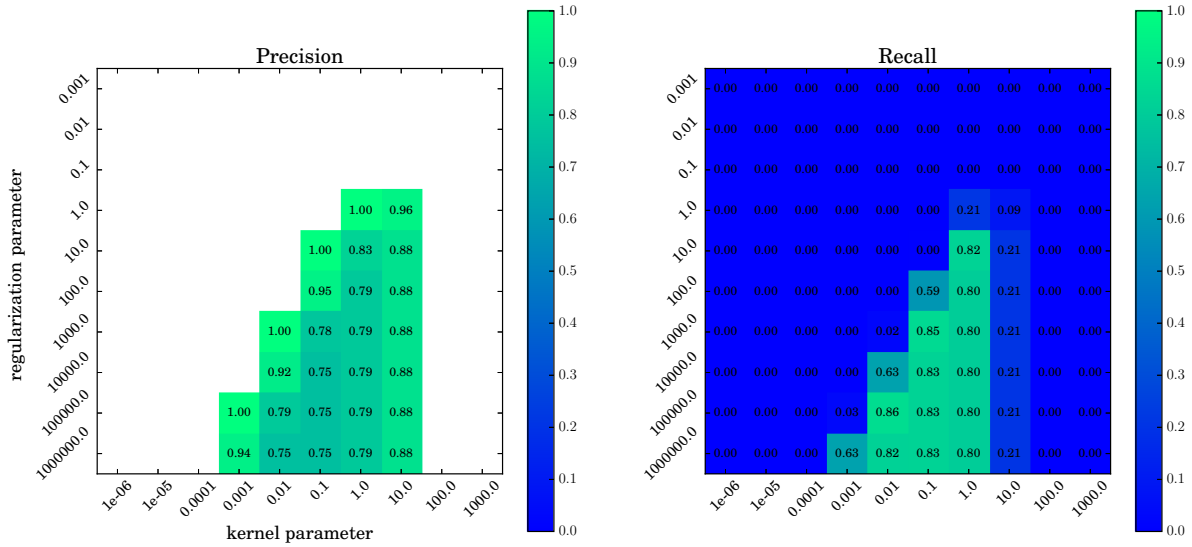


Figure 4.12: Precision (left) and recall (right) metrics as a function of the regularization C and kernel γ parameters for the SVM classifier in Example 25 based on 100000 test data.

Example 26. Suppose instead that $Z = (Z_1, \ldots, Z_{10})$ where: $Z_i = F^{-1}(\Phi(G_i))$, G_i's are zero-mean, unit-variance Gaussian random variables with $E[G_i G_j] = 0.8$ for $1 \leq i, j \leq 10, i \neq j$, $\Phi(\cdot)$ is the standard normal cdf and $F(\cdot)$ is the cdf of $Beta(10, 10)$. The failure region is set to $F = \{z \in [0,1]^{10} \,|\, \|z - (\underbrace{1, \ldots, 1}_{5}, 0, \ldots, 0)\| \leq 0.1.4515\} \cup \{z \in [0,1]^{10} \,|\, \|z - (0, \ldots, 0, \underbrace{1, \ldots, 1}_{5})\| \leq 1.4515\}$ with $Y = 1$ if $Z \in F$ and $Y = -1$, otherwise. As before, the objective is to train an SVM classifier based on 10000 latin hypercube samples to predict if $z \in F$ or $z \notin F$ for any sample z of Z.

The probability of failure is 0.0026, estimated using 100000 Monte Carlo samples of Z. Only 1 set of 10000 LHS was generated since 1886 of these were con-

tained in F, a substantial increase from that of Example 25. The reason for this is that in high dimensions, there is usually an inverse relationship between the Lebesgue and probability measure: sets of low probability occupy a large volume and vice versa. Figure 4.13 compiles the precision and recall rates for various combinations of C and γ computed using a test set of 100000 samples of Z simulated according to the true distribution of Z. A validation set was not utilized due to similar reasons stated above. A good balance between the two metrics is 0.66 for precision and 0.63 for recall attained at $C = 1, \gamma = 10$. The SVM classifier in this case suffers from low precision rates. Since most of the samples in the test set are from the high probability region of Z and outside F, they are concentrated in a small volume of the domain wherein the classifier is incorrectly predicting that they belong to F.

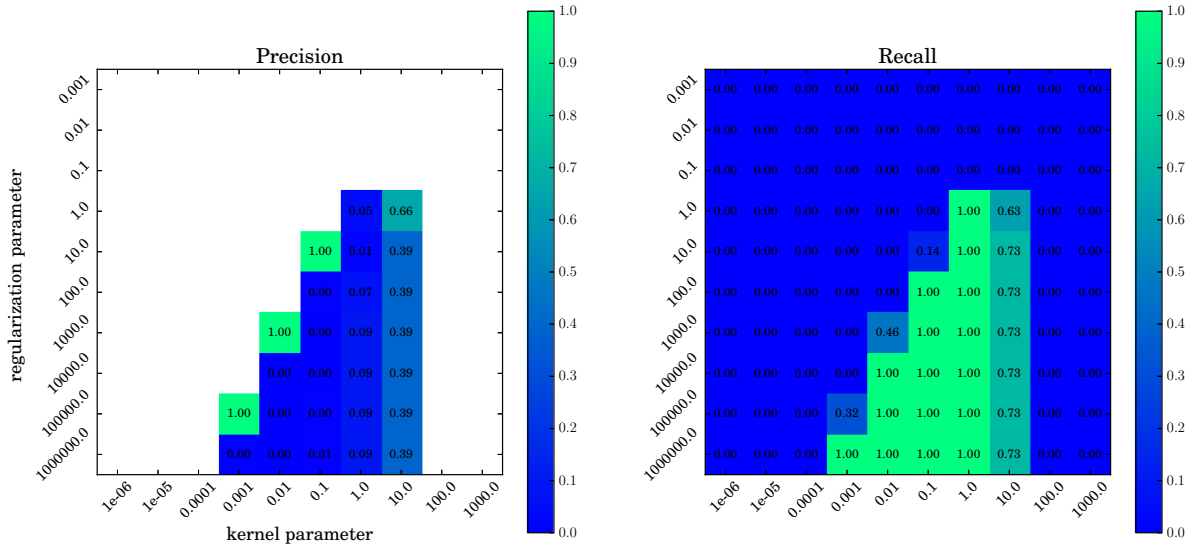


Figure 4.13: Precision (left) and recall (right) metrics as a function of the regularization C and kernel γ parameters for the SVM classifier in Example 26 based on 100000 test data.

