ABSTRACT

HART, JOSEPH LEE. Extensions of Global Sensitivity Analysis: Theory, Computation, and Applications. (Under the direction of Pierre Gremaud.)

Global sensitivity analysis (GSA) seeks to quantify the relative importance of input variables in a model. The classical framework for GSA considers p variables $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ which are inputs of a function $f : \mathbb{R}^p \to \mathbb{R}$. The inputs are assumed to be statistically independent with known marginal distributions. Sensitivity indices are defined in terms of the moments or distribution of $f(\mathbf{X})$, or its derivatives, and are typically computed via sampling based approaches. However, many problems in practice do not satisfy the aforementioned assumptions. This thesis provides extensions of the classical GSA methods to address problems which fail to satisfy one or more of the assumptions in the classical framework. In particular, this thesis considers (i) problems with statistically dependent input variables, (ii) uncertainty in the distribution of $\mathbf{X}$, (iii) models with inherent stochasticity (f depends on $\mathbf{X}$ and another source of stochasticity), (iv) using GSA to analyze statistical model parameters, and (v) using GSA for uncertain parameters in optimization problems constrained by partial differential equations. Other extensions of the classical GSA framework are highlighted and avenues of future development are discussed. Theoretical properties and computational considerations are explored throughout the thesis.

Extensions of Global Sensitivity Analysis: Theory, Computation, and Applications

by
Joseph Lee Hart

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Applied Mathematics

Raleigh, North Carolina

2018

APPROVED BY:

Ralph Smith	Alen Alexanderian
Emil Constantinescu External Member	Brian Reich

Pierre Gremaud
Chair of Advisory Committee

DEDICATION

To Alyson-Grace

BIOGRAPHY

Joseph Lee Hart graduated with a Associate of Arts degree from Cape Fear Community College in July 2012 with membership in Phi Theta Kappa and Mu Alpha Sigma. He transferred to North Carolina State University where we graduated summa cum laude with a Batchelor of Sciences degree in Mathematics in May 2014.

Joseph earned a Master of Sciences degree in Applied Mathematics from North Carolina State University in May 2016 and will receive a Doctorate of Philosophy degree in Applied Mathematics with a interdisciplinary track in Statistics from North Carolina State University in December 2018. During his graduate studies Joseph became a member of Phi Beta Kappa, served various roles in the Society for Industrial and Applied Mathematics, was a graduate fellow in the SAMSI (Statistics and Applied Mathematical Sciences Institute) program on optimization, and had internships at Argonne National Laboratory and Sandia National Laboratories.

ACKNOWLEDGEMENTS

I am grateful to my advisor Pierre Gremaud for all of the time and energy he invested in me. Throughout my graduate studies, Pierre has provided excellent advise, encouragement, ideas, and opportunities to support my research and professional development.

I am grateful to my internship mentors Julie Bessac, Emil Constantinescu, and Bart van Bloemen Waanders for the opportunities and support they provided.

I am grateful to Alen Alexanderian and Ralph Smith for the multitude of helpful conversations and useful feedback they have given me.

Finally, I am grateful to the Math Department at North Carolina State University and the National Science Foundation for financial support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

INTRODUCTION

1.1 Motivation

Mathematical models and computer simulations are ubiquitous in engineering and scientific disciplines. Various systems and processes may be represented by mathematical models such as differential equations, interacting particles, or algebraic equations. These models are used to design structures, make policy decisions, and influence the development of systems in many spheres of society. However, models typically involve variables (or parameters) which are uncertain. These variables come in a variety of forms such as, but not limited to, chemical, material, or mechanical properties, electrical capacitance or resistance, or atmospheric conditions. They may appear as coefficients or boundary conditions in differential equations, parameters defining the simulation of dynamics, or algebraic expressions used to couple states in a system, to name a few. In all of these cases a user is seeking to utilize a mathematical model for scientific insight, design, or decision making, but is limited by uncertainties in the input variables of the model.

A mathematical abstraction of this is to represent uncertain quantities as random variables and the mathematical model as a function with these random variables as inputs. Global sensitivity analysis (GSA) aims to quantify the relative importance of these input variables on the output of the function. Such analysis facilitates, among other things,

- insight into the underlying system processes,

- direction for model development,
- design of experiments and data acquisition,
- engineering risk adverse systems,
- reducing the number of uncertain variables to facilitate further analysis.

In some applications, GSA may not be tractable (or necessary) so practitioners use local sensitivity analysis (LSA) instead. The idea of LSA is to measure the influence of uncertain quantities at some nominal value rather than considering them as random variables. A basic example of this is to use the gradient of a function evaluated at a nominal point to rank the importance of the variables. This thesis focuses on GSA, the reader is directed to [59], and references therein, for a broader introduction to sensitivity analysis (GSA and LSA).

In practice, there are many tools for performing GSA, each of which has its own benefits and challenges. Different goals, model properties, types of uncertain variables, and quantities of interest (QoI) mandate different approaches to GSA. This thesis considers five problems where particular challenges prompted interest in developing a GSA tool relevant to the problem. This chapter introduces two commonly used GSA tools, the Sobol' indices and derivative-based global sensitivity measures, and references other methods of interest that are not immediately relevant to this thesis. The subsequent five chapters present the author's work on five different problems; the final chapter leaves concluding remarks and highlights other problems for future work.

1.2 Review of Sobol' indices

This section reviews the Sobol' indices. The references [57, 96, 104] are noted as useful resources.

1.2.1 Mathematical formulation

Let $(\Theta, \mathcal{F}, \nu)$ be a probability space with sample space Θ, σ-algebra $\mathcal{F}$, and probability measure ν. Let $\mathbf{X} : \Theta \to \mathcal{X}$, $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_p \subseteq \mathbb{R}^p$, be a random vector with distribution function Φ, and $(\mathcal{X}, \mathcal{F}_{\mathbf{X}}, \nu_{\mathbf{X}})$ be the probability space induced by $\mathbf{X}$, i.e. $\mathcal{X}$ is the image of $\mathbf{X}$, $\mathcal{F}_{\mathbf{X}}$ is the Borel σ-algebra, and $\nu_{\mathbf{X}}$ is the measure corresponding to the law of $\mathbf{X}$. Each set $\mathcal{X}_i \subset \mathbb{R}$ corresponds to the image of the i^{th} component of the random vector $\mathbf{X}$. Let $f : \mathcal{X} \to \mathbb{R}$ be a function. The random vector $\mathbf{X}$ represents the uncertain variables in the system and f represents a scalar QoI which is determined from a model output. In this and the subsequent three chapters, $\mathbf{X}$ will be referred to as the *input variables*, the *uncertain variables*, or simply the *variables* when it is clear from the context. For application specific purposes, $\mathbf{X}$ will be referred to as the *parameters* or *uncertain parameters* in Chapters 5 and 6.

Let $u = \{i_1, i_2, \ldots, i_k\}$ be a subset of $\{1, 2, \ldots, p\}$ and $\sim u = \{1, 2, \ldots, p\} \backslash u$ be its complement. The group of variables corresponding to u is referred to as $\mathbf{X}_u = (X_{i_1}, X_{i_2}, \ldots, X_{i_k})$. Assuming $f(\mathbf{X})$ to be square integrable, consider the decomposition

$$f(\mathbf{X}) = f_0 + \sum_{k=1}^{p} \sum_{|u|=k} f_u(\mathbf{X}_u), \tag{1.1}$$

where the f_u's are defined recursively,

$$\begin{aligned} & f_0 = \mathbb{E}[f(\mathbf{X})], \\ & f_i(X_i) = \mathbb{E}[f(\mathbf{X})|X_i] - f_0 \\ & f_{i,j}(X_i, X_j) = \mathbb{E}[f(\mathbf{X})|X_i, X_j] - f_i - f_j - f_0, \\ & \vdots \\ & f_u(\mathbf{X}_u) = \mathbb{E}[f(\mathbf{X})|\mathbf{X}_u] - \sum_{v \subset u} f_v(\mathbf{X}_v) \end{aligned} \tag{1.2}$$

where the sum over $v \subset u$ is summing all subsets of u. The indexing $|u| = k$ indicates that the sum is over all $u \subseteq \{1, 2, \ldots, p\}$ of size k.

If $X_1, X_2, \ldots, X_p$ are independent random variables then the decomposition (1.1) is referred to as the ANOVA (analysis of variance) decomposition of f and

$$\mathbb{E}[f_u(\mathbf{X}) f_v(\mathbf{X})] = 0 \qquad \forall u \neq v. \tag{1.3}$$

Computing the variance of both sides of (1.1) and using (1.3) yields

$$\mathrm{Var}(f(\mathbf{X})) = \sum_{k=1}^{p} \sum_{|u|=k} \mathrm{Var}(f_u(\mathbf{X}_u)). \tag{1.4}$$

This gives the classical definition of the Sobol' indices

$$S_u = \frac{\mathrm{Var}(f_u(\mathbf{X}))}{\mathrm{Var}(f(\mathbf{X}))}. \tag{1.5}$$

The Sobol' index S_u may be interpreted as the relative contribution of $\mathbf{X}_u$ to $\mathrm{Var}(f(\mathbf{X}))$.

For $u \subset \{1, \ldots, p\}$, the *total Sobol' index* T_u is defined as the sum of all indices S_v with $v \cap u \neq \emptyset$, i.e.

$$T_u = \sum_{v \cap u \neq \emptyset} S_v, \tag{1.6}$$

where the sum over $v \cap u \neq \emptyset$ indicates the sum is over all $v \subset \{1, 2, \ldots, p\}$ whose intersection with u is nonempty. The total Sobol' index may be interpreted as the contribution of $\mathbf{X}_u$ to $\mathrm{Var}(f(\mathbf{X}))$ by itself and through interactions with other variables.

1.2.2 Properties

The following are well known properties of the Sobol' indices and total Sobol' indices:

(I) $\sum_{k=1}^{p} \sum_{|u|=k} S_u = 1,$

(II) $S_u, T_u \in [0, 1]\ \forall u,$

(III) $S_u \leq T_u\ \forall u,$

(IV) $\max_{k \in u} T_k \leq T_u \leq \sum_{k \in u} T_k.$

The first two properties support the interpretation of the Sobol' indices as relative contributions to $\mathrm{Var}(f(\mathbf{X}))$. The third property supports the interpretation of S_u as the contribution of $\mathbf{X}_u$ and T_u as the contribution of $\mathbf{X}_u$ along with its interactions. The fourth property allows inferences to be made about subsets of variables using the p total Sobol' indices $\{T_k\}_{k=1}^{p}$.

The indices S_u and T_u provide a clear way to attribute influence to subsets of variables; however, there are $2^p - 1$ possible subsets which makes computing all the indices impractical. Rather, in practice one typically computes $\{S_k, T_k\}_{k=1}^{p}$ and utilizes Property IV to make inferences about subsets. The indices $\{S_k\}_{k=1}^{p}$ are referred to as the *first order Sobol' indices*.

1.2.3 Estimation of Sobol' indices

Their mathematical properties and clear interpretation make Sobol' indices a preferred method for GSA in many applications, at least under the assumption that $X_1, X_2, \ldots, X_p$ are independent. The indices may be estimated via Monte Carlo integration. There are numerous estimators proposed for them, see [89, 96, 102, 103], and references therein, for a overview. This section will present the basic idea of estimating $\{S_k, T_k\}_{k=1}^{p}$ when $X_1, X_2, \ldots, X_p$ are independent.

First note that S_k and T_k may be written probabilistically as

$$S_k = \frac{\mathrm{Var}(\mathbb{E}[f(\mathbf{X})|X_k])}{\mathrm{Var}(f(\mathbf{X}))} \qquad \text{and} \qquad T_k = \frac{\mathbb{E}[\mathrm{Var}(f(\mathbf{X})|\mathbf{X}_{\sim k})]}{\mathrm{Var}(f(\mathbf{X}))}.$$

Using properties of conditional expectation and Fubini's theorem yields

$$\mathrm{Var}(\mathbb{E}[f(\mathbf{X})|X_k]) = \mathbb{E}[f(\mathbf{X})(f(\mathbf{X}'_k, \mathbf{X}_{\sim k}) - f(\mathbf{X}'))]$$

and

$$\mathbb{E}[\mathrm{Var}(f(\mathbf{X})|\mathbf{X}_{\sim k})] = \frac{1}{2}\mathbb{E}[(f(\mathbf{X}) - f(X_k', \mathbf{X}_{\sim k}))^2],$$

where $\mathbf{X}'$ is an independent copy of $\mathbf{X}$.

The indices $\{S_k, T_k\}_{k=1}^p$ may be estimated by Monte Carlo integration using $(p+2)N$ evaluations of f, where N is the number of Monte Carlo samples. Specifically, the samples are generated by constructing 2 matrices of size $N \times p$, call them A and B, where the rows correspond to independent samples of $\mathbf{X}$. Then p additional matrices of size $N \times p$, denote them C_k, $k = 1, 2, \ldots, p$, are generated by setting the i^{th} column of C_k equal to the i^{th} column of A, $i \neq k$, and the k^{th} column of C_k equal to the k^{th} column of B. The set of indices $\{S_k, T_k\}_{k=1}^p$ may be estimated using evaluations of f at the rows of A, B, and C_k, $k = 1, 2, \ldots, p$.

In some cases, the Sobol' indices may be estimated more efficiently using spectral approaches rather than Monte Carlo integration. The fundamental idea of spectral approaches is to represent f by a series expansion using orthogonal basis functions. Then the component functions (1.2) in (1.1) may be identified as linear combinations of the basis functions and the Sobol' indices may be computed using their coefficients. There are a variety of spectral approaches, see [96] for a overview of them, which arise from the plurality of possible basis expansions, series truncations, and coefficient estimation schemes. Spectral approaches are advantageous when f is sufficiently regular so that the coefficients decay quickly.

In many applications, evaluating f is computationally intensive and hence it may only be evaluated sparingly, for instance a few tens or a few hundred times. In such cases, the user cannot estimate Sobol' indices via Monte Carlo integration. On common approach is to approximate f with a surrogate model $\hat{f}$ and compute the Sobol' indices of $\hat{f}$. Polynomial Chaos expansions (a spectral method) are popular for this purpose [1, 6, 22, 114]. They are particularly useful because, as highlighted above, the Sobol' indices of $\hat{f}$ may be computed analytically from the coefficients since it corresponds to an approximation with orthogonal polynomials. Gaussian processes [72, 81] are also commonly used surrogate models because of their approximating properties and flexibility to represent a large class of functions. The Sobol' indices of the Gaussian process must be computed through Monte Carlo integration; however, when the number of uncertain variables is not too large (for instance, $p = 10$), this cost is small compared to the cost of evaluating f. Many other surrogate models may be used, see [113]. There are many factors to consider when choosing a surrogate model; the number of uncertain variables, the smoothness of f, the cost of evaluating f, and the expected level of interactions between variables, are a few such factors. Multivariate Adaptive Regression Splines (MARS) [28, 29] will be used in Chapter 4 because of its internal automation and scalability for large numbers of variables. In general, it is unclear how the surrogate error $f - \hat{f}$ propagates through the computation of Sobol' indices. This will

be revisited in Chapter 7 as an open question.

1.3 Review of derivative-based global sensitivity measures

A challenge frequently encountered in practice is that the number of uncertain variables p is large and this prohibits computation of Sobol' indices. In some case, the gradient of f, ∇f, may be computed independently of the parameter dimension p; a common example of this is a differential equation model with an adjoint. This motivates derivative-based global sensitivity measures. Assume that f is a differentiable function on $\mathcal{X}$ and $\frac{\partial f}{\partial x_k}(\mathbf{X})$ is square integrable, $k = 1, 2, \ldots, p$. The unnormalized derivative-based global sensitivity index [69, 107, 108] is defined as

$$D_k = \mathbb{E}\left[\left(\frac{\partial f}{\partial x_k}(\mathbf{X})\right)^2\right] \tag{1.7}$$

for $k = 1, 2, \ldots, p$.

Unlike the Sobol' indices, the derivative-based global sensitivity measures cannot be interpreted in terms of relative contributions. However, they are related to the Sobol' indices by the following result from [71].

Theorem 1. *If*

- $X_1, X_2, \ldots, X_p$ *are independent,*
- $f(\mathbf{X})$ *is square integrable,*
- $\frac{\partial f}{\partial x_k}(\mathbf{X})$ *is square integrable,*
- X_k *follows a Boltzmann probability measure on* $\mathbb{R}$*,*

then

$$T_k \leq C_k \frac{D_k}{\mathrm{Var}(f(\mathbf{X}))}, \tag{1.8}$$

where

$$C_k = \sup_{x \in \mathbb{R}} \frac{\min\{\Phi_k(x), 1 - \Phi_k(x)\}}{\phi_k(x)},$$

Φ_k *and* ϕ_k *are the cumulative distribution function (CDF) and probability density function (PDF) for* X_k*, respectively.*

Theorem 1 ensures that if the derivative-based global sensitivity measure is sufficiently small then the total Sobol' index for that variable will be small. In other words, the D_k may be used to identify unimportant variables. Note that the special case where each X_k is uniformly distributed on $[0,1]$ does not meet the assumptions of Theorem 1 because a uniform measure is not a Boltzmann measure. This special case is covered in [107] where the constant C_k in (1.8) is π^{-2}.

The shortcoming of derivative-based global sensitivity measures is that they can not reliably identify important variables or rank the order of their importance. Nonetheless, when the gradient of f may be computed independently of p, derivative-based global sensitivity measures may be estimated efficiently with Monte Carlo integration and provide a means to determine unimportant variables. It is common in practice to identify unimportant variables using D_k, $k = 1, 2, \ldots, p$, then perform subsequent analysis with a smaller subset of variables.

1.4 Overview of other GSA methods

This thesis focuses on extensions of Sobol' indices and derivative-based global sensitivity measures. A brief collection of other common GSA methods is given in this section to provide a more complete (though not exhaustive) overview of the GSA landscape.

Shapley values originated in the economics literature [105] and have gained recent interest in the GSA literature [90, 111]. Like the Sobol' indices, the Shapley values use contributions to the model output variance to measure the importance of the variables. It is shown in [90], under the assumption that $X_1, X_2, \ldots, X_p$ are independent, that the Shapley value for variable X_k, denoted by Sh_k, satisfies, $S_k \leq Sh_k \leq T_k$. In [111], the Shapley value is considered as a GSA tool when the input variables are dependent. There is still ongoing research to better understand the Shapley value and its relationship to other GSA methods [91]. One challenge is that estimating the Shapley values generally requires more computational effort than estimating the Sobol' indices or derivative-based global sensitivity indices.

Moment-independent importance measures [9] is an alternative GSA method which seeks to measure importance through changes in the distribution of $f(\mathbf{X})$ rather than its moments (Sobol' indices use the 2nd moment, variance). This is appealing in some applications where the variance of $f(\mathbf{X})$ is not sufficient to characterize uncertainty. Different approaches have been considered, a few examples are [8, 23, 93]. The basic framework shared by the various approaches is to define the importance of X_k by some distance between $f(\mathbf{X})$ and $f(\mathbf{X})|X_k$. The PDF and conditional PDF are used in [8]; the CDF and conditional CDF are used in [93].

Morris screening [86] is a classical method which evaluates f at strategically chosen points to reduce computational cost. It measures the importance of the variables with elementary effects, which may be viewed as coarse finite difference approximations of f's partial derivatives.

Estimating derivative-based global sensitivity indices with finite differences requires $(p+1)N$ evaluations of f to estimate ∇f at N samples from $\mathcal{X}$. In contrast, Morris screening uses an experimental design for sampling trajectories in $\mathcal{X}$ which enables an efficient exploration of $\mathcal{X}$ with relatively few evaluation of f. Morris screening lacks some of the mathematical properties possessed by Sobol' indices or derivative-based global sensitivity measures, but it can be useful in practice when evaluating the model is computationally intensive.