Examples 25 and 26 illustrate that SVMs trained on latin hypercube samples alone may not yield sufficiently high rates of precision and recall for rare

event applications, unless the number of training samples is increased. This is regardless of the number of training samples present in the failure region. Furthermore, this sampling approach is only directly valid in hypercube domains. The objective is now to search for an efficient method to obtain data to train machine learning classifiers. For scenarios in which the data arises from computationally expensive forward models, we accomplish this through a multifidelity surrogate model to produce training data at a low computational cost.

4.4.2 Combining multifidelity surrogates and SVMs

We now introduce the two-stage approach to identify samples of $A(x, Z)$ which yield large QoI. In the first stage, samples of $A(x, Z)$ and their approximate QoI $Q(Z)$ are efficiently generated through a multifidelity surrogate model defined in Algorithm 4. These samples then serve as data to train an SVM that classifies whether $Q > \tau$ in the second stage. This framework enables one to aquire enough samples of $A(x, Z) \,|\, Q(Z) > \tau$ to study its conditional distribution and also offers a scheme for prediction that eliminates the need to evaluate the full model. In view of the examples in Section 4.4.1, the advantage of simulating the training data in this manner is that a sufficiently large number of samples in the failure region can be collected within a specified computational budget while training samples outside the failure region are available at negligible cost. Algorithm 5 details the proposed procedure for generating training data through a multifidelity surrogate. It is assumed that a surrogate approximation $\widetilde{Q}(Z)$ to the QoI $Q(Z)$ has already been constructed. Following the notation in Section 4.4.1, $Y = 1$ if $Z \in F$, the failure region, and $Y = -1$ otherwise.

Algorithm 5 Generating training data using a multifidelity surrogate model

1: Let η_1, η_{-1} be the minimum number of samples needed for (Z, Y) s.t. $Y = 1$ and $Y = -1$, respectively.
2: Set $D = \{\}$ be the set of samples generated.
3: Denote by $i_{\eta_1}, i_{\eta_{-1}}$ the current number of samples in D corresponding to $Y = 1$ and $Y = -1$.
4: **while** $i_{\eta_1} < \eta_1$ **do**
5: Generate a sample z of Z and evaluate $\widetilde{Q}(z)$.
6: **if** $\widetilde{Q}(z) > \tau$ **then**
7: Evaluate $Q(z)$.
8: **if** $Q(z) > \tau$ **then**
9: $D \leftarrow D \cup \{(z, 1)\}$
10: $i_{\eta_1} = i_{\eta_1} + 1$
11: **else**
12: $D \leftarrow D \cup \{(z, -1)\}$
13: $i_{\eta_{-1}} = i_{\eta_{-1}} + 1$
14: **end if**
15: **end if**
16: **end while**
17: **while** $i_{\eta_{-1}} < \eta_{-1}$ **do**
18: Generate a sample z of Z and evaluate $\widetilde{Q}(z)$.
19: **if** $\widetilde{Q}(z) \leq \tau$ **then**
20: $D \leftarrow D \cup \{(z, -1)\}$
21: $i_{\eta_{-1}} = i_{\eta_{-1}} + 1$
22: **end if**
23: **end while**

Succinctly, Algorithm 5 is just an application of Algorithm 4 in which we prioritize generating samples in the failure region until the target number of failure samples is reached because this requires evaluating the full model. A large number of samples outside the failure region can then be produced as desired due to the negligible cost of evaluating $\widetilde{Q}(Z)$ (lines 18-22 of Algorithm 5). Simulating training samples in this manner implies that there are noisy samples since the labels of some samples are flipped from +1 to −1. These false negatives are due to the fact that the full model is not evaluated whenever $\widetilde{Q}(z) \leq \tau$. However, since $P(\widetilde{Q}(Z) \leq \tau, Q(Z) > \tau) \leq P(Q(Z) > \tau)$, we expect the number of samples

with flipped labels to be insignificant especially if the failure probability is very low in magnitude.

The work [33] surveys existing classification schemes that deal with flipped labels in the training data. However, it is usually assumed that samples with flipped labels are independent from each other which is not applicable in our setting. An example of such work for SVMs includes [72] which resolves the noise in the training data by solving a non-convex optimization problem. We prefer to adhere to the standard SVM classifier in Appendix B and argue that the regularization parameter is sufficient in handling the small number of samples with flipped labels as in our case. We now apply the two-stage approach to the following example to supplement the performance of the multifidelity surrogate in Section 4.3.2.

Example 27. We build on the setup in Examples 23 and 24 in which the multifidelity surrogate according to Algorithm 4 was constructed using the full model $Q_{max}(Z)$ and the surrogate model $\widetilde{Q}_{max}^{adapt}(Z)$ following the adaptive construction for the SROM-based surrogate. Samples of (Z, Y) are then generated according to Algorithm 5 to train an SVM classifier $\widetilde{Q}_{max}^{SVM}(Z)$ with $\widetilde{Q}_{max}^{SVM}(z) = 1$ if $z \in F$ and $\widetilde{Q}_{max}^{SVM}(z) = -1$ otherwise. We examine the performance of $\widetilde{Q}_{max}^{SVM}(Z)$ in predicting samples of Z (and consequently $A(x, Z)$) that yield large $Q_{max}(Z)$.

Regardless of how the training data are generated, an imbalance in the number of samples present in and out of the failure region will always exist since the failure domain has low probability. To address this, a common strategy is to oversample the minority class or undersample the majority class [54]. In our experiments, we simulated 5021 samples in the failure region for the training data. Following Algorithm 5, this resulted in 6265 evaluations of the full model

$Q_{max}(Z)$ of which 1244 corresponded to $Q_{max}(Z) \leq \tau$. In addition to these 1244 samples, 68756 samples outside the failure region were produced at negligible cost using $\widetilde{Q}_{max}^{adapt}$ (lines 18-22 of Algorithm 5) so that a total of 70000 samples reside outside F. Recall that an insignificant proportion of the 70000 samples have flipped labels. We therefore have that the proportion of samples in the training data contained in the failure region is $\frac{5021}{70000+5021} \approx 0.0669$ which is substantially larger than $P(Q_{max}(Z) > \tau) \approx 0.00234$. Notice that our oversampling scheme does not entail producing duplicates of existing samples in the training data pertaining to the minority class.

To tune the hyperparameters C and γ for the SVM, we did not pursue k-fold cross validation since a few samples of the training data have flipped labels and more importantly, the distribution of our training set do not reflect the true distribution of Z. Note that the latter reason is also applicable for SVMs trained on latin hypercube samples. Instead, we generated a validation set comprised of 50000 samples $\{z_k^{valid}, y_k^{valid}\}_{k=1}^{50000}$ of (Z, Y) that requires full model evaluations, is simulated according to the true distribution of (Z, Y), and is not contaminated with noise. Our specific validation set includes 112 samples in F. The hyperparameters chosen are the ones that yield a good combination of precision and recall rates when the trained SVM is applied to the validation data. Figure 4.14 summarizes these metrics for certain values of C, γ. It is seen that $C = 7500$ and $\gamma = 0.0005$ offer a good combination with 0.94 for precision and 0.88 for recall and are thus the hyperparameter values we select.

To report the performance of the classifier, the trained SVM with the chosen hyperparameters is applied to test data. This consists of the same set of 10000 and 50000 samples of Z (and the corresponding evaluations of the full model

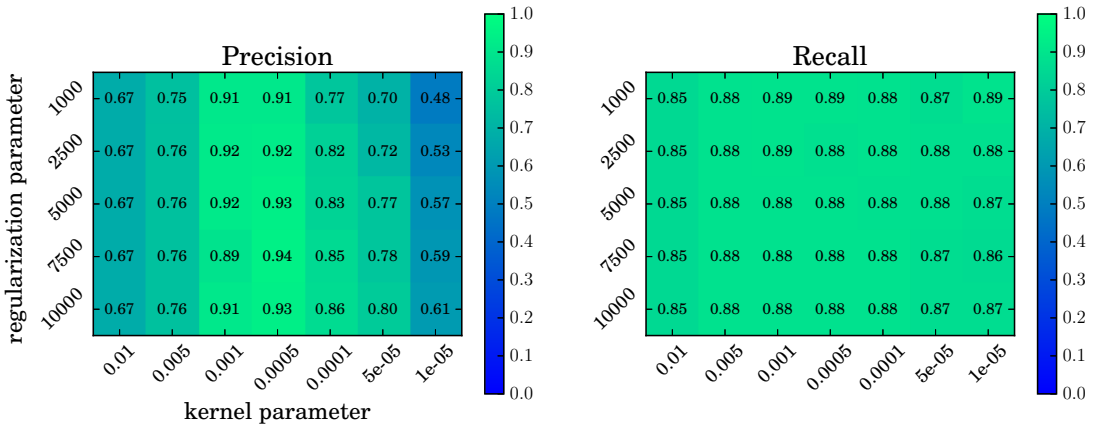


Figure 4.14: Precision (left) and recall (right) metrics of the trained SVM classifier based on the validation data for certain values of C and γ.

$Q_{max}(Z)$) invoked to construct the confusion matrices in Examples 23 and 24. The samples in the test set are distinct from that of the validation set and are also distributed according to the true law of (Z, Y) with no samples possessing flipped labels unlike in the training data.

		$Q_{max}(Z)$		
		Positive	Negative	Total
$\widetilde{Q}^{SVM}_{max}(Z)$	Positive	17	3	20
	Negative	1	9979	9980
	Total	18	9982	10000

Table 4.8: Confusion matrix for $\widetilde{Q}^{SVM}_{max}(Z)$ based on 10000 samples of Z.

		$Q_{max}(Z)$		
		Positive	Negative	Total
$\widetilde{Q}^{SVM}_{max}(Z)$	Positive	118	13	131
	Negative	8	49861	49869
	Total	126	49874	50000

Table 4.9: Confusion matrix for $\widetilde{Q}^{SVM}_{max}(Z)$
based on 50000 samples of Z.

Tables 4.8 and 4.9 summarize the confusion matrices of the SVM classifier based on 10000 and 50000 test samples, respectively. From Table 4.8, we deduce 0.85 for precision, 0.9444 for recall, and 0.1501 for the failure rate while from Table 4.9, we obtain 0.9008 for precision, 0.9365 for recall, and 0.0994 for the failure rate. Observe that the balance in these metrics is a significant improvement compared to those attained in Examples 23 and 24 and that the failure rate alone as introduced in [27] is insufficient and can be misleading as a metric since it does not incorporate recall. Only a total of 10305 evaluations of the full model to gather the training data were required, 4040 of which are due to constructing the surrogate model $\widetilde{Q}_{max}$ in Section 4.3.

We now investigate the effect of the samples with flipped labels present in our training data. Figure 4.15 displays a boxplot of the false negatives, samples whose labels were flipped from +1 to −1 due to the multifidelity surrogate in Algorithm 5. The left panel plots their values according to $\widetilde{Q}^{adapt}_{max}(Z)$ while the right panel plots the $Q_{max}(Z)$ values. These only accounted for 61 out of the 70000 training samples outside the failure region. Notice that these false negatives are mostly concentrated close to the boundary of the failure region as the Q_{max} values on the right panel indicate. This underscores that confusion matrices

alone do not offer the full perspective on a classifier: it is possible that for a sample z, $\widetilde{Q}_{max}^{adapt}(z) \approx Q_{max}(z)$ yet $\widetilde{Q}_{max}^{adapt}(z) > \tau$ and $Q_{max}(z) \leq \tau$ or vice versa. From Appendix B, it is therefore not surprising that the value of the regularization parameter C is relatively large which enables these false negative samples (-1 label) close to the true boundary of the failure region to be located on the side of the boundary of the SVM classifier pertaining to the $+1$ class. Indeed, let $f(z)$ be the separating manifold in Appendix B whose sign indicates whether $\widetilde{Q}_{max}^{SVM}$ is $+1$ or -1. Evaluating $f(z)$ associated to the trained SVM classifier at these 61 false negative samples resulted in 32 of these with $f(z) > 0$, i.e. $\widetilde{Q}_{max}^{SVM}(z) = 1$.