Active subspaces [19] has gained significant interest in recent years as an alternative to GSA for dimension reduction. Rather than measuring the influence of uncertain variables, as all the previous methods sought to do, active subspaces seek to determine important directions in $\mathcal{X}$ such that f varies more in those directions. It uses the gradient of f and finds important directions through the spectral decomposition of the matrix $\mathbb{E}[\nabla f(\mathbf{X})\nabla f(\mathbf{X})^T]$. This may be particularly useful when there are significantly fewer important directions than important variables, though this can only be determined after computing the active subspaces. Along with important directions, activity scores may be derived from the active subspaces [20] as a measure of the importance of each variable.

1.5 Outline of the thesis

The review of GSA in this chapter focused on problems where f is a real-valued function which depends on p uncertain variables that are assumed to be statistically independent and have known probability distributions. In many applications one or more of these assumptions are not valid. This thesis extends the classical GSA methods to address such problems. Theoretical and computational aspects of GSA are studied and applications are considered. Chapter 2 analyzes properties of the Sobol' indices when $X_1, X_2, \ldots, X_p$ possess statistical dependencies. In many applications, the distribution of the inputs and/or dependency structure is not known precisely, Chapter 3 explores the robustness of the Sobol' indices with respect to changes in the distribution of $\mathbf{X}$. Chapter 4 develops a generalization of the Sobol' indices for problems with two different sources of uncertainty, uncertain variables and inherent stochasticity. This work is motivated by an application in chemical reaction networks where the uncertain variables are reaction rates and the inherent stochasticity arises in the system dynamics. Having focused on Sobol' indices in these chapters, the thesis shifts to focus on derivative-based methods in its latter portion. Chapter 5 develops a novel approach using derivative-based GSA to analyze parameters in statistical models. A Gaussian process model for wind speed prediction motivated this research. A general class of PDE-constrained optimization problems prompt interest in Chapter 6 for a derivative-based GSA approach for the solution of a PDE-constrained optimization problem with respect to the uncertain parameters in the PDE. Chapter 7 concludes the thesis by highlighting other extensions and open questions in GSA.

1.6 Overview of the author's work

The content of this thesis corresponds to five articles which are at various stages of preparation, review, or publication. The author is primarily responsible for the contributions of these five articles. Of these five articles, only [51, 52, 53] are published at this time; they correspond to the material in Chapters 4, 5, and 2, respectively. In addition, the author has four other articles whose content is not included in this thesis. One of them, [50], has been published; the other three are in preparation or under review.

CHAPTER

2

SOBOL' INDICES WITH DEPENDENT VARIABLES

2.1 Introduction

In this chapter the Sobol' indices are analyzed for problems with dependent input variables. The content of this chapter is based on the publication [53]. Recall from Chapter 1 that the ANOVA decomposition (1.1) and the orthogonality property (1.3) are the fundamental results used to define the Sobol' indices. These results imply that the Sobol' indices satisfy Property I and Property II, which facilitate their interpretation.

When the input variables are *dependent*, the decomposition (1.1) can still be considered but the resulting orthogonality property (1.3) is not satisfied. The Sobol' indices may then again be defined from (1.1) [73]

$$S_u = \frac{\mathrm{Cov}(f_u(\mathbf{X}_u), f(\mathbf{X}))}{\mathrm{Var}(f(\mathbf{X}))}, \tag{2.1}$$

where (2.1) clearly reverts to (1.5) if the input variables are independent. With dependent variables, the indices (2.1) satisfy Property I but not Property II: while they sum up to 1, some of the S_u's may be negative and thus do not yield a quantitative measure of variable importance. A simple example from [111] is borrowed to illustrate this point.

Example 2.1.1. *Let*

$$f(X_1, X_2) = X_1 + X_2$$

where X_1, X_2 *have a joint normal distribution with* $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 0$, $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = 1$, *and* $\mathrm{Cov}(X_1, X_2) = \rho \in (0, 1)$. *Then, following* (1.1), f *admits the decomposition*

$$f(X_1, X_2) = 0 + (1 + \rho)X_1 + (1 + \rho)X_2 + (-\rho)(X_1 + X_2)$$

and the associated Sobol' indices are $S_1 = S_2 = \frac{1+\rho}{2}$, $S_{1,2} = -\rho$. *These indices sum to 1 but their interpretation as relative contributions to the variance of* $f(\mathbf{X})$ *is lost.*

The total Sobol' indices may be generalized to problems with dependent variables by putting (2.1) into (1.6); however, Property III is lost so their interpretation in terms of interaction effects becomes more challenging.

The issue of GSA with dependent variables has been the object of intense recent research. Regarding Sobol' indices, Xu and Gertner [120] propose a decomposition of the Sobol' indices into a correlated and uncorrelated part for linear models. Li et al. [73] build upon this to decompose the Sobol' indices for a general model. Mara and Tarantola [79] propose to use the Gram-Schmidt process to decorrelate the inputs variables then define new indices through the Sobol' indices of the decorrelated problem. Building off the older work of [55] and [112], Chastaing et al. [16] provide a theoretical framework to generalize the ANOVA decomposition to problems with dependent variables. In a subsequent article [17], they also provide a computational algorithm to accompany their theoretical work. In contrast to the other works which focus on generalizing the ANOVA decomposition, Kucherenko et al. [68] develop the Sobol' indices via the law of total variance. Recent work of Mara and Tarantola [115] considered estimating Sobol' indices with dependent variables using the Fourier Amplitude Sensitivity Test.

As highlighted in Chapter 1, there are various methods in GSA and many applications of them. To facilitate analysis for problems with dependent input variables, this chapter focuses on applying Sobol' indices for dimension reduction, i.e. approximating $f(\mathbf{X})$ by a function which depends on fewer variables. To this end, the total Sobol' index, with independent or dependent variables, is characterized in terms of approximation error rather than variance analysis. This approximation theoretic characterization is used to analyze the error introduced through different approaches to dimension reduction.

2.2 An approximation theoretic perspective of total Sobol' indices

As in Chapter 1, $f : \mathcal{X} \to \mathbb{R}$ is a square integrable function with mean $f_0 = \mathbb{E}[f(\mathbf{X})]$. Given $u \subset \{1, 2, \ldots, p\}$, a natural question is "how accurately can $f(\mathbf{X}) - f_0$ can be approximated (in $L^2(\mathcal{X})$) without the variables $\mathbf{X}_u$?" In other words, what is the error associated with the approximation

$$f(\mathbf{X}) - f_0 \approx \mathcal{P}_{\sim u} f(\mathbf{X}_{\sim u}), \tag{2.2}$$

where $\mathcal{P}_{\sim u} f(\mathbf{X}_{\sim u})$ is the optimal $L^2(\mathcal{X})$ approximation of $f(\mathbf{X}) - f_0$ which does not depend on $\mathbf{X}_u$? It is shown that this error,

$$\frac{\|(f(\mathbf{X}) - f_0) - \mathcal{P}_{\sim u} f(\mathbf{X}_{\sim u})\|_2^2}{\|f(\mathbf{X}) - f_0\|_2^2}, \tag{2.3}$$

coincides with the classical definition of the total Sobol' index (1.6). This requires a few technical considerations.

Definition 1. *For $f \in L^2(\mathcal{X})$, f does not depend on X_k if and only if there exists $N \in \mathcal{F}_{\mathbf{X}}$ with $\nu_{\mathbf{X}}(N) = 0$ such that*

$$f(\mathbf{x}) = f(\mathbf{y}) \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{X} \setminus N \text{ with } \mathbf{x}_{\sim k} = \mathbf{y}_{\sim k}.$$

Otherwise, f is said to depend on X_k.

Definition 2. *For any $v \subset \{1, \ldots, p\}$, define M_v as the set of all functions in $L^2(\mathcal{X})$ that do not depend on any variables in $\mathbf{X}_{\sim v}$.*

Roughly speaking, M_v is the set of those functions in $L^2(\mathcal{X})$ that only depend on $\mathbf{X}_v$. Theorem 2 gives that M_v is a closed subspace of $L^2(\mathcal{X})$.

Theorem 2. *M_u is a closed subspace of $L^2(\mathcal{X})$.*

Proof. M_u is clearly a subset of $L^2(\mathcal{X})$. To show that it is closed, let $\{f_n\}$ be a sequence in M_u which converges to $f \in L^2(\mathcal{X})$. It is enough to show that $f \in M_u$. Suppose by contradiction that $f \notin M_u$. Then $\exists i \in \{1, 2, \ldots, p\}$ such that $i \notin u$ and f depends on X_i. Then $\exists A \in \mathcal{F}_{\mathbf{X}}$ and $\mathbf{x}, \mathbf{y} \in A$ such that $\nu_{\mathbf{X}}(A) > 0$ with $\mathbf{x}_{\sim i} = \mathbf{y}_{\sim i}$ and $f(\mathbf{x}) \neq f(\mathbf{y})$. Since $f_n \to f$ in $L^2(\mathcal{X})$ then $\exists \{f_{n_k}\}$, a subsequence of $\{f_n\}$, such that $f_{n_k} \to f$ almost surely. Since $\mathbf{x}, \mathbf{y} \in A$ and $\nu_{\mathbf{X}}(A) > 0$ then $f_{n_k}(\mathbf{x}) \to f(\mathbf{x})$ and $f_{n_k}(\mathbf{y}) \to f(\mathbf{y})$. But f_{n_k} does not depend on X_i so $f_{n_k}(\mathbf{x}) = f_{n_k}(\mathbf{y})$ $\forall k \in \mathbb{N} \implies f(\mathbf{x}) = f(\mathbf{y})$. This is a contradiction so $f \in M_u$ and hence M_u is closed. □

Using classical results on orthogonal decompositions, $L^2(\mathcal{X})$ can be decomposed as a direct sum of M_v and $M_v^{\perp}$, the orthogonal complement of M_v, i.e.

$$L^2(\mathcal{X}) = M_v \oplus M_v^{\perp}. \tag{2.4}$$

It is worth noting here that $M_v^{\perp} \neq M_{\sim v}$.

Setting $v = \sim u$, rewrite (2.2) more explicitly as

$$f(\mathbf{X}) = f_0 + \mathcal{P}_{\sim u} f(\mathbf{X}_{\sim u}) + \mathcal{P}_{\sim u}^{\perp} f(\mathbf{X}), \tag{2.5}$$

where $\mathcal{P}_{\sim u} f(\mathbf{X}_{\sim u}) = \mathbb{E}[f(\mathbf{X}) - f_0 | \mathbf{X}_{\sim u}] = \mathbb{E}[f(\mathbf{X}) | \mathbf{X}_{\sim u}] - f_0$ is the projection of $f(\mathbf{X}) - f_0$ onto $M_{\sim u}$. This orthogonal decomposition yields

$$\frac{\|(f(\mathbf{X}) - f_0) - \mathcal{P}_{\sim u} f(\mathbf{X}_{\sim u})\|_2^2}{\|f(\mathbf{X}) - f_0\|_2^2} = \frac{\|\mathcal{P}_{\sim u}^{\perp} f(\mathbf{X})\|_2^2}{\|f(\mathbf{X}) - f_0\|_2^2} = 1 - \frac{\|\mathcal{P}_{\sim u} f(\mathbf{X}_{\sim u})\|_2^2}{\|f(\mathbf{X}) - f_0\|_2^2}. \tag{2.6}$$

Theorem 3 shows that the total Sobol' index T_u equals (2.3), thus providing a new characterization of the total Sobol' indices, with independent or dependent variables, and giving a clear interpretation of these indices in terms of relative approximation error.

Theorem 3. *For $u \subset \{1, 2, \ldots, p\}$,*

$$T_u = \frac{\|(f(\mathbf{X}) - f_0) - \mathcal{P}_{\sim u} f(\mathbf{X}_{\sim u})\|_2^2}{\|f(\mathbf{X}) - f_0\|_2^2}.$$

Proof. Rearranging (1.1) gives,

$$f(\mathbf{X}) - f_0 = \sum_{v \cap u = \emptyset} f_v(\mathbf{X}_v) + \sum_{v \cap u \neq \emptyset} f_v(\mathbf{X}_v).$$

Taking into account (1.6), (2.5), (2.6), and the linearity of the covariance operator, it follows that

$$\begin{aligned} T_u &= \sum_{v \cap u \neq \emptyset} \frac{\mathrm{Cov}(f_v(\mathbf{X}_v), f(\mathbf{X}))}{\mathrm{Var}(f(\mathbf{X}))} \\ &= \frac{1}{\|f(\mathbf{X}) - f_0\|_2^2} \mathrm{Cov}\left(\sum_{v \cap u \neq \emptyset} f_v(\mathbf{X}_v), \mathcal{P}_{\sim u} f(\mathbf{X}_{\sim u}) + \mathcal{P}_{\sim u}^{\perp} f(\mathbf{X}) \right) \\ &= \frac{\mathrm{Cov}(\mathcal{P}_{\sim u}^{\perp} f(\mathbf{X}), \mathcal{P}_{\sim u} f(\mathbf{X}_{\sim u}))}{\|f(\mathbf{X}) - f_0\|_2^2} + \frac{\mathrm{Cov}(\mathcal{P}_{\sim u}^{\perp} f(\mathbf{X}), \mathcal{P}_{\sim u}^{\perp} f(\mathbf{X}))}{\|f(\mathbf{X}) - f_0\|_2^2} \end{aligned}$$

$$
\begin{aligned}
&= \frac{0}{\|f(\mathbf{X}) - f_0\|_2^2} + \frac{\|\mathcal{P}_{\sim u}^{\perp} f(\mathbf{X})\|_2^2}{\|f(\mathbf{X}) - f_0\|_2^2} \\
&= \frac{\|(f(\mathbf{X}) - f_0) - \mathcal{P}_{\sim u} f(\mathbf{X}_{\sim u})\|_2^2}{\|f(\mathbf{X}) - f_0\|_2^2}
\end{aligned}
$$

□

Corollary 1 shows that the total Sobol' index defined by using (2.1) and (1.6) is equivalent to the total Sobol' index defined in [68], thus equipping it with both an approximation theoretic and probabilistic interpretation. The function decomposition (2.5) is the approximation theoretic analogue of the law of total variance approach in [68].

Corollary 1.

$$
T_u = 1 - \frac{\mathrm{Var}(\mathbb{E}[f(\mathbf{X})|\mathbf{X}_{\sim u}])}{\mathrm{Var}(f(\mathbf{X}))} = \frac{\mathbb{E}[\mathrm{Var}(f(\mathbf{X})|\mathbf{X}_{\sim u})]}{\mathrm{Var}(f(\mathbf{X}))}
$$

Proof. Recall from the proof of Proposition 3 that

$$
\mathcal{P}_{\sim u} f(\mathbf{X}_{\sim u}) = \mathbb{E}[f(\mathbf{X}) - f_0|\mathbf{x}_{\sim u}] = \sum_{v \cap u = \emptyset} f_v(\mathbf{X}_v).
$$

Using this equality along with (2.1), (1.6), and the orthogonal decomposition (2.5) gives

$$
\begin{aligned}
T_u &= \sum_{v \cap u \neq \emptyset} S_v \\
&= \sum_{v \cap u \neq \emptyset} \frac{\mathrm{Cov}(f_v(\mathbf{X}_v), f(\mathbf{X}))}{\mathrm{Var}(f(\mathbf{X}))} \\
&= \frac{1}{\mathrm{Var}(f(\mathbf{X}))} \mathrm{Cov}\left(\sum_{v \cap u \neq \emptyset} f_v(\mathbf{X}_v), f(\mathbf{X}) \right) \\
&= \frac{1}{\mathrm{Var}(f(\mathbf{X}))} \mathrm{Cov}\left(f(\mathbf{X}) - \sum_{v \cap u = \emptyset} f_v(\mathbf{X}_v) - f_0, f(\mathbf{X}) \right) \\
&= \frac{1}{\mathrm{Var}(f(\mathbf{X}))} \left(\mathrm{Cov}\left(f(\mathbf{X}), f(\mathbf{X})\right) - \mathrm{Cov}\left(\mathcal{P}_{\sim u} f(\mathbf{X}_{\sim u}), \mathcal{P}_{\sim u} f(\mathbf{X}_{\sim u}) + \mathcal{P}_{\sim u}^{\perp} f(\mathbf{X}) \right) \right) \\
&= 1 - \frac{1}{\mathrm{Var}(f(\mathbf{X}))} \mathrm{Cov}\left(\mathcal{P}_{\sim u} f(\mathbf{X}_{\sim u}), \mathcal{P}_{\sim u} f(\mathbf{X}_{\sim u})\right) \\
&= 1 - \frac{1}{\mathrm{Var}(f(\mathbf{X}))} \mathrm{Var}(\mathbb{E}[f(\mathbf{X}) - f_0|\mathbf{X}_{\sim u}]) \\
&= 1 - \frac{1}{\mathrm{Var}(f(\mathbf{X}))} \mathrm{Var}(\mathbb{E}[f(\mathbf{X})|\mathbf{X}_{\sim u}])
\end{aligned}
$$

□

Because of their approximation theoretic interpretation, this chapter focuses on the total Sobol' indices.

2.3 Applying the approximation theoretic perspective for dimension reduction

One common use of the total Sobol' indices is dimension reduction, i.e., approximating f by a function which depends on fewer variables. There are several ways to do this, three examples are:

1. projecting f onto a subspace of functions which only depend on a subset of the input variables,

2. constructing a surrogate model (from sample data) using only a subset of the variables,

3. fixing some of the input variables to nominal values, or possibly a function of the other input variables.

The approximation theoretic perspective in Section 2.2 provides useful insights for all three of these possible approaches. For the first approach, the total Sobol' index T_u is the relative $L^2(\Omega)$ error squared when f is approximated by its orthogonal $L^2(\Omega)$ projection onto the subspace of functions which only depend on $\mathbf{X}_{\sim u}$. However, acquiring this projection is computationally costly since it requires computing many high dimensional integrals, so the applicability of this approach is limited in practice. The second approach is practical in cases where the user wishes to use existing evaluations of f to train a surrogate model. The unimportant variables may be considered as latent and the surrogate model may be trained using only a subset of input variables. Since T_u is the error for the optimal $L^2(\Omega)$ approximation, it provides a lower bound on the $L^2(\Omega)$ error of a surrogate model approximation. Hence T_u is useful for making decisions about which variables to use when constructing a surrogate model in the second approach. The third approach, fixing inputs, is commonly used in practice because of its simplicity. As demonstrated below, the approximation theoretic perspective of total Sobol' indices is useful for analyzing approximation error in this setting as well.

With the assumption of independent variables, classical results exist [110] which use the total Sobol' indices to bound the error incurred when fixing variables to nominal values. Partition $\mathbf{X} = (\mathbf{X}_u, \mathbf{X}_{\sim u})$ and assume that T_u is small. The goal is to approximate f by replacing $\mathbf{X}_u$ with a function of $\mathbf{X}_{\sim u}$. Specifically, $f(\mathbf{X})$ is approximated by $f(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u})$, where $g(\mathbf{X}_{\sim u})$

is an approximation of $\mathbf{X}_u$. It is common to take the constant approximation $g(\mathbf{X}_{\sim u}) = \mathbb{E}[\mathbf{X}_u]$ when the variables are independent. The subsequent analysis considers a general g.

The relative error incurred by replacing $\mathbf{X}_u$ with $g(\mathbf{X}_{\sim u})$ is

$$\delta_u = \frac{||f(\mathbf{X}) - f(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u})||_2^2}{||f(\mathbf{X}) - f_0||_2^2}. \tag{2.7}$$

Theorem 4 extends a result in [110] to the case with dependent variables.