Finally, we remark that metrics we obtained in Table 4.8 and 4.9 are not necessarily the most optimal that could possibly be found. This is because many parameters in Example 27 could be adjusted such as the ratio of the samples in the failure region to the samples outside it in the training data. The former can be increased at additional computational expense while the latter can be increased or decreased at negligible cost since they are produced using the surrogate model.

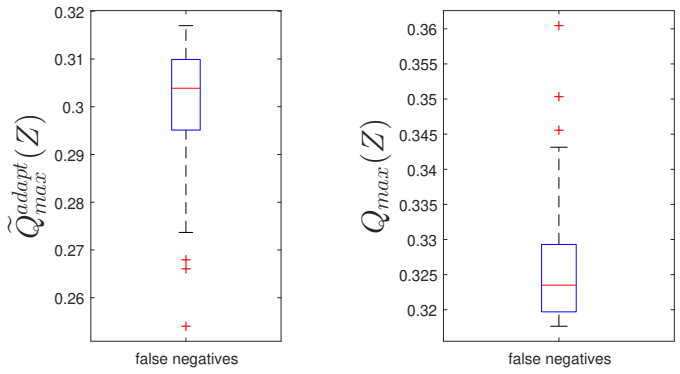


Figure 4.15: Box plot of the false negatives in the training data in Example 27 for their corresponding $\widetilde{Q}_{max}^{adapt}(Z)$ (left panel) and $Q_{max}(Z)$ (right panel) values.

We conclude this section by underscoring the disparity in the performance of the SVM classifier if the training data were generated using standard Monte

Carlo simulation as Example 28 demonstrates.

Example 28. We revisit Example 27 except that instead of generating training data according to the proposed scheme in Algorithm 5, they are produced simply through Monte Carlo simulation based on the true law of Z. The effect of the scarcity of training samples in the failure region on the SVM classifier is examined.

The training data for the SVM was gathered from 50000 samples of Z and $Q_{max}(Z)$ of which 115 resided in the failure region. Only 1 set of Monte Carlo samples was generated for the purposes of illustration. The hyperparameter values were varied in the ranges $1e^{-3} \leq C \leq 1e^{6}$ and $1e^{-6} \leq \gamma \leq 1e^{3}$ with 10 evenly spaced values for each on a logarithmic scale. For each combination of C and γ values, the SVM was applied to the same test set of 50000 samples of Z utilized in Examples 23, 24, and 27. We did not make use of a validation set in order to inspect the metrics for a spectrum of hyperparameter values. The precision and recall rates are presented in Figure 4.16 with the white squares indicating `nan` values. A good balance between the two rates is attained at $C = 100, \gamma = 0.001$ with 0.86 for precision and 0.71 for recall. The fact that the classifier achieves high rates of precision but low rates of recall for various combinations of hyperparameter values signifies that it is unable to capture a large scope of the failure region: it is likely that the high probability regions of the failure domain identified by the SVM is mostly contained in the true failure domain but that the high probability region of the true failure domain is largely absent in the failure domain identified by the SVM.

The above examples serves to highlight the synergy that can be obtained by combining a multifidelity surrogate to generate training data efficiently and a

machine learning classifier to identify samples of $A(x, \omega)$ that yield large QoI. Implicit in them is the availability of validation data that have no flipped labels and that are distributed according to its true distribution to tune the SVM hyperparameters. Obtaining such data efficiently will be considered in future work.

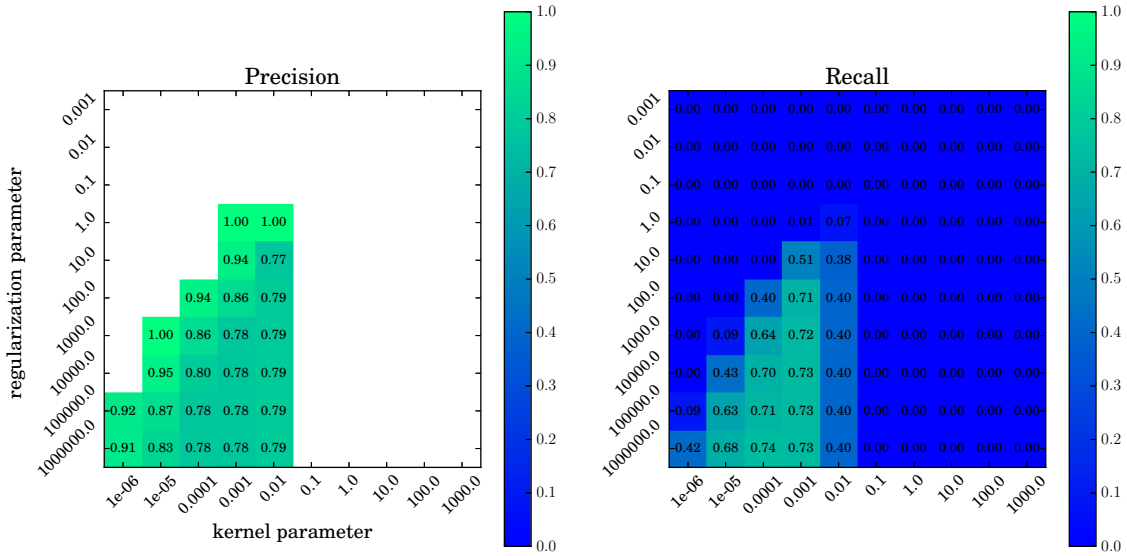


Figure 4.16: Precision (left) and recall (right) as a function of the regularization C and kernel γ parameters for the SVM classifier in Example 28.

4.5 Conclusion

In this work, we studied the problem of identifying samples of the input random field which yield large quantities of interest. This presents a shift from the traditional objectives in reliability engineering which is mostly concerned with computing probabilities of failure. While such quantities are useful in applications, they serve as global measures in that they do not offer clues as to what

types of inputs cause failure nor how to design inputs that do not lead to such events.

Several classification schemes were investigated to achieve our objectives. These included physics-based indicators which are quantities computed from the input random field that signal the occurrence of extreme events. However, in practice, indicators may be difficult to derive unless a strong physical intuition about the system is available. In addition, a large number of samples of the rare event and hence full model evaluations is needed to verify and calibrate the indicator. In search of an alternative, we examined a multifidelity approximation to the quantity of interest which is comprised of the surrogate model and infrequent evaluations of the full model. Functioning as a classifier that determines if an input random field sample leads to large QoI, the multifidelity surrogate in our specific example achieved high rates of precision yet low rates of recall. Since recurring to infrequent full model evaluations may still be undesirable, we proposed a two-stage approach to identify the samples of interest where in the first step, training data are efficiently gathered using the multifidelity approximation and a machine learning classifier is trained on these samples in the second step. We demonstrated the synergy of these 2 methods in obtaining high rates of precision and recall compared to other options for producing training data such as Monte Carlo simulation and latin hypercube sampling.

APPENDIX A

COMPUTING THE PDF ON 1-DIMENSIONAL CONTOURS

Fix $x_{\mathcal{L}}$ and consider the contour $\pi^{-1}(x_{\mathcal{L}})$. Let x_C parameterize the arc length of $\pi^{-1}(x_{\mathcal{L}})$ and denote by $\mu(\pi^{-1}(x_{\mathcal{L}}))$ the arc length of $\pi^{-1}(x_{\mathcal{L}})$ so that $0 \leq x_C \leq \mu(\pi^{-1}(x_{\mathcal{L}}))$. The conditional pdf $f_{X_C|X_{\mathcal{L}}}(x_C|x_{\mathcal{L}})$ along the contour can be approximated as follows:

- Select equidistant points $\{x_C^{(i)}\}_{i=0}^N \subset [0, \mu(\pi^{-1}(x_{\mathcal{L}}))]$ such that $x_C^{(0)} = 0, x_C^{(N)} = \mu(\pi^{-1}(x_{\mathcal{L}}))$ and $x_C^{(i+1)} - x_C^{(i)} = \frac{\mu(\pi^{-1}(x_{\mathcal{L}}))}{N}$ for $i = 0, \ldots, N-1$.
- From $P(X_C \in (x_C^{(i)}, x_C^{(i+1)}) \,|\, X_{\mathcal{L}} = x_{\mathcal{L}}) = \int_{x_C^{(i)}}^{x_C^{(i+1)}} f_{X_C \,|\, X_{\mathcal{L}}}(x_C|x_{\mathcal{L}})\, dx_C$, we deduce the approximation

$$f_{X_C|X_{\mathcal{L}}}(x_C|x_{\mathcal{L}}) \simeq \frac{P(X_C \in (x_C^{(i)}, x_C^{(i+1)}) \,|\, X_{\mathcal{L}} = x_{\mathcal{L}})}{x_C^{(i+1)} - x_C^{(i)}} \tag{A.0.1}$$

  for $x_C \in (x_C^{(i)}, x_C^{(i+1)})$ assuming that N is sufficiently large.

- To approximate the numerator in (A.0.1), we construct an infinitesimal region R in $\Gamma = [0,1]^2$ bounded by the contours $\pi^{-1}(x_{\mathcal{L}})$ and $\pi^{-1}(x_{\mathcal{L}} + \epsilon)$ for ϵ sufficiently small and partition R into regions $\{R_i\}_{i=1}^N \subset \Gamma$ with R_i corresponding to the arc lengths between $x_C^{(i-1)}$ and $x_C^{(i)}$ for $i = 1, \ldots, N$. It then follows that

$$P(X_C \in (x_C^{(i)}, x_C^{(i+1)}) \,|\, X_{\mathcal{L}} = x_{\mathcal{L}}) \simeq \iint_{R_{i+1}} f_Z(z_1, z_2)\; dz_1 dz_2 \tag{A.0.2}$$

  where f_Z is the joint pdf of Z_1, Z_2.

APPENDIX B

REVIEW OF SUPPORT VECTOR MACHINES

In this section, the important concepts of support vector machines (SVM) as referred to in Section 4.4 are reviewed. Additional information such as theoretical aspects about errors can be consulted in [20,42,58]. Let $\{(z_i, y_i)\}_{i=1}^M$ be samples of the random vector (Z, Y) which is distributed according to $(Z, Y) \sim \mathcal{D}$ with $Z \in \mathbb{R}^d$ and $Y = \pm 1$. SVMs attempt to discover a function $f(z)$ that separates $\{z_i\}_{i=1}^M$ into 2 classes according to the value of y_i. Once found, it can be used to predict to which class an arbitrary sample z of Z belongs.

Suppose that the data is linearly separable, i.e. $\{z_i\}_{i=1}^M$ can be partitioned into two disjoint sets by a $(d-1)$-dimensional hyperplane $f(z) = \beta_0 + \beta^T z$ where $\beta_0 \in \mathbb{R}$, $\beta, z \in \mathbb{R}^d$. If the constants β_0, β are known, SVMs classify data by evaluating $f(z)$ and assigning z to the positive class ($y = 1$) if $f(z) > 0$ or to the negative class ($y = -1$) otherwise. These constants are chosen to maximize the Euclidean distance between the hyperplane $f(z)$ and the sample z_i closest to it. This requirement can be recast into a convex optimization problem given by

$$\begin{aligned} \underset{\beta_0, \beta}{\text{minimize}} \quad & \frac{1}{2}\|\beta\|^2 \\ \text{subject to} \quad & y_i(\beta^T z_i + \beta_0) \geq 1, \ i = 1, \ldots, M. \end{aligned} \tag{B.0.1}$$

where the constraints signify that 1) the separating hyperplane $f(z)$ correctly segregates the data into 2 classes and that 2) each sample z_i is at least a distance of $\frac{1}{\|\beta\|}$ from the resulting $f(z)$. However, if the data is not linearly separable, i.e. no such hyperplane exists, (B.0.1) can be modified by introducing slack variables $\{\epsilon_i\}_{i=1}^M$ so that $f(z)$ only correctly classifies most of the samples. The