Theorem 4. *For any $u \subset \{1, 2, \ldots, p\}$ and any $g : \Omega_{\sim u} \to \Omega_u$ such that $f(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u}) \in L^2(\Omega)$,*

$$\delta_u \geq T_u.$$

Proof. The result follows since T_u is the squared relative $L^2(\mathcal{X})$ error of the the orthogonal projection of $f(\mathbf{X}) - f_0$ onto $M_{\sim u}$, i.e. the optimal approximation in $M_{\sim u}$, and $f(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u}) \in M_{\sim u}$. □

An upper bound on δ_u is more useful than a lower bound in most cases; however, a tight upper bound is difficult to attain. Substituting (2.5) into (2.7) yields

$$\delta_u = \frac{||\mathcal{P}_{\sim u}^{\perp} f(\mathbf{X}) - \mathcal{P}_{\sim u}^{\perp} f(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u})||_2^2}{||f(\mathbf{X}) - f_0||_2^2}. \tag{2.8}$$

Recall that the Sobol' index T_u, which is assumed to be small, is given by

$$T_u = \frac{||\mathcal{P}_{\sim u}^{\perp} f(\mathbf{X})||_2^2}{||f(\mathbf{X}) - f_0||_2^2}.$$

Hence, $\mathcal{P}_{\sim u}^{\perp} f(\mathbf{X})$ is small relative to $f(\mathbf{X}) - f_0$.

Theorem 5 provides a loose, but informative (see below), upper bound on δ_u.

Theorem 5.

$$\delta_u \leq T_u + \frac{||\mathcal{P}_{\sim u}^{\perp} f(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u})||_2^2}{||f(\mathbf{X}) - f_0||_2^2} + 2T_u \frac{||\mathcal{P}_{\sim u}^{\perp} f(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u})||_2}{||\mathcal{P}_{\sim u}^{\perp} f(\mathbf{X})||_2}$$

Proof. Notice,

$$\begin{aligned} ||\mathcal{P}_{\sim u}^{\perp} f(\mathbf{X}) - \mathcal{P}_{\sim u}^{\perp} f(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u})||_2^2 = &||\mathcal{P}_{\sim u}^{\perp} f(\mathbf{X})||_2^2 \\ &+ ||\mathcal{P}_{\sim u}^{\perp} f(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u})||_2^2 \\ &- 2\mathbb{E}[\mathcal{P}_{\sim u}^{\perp} f(\mathbf{X}) \mathcal{P}_{\sim u}^{\perp} f(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u})]. \end{aligned}$$

Applying the Triangle inequality and Cauchy-Schwarz inequality gives

$$\begin{aligned}||\mathcal{P}^{\perp}_{\sim u} f(\mathbf{X}) - \mathcal{P}^{\perp}_{\sim u} f(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u})||_2^2 \leq & ||\mathcal{P}^{\perp}_{\sim u} f(\mathbf{X})||_2^2 \\ & + ||\mathcal{P}^{\perp}_{\sim u} f(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u})||_2^2 \\ & + 2||\mathcal{P}^{\perp}_{\sim u} f(\mathbf{X})||_2 ||\mathcal{P}^{\perp}_{\sim u} f(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u})||_2.\end{aligned}$$

Multiplying and dividing $2||\mathcal{P}^{\perp}_{\sim u} f(\mathbf{X})||_2 ||\mathcal{P}^{\perp}_{\sim u} f(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u})||_2$ by $||\mathcal{P}^{\perp}_{\sim u} f(\mathbf{X})||_2$ and dividing both sides of the inequality by $||f(\mathbf{X}) - f_0||_2^2$ completes the proof. □

Observe that if

$$||\mathcal{P}^{\perp}_{\sim u} f(\mathbf{X})||_2 = ||\mathcal{P}^{\perp}_{\sim u} f(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u})||_2$$

then

$$T_u \leq \delta_u \leq 4T_u.$$

This assumption typically does not hold in practice, nor is it easily verifiable; however, it provides some intuition about the behavior of the error. In particular, δ_u will be small when T_u is small and $||\mathcal{P}^{\perp}_{\sim u} f(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u})||_2$ is approximately $||\mathcal{P}^{\perp}_{\sim u} f(\mathbf{X})||_2$. The error, δ_u, will be large when the magnitude of $\mathcal{P}^{\perp}_{\sim u} f(\mathbf{X})$ increases dramatically on subsets of $\mathcal{X}$ which have a small probability under $\mathbf{X}$ and a larger probability under $(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u})$. The magnitude of δ_u is closely linked to how well the distribution of $(g(\mathbf{X}_{\sim u}), \mathbf{X}_{\sim u})$ approximates the distribution of $\mathbf{X}$, and the robustness of the Sobol' index which respect to changes in the distribution of $\mathbf{X}$, i.e. how much T_u changes when the distribution of $\mathbf{X}$ is changed. The method proposed in Chapter 3 may be used to test robustness.

Three conclusions may be drawn from the arguments above:

1. Dependencies between the variables can help reduce δ_u if $g(\mathbf{X}_{\sim u}) \approx \mathbf{X}_u$.
2. A tight upper bound will be difficult attain without placing additional assumptions on the behavior of f on sets of small probability.
3. Testing the robustness of T_u with respect to changes in the distribution of $\mathbf{X}$ provides a heuristic to asses when δ_u will be small.

2.4 Practical and computational considerations

The total Sobol' indices with dependent variables may be estimated via Monte Carlo integration [68] or the Fourier Amplitude Sensitivity Test [115]. In what follows, $\{T_k\}_{k=1}^p$ is estimated via

Monte Carlo integration using $(p+1)N$ evaluations of f, where N is the number of Monte Carlo samples [68].

When the variables are independent, Property IV bounds T_u using $\{T_k\}_{k=1}^p$, so it is typically sufficient to compute $\{T_k\}_{k=1}^p$ as inferences about T_u may be made using $\{T_k\}_{k=1}^p$. This does not generalize when the variables are dependent. The example in Subsection 2.5.1 provides a case where T_1 and T_2 are small, but $T_{1,2}$ is large. The approximation theoretic framework is helpful for interpreting this. When f is sensitive to two variables which are dependent on one another then one variable may be projected out with little error because the remaining variable can approximate its influence on f; however, when both are projected out a large error is incurred.

A practical strategy with dependent variables is to estimate $\{T_k\}_{k=1}^p$, which requires $(p+1)N$ evaluations of f. Then $\{T_k\}_{k=1}^p$ may be analyzed, along with information about the dependencies in $\mathbf{X}$ (known analytically or from the samples), and the user may select particular subsets $u \subset \{1, 2, \ldots, p\}$ for which to compute T_u. Using the estimator from [68], the additional cost to compute T_u, for a given u, will be N evaluations of f.

As shown in Chapter 3, the robustness of T_k to changes in the distribution of $\mathbf{X}$ may be computed as a by-product of computing $\{T_k\}_{k=1}^p$. If T_k, $k \in u$, is not robust to changes in the distribution of $\mathbf{X}$, then δ_u may be significantly larger than T_u. For a particular function g and subset u, the user may compute δ_u directly; this also requires N evaluations of f.

Choosing g is a challenge in practice. The simplest choice, $g(\mathbf{X}_{\sim u}) = \mathbb{E}[\mathbf{X}_u]$, fails to exploit dependency information and is not suggested. Rather, $g(\mathbf{X}_{\sim u}) = \mathbb{E}[\mathbf{X}_u|\mathbf{X}_{\sim u}]$ is suggested since (i) linear dependencies are common in practice (normal distributions and copula models are two common examples), and (ii)$\mathbb{E}[\mathbf{X}_u|\mathbf{X}_{\sim u}]$ is easily computed (either analytically or through linear regression with the existing samples). If the dependencies in $\mathbf{X}$ are known to be nonlinear then g may be estimated by nonlinear regression (using the existing samples). The challenge in this case is determining an appropriate nonlinear model for g.

2.5 Illustrative examples

This section provides two illustrative examples to highlight properties of the total Sobol' indices and their association with approximation error.

2.5.1 A linear function

Let

$$f(\mathbf{X}) = 20X_1 + 16X_2 + 12X_3 + 10X_4 + 4X_5 \tag{2.9}$$

and $\mathbf{X}$ follow a multivariate normal distribution with mean μ and covariance matrix Σ given by

$$\mu = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 1 & .5\rho & .5\rho & 0 & .8\rho \\ .5\rho & 1 & 0 & 0 & 0 \\ .5\rho & 0 & 1 & 0 & .3\rho \\ 0 & 0 & 0 & 1 & 0 \\ .8\rho & 0 & .3\rho & 0 & 1 \end{bmatrix}, \qquad 0 \leq \rho \leq 1.$$

The total Sobol' indices T_k, $k = 1, \ldots, 5$, are computed analytically and displayed in Figure 2.1 as a function of ρ. Observe that the ordering of importance changes as the correlations become stronger. This underscores the significance of accounting for dependencies. Also notice that the total Sobol' indices are decreasing as a function of ρ. The approximation theoretic perspective provides a nice interpretation of this. As the correlations are strengthened, the error associated with projecting out a variable decreases because its influence on $f(\mathbf{X})$ may be approximated by the other variables.

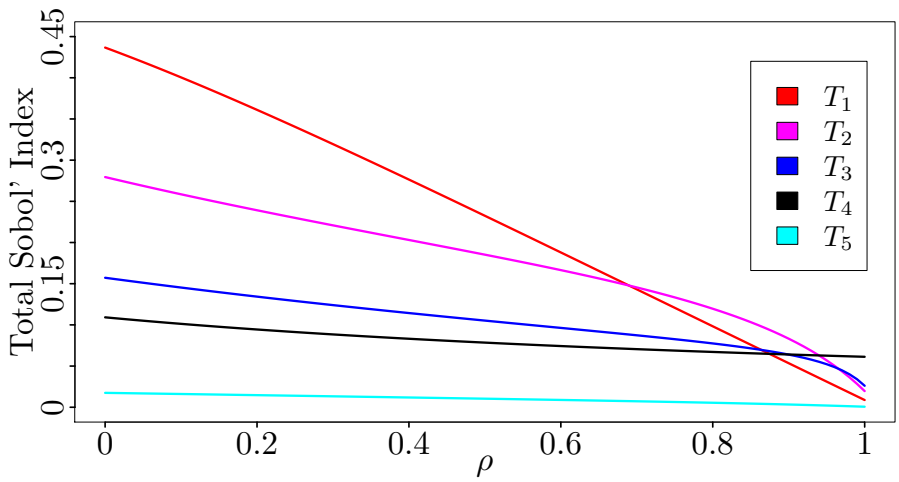

Figure 2.1 Total Sobol' indices for (2.9) with increasing correlation strength as ρ varies from 0 to 1.

Table 2.1 displays the total Sobol' indices T_1, T_2, and $T_{1,2}$ when $\rho = 1$. This demonstrates that two variables may have small total Sobol' indices individually (T_1 and T_2), but their joint total Sobol' index ($T_{1,2}$) may be significantly larger. This phenomenon, which does not occur when the variables are independent, is important when analyzing results with dependent variables; inference about subsets cannot be made with $\{T_k\}_{k=1}^p$ alone.

Table 2.1 Total Sobol' indices of (2.9) for variables X_1, X_2, and (X_1, X_2) when $\rho = 1$.

T_1	T_2	$T_{1,2}$
0.0087	0.0196	0.4228

2.5.2 A nonlinear function

Let f be the g-function of [109] with $p = 10$ variables; more precisely, f is given by

$$f(\mathbf{X}) = \prod_{k=1}^{10} \frac{|4X_k - 2| + a_k}{1 + a_k}, \tag{2.10}$$

where the parameters a_k, $k = 1, 2, \ldots, 10$, is given by $\mathbf{a} = (1, 2, 3, 9, 11, 13, 20, 25, 30, 35)$. Let $\mathbf{X}$ follow a multivariate normal distribution with mean $\mu \in \mathbb{R}^{10}$,

$$\mu_k = \frac{1}{2}, \qquad k = 1, 2, \ldots, 10,$$

and covariance matrix $\Sigma \in \mathbb{R}^{10\times 10}$,

$$\Sigma_{k,k} = \frac{1}{6}, \qquad k = 1, 2, \ldots, 10, \qquad \text{and} \qquad \Sigma_{i,j} = \frac{\rho}{6|i - j + 1|^{\frac{1}{\gamma}}}, \qquad i \neq j.$$

The covariance matrix is parameterized so that the magnitude of the covariances are large near the diagonal of Σ and decrease as they move away from the diagonal. The parameter γ determines the rate at which they decrease, as $\gamma \to \infty$, the off diagonal elements of Σ all converge to $\rho/6$. Hence γ tunes how many variables are strongly correlated with one another. The parameter ρ scales the strength of the correlations.

Direct calculations yield that variables X_i, $i = 7, 8, 9, 10$, are not influential for any ρ, γ, though $T_{7,8,9,10}$ does depend on ρ and γ. Figure 2.2 demonstrates how the Sobol' index $T_{7,8,9,10}$ and the approximation error $\delta_{7,8,9,10}$ vary with respect to ρ and γ. On the left panel $\gamma = 1$ is fixed and ρ is varied from 0 to 1; on the right panel $\gamma = 6$ is fixed and ρ is varied from 0 to 1.

Figure 2.2 shows that linear dependencies, as in the case of a multivariable normal random vector, aid in approximating f by fixing unimportant variables. In particular, observe that the error from replacing $\mathbf{X}_{7,8,9,10}$ with its conditional expectation decreases as ρ increases. Taking a larger γ, as in the right panel, corresponds to having more variables which are strongly correlated. Having $T_{7,8,9,10} \approx \delta_{7,8,9,10}$, as in the right panel with $\rho \geq 0.7$, demonstrates that the optimal approximation (which corresponds to computing a $L^2(\Omega)$ projection) may be accurately approximated by replacing $\mathbf{X}_{7,8,9,10}$ with its conditional expectation.

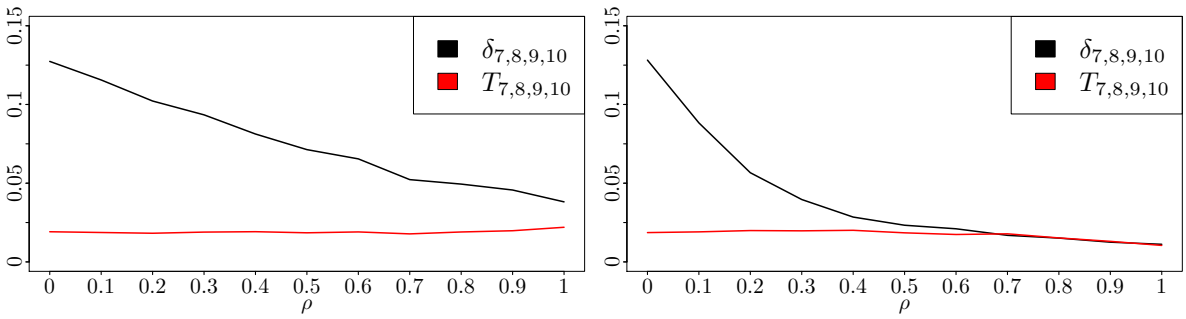

Figure 2.2 Sobol' index $T_{7,8,9,10}$ and approximation error $\delta_{7,8,9,10}$ for (2.10) as ρ varies from 0 to 1. Left: $\gamma = 1$; right: $\gamma = 6$.

2.6 Conclusion

This chapter provides a framework to analyze dimension reduction with dependent variables and highlights how dependencies may aid the user in dimension reduction. The approximation theoretic characterization of total Sobol' indices is useful as it demonstrates how total Sobol' indices are linked to optimal approximation and how they may be used to analyze the error when replacing variables $\mathbf{X}_u$ with a function $g(\mathbf{X}_{\sim u})$, a practical approach. An important factor in this analysis is the robustness of the total Sobol' indices to changes in the distribution of $\mathbf{X}$. Further analysis is needed to (i) connect the robustness studies in Chapter 3 with dimension reduction and (ii) determine an optimal g, particularly in the presence of nonlinear dependencies.

As mentioned in Chapter 1, there has been progress with derivative-based global sensitivity indices [100, 107, 108] and active subspaces [19] as alternative approaches for dimension reduction with independent variables. Extending analysis of these methods for dimension reduction with dependent variables is another avenue of future research.

The approximation theoretic perspective provides a useful characterization of Sobol' indices for dimension reduction with dependent variables. However, this characterization does not address difficulties which arise in other applications of GSA. For instance, when GSA is used for model development the user wants to identify the most important variables. If the most important variables have strong dependencies then their Sobol' indices may be small which hides the information the user desires. Future work may consider alternative characterizations which are focused toward other applications. This may involve the Sobol' indices, or possibly other tools in GSA.

CHAPTER

3

ROBUSTNESS OF SOBOL' INDICES TO INPUT DISTRIBUTION UNCERTAINTY

3.1 Introduction

As mentioned in Chapter 1, the Sobol' indices, and most other GSA methods, assume that the distribution of $\mathbf{X}$ is known and may be sampled. In many applications the distribution of $\mathbf{X}$ is not known and practitioners are left to estimate it. A common approach is to assume $X_1, X_2, \ldots, X_p$ are independent and define the marginal distribution of X_k to be uniform or normal centered at some nominal value. Defining a probability distribution for $\mathbf{X}$ is a challenging task which carries uncertainties itself. This raises the question, "how robust is a GSA method to changes in the distribution of $\mathbf{X}$?" This question is addressed for the Sobol' indices in this chapter.

The robustness of the Sobol' indices with respect to changes in the marginal distributions of X_k, $k = 1, 2, \ldots, p$, is considered in the Life Cycle Analysis literature [70]. Their approach requires the user to specify various admissible marginal distributions for the variables and compute the Sobol' indices for each possibility. This requires many evaluations of f thus limiting its use more broadly. The robustness of the Sobol' indices with respect to changes in the correlation structure is highlighted in [45] where the authors seek to quantify the risk of ignoring correlations between the variables. Imprecise probabilities are used in [49] to quantify uncertainty in the sensitivity

indices. This approach requires the user to parameterize admissible distributions and collect additional samples, i.e. additional evaluations of f. In the ecology literature [92], the robustness of Sobol' indices to changes in the means and standard deviations defining normally distributed inputs is examined. The challenges of deep uncertainty, i.e. uncertainty in the distributions of X_k, $k = 1, 2, \ldots, p$, are identified in [32] where the authors compute Sobol' indices for different distributions and define "robust sensitivity indicators" as a function of the Sobol' indices from different distributions. Similar questions about the robustness of computed quantities with respect to distributional uncertainty may be found in [3, 18, 56, 85, 88].

In [10], the ANOVA decomposition is analyzed when $\mathbf{X}$ does not have a unique distribution but rather multiple possible distributions. The authors provide a framework for analyzing the robustness of the Sobol' indices which depends upon the user specifying a prior on the space of possible distributions of $\mathbf{X}$. In line with [7], the robustness of the Sobol' indices may be determined by considering a set of possible distributions, sampling from their mixture distribution, and computing the Sobol' indices with respect to each distribution using a weighting scheme. This approach does not require additional evaluations of f, but the user must specify the set of possible distributions, which is challenging in practice.

All of the aforementioned approaches require user specification of possible distributions, additional evaluations of f, or both. In this chapter, a method is presented to measure the robustness of the Sobol' indices to distributional uncertainties without requiring either. In particular, the PDF of $\mathbf{X}$ is perturbed and the Sobol' indices are computed using the perturbed PDF. A judicious formulation facilitates a closed form solution to an optimization problem which determines PDF perturbations yielding maximum change in the Sobol' indices. The Sobol' indices with a perturbed PDF are then computed using weighted averages. The proposed method is a post-processing step which requires minimal user specification and no evaluations of f beyond those already taken to compute the Sobol' indices.

3.2 Robustness of the Sobol' index to PDF perturbations

As in Chapter 1, let $f : \mathcal{X} \to \mathbb{R}$, $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_p \subset \mathbb{R}^p$, be a function or model and let $\mathbf{X} = (X_1, X_2, \ldots, X_p) \in \mathcal{X}$ be the input variables of that model; assume $\mathbf{X}$ admits a PDF ϕ. For conciseness, this chapter focuses on the robustness of the total Sobol' index T_u to changes in ϕ; the Sobol' index S_u may be analyzed in a similar fashion.