modified optimization problem now reads as

$$\begin{aligned} \underset{\beta_0,\beta}{\text{minimize}} \quad & \frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{M}\epsilon_i \\ \text{subject to} \quad & y_i(\beta^T z_i + \beta_0) \geq 1 - \epsilon_i,\ i = 1,\ldots,M, \\ & \epsilon_i \geq 0,\ i = 1,\ldots,M, \end{aligned} \tag{B.0.2}$$

where C is a regularization parameter that must be specified. Depending on the value of ϵ_i, the updated constraints permit the hyperplane to misclassify some samples ($\epsilon_i > 1$) or allow samples to be very near the hyperplane, i.e. a distance less than $\frac{1}{\|\beta\|}$, despite being classified correctly ($\epsilon_i \leq 1$). The parameter C controls the number of such samples: in general, large C means a smaller penalty term $\sum_{i=1}^{M}\epsilon_i$ or smaller values for ϵ_i, $i = 1,\ldots,M$ so that misclassified samples in $\{(z_i, y_i)\}_{i=1}^{M}$ mostly occur near the separating hyperplane $f(z)$ and vice versa [42, p. 418, Figure 12.1].

In practice, it may be more convenient to solve the dual problem of (B.0.2) which lends additional geometric interpretation. The dual problem is expressed as

$$\begin{aligned} \underset{\alpha_1,\ldots,\alpha_M\in\mathbb{R}}{\text{maximize}} \quad & \sum_{i=1}^{M}\alpha_i - \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} y_i y_j \alpha_i \alpha_j z_i^T z_j \\ \text{subject to} \quad & 0 \leq \alpha_i \leq C,\ i = 1,\ldots,M, \\ & \sum_{i=1}^{M}\alpha_i y_i = 0. \end{aligned} \tag{B.0.3}$$

From the Karush-Kuhn-Tucker conditions, it can be concluded that $\beta = \sum_{i=1}^{M}\alpha_i y_i z_i$ so that

$$f(z) = \beta_0 + \beta^T z = \beta_0 + \sum_{i=1}^{M}\alpha_i y_i z_i^T z. \tag{B.0.4}$$

Furthermore, if $\alpha_i = 0$, $y_i f(z_i) \geq 1$, whereas if $\alpha_i = C$, $y_i f(z_i) \leq 1$, while $0 < \alpha_i < C$ implies that $y_i f(z_i) = 1$. These conditions demonstrate that $f(z)$ only depends on

the samples z_i for which $\alpha_i \neq 0$ or whose distance from $f(z)$ is at most $\frac{1}{\|\beta\|}$, i.e. the support vectors, and that it is not altered by the remaining samples in the data.

In more complex classification problems, the data may often be separated more adequately by a nonlinear manifold. SVMs handle this by mapping the input z into a feature space $\psi(z) = (\psi_1(z), \ldots, \psi_D(z)) \in \mathbb{R}^D$ in which the data is now linearly separable. It turns out that the mapping $\psi(z)$ need not be known explicitly; it suffices to define a kernel $K(x, z)$ where $\psi(z)$ must satisfy $\psi(x)^T\psi(z) = K(x, z)$. The above discussion on separating data using hyperplanes still holds except that (B.0.4) becomes $f(z) = \beta_0 + \sum_{i=1}^{M} \alpha_i y_i K(z_i, z)$ while the objective function in (B.0.3) is rewritten into $\sum_{i=1}^{M} \alpha_i - \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} y_i y_j \alpha_i \alpha_j K(z_i, z_j)$. A commonly used kernel is the Gaussian kernel $K(x, z) = \exp\left(-\frac{\|x-z\|^2}{2\gamma^2}\right)$ for specified $\gamma > 0$.

In summary, constructing an SVM entails solving a convex optimization problem to obtain the parameters β, β_0 assuming that the hyperparameter values such as C and the kernel parameter γ are fixed. A standard procedure is to split the data into 3 disjoint sets: the training data from which β, β_0 are estimated for fixed hyperparameter values, the validation data for tuning the hyperparameters, and the test data for reporting the performance of the SVM beyond the training data. An alternative is to combine the training and validation data and employ cross-validation to identify the parameters and hyperparameters.

BIBLIOGRAPHY

[1] Nitin Agarwal and N.R. Aluru. Weighted smolyak algorithm for solution of stochastic differential equations on non-uniform probability measures. *Int. J. Numer. Meth. Engng.*, 85:1365–1389, 2011.

[2] Umberto Alibrandi, Amir M. Alani, and Giuseppe Ricciardi. A new sampling strategy for SVM-based response surface for structural reliability analysis. *Probabilistic Engineering Mechanics*, 41:1–12, 2015.

[3] Tom M. Apostol. *Mathematical analysis*. Addison-Wesley Publishing Co., Reading, Mass., 2nd edition, 1974.

[4] I. Babuška, R. Tempone, and G.E. Zouraris. Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation. *Comput. Methods in Appl. Mech. Eng.*, 195:1251–1294, 2005.

[5] Anirban Basudhar, Samy Missoum, and Antonio Harrison Sanchez. Limit state function identification using support vector machines for discontinuous responses and disjoint failure domains. *Probabilistic Engineering Mechanics*, 23(1):1–11, 2008.

[6] Jean-Paul Berrut and Lloyd N. Trefethen. Barycentric lagrange interpolation. *SIAM Rev.*, 46(3):501–517, 2004.

[7] Martin Bertram, Shirley E. Konkle, Hans Hagen, Bernd Hamann, and Kenneth I. Joy. Terrain modeling using voronoi hierarchies. In G. Farin, H. Hagen, , and B. Hamann, editors, *Hierarchical and Geometrical Methods in Scientific Visualization*, pages 89–97. Springer, 2003.

[8] Tom Bobach, Gerald Farin, Dianne Hansford, and Georg Umlauf. Natural neighbor extrapolation using ghost points. *Computer-Aided Design*, 41:350–365, 2009.

[9] Jean-Daniel Boissonnat and Julia Flötotto. A coordinate system associated with points scattered on a surface. *Computer-Aided Design*, 36:161–174, 2004.

[10] J. Borggaard and H. van Wyk. Gradient-based estimation of uncertain parameters for elliptic partial differential equations. *Inverse Problems*, 31, 2015.

[11] Brad L. Boyce, Bradley C. Salzbrenner, Jeffrey M. Rodelas, Laura P. Swiler, Jonathan D. Madison, Bradley H. Jared, and Yu-Lin Shen. Extreme-value statistics reveal rare failure-critical defects in additive manufacturing. *Advanced Engineering Materials*, 19(8):1700102, 2017.

[12] J. Breidt, T. Butler, and D. Estep. A measure-theoretic computational method for inverse sensitivity problems i: method and analysis. *SIAM J. Numerical Analysis*, 49:1836–1859, 2011.

[13] Ph. Bressolette, M. Fogli, and C. Chauvière. A stochastic collocation method for large classes of mechanical problems with uncertain parameters. *Probabilist. Eng. Mech.*, 25:255–270, 2010.

[14] T. Butler, D. Estep, S. Tavener, C. Dawson, and J. J. Westerink. A measure-theoretic computational method for inverse sensitivity problems iii: multiple quantities of interest. *SIAM/ASA J. Uncertainty Quantification*, 2:174–202, 2014.

[15] T. Butler, L. Graham, D. Estep, C. Dawson, and J. J. Westerink. Definition and solution of a stochastic inverse problem for the manning's *n* parameter field in hydrodynamic models. *Advances in Water Resources*, 78:60–79, 2015.

[16] T. Butler, J. Jakeman, and T. Wildey. Combining push-forward measures and bayes' rule to construct consistent solutions to stochastic inverse problems. *SIAM J. Sci. Comput.*, 40(2):A984–A1011, 2018.

[17] Troy Butler and Timothy Wildey. Utilizing adjoint-based error estimates for surrogate models to accurately predict probabilities of events. *International Journal for Uncertainty Quantification*, 8(2):143–159, 2018.

[18] A.A. Chojaczyk, A.P. Teixeira, L.C. Neves, J.B. Cardoso, and C. Guedes Soares. Review and application of artificial neural networks models in reliability analysis of steel structures. *Structural Safety*, 52:78–89, 2015.

[19] T.M. Cover and J.A. Thomas. *Elements of information theory*. Wiley, New York, 2nd edition, 2006.

[20] Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.

[21] G. Dematteis, T. Grafke, and E. Vanden-Eijnden. Extreme event

quantification in dynamical systems with random components, 2018. arXiv:1808.10764.

[22] C. Desceliers, R. Ghanem, and C. Soize. Maximum likelihood estimation of stochastic chaos representations from experimental data. *Int. J. Numer. Meth. Engng*, 66:978–1001, 2006.

[23] R. Durrett. *Probability: theory and examples*. Cambridge University Press, New York, 4th edition, 2010.

[24] Howard C. Elman and Christopher W. Miller. Stochastic collocation with kernel density estimation. *Comput. Methods Appl. Mech. Engrg.*, 25:34–46, 2012.

[25] J.M. Emery, M.D. Grigoriu, and R.V. Field Jr. Bayesian methods for characterizing unknown parameters of material models. *Applied Mathematical Modelling*, 40(13):6395–6411, 2016.

[26] M. Farazmand and T.P. Sapsis. A variational approach to probing extreme events in turbulent dynamical systems. *Science Advances*, 3(9):e1701533, 2017.

[27] M. Farazmand and T.P. Sapsis. Extreme events: Mechanisms and prediction, 2018. arXiv:1803.06277v1.

[28] Gerald Farin. Surfaces over dirichlet tessellations. *Computer Aided Geometric Design*, 7:281–292, 1990.

[29] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

[30] R. V. Field and M. Grigoriu. Convergence properties of polynomial chaos approximations for l^2-random variables. *Sandia Report SAND2007-1262*, 2007.

[31] R.V. Field and M. Grigoriu. On the accuracy of the polynomial chaos approximation. *J. Comput. Phys.*, 209:617–642, 2005.

[32] Jasmine Foo, Xiaoliang Wan, and George Em Karniadakis. The multi-element probabilistic collocation method (me-pcm): error analysis and applications. *J. Comput. Phys.*, 227:9572–9595, 2008.

[33] Benoit Frenay and Michel Verleysen. Classification in the presence of label noise: a survey. *IEEE Transactions on Neural Networks and Learning Systems*, 25(5):845–869, 2014.

[34] J. Gerbeau, D. Lombardi, and E. Tixier. A moment-matching method to study the variability of phenomena described by partial differential equations. *SIAM J. Sci. Comput.*, 40(3):B743–B765, 2018.

[35] M. Grigoriu. *Stochastic calculus. Applications in science and engineering.* Birkhäuser, Boston, 2002.

[36] M. Grigoriu. *Stochastic systems. Uncertainty quantification and propagation*. Springer Ser. Reliab. Eng., Springer, London, 2012.

[37] M. Grigoriu. Material responses at micro- and macro-scales. *Computational Materials Science*, 107:190–203, 2015.

[38] M. Grigoriu. Parametric models for samples of random functions. *Journal of Computational Physics*, 297:47–71, 2015.

[39] M.D. Grigoriu and R.V. Field Jr. A solution to the static frame validation challenge problem using bayesian model selection. *Comput. Methods. Appl. Mech. Engrg.*, 197:2540–2549, 2008.

[40] Mircea Grigoriu. Response statistics for random heterogeneous microstructures. *SIAM/ASA J. Uncertainty Quantification*, 2:252–275, 2014.

[41] P.C. Hansen. *Discrete inverse problems: Insight and algorithms*. SIAM, Philadelphia, 2010.

[42] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer New York, 2009.

[43] J.D. Jakeman and T. Wildey. Enhancing adaptive sparse grid approximations and improving refinement strategies using adjoint-based a posteriori error estimates. *J. Comput. Phys.*, 280:54–71, 2015.

[44] R.V. Field Jr., M. Grigoriu, and J.M. Emery. On the efficacy of stochastic collocation, stochastic galerkin, and stochastic reduced order models for solving stochastic problems. *Probabilist. Eng. Mech.*, 41:60–72, 2015.