There are multiple ways to express and estimate T_u; a useful expression from [68] is

$$T_u = \frac{\frac{1}{2}\int_{\mathcal{X}\times\mathcal{X}_u} (f(\mathbf{x}) - f(\mathbf{x}'))^2 \phi(\mathbf{x})\phi_{\mathbf{x}|\mathbf{x}_{\sim u}}(\mathbf{x}'|\mathbf{x}_{\sim u})\, d\mathbf{x}\, d\mathbf{x}'_u}{\int_{\mathcal{X}} f(\mathbf{x})^2\phi(\mathbf{x})d\mathbf{x} - \left(\int_{\mathcal{X}} f(\mathbf{x})\phi(\mathbf{x})d\mathbf{x}\right)^2} \tag{3.1}$$

where $\mathbf{x} = (\mathbf{x}_u, \mathbf{x}_{\sim u})$, $\mathbf{x}' = (\mathbf{x}'_u, \mathbf{x}_{\sim u})$, $\phi_{\mathbf{x}|\mathbf{x}_{\sim u}}$ is the conditional density for $\mathbf{X}|\mathbf{X}_{\sim u}$, and $\mathcal{X}_u$ is the Cartesian product of each $\mathcal{X}_k$, $k \in u$. Note that $\mathbf{x} = (\mathbf{x}_u, \mathbf{x}_{\sim u})$ is not a permutation of the entries of $\mathbf{x}$ but rather a partitioning of them. Then T_u may be estimated by drawing samples from $\mathbf{X}$, drawing a second set of samples from $\mathbf{X}|\mathbf{X}_{\sim u}$, and estimating (3.1) via Monte Carlo integration of the numerator and denominator separately.

The basic idea of the proposed approach is to view T_u as an operator which takes a PDF as input and returns the total Sobol' index. The Fréchet derivative of this operator is computed at ϕ and used to analyze the robustness of T_u. To this end, the following assumptions are made throughout the chapter:

1. $\mathcal{X}$ is a Cartesian product of compact intervals,
2. $\phi(\mathbf{x}) > 0 \ \forall \mathbf{x} \in \mathcal{X}$,
3. ϕ is continuous on $\mathcal{X}$,
4. f is bounded on $\mathcal{X}$.

Some of the results below may be shown with weaker assumptions, these overarching assumptions are made now for conciseness and simplicity. Without loss of generality, under the assumptions above, assume that $\mathcal{X} = [0,1]^p$.

The proposed method seeks to perturb the PDF, so it is essential that the perturbations preserve properties of PDFs, specifically, that every PDF is nonnegative and its integral over $\mathcal{X}$ equals one.

Since $\phi > 0$ is continuous and $\mathcal{X}$ is compact, ϕ is bounded above and below by positive real numbers. Define the Banach space V as the set of all bounded functions on $\mathcal{X}$ equipped with the norm

$$||\psi||_V = \left|\left|\frac{\psi}{\phi}\right|\right|_{L^\infty(\mathcal{X})},$$

where $||\cdot||_{L^\infty(\mathcal{X})}$ is the supremum norm on $L^\infty(\mathcal{X})$, the set of bounded function on $\mathcal{X}$. This norm ensures that $\phi + \psi \geq 0$ for all $\psi \in V$ with $||\psi||_V \leq 1$, the non negativity property of PDFs.

To ensure that the integral over $\mathcal{X}$ equals one, a normalization operator is introduced which takes $\eta \in V$ and returns $\frac{\eta}{\int_{\mathcal{X}} \eta(\mathbf{x}) d\mathbf{x}}$. Composing this normalization operator with (3.1) yields the total Sobol' index as an operator on V. Define $F, G, T_u : V \to \mathbb{R}$ by

$$F(\eta) = \frac{1}{2} \int_{\mathcal{X} \times \mathcal{X}_u} (f(\mathbf{x}) - f(\mathbf{x}'))^2 \eta(\mathbf{x}) \eta(\mathbf{x}') \frac{1}{\int_{\mathcal{X}_u} \eta(\mathbf{x}) d\mathbf{x}_u} d\mathbf{x} d\mathbf{x}'_u, \tag{3.2}$$

$$G(\eta) = \int_{\mathcal{X}} f(\mathbf{x})^2 \eta(\mathbf{x}) d\mathbf{x} - \frac{1}{\int_{\mathcal{X}} \eta(\mathbf{x}) d\mathbf{x}} \left(\int_{\mathcal{X}} f(\mathbf{x}) \eta(\mathbf{x}) d\mathbf{x} \right)^2, \tag{3.3}$$

$$T_u(\eta) = \frac{F(\eta)}{G(\eta)}. \tag{3.4}$$

It is easily observed that multiplying the numerator and denominator of (3.4) by $\frac{1}{\int_{\mathcal{X}} \eta(\mathbf{x}) d\mathbf{x}}$ yields that (3.4) and (3.1) coincide with ϕ replaced by $\frac{\eta}{\int_{\mathcal{X}} \eta(\mathbf{x}) d\mathbf{x}}$. In this framework, every $\eta \in V$ such that $||\phi - \eta||_V \leq 1$ is nonnegative and $T_u(\eta)$ corresponds to the total Sobol' index computed with respect to the PDF $\frac{\eta}{\int_{\mathcal{X}} \eta(\mathbf{x}) d\mathbf{x}}$.

Having defined the total Sobol' index as an operator which inputs bounded PDFs, Theorem 6 below gives the Fréchet derivative of the total Sobol' index at ϕ.

Theorem 6. *The operator T_u is Fréchet differentiable at ϕ with Fréchet derivative $\mathcal{D}T_u(\phi) : V \to \mathbb{R}$ given by the bounded linear operator*

$$\mathcal{D}T_u(\phi)\psi = \frac{\mathcal{D}F(\phi)\psi}{G(\phi)} - T_u(\phi)\frac{\mathcal{D}G(\phi)\psi}{G(\phi)}, \tag{3.5}$$

where

$$\begin{aligned}
\mathcal{D}F(\phi)\psi =& \frac{1}{2} \int_{\mathcal{X}\times\mathcal{X}_u} (f(\mathbf{x}) - f(\mathbf{x}'))^2 \frac{\psi(\mathbf{x}')}{\phi(\mathbf{x}')} \phi(\mathbf{x}) \phi_{\mathbf{x}|\mathbf{x}_{\sim u}}(\mathbf{x}'|\mathbf{x}_{\sim u}) d\mathbf{x} d\mathbf{x}'_u \\
&+ \frac{1}{2} \int_{\mathcal{X}\times\mathcal{X}_u} (f(\mathbf{x}) - f(\mathbf{x}'))^2 \frac{\psi(\mathbf{x})}{\phi(\mathbf{x})} \phi(\mathbf{x}) \phi_{\mathbf{x}|\mathbf{x}_{\sim u}}(\mathbf{x}'|\mathbf{x}_{\sim u}) d\mathbf{x} d\mathbf{x}'_u \\
&- \frac{1}{2} \int_{\mathcal{X}\times\mathcal{X}_u} (f(\mathbf{x}) - f(\mathbf{x}'))^2 \frac{\int_{\mathcal{X}_u} \psi(\mathbf{x}) d\mathbf{x}_u}{\int_{\mathcal{X}_u} \phi(\mathbf{x}) d\mathbf{x}_u} \phi(\mathbf{x}) \phi_{\mathbf{x}|\mathbf{x}_{\sim u}}(\mathbf{x}'|\mathbf{x}_{\sim u}) d\mathbf{x} d\mathbf{x}'_u
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{D}G(\phi)\psi = & \int_{\mathcal{X}} f(\mathbf{x})^2 \frac{\psi(\mathbf{x})}{\phi(\mathbf{x})} \phi(\mathbf{x}) d\mathbf{x} \\
& - 2 \int_{\mathcal{X}} f(\mathbf{x}) \phi(\mathbf{x}) d\mathbf{x} \int_{\mathcal{X}} f(\mathbf{x}) \frac{\psi(\mathbf{x})}{\phi(\mathbf{x})} \phi(\mathbf{x}) d\mathbf{x} \\
& + \left(\int_{\mathcal{X}} \frac{\psi(\mathbf{x})}{\phi(\mathbf{x})} \phi(\mathbf{x}) d\mathbf{x} \right) \left(\int_{\mathcal{X}} f(\mathbf{x}) \phi(\mathbf{x}) d\mathbf{x} \right)^2.
\end{aligned}$$

Proof. One may easily observe that $G(\eta) > 0$ in a neighborhood of ϕ (assuming $f(\mathbf{X})$ is non constant). It is sufficient to compute the Fréchet derivatives of F and G, the Fréchet derivative

of T_u follows from the quotient rule. The Fréchet derivatives of

$$\int_{\mathcal{X}} f(\mathbf{x})\phi(\mathbf{x})d\mathbf{x}, \qquad \int_{\mathcal{X}} f(\mathbf{x})^2\phi(\mathbf{x})d\mathbf{x}, \qquad \text{and} \qquad \int_{\mathcal{X}} \phi(\mathbf{x})d\mathbf{x},$$

when considered as operators from V to $\mathbb{R}$, acting on ψ, are easily shown to be

$$\int_{\mathcal{X}} f(\mathbf{x})\psi(\mathbf{x})d\mathbf{x}, \qquad \int_{\mathcal{X}} f(\mathbf{x})^2\psi(\mathbf{x})d\mathbf{x}, \qquad \text{and} \qquad \int_{\mathcal{X}} \psi(\mathbf{x})d\mathbf{x},$$

respectively, using the definition of the Fréchet derivative. The Fréchet derivative of G follows from the sum/difference, product, and chain rule of differentiation.

The Fréchet derivative of F may be computed by first defining an operator $H : V \to L^\infty(\mathcal{X} \times \mathcal{X}_u)$,

$$H(\eta) = \eta(\mathbf{x})\eta(\mathbf{x}')\frac{1}{\int_{\mathcal{X}_u} \eta(\mathbf{x})d\mathbf{x}_u},$$

where $\mathbf{x}'_{\sim u} = \mathbf{x}_{\sim u}$. The Fréchet derivatives of

$$\eta(\mathbf{x}), \qquad \eta(\mathbf{x}'), \qquad \text{and} \qquad \int_{\mathcal{X}_u} \eta(\mathbf{x})d\mathbf{x}_u,$$

when considered as operators from V to $L^\infty(\mathcal{X} \times \mathcal{X}_u)$, acting on ψ, are easily shown to be

$$\psi(\mathbf{x}), \qquad \psi(\mathbf{x}'), \qquad \text{and} \qquad \int_{\mathcal{X}_u} \psi(\mathbf{x})d\mathbf{x}_u,$$

respectively, using the definition of the Fréchet derivative. The Fréchet derivative of H follows from the product and quotient rules of differentiation. The Fréchet derivative of F may be easily computed using the Fréchet derivative of H, the boundedness of f, and the chain rule of differentiation. □

If the total Sobol' index is computed using a Monte Carlo estimator for the numerator and denominator of (3.1), then $\mathcal{D}T_u(\phi)\psi$ may be estimated using these samples and evaluations of f; the only additional work is evaluating ϕ and ψ at the sample points. Hence $\mathcal{D}T_u(\phi)\psi$ may be estimated at any $\psi \in V$ with negligible computational cost. This is why, as previously mentioned, the proposed method is a post-processing step which requires no additional evaluations of f beyond those taken to compute the total Sobol' indices.

An "optimal" perturbation of ϕ is sought in the sense that it causes the greatest change in the total Sobol' index. The locally optimal perturbation is the $\psi \in V$, $||\psi||_V \leq 1$, which maximizes $|\mathcal{D}T_u(\phi)\psi|$. To estimate this ψ, define a finite dimensional subspace $V_M \subset V$ and compute the operator norm of the restriction of $\mathcal{D}T_u(\phi)$ to V_M. When choosing V_M, there is a

tradeoff to consider between the approximating properties of functions from V_M, the ability to use existing samples to estimate the action of $\mathcal{D}T_u(\phi)$ on functions from V_M, and the ease of computing the operator norm of $\mathcal{D}T_u(\phi)$ restricted to V_M. In what follows, V_M is chosen to be a subspace generated by the span of a set of locally supported piecewise constant functions.

Let R_i, $i = 1, 2, \ldots, M$, be a partition of $\mathcal{X}$ into open hyperrectangles, i.e. $\mathcal{X} = \cup_{i=1}^{M} \overline{R_i}$ and $R_i \cap R_j = \emptyset$ for $i \neq j$; $\overline{R_i}$ denotes the closure of R_i. Define

$$\psi_i(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in R_i \\ 0 & \mathbf{x} \notin R_i \end{cases}$$

to be the indicator function of R_i, $i = 1, 2, \ldots, M$, and $V_M = \text{span}\{\psi_1, \psi_2, \ldots, \psi_M\}$, a M dimensional subspace of V. A partition may be efficiently constructed using Regression Trees [12]; this will be elaborated on in Section 3.4.

The operator norm of $\mathcal{D}T_u(\phi)$ restricted to V_M is given by

$$\begin{aligned} ||\mathcal{D}T_u(\phi)||_{\mathcal{L}(V_M,\mathbb{R})} &= \max_{\substack{\psi \in V_M \\ ||\psi||_V \leq 1}} |\mathcal{D}T_u(\phi)\psi| \\ &= \max_{\substack{\mathbf{a} \in \mathbb{R}^M \\ ||\sum_{i=1}^{M} a_i \psi_i||_V \leq 1}} \left| \mathcal{D}T_u(\phi) \left(\sum_{i=1}^{M} a_i \psi_i \right) \right| \\ &= \max_{\substack{\mathbf{a} \in \mathbb{R}^M \\ ||\sum_{i=1}^{M} a_i \psi_i||_V \leq 1}} \left| \sum_{i=1}^{M} a_i \mathcal{D}T_u(\phi)\psi_i \right|. \end{aligned}$$

Since the basis functions have disjoint support, it follows that

$$\left|\left| \sum_{i=1}^{M} a_i \psi_i \right|\right|_V = \left|\left| \sum_{i=1}^{M} a_i \frac{1}{\phi} \psi_i \right|\right|_{L^\infty(\mathcal{X})} = \max_{i=1,2,\ldots,M} |a_i| \left|\left| \frac{1}{\phi} \right|\right|_{L^\infty(R_i)},$$

which implies

$$\left|\left| \sum_{i=1}^{M} a_i \psi_i \right|\right|_V \leq 1$$

is equivalent to $|a_i| \leq b_i$, where b_i is the infimum of ϕ on R_i, $i = 1, 2, \ldots, M$.

Let $\mathbf{d} \in \mathbb{R}^M$ be defined by $d_i = \mathcal{D}T_u(\phi)\psi_i$, $i = 1, 2, \ldots, M$. Then

$$||\mathcal{D}T_u(\phi)||_{\mathcal{L}(V_M,\mathbb{R})} = \max_{\substack{\mathbf{a} \in \mathbb{R}^M \\ |a_i| \leq b_i \\ i=1,2,\ldots,M}} |\mathbf{d}^T \mathbf{a}|.$$

This problem may be solved in closed form to get

$$a_i = \text{sign}(d_i) b_i$$

and

$$||\mathcal{D}T_u(\phi)||_{\mathcal{L}(V_M,\mathbb{R})} = ||\mathbf{d}||_1.$$

In what follows, $\psi \in V_M$, $||\psi||_V \leq 1$, which maximizes the Fréchet derivative will be referred to as the *optimal perturbation*. Finding the optimal perturbation and the corresponding operator norm simplifies to evaluating $\mathcal{D}T_u(\phi)\psi_i$ for $i = 1, 2, \ldots, M$, which may be estimated with negligible additional computation. However, estimating $\mathcal{D}T_u(\phi)\psi_i$ is typically more challenging than estimating the total Sobol' index. Rather than inferring robustness with $||\mathcal{D}T_u(\phi)||_{\mathcal{L}(V_M,\mathbb{R})}$, it is proposed to:

i. estimate $a_i = \text{sign}(d_i) b_i$, $i = 1, 2, \ldots, M$,

ii. use weighted averaging with the existing evaluations of f and ϕ to estimate the total Sobol' indices with respect to the *optimally perturbed PDF*, which is defined as

$$\frac{\phi + \delta \sum_{i=1}^{M} a_i \psi_i}{1 + \delta \sum_{i=1}^{M} a_i \text{vol}(R_i)}, \tag{3.6}$$

where $\delta \in [-1, 1]$ is a parameter to scale the size of the perturbation and $\text{vol}(R_i)$ is the volume of the set R_i; the determination of δ will be discussed in Section 3.4. The total Sobol' indices computed with ϕ will be referred to as the *nominal total Sobol' indices* and the total Sobol' indices computed with the optimally perturbed PDF (3.6) will be referred to as the *perturbed total Sobol' indices*.

In practice, it is suggested to estimate the terms $\text{vol}(R_i)$, $i = 1, 2, \ldots, M$, in (3.6) with a Monte Carlo estimator from the existing data. They may be computed analytically since R_i is known; however, if they are computed exactly then the weights used to estimate perturbed Sobol' indices may not sum to one because of Monte Carlo error in the estimate. This can bias the resulting analysis. Estimating $\text{vol}(R_i)$, $i = 1, 2, \ldots, M$, from the existing data diminishes this potential bias.

Computing perturbed total Sobol' indices with weighted averages is an improvement from traditional derivative based robustness (or stability) analysis in several ways:

- Estimating a_i is easy. Since ϕ is known, b_i may be computed numerically by querying the

existing evaluations of ϕ (or possibly analytically, for instance if $\phi \equiv 1$ then $b_i = 1$ for all i), this computation is negligible. Assuming that enough samples have been taken for the total Sobol' index estimation to converge, determining the sign of d_i with these samples is relatively easy. Additionally, when the sign of d_i is not determined correctly it is frequently because $\mathcal{D}T_u(\phi)\psi_i \approx 0$, in which case this error is benign in the scope of the analysis.

- The user must determine δ; however, various values of $\delta \in [-1, 1]$ may be tested at negligible computation cost. The sample standard deviation of the weighted average may be compared with the sample standard deviation in the original estimator to determine admissible values of δ. Additional details are given in Section 3.4.

- Computing the perturbed Sobol' indices estimates a realized worst case. This is superior to worst case bounds, error bars, or confidence intervals, which in many cases are overly pessimistic. Further, computing error bars for each Sobol' index individually may yield misleading results. For instance, error bars for two variables may yield large intervals for each Sobol' index, but their magnitude relative to one another is nearly constant for any PDF perturbation. In this case the user would incorrectly conclude that the relative importance of the variables to one another is uncertain.

As previously highlighted, one way to test for robustness is to use weighted averages to estimate the total Sobol' indices with different PDFs. The challenge with this approach is that the user must specify the perturbed PDFs. The proposed method may be viewed as an improvement on this idea by automating the choice of perturbed PDFs. The Fréchet derivative operator norm yields a locally optimal perturbation, which will likely reveal greater changes in the total Sobol' indices when compared with a user manually selecting a small set of perturbed PDFs. However, the proposed method does not have the danger of finding unrealistic worst cases since it only seeks perturbations in a neighborhood of the existing PDF and is constrained to use the existing samples.

3.3 Robustness of the normalized total Sobol' index to PDF perturbations

Section 3.2 provides the necessary tools to analyze the robustness of the total Sobol' indices. As highlighted in Chapter 2, the total Sobol' indices may be smaller in magnitude when $\mathbf{X}$ possesses stronger dependencies. When the PDF is perturbed it may result in all of total Sobol' indices decreasing on a relative scale, that is, they decrease but their relative size does not change significantly. In this case, the inference of the relative importance of the variables is unchanged by the perturbed PDF. Rather than considering the robustness of the total Sobol' indices, it

is also advantageous to analyze the robustness of the relative importance of the variables. For clarity and notational simplicity, the remainder of the chapter will focus on the total Sobol' indices when $u = \{k\}$ is a singleton, i.e. the set of total Sobol' indices $\{T_k\}_{k=1}^p$.

In order to measure the relative importance of the variables as the PDF varies, define the normalized total Sobol' index $\mathcal{T}_k : V \to \mathbb{R}$ as

$$\mathcal{T}_k(\phi) = \frac{T_k(\phi)}{\sum_{i=1}^{p} T_i(\phi)} \tag{3.7}$$

for $k = 1, 2, \ldots, p$. Applying Theorem 6 and the quotient rule to (3.7) yields that $\mathcal{T}_k$ is Fréchet differentiable with Fréchet derivative

$$\mathcal{D}\mathcal{T}_k(\phi)\psi = \frac{\left(\sum_{i=1}^{p} T_i(\phi)\right) \mathcal{D}T_k(\phi)\psi - T_k(\phi)\left(\sum_{i=1}^{p} \mathcal{D}T_i(\phi)\psi\right)}{\left(\sum_{i=1}^{p} T_i(\phi)\right)^2}.$$

Since $\mathcal{D}\mathcal{T}_k(\phi)$ is a linear combination of the operators $\mathcal{D}T_k(\phi)$ from Section 3.2, $\mathcal{D}\mathcal{T}_k(\phi)\psi$ may be easily estimated using the same results previously presented. In fact, in Section 3.2 a subspace $V_M = \text{span}\{\psi_1, \psi_2, \ldots, \psi_M\}$ is defined and $\mathcal{D}T_k(\phi)\psi_i$ is computed for $i = 1, 2, \ldots, M$. Using this computation, $\mathcal{D}\mathcal{T}_k(\phi)\psi_i$, $i = 1, 2, \ldots, M$, can be computed at no additional cost. The same procedure from Section 3.2 may be adopted to compute perturbed PDFs and perturbed total Sobol' indices via weighted averages. Since the cost is negligible, it is suggested to compute the optimal perturbation using $\mathcal{D}T_k(\phi)$ and $\mathcal{D}\mathcal{T}_k(\phi)$, and estimate the perturbed total Sobol' indices for each perturbation.

Definition 3 below aids to identify perturbations which change the total Sobol' indices but not the relative importance of the variables.

Definition 3. *Let $\tilde{T}_k$ and $\tilde{\mathcal{T}}_k$ denote the perturbed total Sobol' indices and perturbed normalized total Sobol' indices, respectively, for some perturbation of ϕ. The absolute change in the total Sobol' indices is*

$$\sum_{k=1}^{p} |T_k - \tilde{T}_k|$$

and the relative change in the total Sobol' indices is

$$\sum_{k=1}^{p} |\mathcal{T}_k - \tilde{\mathcal{T}}_k|.$$

It is suggested to consider two sets of perturbed total Sobol' indices, those which yield the largest absolute change and those which yield the largest relative change. This will be further described in Section 3.4 and demonstrated in Section 3.5.

3.4 Algorithmic description

Algorithm 1 below summarizes the proposed method. In this section, the user inputs of Algorithm 1 are discussed in detail, important algorithmic features are highlighted, and the visualization and interpretation of the results are considered.

It was previously suggested to generate the partition R_i, $i = 1, 2, \ldots, M$, with a Regression Tree [12]. This is a judicious choice because the minimum number of samples in the sets R_i is easily specified. An integer L may be input and the Regression Tree will recursively partition $\mathcal{X}$ ensuring that each set of the partition contains at least L samples. This simplicity makes Regression Trees attractive. Taking small values of L typically results in V_M being a larger subspace, but will create error when estimating $\mathcal{D}T_u(\phi)\psi_i$ (since there will be fewer samples to estimate the integrals). The determination of L is discussed below. The relationship between L and M depends on the algorithm used to generate the partition; a Regression Tree will uniquely determine M as a function L, typically a decreasing function of L.

A Regression Tree may be trained using the existing samples and evaluations of f. It will pursue a partition of $\mathcal{X}$ for which f is approximately constant on the sets R_i. In some cases, as illustrated in Subsection 3.5.3, this may result in sets R_i, $i = 1, 2, \ldots, M$, where most of the b_i's are small. This is problematic because it limits the size of admissible perturbations. To mitigate this, a Regression Tree may be trained to generate a coarser partition which can be refined by the user to ensure that only a few b_i's are small. This is demonstrated in Subsection 3.5.3.

The norm of the perturbed PDF in (3.6) depends on δ. It was suggested to try various values of $\delta \in [-1, 1]$ (equally spaced points in $[-1, 1]$) and accept those which meet a convergence tolerance. If Monte Carlo integration is used to estimate the total Sobol' indices $\{T_k\}_{k=1}^p$, then the sample standard deviation may be used as a metric for convergence. Let σ_j and $\tilde{\sigma}_j$, $j = 1, 2, \ldots, p$, denote the sample standard deviation for the nominal and perturbed total Sobol' indices, respectively. For the results presented in this Chapter, the sample standard deviation is estimated by computing the standard derivation of 50 estimates generated by randomly subsampling half of the function evaluations. Assuming that σ_j, $j = 1, 2, \ldots, p$, are sufficiently small to ensure convergence of the nominal total Sobol' indices, it is required that $(\tilde{\sigma}_j/\tilde{T}_j)/(\sigma_j/T_j)$ be less than a threshold. Define

$$t = \max_{j=1,2,\ldots,p} \frac{(\tilde{\sigma}_j/\tilde{T}_j)}{(\sigma_j/T_j)}$$

and specify a threshold $\tau > 1$. The perturbed total Sobol' indices are accepted if $t \leq \tau$.

The inputs of Algorithm 1 are:

- N, the number of Monte Carlo samples,
- L, the minimum number of samples in each set of the partition,
- r, an integer denoting how many values of $\delta \in [-1, 1]$ to consider,
- and τ, the acceptance threshold for the perturbed total Sobol' indices.

The results in Section 3.5 use $L = 50$, $r = 60$, and $\tau = 1.5$; the number of Monte Carlo samples required depends on the problem. Numerical evidence, and intuition, indicate that t is approximately a quadratic function of δ centered at $\delta = 0$. To determine δ, the scalar nonlinear equation $t(\delta) = \tau$ may be solved by evaluating $t(\delta)$ at r equally spaced points in $[-1, 1]$. It is not necessary to take large values for r; the choice $r = 60$ introduces negligible computation and provides sufficient resolution. The choice $\tau = 1.5$ is considered a reasonable threshold to permit non trivial perturbations without introducing significant numerical errors. The choice of $L = 50$ is the least intuitive of the inputs. To justify this choice, a numerical experiment was performed varying $L = 25 + 5\ell$, $\ell = 0, 1, \ldots, 10$. The results, omitted from this chapter for conciseness, indicate that the proposed method is robust to changes in L. If necessary, the user may easily verify the particular choice of inputs used in their application by varying them. The computational cost of this numerical experiment is small.

Lines 2-5 of Algorithm 1 below is the total Sobol' index estimation and Lines 6-17 is the robustness analysis. In many applications, Line 4 dominates the computational cost and hence the cost of robustness analysis is negligible. Lines 6 and 8 may be done analytically in many applications. The computation in Lines 9-18 is primarily taking sample averages of data on memory so its cost is small. In particular, the nested for loops may appear burdensome, but the operations inside of them are sufficiently simple that they may be executed quickly.

Algorithm 1 returns a collection of $2p$ sets perturbed Sobol' indices. It is suggested to extract the perturbed total Sobol' indices with the largest absolute and relative changes to visualize alongside the nominal total Sobol' indices, denote them as $\{\tilde{T}_k^a, \tilde{T}_k^r, T_k\}_{k=1}^p$ where the superscripts a and r identify the total Sobol' indices with largest absolute and relative changes, respectively. This may be done by querying the collection of perturbed total Sobol' indices and creating a bar plot of $\{\tilde{T}_k^a, \tilde{T}_k^r, T_k\}_{k=1}^p$, see Figure 3.2 for an illustration of this. There are several possible scenarios the user may observe:

- If $\tilde{T}_k^a \approx T_k$, $k = 1, 2, \ldots, p$, then the user may confidently make inferences with the total Sobol' indices.

- If $\tilde{T}_k^a \not\approx T_k$, $k = 1, 2, \ldots, p$, but $\tilde{T}_k^r \approx T_k$, $k = 1, 2, \ldots, p$, then the user may confidently make inferences about the relative importance of the variables but not the magnitude of the total Sobol' indices.
- If there are variables such that $T_k \approx \tilde{T}_k^a \approx 0$ then they may be considered unimportant.
- If $T_k \approx 0$ but $\tilde{T}_k^a \not\approx 0$ then the user should excise caution treating X_k as unimportant.
- If $T_i > T_j$ but $\tilde{T}_i^r < \tilde{T}_j^r$ then the user may not be certain of the importance of X_i and X_j relative to one another.

If a particular Sobol' index T_k is of interest, the collection of perturbed Sobol' indices may be queried to assess its robustness. The user may easily visualize all $2p$ of the perturbed indices $\tilde{T}_k$ in a histogram.

Algorithm 1 Computation of total Sobol' indices with robustness post-processing

Input: N, L, r, τ
Draw N samples of $\mathbf{X}$, store them in $X_0 \in \mathbb{R}^{N \times p}$
Draw N samples of $\mathbf{X}|\mathbf{X}_{\sim k}$, store them in $X_k \in \mathbb{R}^{N \times p}$, $k = 1, 2, \ldots, p$
Evaluate $f(X_j)$, $j = 0, 1, \ldots, p$
Compute T_k, $k = 1, 2, \ldots, p$
Evaluate $\phi(X_j)$, $j = 0, 1, \ldots, p$
Generate a partition $\{R_i\}_{i=1}^M$ by using the data $(X_0, f(X_0))$ to train a Regression Tree with a minimum of L data points in each terminal node
Determine $b_i = \inf_{\mathbf{x} \in R_i} \phi(\mathbf{x})$, $i = 1, 2, \ldots, M$
Compute $\mathcal{D}T_k(\phi)\psi_i$, $i = 1, 2, \ldots, M$, $k = 1, 2, \ldots, p$
Compute $\mathcal{D}\mathcal{T}_k(\phi)\psi_i$, $i = 1, 2, \ldots, M$, $k = 1, 2, \ldots, p$
for k from 1 to p **do**
    Determine $\psi^{(k,1)} \in V_M$, $||\psi^{(k,1)}||_V \leq 1$, to maximize $|\mathcal{D}T_k(\phi)|$
    Determine $\psi^{(k,2)} \in V_M$, $||\psi^{(k,2)}||_V \leq 1$, to maximize $|\mathcal{D}\mathcal{T}_k(\phi)|$
    for ℓ from 0 to r **do**
        Compute $\{\tilde{T}_j^{(k,\ell,1)}\}_{j=1}^p$ and $t^{(k,\ell,1)}$ with perturbation $(\phi + \left(-1 + \frac{2\ell}{r}\right)\psi^{(k,1)})/C^{(k,\ell,1)}$
        Compute $\{\tilde{T}_j^{(k,\ell,2)}\}_{j=1}^p$ and $t^{(k,\ell,2)}$ with perturbation $(\phi + \left(-1 + \frac{2\ell}{r}\right)\psi^{(k,2)})/C^{(k,\ell,2)}$
    end for
end for
Output: $2p$ sets of perturbed Sobol' indices with largest admissible $t^{(k,\ell,I)} \leq \tau$
Note: $C^{(k,\ell,1)}$, $C^{(k,\ell,2)}$ are constants ensuring the perturbed PDF integrates to one.

3.5 Numerical results

In this section, three examples are presented to highlight different properties of the proposed method. The first example analyzes how the robustness analysis changes as more samples are taken. The second example expands on Section 3.3 by highlighting a case when the largest absolute change in the total Sobol' indices yields a small relative change. The final example is an application of the proposed method to the Lorenz system [74]. Two cases are considered in this example to demonstrate the effect of the partitioning on the robustness analysis.

3.5.1 Synthetic example to demonstrate convergence in samples

Let

$$f(\mathbf{X}) = \prod_{k=1}^{10} \frac{|4X_k - 2| + a_k}{1 + a_k}, \tag{3.8}$$

where each X_k is independent and uniformly distributed on $[0, 1]$, and $a_k = k-1$ for $k = 1, 2, ..., 10$. This is the "g-function" [109] commonly used in the GSA literature.

The nominal total Sobol' indices and perturbed total Sobol' indices of (3.8) are computed. The number of Monte Carlo samples is varied to analyze the convergence behavior of the robustness estimation, specifically, 1,000, 5,000, 10,000, and 50,000 Monte Carlo samples are used. For each fixed sample size, 32 repetitions of the calculation is performed to understand sampling variability. Figure 3.1 below displays box plots for the estimation of the largest total Sobol' index, T_1. The center panel is the estimation of T_1; the median estimation is nearly constant and the quantiles shrink as the number of samples increases, this reflects convergence of the estimator. The perturbation size δ is varied between -1 and 1 and it is determined that $|\delta| = .33$ is the maximum admissible perturbation size for the threshold $\tau = 1.5$. The left and right panels show the convergence of $\tilde{T}_1$ with perturbations $\delta = -.33$ and $\delta = .33$, respectively. The shrinking quantiles are very similar to those in the center panel demonstrating that the estimation error in $\tilde{T}_1$ is comparable to the estimation error in T_1. The left and right panels have slight decreasing and increasing trends, respectively. This is because the subspace V_M is larger when more samples are taken, thus the perturbations yield larger changes in the total Sobol' indices. For this example, the trend is relatively small reflecting the fact that taking a larger subspace does not yield significant changes in the perturbed total Sobol' index.

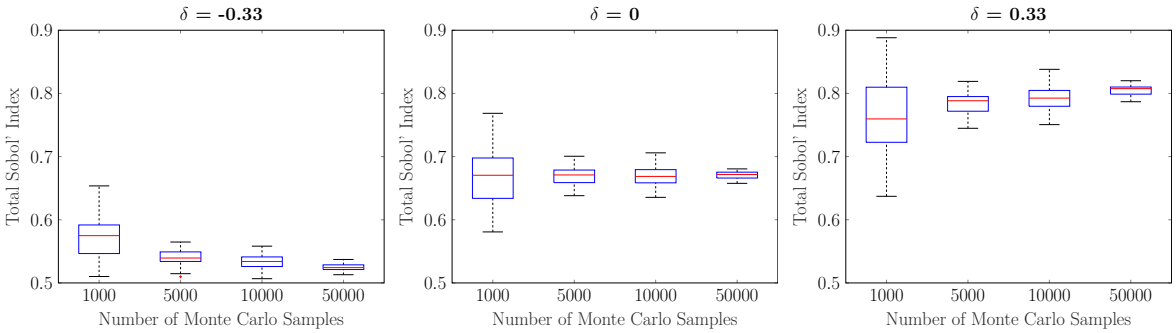

Figure 3.1 Convergence of the total Sobol' index for variable X_1 as the number of Monte Carlo samples vary. Left: perturbed total Sobol' index with $\delta = -.33$, center: nominal total Sobol' index, right: perturbed total Sobol' index with $\delta = .33$.

3.5.2 Synthetic example to demonstrate the normalized total Sobol' indices

This example illustrates the difference in the largest absolute and relative perturbations of the total Sobol' indices. Let

$$f(\mathbf{X}) = 1.5X_1 + 1.25X_2 + X_3 \tag{3.9}$$

where each X_k is independent and uniformly distributed on $[0, 1]$, $k = 1, 2, 3$. The total Sobol' indices are estimated with 5,000 Monte Carlo samples. Figure 3.2 displays the nominal total Sobol' indices of (3.9) in blue, the perturbed total Sobol' indices with the largest absolute differences change in cyan, and the perturbed total Sobol' indices with the largest relative change in yellow.

The largest absolute change of the total Sobol' indices corresponds to the case when they are all shifted down but their relative importance does not change. The largest relative change identifies a case where T_1 decreases while T_2 and T_3 increase. The relative importance of the variables change with this perturbation, demonstrating the benefit of considering the largest absolute and relative perturbations.

3.5.3 Lorenz system

This example applies the proposed method to the well known Lorenz system [74], a model for atmospheric convection. Sobol' indices were considered for this system in [84]. The Lorenz system is described by the system of ordinary differential equations

$$\frac{dy_1}{dt} = \sigma(y_2 - y_1)$$

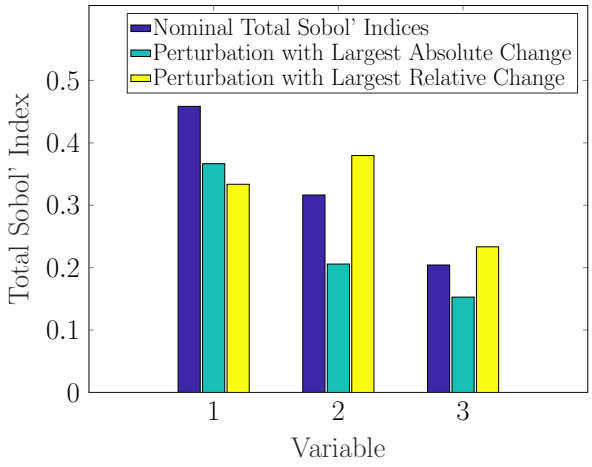

Figure 3.2 Total Sobol' indices for (3.9), the height of each bar indicates the total Sobol' index. The blue bars indicate the nominal total Sobol' indices; the cyan and yellow bars indicate the total Sobol' indices when the PDF of $\mathbf{X}$ was perturbed in extreme cases; cyan: the largest absolute change; yellow: the largest relative change.

$$\frac{dy_2}{dt} = y_1(\rho - y_3) - y_2$$
$$\frac{dy_3}{dt} = y_1 y_2 - \beta y_3$$

with initial conditions $y_i(0) = \alpha_i$, $i = 1, 2, 3$. Letting $\mathbf{X} = (\sigma, \rho, \beta, \alpha_1, \alpha_2, \alpha_3)$ denote the uncertain variables, the total Sobol' indices of the function

$$f(\mathbf{X}) = \frac{y_3(1)}{y_2(1)}$$

are computed. This choice of f corresponds to a ratio of temperature variations after a duration of 1 time unit.

The distribution of $\mathbf{X}$ is chosen to reflect uncertainty about nominal values of the variables. Two different cases, in the sub-subsections below, are considered to highlight different features of the proposed method. For each case, $10,000$ Monte Carlo samples are taken for the total Sobol' index estimation.

3.5.3.1 Lorenz system case 1

In this first case, assume the variables are independent with the uniform distributions given in Table 3.1 below. Figure 3.3 displays the nominal total Sobol' indices in blue, the perturbed total Sobol' indices with the largest absolute change in cyan, and the perturbed total Sobol' indices with the largest relative change in yellow. Several inferences may be drawn from this result,

- ρ and β are the most influential variables, although their total Sobol' indices and relative importance is uncertain,
- the total Sobol' indices for σ, α_1, and α_2 and their importance relative to one another is robust,
- α_3 has little influence and its small total Sobol' index is robust, it may be considered a non-influential variable.

Table 3.1 Marginal distribution for uncertain variables in Lorenz system Case 1. The means of σ, ρ, and β are the nominal values in [84].