[45] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*. Springer, New York, 2005.

[46] A. Klimke. *Uncertainty modeling using fuzzy arithmetic and sparse grids*. PhD thesis, Universität Stuttgart, Shaker Verlag, Aachen, 2006.

[47] Andreas Klimke. Sparse Grid Interpolation Toolbox – user's guide. Technical Report IANS report 2007/017, University of Stuttgart, 2007.

[48] Andreas Klimke and Barbara Wohlmuth. Algorithm 847: spinterp: Piecewise multilinear hierarchical sparse grid interpolation in MATLAB. *ACM Transactions on Mathematical Software*, 31(4), 2005.

[49] Jan Kremer, Kim Steenstrup Pedersen, and Christian Igel. Active learning with support vector machines. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(4):313–326, 2014.

[50] F. Legoll, W. Minvielle, A. Obliger, and M. Simon. A parameter identification problem in stochastic homogenization. *ESAIM Proc.*, 48:190–214, 2015.

[51] Jing Li and Dongbin Xiu. Evaluation of failure probability via surrogate models. *Journal of Computational Physics*, 229(23):8966–8980, 2010.

[52] Jingchen Liu, Jianfeng Lu, and Xiang Zhou. Efficient rare event simulation for failure problems in random media. *SIAM Journal on Scientific Computing*, 37(2):A609–A624, 2015.

[53] Jingchen Liu and Xiang Zhou. Extreme analysis of a random ordinary differential equation. *Journal of Applied Probability*, 51(04):1021–1036, 2014.

[54] Victoria López, Alberto Fernández, Salvador García, Vasile Palade, and Francisco Herrera. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences*, 250:113–141, November 2013.

[55] Xiang Ma and Nicholas Zabaras. An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *J. Comput. Phys.*, 228:3084–3113, 2009.

[56] S.A. Mattis, T.D. Butler, C.N. Dawson, D. Estep, and V.V. Vesselinov. Parameter estimation and prediction for groundwater contamination based on measure theory. *Water Resour Res*, 51(9):7608–7628, 2015.

[57] Mustafa A. Mohamad and Themistoklis P. Sapsis. Sequential sampling strategy for extreme event statistics in nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 115(44):11138–11143, 2018.

[58] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning (Adaptive Computation and Machine Learning series)*. The MIT Press, 2018.

[59] Jérôme Morio, Mathieu Balesdent, Damien Jacquemart, and Christelle Vergé. A survey of rare event simulation methods for static input–output models. *Simulation Modelling Practice and Theory*, 49:287–304, 2014.

[60] V.A.B. Narayanan and N. Zabaras. Stochastic inverse heat conduction using a spectral approach. *Int. J. Numer. Meth. Engng*, 60:1–24, 2004.

[61] F. Nobile, R. Tempone, and C. G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2309–2345, 2008.

[62] J. Nolen and G. Papanicolaou. Fine scale uncertainty in parameter estimation for elliptic equations. *Inverse Problems*, 25, 2009.

[63] Qiujing Pan and Daniel Dias. An efficient reliability method combining adaptive support vector machine and monte carlo simulation. *Structural Safety*, 67:85–95, 2017.

[64] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[65] B. Piper. Properties of local coordinates based on dirichlet tessellations. *Computing Suppl.*, 8:227–239, 1993.

[66] Benjamin Richard, Christian Cremona, and Lucas Adelaide. A response surface method based on support vector machines trained with an adaptive experimental design. *Structural Safety*, 39:14–21, 2012.

[67] R. Sibson. A vector identity for the dirichlet tessellaions. *Math. Proc. Cambridge Philos. Soc.*, 87:151–155, 1980.

[68] R. Sibson. A brief description of natural neighbor interpolation. In V. Barnett, editor, *Interpreting Multivariate Data*, chapter 2, pages 21–36. John Wiley, Chichester, 1981.

[69] Hyeongjin Song, K. K. Choi, Ikjin Lee, Liang Zhao, and David Lamb. Adaptive virtual support vector machine for reliability analysis of high-dimensional problems. *Structural and Multidisciplinary Optimization*, 47(4):479–491, 2012.

[70] T.T. Soong and M. Grigoriu. *Random vibrations of structural and mechanical systems*. Prentice Hall, New Jersey, 1993.

[71] Michael Stein. Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29(2):143–151, 1987.

[72] Guillaume Stempfel and Liva Ralaivola. Learning SVMs from sloppily labeled data. In *Artificial Neural Networks – ICANN 2009*, pages 884–893. Springer Berlin Heidelberg, 2009.

[73] Wayne Isaac T. Uy and Mircea D. Grigoriu. An adaptive method for solving stochastic equations based on interpolants over voronoi cells. *Probabilistic Engineering Mechanics*, 51:23–41, 2018.

[74] Xiaoliang Wan and George Em Karniadakis. An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. *J. Comput. Phys.*, 209:617–642, 2005.

[75] Xiaoliang Wan and George Em Karniadakis. Multi-element generalized polynomial chaos for arbitrary probability measures. *SIAM J. Sci. Comput.*, 28(3):901–928, 2006.

[76] I. Weissman. Sum of squares of uniform random variables. *Statistics and Probability Letters*, 129:147–154, 2017.

[77] Jeroen A. S. Witteveen and Gianluca Iaccarino. Refinement criteria for simplex stochastic collocation with local extremum diminishing robustness. *SIAM J. Sci. Comput.*, 34(3):A1522–A1543, 2012.

[78] Jeroen A. S. Witteveen and Gianluca Iaccarino. Simplex stochastic collocation with random sampling and extrapolation for nonhypercube probability spaces. *SIAM J. Sci. Comput.*, 34(2):A814–A838, 2012.

[79] E. Wong and B. Hajek. *Stochastic processes in engineering systems*. Springer, New York, 1985.

[80] Dongbin Xiu and Jan S. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM J. Sci. Comput.*, 27(3):1118–1139, 2005.

[81] N. Zabaras and B. Ganapathysubramanian. A scalable framework for the solution of stochastic inverse problems using a sparse grid collocation approach. *Journal of Computational Physics*, 227:4697–4735, 2008.