Variable	Distribution	Support
σ	Uniform	$\left[\frac{97}{10}, \frac{103}{10}\right]$
ρ	Uniform	$\left[\frac{2716}{100}, \frac{2884}{100}\right]$
β	Uniform	$\left[\frac{194}{75}, \frac{206}{75}\right]$
α_1	Uniform	$\left[\frac{4}{5}, \frac{6}{5}\right]$
α_2	Uniform	$\left[\frac{4}{5}, \frac{6}{5}\right]$
α_3	Uniform	$\left[\frac{4}{5}, \frac{6}{5}\right]$

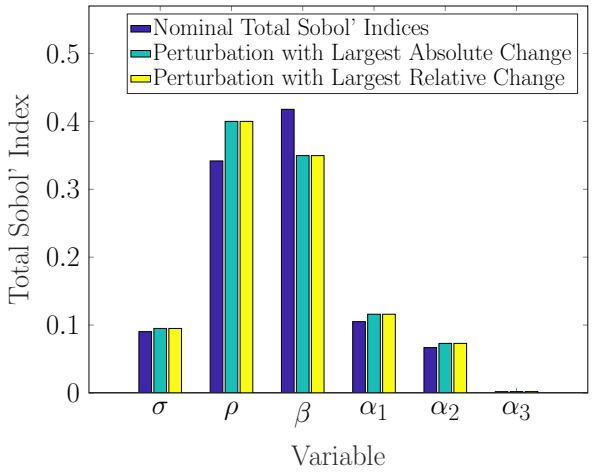

Figure 3.3 Total Sobol' indices for the Lorenz System Case 1 example, the height of each bar indicates the total Sobol' index. The blue bars indicate the nominal total Sobol' indices; the cyan and yellow bars indicate the total Sobol' indices when the PDF of $\mathbf{X}$ was perturbed in extreme cases; cyan: the largest absolute change; yellow: the largest relative change.

3.5.3.2 Lorenz system case 2

In this second case, assume the variables are independent and that all variables have the same marginal distribution given in Table 3.1 with the exception of α_3. Instead of being uniformly distributed on $\left[\frac{4}{5}, \frac{6}{5}\right]$ as in Case 1, α_3 is taken to have a Beta distribution on $\left[\frac{4}{5}, \frac{6}{5}\right]$ with shape parameters [1] $(1, 4)$. This corresponds to giving greater probability to $\alpha_3 < 1$.

A partition is generated by training a Regression Tree to predict f. The left panel of Figure 3.4 displays the nominal total Sobol' indices in blue, the perturbed total Sobol' indices with the largest absolute change in cyan, and the perturbed total Sobol' indices with the largest relative change in yellow. The results indicate that the total Sobol' indices are robust, a different conclusion than was reached in Case 1. This occurred because the Regression Tree never partitioned on α_3 so each set R_i contained the entire support of α_3. Because the marginal PDF for α_3 takes small values on part of its support, namely near $\frac{5}{4}$, the infimum of ϕ on each R_i is small. The partition generated by the Regression Tree yielded very small perturbations and as a result did not produce significant changes in the total Sobol' indices.

To alleviate this problem, a partition is generated by a Regression Tree trained to predict f using all of the variables except α_3. A minimum of $4L$ samples are requested in each hyperrectangle rather than L, as requested previously. This yields a coarser discretization of the other 5 variables. The resulting partition is refined by splitting each set into 4 subsets defined by partitioning α_3 at its quantiles. This yields a partition with approximately L samples per subset and a sufficient discretization of α_3 to enable larger perturbations. Figure 3.4 displays the nominal total Sobol' indices in blue, the perturbed total Sobol' indices with the largest absolute change in cyan, and the perturbed total Sobol' indices with the largest relative change in yellow. Larger changes in the total Sobol' indices are observed, as is expected. However the changes are smaller than what was observed in Case 1. This is because the partition used in Case 1 was generated by a Regression Tree which better approximated f, and hence allowed for larger perturbations of the total Sobol' indices. The general conclusion from this example is that the partition should be generated so that the Regression Tree approximates f as well as possible. If small values of ϕ prohibit taking large perturbations, then the partition may be generated with fewer hyperrectangles, followed by a refining of this coarse partition to sufficiently discretize the necessary regions. This may result in a failure to discover the largest possible perturbations, as demonstrated by comparing Case 1 and Case 2.

[1] For shape parameters (a, b), a Beta random variable x on $[0, 1]$ has a unnormalized PDF given by $x^a(1-x)^b$.

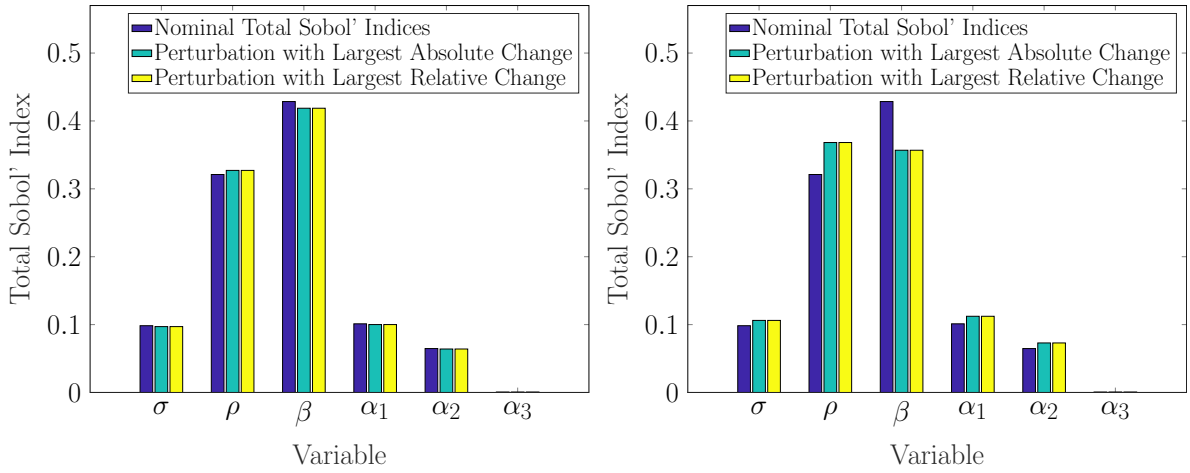

Figure 3.4 Total Sobol' indices for the Lorenz System Case 2 example, the height of each bar indicates the total Sobol' index. The blue bars indicate the nominal total Sobol' indices; the cyan and yellow bars indicate the total Sobol' indices when the PDF of $\mathbf{X}$ was perturbed in extreme cases; cyan: the largest absolute change; yellow: the largest relative change. The left and right panel correspond to generating the partition by training a Regression Tree to: predict f with a minimum of L samples per hyperrectangle (left) and predict f with a minimum of $4L$ samples per hyperrectangle, followed by additional partitioning of α_3 (right).

3.6 Conclusion

This chapter presents a novel framework in which robustness of the total Sobol' indices with respect to the input variables distribution may be assessed. The proposed method permits such analysis to be done at negligible computational cost. For a modeler using total Sobol' indices, this robustness analysis may be obtained as a by-product of computing total Sobol' indices and may be easily visualized along with the indices themselves. Understanding the robustness of the Sobol' indices to distributional uncertainty prevents the user from making incorrect inferences which have significant consequences. For instance, reducing dimensions by fixing variables with small Sobol' indices—which are not robust—may result in model variations which are not explained in the lower dimensional space.

The proposed method suffers three primary limitations, namely,

1. the nominal PDF must be compactly supported,
2. perturbations may not change the support of the nominal PDF,
3. the perturbation direction is only locally optimal.

The first limitation prohibits a direct application of the method to many commonly used PDFs. This occurs because the Fréchet derivative is not well defined if perturbations in the tail

of the distribution are allowed. However, one may define a compact subset of the domain where the PDF assumes most of its mass and allow perturbations on this subset while keeping the tail fixed. A theory analogous to what is presented in the chapter may be developed for this scenario. In practice, the modeler will take a compact subset which contains all of the existing samples and the robustness analysis will be identical to what is presented in this chapter.

The second limitation occurs because the method is formulated to work with existing samples. If the support of the PDF increases, then additional evaluations of f in these unexplored regions are required.

The third limitation is more significant in practice than the previous two; the proposed method only considers locally optimal perturbations about the nominal PDF. This limitation is inherent to derivative-based approaches, but may be mitigated by computing higher order Fréchet derivatives, or computing the Fréchet derivative at other nominal PDFs.

CHAPTER

4

SOBOL' INDICES FOR STOCHASTIC MODELS

4.1 Introduction

In this chapter the Sobol' indices are extended from their classical framework, presented in Chapter 1, to stochastic models of the form

$$f : \mathcal{X} \times \Omega \to \mathbb{R} \tag{4.1}$$

where $(\Omega, \mathcal{E}, \lambda)$ is a probability space. The content of this chapter is based on the publication [51]. Here $f(\mathbf{X}, \omega)$ depends on two sources of uncertainty, namely, uncertain input variables $\mathbf{X}$ and some intrinsic stochasticity ω. Examples of such models include agent-based models, queuing models, Monte Carlo based numerical models, and models of intrinsically stochastic phenomena such as those found in biological systems [118] or chemical reaction networks [36].

Stochastic models of the form (4.1) present theoretical and computational challenges to the traditional GSA paradigm. There are multiple ways in which the framework of Chapter 1 may be extended to stochastic models. A natural approach is to first marginalize over ω, and then

evaluate the Sobol' indices of the resulting function of $\mathbf{X}$. Specifically, the Sobol' indices of

$$g(\mathbf{X}) = \mathbb{E}_\omega(f(\mathbf{X}, \omega))$$

may be defined and computed using the classical theory presented in Chapter 1. This approach is used in [58, 80]. Though appealing for its simplicity, marginalizing over ω a priori may yield misleading results; this is demonstrated in Example 4.1.1 below.

This chapter presents a new approach to extend the Sobol' indices to stochastic models. It avoids the misleading results which may arise from marginalizing over ω a priori. The proposed approach defines the Sobol' index as a random variable and considers its statistical properties.

The two aforementioned approaches are considered in Example 4.1.1 below; a fuller description of the proposed method is given later in the chapter.

Example 4.1.1. *Let $(\Theta, \mathcal{F}, \nu)$, $(\Omega, \mathcal{E}, \lambda)$, be probability spaces, and $\mathbf{X} : \Theta \to \mathbb{R}^2$, $W : \Omega \to \mathbb{R}$ be random vectors/variables defined by $\mathbf{X} = (\mu, \sigma)$ where $\mu \sim \mathcal{U}(0, 1)$ and $\sigma \sim \mathcal{U}(1, L + 1)$, $L \geq 0$, are independent random variables, and $W \sim \mathcal{N}(0, 1)$; $\mathcal{U}$ and $\mathcal{N}$ denote uniform and normal distributions, respectively. Then*

$$Y = f(\mathbf{X}, \omega) = \mu + \sigma W(\omega), \tag{4.2}$$

is a stochastic model of the form (4.1). For $L = 0$, σ is deterministic; as the value of L increases, so does the uncertainty on σ. One expects the importance of σ to increase with L. This is confirmed by direct calculations. The first order Sobol' indices of Y with respect to both μ and σ can be found analytically

$$S_\mu(Y)(\omega) = \frac{1}{1 + L^2 W(\omega)^2} \quad \text{and} \quad S_\sigma(Y)(\omega) = \frac{L^2 W(\omega)^2}{1 + L^2 W(\omega)^2},$$

and the corresponding expected values are given by

$$\begin{aligned} \mathbb{E}_\omega\left(S_\mu(Y)\right) &= \frac{1}{L}\sqrt{\frac{\pi}{2}} \exp\left(\frac{1}{2L^2}\right) \operatorname{erfc}\left(\frac{1}{\sqrt{2}L}\right), \\ \mathbb{E}_\omega\left(S_\sigma(Y)\right) &= 1 - \mathbb{E}_\omega\left(S_\mu(Y)\right), \end{aligned}$$

where erfc *is the complementary error function. Figure 4.1 illustrates the behavior of $\mathbb{E}_\omega\left(S_\sigma(Y)\right)$ as a function of L confirming the increasing importance of σ with L.*

Reversing the order of operations between averaging and computing the Sobol' indices leads to an entirely different picture which is at odds with the very nature of (4.2). Indeed, the expected value of Y with respect to ω is simply $\mathbb{E}_\omega(Y) = \mu$ and therefore the first order Sobol' indices are

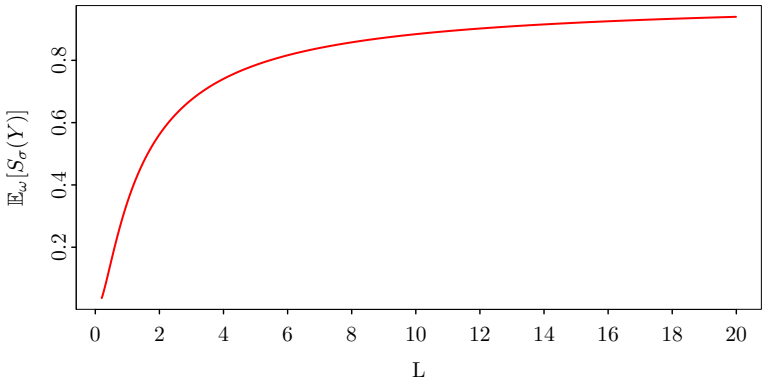

Figure 4.1 Expected first order Sobol' index of Y from (4.2) with respect to the uncertain variable σ as a function of L.

given by

$$S_\mu(\mathbb{E}_\omega(Y)) = 1 \quad \textit{and} \quad S_\sigma(\mathbb{E}_\omega(Y)) = 0.$$

In other words, σ appears insignificant regardless of L.

As highlighted in Example 4.1.1, *computing ω-moments* and *evaluating Sobol' indices* for (4.1) are two operations that do not commute. This simple observation has significant consequences as averaging over ω before computing sensitivity indices significantly reduces the amount of information available for analysis. Motivated by this, theoretical and computation analysis is presented in this chapter to demonstrate the validity and practical value of the proposed approach.

A fair amount of recent work on GSA for stochastic models has focused on analyzing stochastic chemical systems [24, 75, 76]. In [75], for instance, the authors develop a method, based on polynomial chaos expansion and a stochastic Galerkin formalism, to analyze the variance of stochastic differential equations driven by additive or multiplicative Wiener noise. In [76], a method is proposed for sensitivity analysis of stochastic chemical systems to the different reaction channels and channel interactions. Local derivative-based sensitivity analysis for stochastic systems has also seen significant progress. References [46, 63, 87, 95, 98] provide a sample of such efforts. Finally, other approaches for sensitivity analysis of stochastic systems rely on information theory [2, 66, 77, 78].

4.2 Formulation and method

Given the probability spaces $(\Theta, \mathcal{F}, \nu)$ and $(\Omega, \mathcal{E}, \lambda)$, the product probability space

$$(\Theta, \mathcal{F}, \nu) \otimes (\Omega, \mathcal{E}, \lambda) = (\Theta \times \Omega, \mathcal{F} \otimes \mathcal{E}, \nu \otimes \lambda),$$

may be defined in the customary way. Its sample space is $\Theta \times \Omega$, $\mathcal{F} \otimes \mathcal{E}$ is the σ-algebra generated by the sets $\mathcal{F} \times \mathcal{E}$, and $\nu \otimes \lambda$ is the product measure.

If $f : \mathcal{X} \times \Omega \to \mathbb{R}$, then $f(\mathbf{X}(\cdot), \cdot) : \Theta \times \Omega \to \mathbb{R}$ is a random variable on $(\Theta \times \Omega, \mathcal{F} \otimes \mathcal{E}, \nu \otimes \lambda)$, it is assumed to be square integrable. For a fixed $\omega \in \Omega$, $f(\mathbf{X}, \omega)$ is a deterministic function of uncertain variables for which the Sobol' indices may be computed in the classical fashion. Doing this for each $\omega \in \Omega$ defines functions

$$\begin{aligned} S_u : \Omega &\to [0, 1], \\ \omega &\mapsto S_u(f(\mathbf{X}, \omega)), \qquad u \subset \{1, 2, \ldots, p\}. \end{aligned}$$

Invoking the remarks in Theorem 1.7.2 of [25] and elementary properties of measurable functions, it can be seen that the S_u's are $\mathcal{E}$-measurable functions, i.e., they are random variables. As illustrated by Example 4.1.1, the statistical properties of the indices $\{S_u(\omega)\}_{u \subset \{1,2,\ldots,p\}}$ contains significant information needed for sensitivity analysis. The total Sobol' indices may likewise be defined as random variables using this stochastic formulation of S_u along with (1.6). This chapter focuses on the first order Sobol' indices but total Sobol' indices may be treated analogously.

Computing these sensitivity indices is, in general, costly. For instance, to compute all p first order indices $\{S_k(\omega)\}_{k=1}^p$ through the sampling based method from Chapter 1 with N Monte Carlo samples requires $(p+1)N$ evaluations of $f(\mathbf{X}(\cdot), \omega)$ for each fixed $\omega \in \Omega$. To characterize the statistical properties of the indices, an additional Monte Carlo sampling over Ω has to be performed. Assuming a sample size of m in Ω leads to a total of

$$\ell_{\text{samp}} = m(p+1)N \tag{4.3}$$

evaluations of f. Such a cost is prohibitive in many applications where N might be of the order of thousands or tens of thousands.

To mitigate this challenge, it is proposed to use a surrogate model for f. Specifically, for each fixed ω, a surrogate $\hat{f}(\mathbf{X}, \omega) \approx f(\mathbf{X}, \omega)$ may be constructed and its Sobol' indices computed. The construction of the surrogate $\hat{f}$ requires an ensemble of function evaluations, $\{f(\mathbf{X}_j, \omega)\}_{j=1}^n$, where $\{\mathbf{X}_j\}_{j=1}^n$ are realizations of the uncertain variables.

The construction of an efficient surrogate only requires n function evaluations, where n is much smaller than the number of Monte Carlo samples, i.e., $n \ll N$. Assuming the cost

to compute Sobol' indices is negligible relative to evaluations of f, the total cost is reduced from (4.3) to

$$\ell_{\text{surrogate}} = mn \tag{4.4}$$

evaluations of f, where n is significantly smaller than $(p+1)N$. While the computational cost (4.4) appears to be independent of the uncertain variable dimension p, it should be noted that the choice of n depends on p. This dependence is linked to the surrogate model itself.

Other indices, S_u or T_u, $u \subset \{1, 2, \ldots, p\}$, may also be computed from the surrogate model. The main steps of the method for computing S_u for stochastic models is given in Algorithm 2 below.

Algorithm 2 Efficient approximation of $\{S_u\}_{u \subset \{1,2,\ldots,p\}}$

for i from 1 to m **do**
  Randomly generate $\omega_i \in \Omega$
  Generate realizations $\{\mathbf{X}_j\}_{j=1}^n$ of the uncertain variables
  Evaluate $f(\mathbf{X}_j, \omega_i)$, for each $j = 1, \ldots, n$
  Construct surrogate $\hat{f}(\mathbf{X}, \omega_i) \approx f(\mathbf{X}, \omega_i)$ using data set $(\mathbf{X}_j, f(\mathbf{X}_j, \omega_i))_{j=1}^n$
  Compute $\hat{S}_u(\omega_i)$ using the surrogate $\hat{f}(\mathbf{X}, \omega_i)$ for $u \subset \{1, 2, \ldots, p\}$
end for
Approximate statistical properties of S_u using $\{\hat{S}_u(\omega_i)\}_{i=1}^m$ for $u \subset \{1, 2, \ldots, p\}$

The algorithm returns m realizations of Sobol' indices of $\hat{f}(\mathbf{X}, \omega)$, i.e., $\hat{S}_u(\omega)$, $u \subset \{1, 2, \ldots, p\}$. Denote these realizations $\hat{S}_u^i \overset{iid}{\sim} \hat{S}_u$, $i = 1, \ldots, m$, and consider the sample r-th moment

$$\hat{\mu}_u^{[r]}(\omega) = \frac{1}{m} \sum_{i=1}^{m} (\hat{S}_u^i(\omega))^r. \tag{4.5}$$

Clearly,

$$\mathbb{E}_\omega\left(\hat{\mu}_u^{[r]}\right) = \mathbb{E}_\omega\left((\hat{S}_u)^r\right) \quad \text{and} \quad \text{Var}_\omega\left(\hat{\mu}_u^{[r]}\right) = \frac{\text{Var}_\omega((\hat{S}_u)^r)}{m}, \quad u \subset \{1, 2, \ldots, p\}.$$

The error in approximating $\mathbb{E}_\omega\left((S_u)^r\right)$ can be decomposed into Monte Carlo error using m samples from Ω and surrogate approximation error using n samples from $\mathcal{X}$.

Theorem 7. *Let $\hat{\mu}_u^{[r]}$, S_u, and $\hat{S}_u$ be as defined above. Then,*

1. $\mathbb{E}_\omega\left(\hat{\mu}_u^{[r]} - \mathbb{E}_\omega\left((S_u)^r\right)\right) = \mathbb{E}_\omega\left((\hat{S}_u)^r - (S_u)^r\right)$,

2. $\mathrm{Var}_\omega\left(\hat{\mu}_u^{[r]} - \mathbb{E}_\omega\left((S_u)^r\right)\right) \leq \dfrac{\mathbb{E}_\omega\left((\hat{S}_u)^r\right)\left(1 - \mathbb{E}_\omega\left((\hat{S}_u)^r\right)\right)}{m} \leq \dfrac{1}{4m}.$

Proof. The first statement follows from

$$\begin{aligned}\mathbb{E}_\omega\left(\hat{\mu}_u^{[r]} - \mathbb{E}_\omega\left((S_u)^r\right)\right) &= \mathbb{E}_\omega\left(\hat{\mu}_u^{[r]} - \mathbb{E}_\omega\left((\hat{S}_u)^r\right) + \mathbb{E}_\omega\left((\hat{S}_u)^r\right) - \mathbb{E}_\omega\left((S_u)^r\right)\right) \\ &= \left(\mathbb{E}_\omega\left(\hat{\mu}_u^{[r]}\right) - \mathbb{E}_\omega\left((\hat{S}_u)^r\right)\right) + \mathbb{E}_\omega\left((\hat{S}_u)^r - (S_u)^r\right) = \mathbb{E}_\omega\left((\hat{S}_u)^r - (S_u)^r\right).\end{aligned}$$

For the second statement,

$$\begin{aligned}\mathrm{Var}_\omega\left(\hat{\mu}_u^{[r]} - \mathbb{E}_\omega\left((S_u)^r\right)\right) &= \mathrm{Var}_\omega\left(\hat{\mu}_u^{[r]}\right) \\ &= \frac{\mathrm{Var}_\omega\left((\hat{S}_u)^r\right)}{m} \leq \frac{\mathbb{E}_\omega\left((\hat{S}_u)^r\right)\left(1 - \mathbb{E}_\omega\left((\hat{S}_u)^r\right)\right)}{m} \leq \frac{1}{4m},\end{aligned}$$

where the inequalities follow from the Theorem 2 in [5] and the fact that $\hat{S}_u$ is supported on $[0, 1]$. □

To understand the implication of the above result, consider the point estimator for the expected value of $S_u(\omega)$ given by the sample mean:

$$\hat{\mu}_u^{[1]}(\omega) = \frac{1}{m}\sum_{i=1}^{m} \hat{S}_u^i(\omega). \tag{4.6}$$

Theorem 7 characterizes the bias of this estimator as the approximation error due to the surrogate, i.e., $\mathbb{E}_\omega\left(\hat{S}_u - S_u\right)$. Further, since, $\mathbb{E}_\omega\left((\hat{S}_u)^r\right) \in [0, 1]$, for every $r \geq 1$, the second statement of Theorem 7 indicates that even a modest value of m, the order of a few hundreds for instance, can be very effective in obtaining an estimator with small variance. Finally, an estimate of the error can be obtained in the $L^2(\Omega)$ norm by using the elementary definition of the variance and Theorem 7,

$$E_\omega\left((\hat{\mu}_u^{[1]} - \mathbb{E}_\omega\left(S_u\right))^2\right) \leq \mathbb{E}_\omega\left(\hat{S}_u - S_u\right)^2 + \frac{1}{4m}.$$

4.3 Numerical results

The following two examples illustrate some of the points raised in the previous section: (i) the convergence of the estimators as a function of n (number of samples to build the surrogate) and m (number of samples over Ω) and, (ii) the effect of the surrogate bias on the statistical distribution of the Sobol' indices. Multivariate Adaptive Regression Splines (MARS) [28, 29]

is used as the surrogate model in these two examples because of its internal adaptivity which minimizes the users responsibility to tune parameters.

4.3.1 The stochastic g-function

Let $(\Theta, \mathcal{F}, \nu)$ and $(\Omega, \mathcal{E}, \lambda)$ be two probability spaces and let $\mathbf{X} : \Theta \to \mathbb{R}^{15}$ and $W : \Omega \to \mathbb{R}$ be random vectors/variables such that

$$\begin{aligned}&\mathbf{X} = (X_1, \ldots, X_p) \text{ with } X_k \overset{iid}{\sim} \mathcal{U}(0,1),\ k = 1, \ldots, p,\\ &W \sim \text{Beta}(5,3);\end{aligned}$$

in other words,

$$\nu\big(\mathbf{X} \in (c_1, d_1) \times \cdots \times (c_{15}, d_{15})\big) = \prod_{k=1}^{15} (d_k - c_k) \text{ for any } 0 \le c_k \le d_k \le 1, k = 1, \ldots, p,$$

$$\lambda(W \in (a,b)) = K \int_a^b t^4 (1-t)^2 \, dt \text{ for any } a,\ b, 0 \le a \le b \le 1,$$

with a normalization factor $K = \Gamma(5+3)/\big(\Gamma(5)\Gamma(3)\big) = 105$.

Then define a stochastic version of the g-function [109]

$$f(\mathbf{X}, \omega) = \prod_{k=1}^{15} \frac{|4X_k - 2| + a_k(W(\omega))}{1 + a_k(W(\omega))}, \tag{4.7}$$

where the parameters $a_k : [0,1] \to \mathbb{R}$, $k = 1, \ldots, 15$, are chosen to create a variety of means and variances for the Sobol' indices. Analytic expressions for the a_k's are given in Table 4.1.

Table 4.1 Expressions of the parameters a_k, $k = 1, \ldots, 15$, for the stochastic g-function example (4.7).

$a_1(t) = (1-t)^5$	$a_2(t) = t^5$	$a_3(t) = \sin^2(8t)$
$a_4(t) = \sin^2(10(1-t))$	$a_5(t) = \cos^2(10(1-t))$	$a_6(t) = \cos^2(8t)$
$a_7(t) = (1.5-t)^2$	$a_8(t) = (.5+t)^2$	$a_9(t) = (3-t)^2$
$a_{10}(t) = (2+t)^2$	$a_{11}(t) = (3.5-t)^2$	$a_{12}(t) = (2.5+t)^2$
$a_{13}(t) = (4-t)^2$	$a_{14}(t) = (3+t)^2$	$a_{15}(t) = (4+t)^2$

The S_k's are computed analytically and $\mathbb{E}_\omega(S_k)$, $k = 1, \ldots, 15$, is evaluated using numerical quadratures. The trapezoidal rule with 10^6 quadrature nodes is used to ensure accurate

computation of the expectations.

In the first test, the error approximating $\mathbb{E}_\omega(S_k)$, $k = 1, \ldots, 15$, is considered as a function of the number of surrogate samples n. For $n = 100, 150, \ldots, 950, 1000$, the $\hat{\mu}_k^{[1]}$'s from (4.6) are obtained from Algorithm 2. Samples from $\mathcal{X}$ are taken using a Latin Hypercube design. To remove dependence upon sampling, 500 different datasets are generated for each fixed n and the error is defined as the average errors over these datasets. Since higher order interactions are difficult for surrogate models to represent in high dimensions, the normalized first order Sobol' indices $\mathbb{E}_\omega(S_k) / \sum_{i=1}^p \mathbb{E}_\omega(S_i)$ are considered. This facilitates a more useful comparison of the exact Sobol' indices and the Sobol' indices of the surrogate by ensuring that they both sum to 1. Define the error

$$\bar{E} = \frac{1}{500} \sum_{\ell=1}^{500} \left\| \frac{\mathbb{E}_\omega(\mathbf{S})}{\sum_{k=1}^p \mathbb{E}_\omega(S_k)} - \frac{\hat{\boldsymbol{\mu}}(\ell, n)}{\sum_{k=1}^p \hat{\mu}_k(\ell, n)} \right\|_\infty , \tag{4.8}$$

where $\mathbf{S} = (S_1, S_2, \ldots, S_p)$, $\hat{\boldsymbol{\mu}}(\ell, n) = (\hat{\mu}_1(\ell, n), \hat{\mu}_2(\ell, n), \ldots, \hat{\mu}_p(\ell, n))$, and each $\hat{\mu}_k(\ell, n)$ is a realization of the random variable $\hat{\mu}_k^{[1]}$ using the ℓ^{th} dataset of size n. Figure 4.2 (left) shows the error $\bar{E}$ as a function of n while Figure 4.2 (right) compares the normalized expected Sobol' indices of MARS with the normalized exact indices for a representative sample of size $n = 600$.

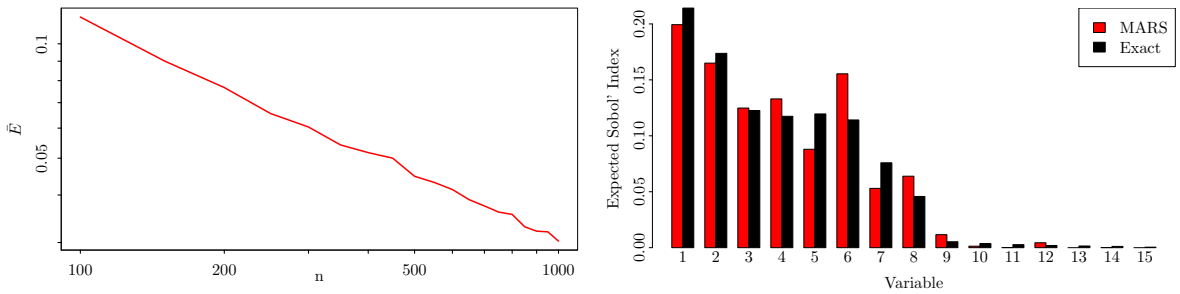

Figure 4.2 Convergence of the expected Sobol' indices for the stochastic g-function (4.7). Left: average error $\bar{E}$ (4.8) in the normalized expectation of the indices as the surrogate sampling size n varies; the empirical convergence rate is 0.59. Right: comparison of normalized expectation of the indices for a representative sample size of $n = 600$.

The effect of m in (4.6) is studied in Figure 4.3, which shows the convergence of $\hat{\mu}_1^{[1]}$ and $\hat{\mu}_3^{[1]}$ as the number of samples m increases. The results in Figure 4.3 are computed using the sample of size $n = 600$ from Figure 4.2 (right). The first and third variables are chosen because they have the largest expectation and variance, respectively. These results confirm both the efficiency

of MARS as a surrogate and the fast convergence of the expectation of the indices with only 200 samples in Ω.

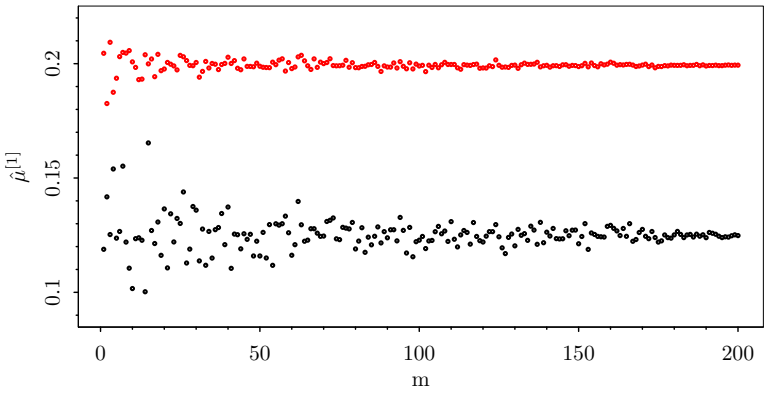

Figure 4.3 Dependency of $\hat{\mu}_1^{[1]}$ (red) and $\hat{\mu}_3^{[1]}$ (black) on the number of samples m in Ω.

Accurate approximations of the distributions of the Sobol' indices can be obtained from sampling their analytical expressions; as above, 10^6 samples are taken from Ω. This provides highly accurate approximations to assess convergence in distribution of the Sobol's indices computed through the proposed method.

For each ω_i, $i = 1, \ldots, 200$, one thousand samples are drawn from $\mathcal{X}$; these are subsampled for $n = 100, 150, \ldots, 950, 1000$ and the resulting histograms are evaluated. Figure 4.4 illustrates convergence in distribution of both S_1, the index with largest expectation, and S_3, the index with largest variance.

For S_1, the histograms appear to converge, see Figure 4.4, top left; however, Figure 4.4, top right, shows that they converge to a distribution that is biased high. This results from the built-in adaptivity of MARS which causes an inherent bias toward the more important variables. This is a useful feature for the purpose of screening unimportant variables. The bottom row of Figure 4.4 illustrates the unbiased convergence in distribution of S_3. Figure 4.5 further illustrates the bias from MARS with a QQ plot of the eight most important variables. The "exact" distribution was generated using 10^6 samples from the analytic expressions of the Sobol' indices. Lying along the line indicates being unbiased; lying below or above the line indicates being biased low or high respectively. This plot demonstrates the general trend that the most important variables are biased high while the less important variables are biased low; S_6 and S_8 fail to follow this trend.

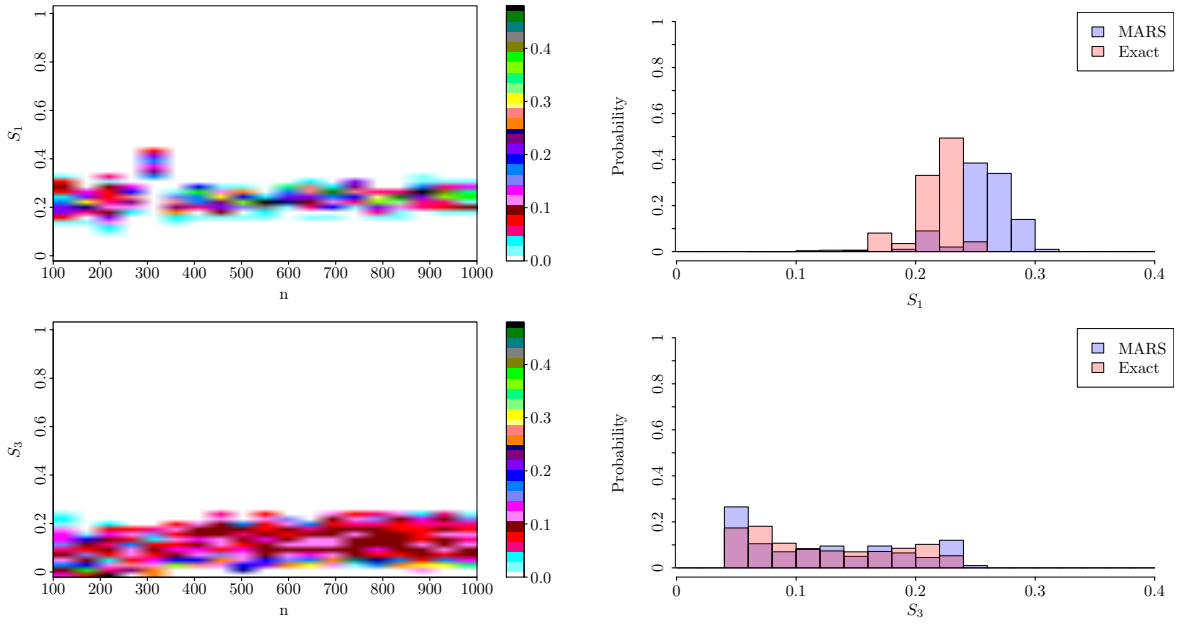

Figure 4.4 Convergence in distribution of the Sobol' indices for the g-function (4.7). Top row; S_1, index with the largest expectation; bottom row: S_3, index with the largest variance. Left: heat map of the histograms as the surrogate sampling size n varies. Each vertical slice is a histogram for a fixed n; right: comparison of the "exact" (see text) and approximation distributions using $n = 1000$.

4.3.2 Genetic oscillator

The proposed stochastic sensitivity analysis method is applied to study a circadian oscillator mechanism from biochemistry. The problem is detailed in [117] and is commonly referred to as the genetic oscillator. It corresponds to a biochemical reaction network consisting of nine species and sixteen reactions. The nine system species are described in Table 4.2.

Table 4.2 Nine species of the genetic oscillator problem from [117].

D_A, D'_A	activator genes
D_R, D'_R	repressor genes
A, R	activator and repressor proteins
M_A, M_R	mRNA of A and R
C	complex species

Denoting the number of molecules of each of the species with the corresponding symbol, the

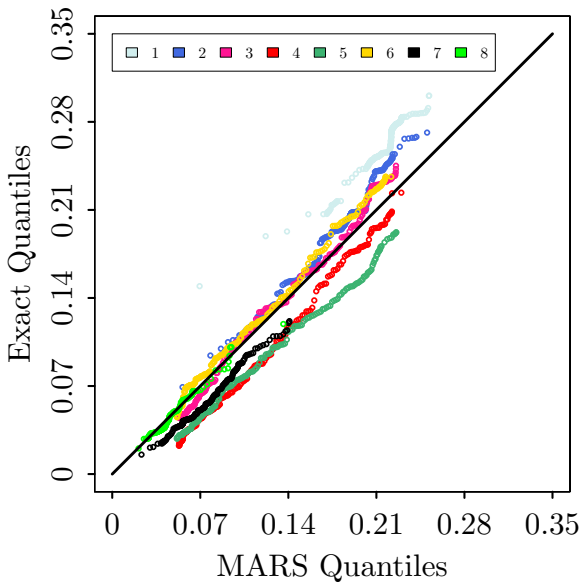

Figure 4.5 QQ plot of the Sobol' indices of the eight most important variables. Lying above or below the line indicates being biased high or low respectively.

state vector of the system is given by

$$\left(\mathrm{D_A}, \mathrm{D'_A}, \mathrm{A}, \mathrm{D'_R}, \mathrm{D_R}, \mathrm{R}, \mathrm{M_A}, \mathrm{M_R}, \mathrm{C}\right) \in \mathbb{R}^9.$$

The initial state is taken as

$$\left(1, 0, 0, 0, 1, 0, 0, 0, 0\right),$$

that is, initially, $\mathrm{D_A} = \mathrm{D_R} = 1$, and all the other variables are set to zero.

The reactions and reaction rates are listed in Table 4.3 which also includes the nominal reaction rates from [117].

Uncertainties in the reaction rate constants are considered; that is, the uncertain variables for the system are given by

$$\mathbf{X} = \left(\gamma_A, \theta_A, \gamma_R, \theta_R, \gamma_C, \alpha_A, \alpha_R, \alpha'_A, \alpha'_R, \beta_A, \beta_R, \delta_{MA}, \delta_A, \delta_{MR}, \delta_R\right) \in \mathbb{R}^{15}.$$

It is assumed that the coordinates of $\mathbf{X}$, i.e., the reaction rates, are iid uniform random variables centered at their respective nominal values given in Table 4.3, and with a 10% perturbation around the mean.

With fixed reaction rates the time evolution of the state vector is stochastic. The sequence of reactions is random with probabilities parameterized by the reaction rates and state vector;

Table 4.3 Reactions and reaction rates for the genetic oscillator system [117].

reaction #	reaction		rate (nominal value)
1	$\mathrm{D_A} + \mathrm{A}$	$\rightarrow \mathrm{D'_A}$	γ_A (1.0)
2	$\mathrm{D'_A}$	$\rightarrow \mathrm{D_A} + \mathrm{A}$	θ_A (50.0)
3	$\mathrm{D_R} + \mathrm{A}$	$\rightarrow \mathrm{D'_R}$	γ_R (1.0)
4	$\mathrm{D'_R}$	$\rightarrow \mathrm{D_R} + \mathrm{A}$	θ_R (100.0)
5	$\mathrm{A} + \mathrm{R}$	$\rightarrow \mathrm{C}$	γ_C (2.0)
6	$\mathrm{D_A}$	$\rightarrow \mathrm{D_A} + \mathrm{M_A}$	α_A (50.0)
7	$\mathrm{D_R}$	$\rightarrow \mathrm{D_R} + \mathrm{M_R}$	α_R (0.01)
8	$\mathrm{D'_A}$	$\rightarrow \mathrm{D'_A} + \mathrm{M_A}$	α'_A (500.0)
9	$\mathrm{D'_R}$	$\rightarrow \mathrm{D'_R} + \mathrm{M_R}$	α'_R (50.0)
10	$\mathrm{M_A}$	$\rightarrow \mathrm{M_A} + \mathrm{A}$	β_A (50.0)
11	$\mathrm{M_R}$	$\rightarrow \mathrm{M_R} + \mathrm{R}$	β_R (5.0)
12	$\mathrm{M_A}$	$\rightarrow \emptyset$	δ_{MA} (10.0)
13	A	$\rightarrow \emptyset$	δ_A (1.0)
14	$\mathrm{M_R}$	$\rightarrow \emptyset$	δ_{MR} (0.5)
15	R	$\rightarrow \emptyset$	δ_R (0.2)
16	C	$\rightarrow \mathrm{R}$	δ_A (same as react. 13)

see e.g., [26]. Realizations of the genetic oscillator are computed through Gillespie's stochastic simulation algorithm (SSA) [26, 36, 37]. Fixing a realization of the inherent stochasticity and sampling the reaction rates requires generating and saving a sequence of random numbers to input to SSA for each sample of the reaction rates. This corresponds to evaluating $f(\mathbf{X}_j, \omega_i)$, for each $j = 1, \ldots, n$, in Algorithm 2; ω_i corresponds to the fixed sequence of random numbers.

The goal is to evaluate the sensitivity of the number of C molecules to the uncertain reaction rates. Letting the probability spaces $(\Theta, \mathcal{F}, \nu)$ and $(\Omega, \mathcal{E}, \lambda)$ carry the input variable uncertainty and intrinsic stochasticity of the system, respectively, and $\mathbf{X} : \Theta \to \mathcal{X}$ model the uncertain variables, note that $\mathrm{C} : \mathcal{X} \times \Omega \times [0, T_{\text{final}}] \to \mathbb{R}$ is a stochastic process. Figure 4.6 illustrates the dynamics of $\mathrm{C} = \mathrm{C}(\mathbf{X}, \omega, t)$ by displaying four typical realizations of the stochastic process when the uncertain variables are fixed at their nominal values. The differing periods and small oscillations are a result of the inherent stochasticity of the system.

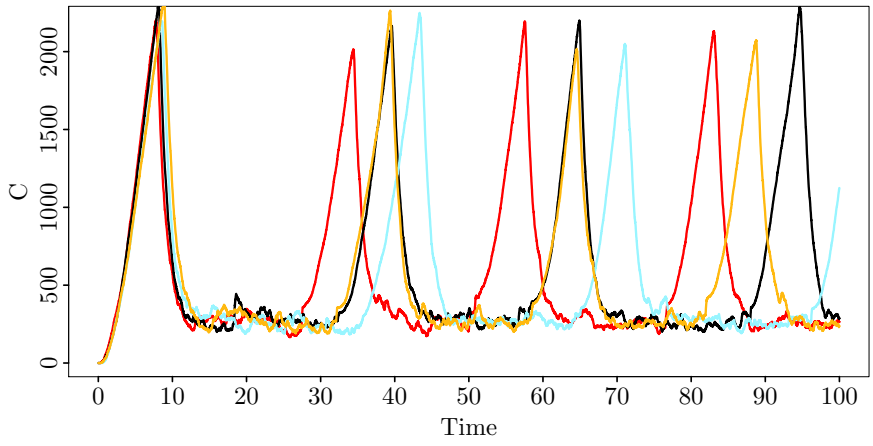

Figure 4.6 Genetic oscillator: four realizations of the evolution of the complex C.

Algorithm 2 is used with $n = 400$, $m = 200$, and a MARS surrogate model constructed point-wise in time. By subsampling $\mathbf{X}$ from existing data, convergence in n is assessed and it is determined that $n = 400$ is an adequate sample size for this application. Moreover, thanks to Theorem 7, a relatively small m is sufficient to ensure a small variance of the estimators of the Sobol' indices.

Figure 4.7, left, shows the time evolution of the expectation of the Sobol' indices; for each index, the expectation becomes periodic after an initial transient. Moreover, observe that the reaction rates β_R and α'_R have the most notable contribution to the model variance during the transient regime. After the transient regime, the degradation rates for the proteins A and R, i.e.,

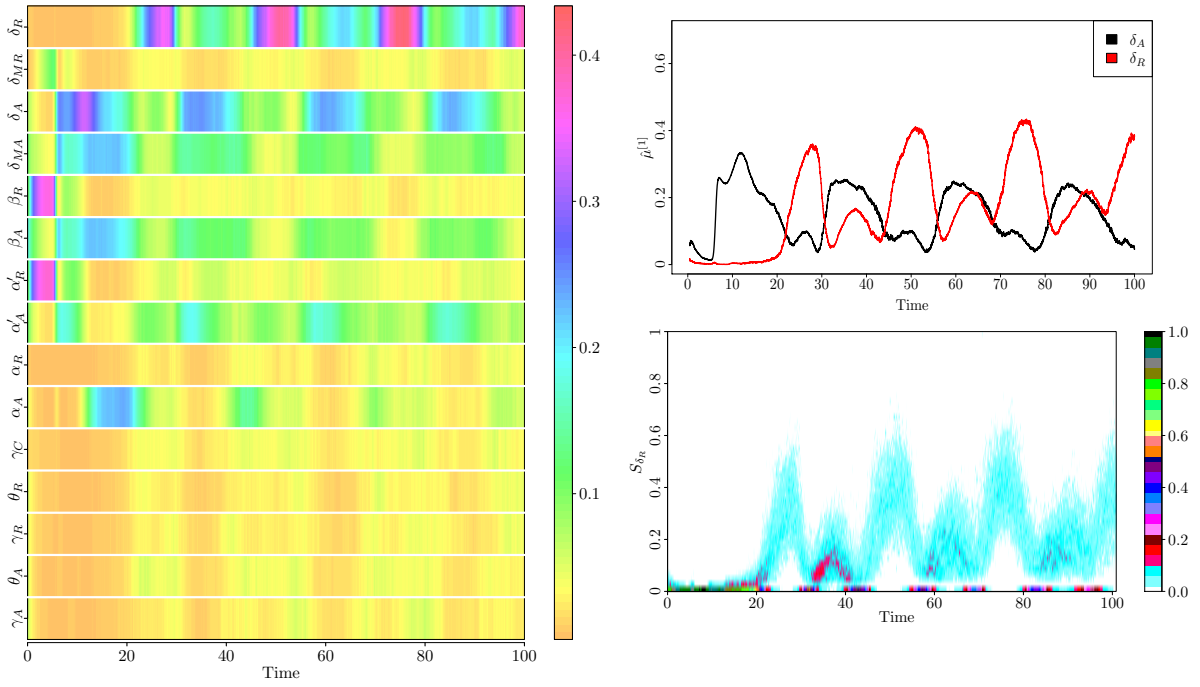

Figure 4.7 Evolution of the Sobol' indices for the genetic oscillator. Left: expectation of the Sobol' indices. Each row corresponds to a specific reaction rate. Top right: time evolution of the expectation of the Sobol' indices for the two most important reaction rates. Bottom right: time evolution of the histogram of the Sobol' index of δ_R. Each vertical slice is a histogram for the Sobol' index of δ_R at a given time.

δ_A and δ_R, are the most important factors. Figure 4.7, top right, displays the time evolution of the expectation of the Sobol' indices for these two reaction rates. Note in particular that δ_A and δ_R periodically swap role of the most dominant contributor to variance of C after the intial transient regime. The time-dependent behavior of the statistical distribution of sensitivity index for δ_R is illustrated in Figure 4.7, bottom right.

The computational cost of the above analysis is $n \times m = 400 \times 200 = 80,000$ SSA simulations. While this is still a significant number of function evaluations, it should be contrasted with the complexity of traditional sampling based method for computing the Sobol' indices, given in (4.3), which, for such a stochastic model, would be orders of magnitudes larger.

4.4 Conclusion

A strategy for GSA of stochastic models has been proposed and investigated. Within this framework, a thorough analysis of variable importance is obtained by computing the statistical properties of the Sobol' indices. The proposed approach requires sampling in the product of

probability spaces carrying the model stochasticity and uncertain variables. The number of samples in the uncertain variable space is driven by the choice of surrogate model construction, and can be controlled via the use of adaptive surrogates such as MARS. Theoretical and numerical evidence indicates that the moments of the indices can be evaluated with only a modest number of samples in the stochastic space.

As mentioned in Section 4.1, one may also consider performing sensitivity analysis on $\mathbb{E}_\omega\left(f(\mathbf{X},\omega)\right)$ directly. However, as illustrated in Example 4.1.1, this approach may result in a significant loss of information. The variance of the Monte Carlo estimator for $E_\omega\left(f(\mathbf{X},\omega)\right)$ is $\mathrm{Var}_\omega\left(f(\mathbf{X},\omega)\right)/m$, which in general is not known a priori. In contrast, the variance of the Monte Carlo estimator for $\mathbb{E}_\omega\left(S_u(\omega)\right)$ in the proposed framework is bounded above by $1/(4m)$ independently of f. This gives the proposed method theoretical and computational advantages.

The numerical results focus on computing first order Sobol' indices and MARS was shown to be an efficient surrogate for this task. Other indices may be computed in the proposed framework as well, provided a sufficiently accurate surrogate model is available. Further study of surrogate models and their properties which facilitate accurate computation of Sobol' indices remains an important question for future work.

CHAPTER

5

DERIVATIVE-BASED GSA FOR STATISTICAL MODEL PARAMETERS

5.1 Introduction

This chapter considers GSA for parameters in statistical models. The content of this chapter is based on the publication [52]. To the author's knowledge, the GSA tools developed for mathematical models and simulations have not been systematically developed for analysis of statistical models. Challenges arise with statistical model parameters which prohibit a direct application of GSA tools to statistical models. Nevertheless, problem structure in statistical models may be exploited to enable efficient GSA. This chapter provides a framework to use existing GSA tools along with methods from statistics to address these challenges and yield a new GSA approach for analysis of statistical models.

This work is motivated by a statistical model that fuses two datasets of atmospheric wind speed in order to provide statistical prediction of wind speed in space and time [4]. The predictions are generated from a Gaussian process whose mean and covariance are parameterized through a large number of parameters, which are determined by numerical optimization. Because of changing weather patterns, the parameters must be re-optimized on a regular basis. The optimization procedure to fit parameters is an important problem feature influencing the approach to GSA. The method is developed in an abstract setting and subsequently used to

analyze the Gaussian process wind speed model.

Recall the assumptions in Chapter 1:

- the inputs are independent,
- the inputs have a known probability distribution,
- the output is scalar-valued.

None of these assumptions hold for statistical models such as the motiving Gaussian process. The parameters will typically have strong dependencies because of the way they appear in the model. A probability distribution for the parameters may be acquired by inversion, but it is not readily available from problem specific information a priori. In general, the model output is a set of random vectors (or a random field). Results from the previous chapters may be applicable to address these challenges; however, they fail to exploit problem structure and hence applying them directly is not a desirable approach.

A derivative-based strategy for GSA of statistical models is proposed. In principle, any of the methods discussed in Chapter 1 may be generalized for use with statistical models. Computational considerations make derivative-based methods preferable. In particular, the number of gradient evaluations is independent of the parameter space dimension and it is not necessary to sample from conditional distributions (which is challenging in the motivating application). Further, gradients are readily available in the motivating application and many other applications arising in statistics.

This chapter provides a framework to connect the mathematical and statistical tools needed to efficiently extend GSA to statistical models. A loss function associated with the statistical model parameter estimation is used to define a joint probability distribution which respects the dependency structure in the problem. This distribution is sampled using a Markov Chain Monte Carlo (MCMC) method, and derivative-based sensitivity indices are computed from these samples. In this framework, both sensitivities and dependency structures are discovered without requiring a priori knowledge about the parameters.

5.2 Preliminaries

Let f be a statistical model defined through parameters $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ which assume values in $\mathcal{X} \subseteq \mathbb{R}^p$. In general, f may be considered a function whose domain is $\mathcal{X}$ and whose range is a set of random vectors (or random fields). Unlike the previous chapters, $\mathbf{X}$ will be referred to as "parameters" instead of "variables" since they correspond to parameters in the statistical model.

Let $\mathcal{L}(.;y) : \mathcal{X} \to \mathbb{R}$ be a loss function (or cost function) associated with f; that is, to fit f to the data y, one computes $\arg\min_{\mathbf{x}\in\mathcal{X}} \mathcal{L}(\mathbf{x};y)$. In the model-oriented context of statistics, parameters of the model are usually estimated by minimizing $\mathcal{L}(\cdot;y)$ computed on observed data y. The loss function is chosen given the model and its intended use. Common examples are least-squares and maximum likelihood. For simplicity, $\mathcal{L}(\cdot)$ will be used instead of $\mathcal{L}(\cdot;y)$ for the remainder of the chapter. In the motivating application, f is a Gaussian process with mean $\mu(\mathbf{X})$ and covariance $\Sigma(\mathbf{X})$ depending on the vector of parameters $\mathbf{X}$ and $\mathcal{L}$ is the the negative log likelihood. Throughout the chapter, it is assumed that $\nabla\mathcal{L}$ may be computed efficiently, as is the case in the motivating application.

The goal is discovering the the sensitivity of f to $\mathbf{X}$. Since f may be a complex mathematical object, the sensitivity of f to $\mathbf{X}$ is analyzed through the sensitivity of $\mathcal{L}$ to $\mathbf{X}$. This makes the analysis dependent on the choice of loss function, which is consistent with a goal oriented choice of $\mathcal{L}$. This is appropriate since $\mathcal{L}$ encodes the dependence of f on $\mathbf{X}$. Further, since $\mathcal{L}$ is a deterministic real-valued function of $\mathbf{X}$, its analysis is more tractable than f.

One challenge is that the statistical model f may be mathematically well defined for parameters $\mathbf{X} \in \mathcal{X}$ but yield a practically irrelevant solution in the context of a given application. To avoid this scenario, let $\mathcal{X}_r \subseteq \mathcal{X}$ be the subset that restricts $\mathcal{X}$ to parameters yielding relevant solutions. For instance, a quantity in f may be required to be nonnegative so $\mathcal{X}_r$ restricts to parameters respecting this constraint. Assume that $\mathcal{X}_r$ is a Lebesgue measurable set; this is easily verified in most applications.

The following three assumptions are made about $\mathcal{L}$:

(I) $\mathcal{L}$ is differentiable.

(II) $\exists\ \delta_{\min} \geq 0$ such that $\int_{\mathcal{X}_r} e^{-\delta_{\min}\mathcal{L}(\mathbf{x})}\mathrm{d}\mathbf{x}$, $\int_{\mathcal{X}_r} |x_k| e^{-\delta_{\min}\mathcal{L}(\mathbf{x})}\mathrm{d}\mathbf{x}$, and $\int_{\mathcal{X}_r} |\frac{\partial\mathcal{L}}{\partial x_k}(\mathbf{x})| e^{-\delta_{\min}\mathcal{L}(\mathbf{x})}\mathrm{d}\mathbf{x}$, $k = 1, 2, \ldots, p$, exist and are finite.

(III) $\exists\ L_{\min} \in \mathbb{R}$ such that $L_{\min} \leq \mathcal{L}(\mathbf{x})$, $\forall \mathbf{x} \in \mathcal{X}_r$.

Assumption 1 is necessary since a derivative-based GSA is used. This assumption is easily verifiable in most cases. Assumption 2 is needed so that global sensitivity indices (5.3) are well defined. Assumption 3 is a needed technical assumption requiring that the loss function be bounded below. Note that if $\mathcal{L}$ is continuously differentiable and $\mathcal{X}_r$ is compact, then all three assumptions follow immediately.

Thus far, a probability distribution has not been specified for $\mathbf{X}$. GSA may not be performed until a distribution is specified for $\mathbf{X}$. To this end, define

$$\eta(\mathbf{x}) = \begin{cases} e^{-\delta(\mathcal{L}(\mathbf{x})+\lambda||\mathbf{x}||_2^2)} & \mathbf{x} \in \mathcal{X}_r \\ 0 & \mathbf{x} \notin \mathcal{X}_r \end{cases} \tag{5.1}$$

for some $\delta \geq \delta_{\min}$ and $\lambda \geq 0$; $||\cdot||_2$ is the Euclidean norm. Note that $\mathcal{X}_r$ is defined through constraints on f so it is generally difficult to express $\mathcal{X}_r$ in terms of simple algebraic constraints. In most cases, however, the constraints may be checked when $\mathcal{L}$ is evaluated, and hence η is easily evaluated through evaluating $\mathcal{L}$.

Using Assumption 2 and the fact that $e^{-\delta\lambda||\mathbf{x}||_2^2} \leq 1$, it follows that η is integrable. Define the PDF

$$\phi(\mathbf{x}) = \frac{\eta(\mathbf{x})}{\int_{\mathcal{X}} \eta(\tilde{\mathbf{x}})\mathrm{d}\tilde{\mathbf{x}}} = \frac{e^{-\delta(\mathcal{L}(\mathbf{x})+\lambda||\mathbf{x}||_2^2)}}{\int_{\mathcal{X}_r} e^{-\delta(\mathcal{L}(\tilde{\mathbf{x}})+\lambda||\tilde{\mathbf{x}}||_2^2)}\mathrm{d}\tilde{\mathbf{x}}}. \tag{5.2}$$

Then ϕ is supported on $\mathcal{X}_r$ and gives the greatest probability to regions where $\mathcal{L}$ is close to its minimum namely, where $\mathbf{x}$ is a good fit. A PDF of this form corresponds to a Gibbs measure [64] with temperature δ; the temperature determines how the probability mass disperses from the modes. The scalar $\lambda \geq 0$ is a regularization factor that aids when ϕ is too heavy tailed; this is illustrated in Section 5.4. The determination of δ and λ is considered in Section 5.3. This formulation shares similar characteristics to Bayesian inference. For instance, if $\mathcal{L}$ is a negative log likelihood and $\delta = 1$ then (5.2) is the posterior PDF of $\mathbf{X}$ using a Gaussian prior truncated to $\mathcal{X}_r$ (or when $\lambda = 0$, the prior is simply a uniform distribution on $\mathcal{X}_r$).

Definition 4. *Define the sensitivity index of f with respect to X_k as*

$$\mathfrak{S}_k = \mathbb{E}\left(|X_k|\right)\mathbb{E}\left(\left|\frac{\partial \mathcal{L}}{\partial x_k}(\mathbf{X})\right|\right) = \int_{\mathcal{X}_r} |x_k|\phi(\mathbf{x})\mathrm{d}\mathbf{x}\int_{\mathcal{X}_r}\left|\frac{\partial \mathcal{L}}{\partial x_k}(\mathbf{x})\right|\phi(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{5.3}$$

Unlike (1.7), the sensitivity index (5.3) takes the absolute value instead of squaring the partial derivative. This is to prevent giving too much weight to outliers. The partial derivative depends on the units of X_k, which creates challenges comparing sensitivities for parameters on different scales. Multiplying by $\mathbb{E}\left(|X_k|\right)$ seeks to relieve this problem, yielding a sensitivity index which only depends on the units of $\mathcal{L}$.

Dependencies in $\mathbf{X}$ make Monte Carlo integration with uniform sampling intractable for computing the $\mathfrak{S}_k$'s. Importance sampling may be used if an efficient proposal distribution is found; however, this is also challenging in most cases. Therefore, the $\mathfrak{S}_k$'s will be estimated with MCMC methods.

In summary, the global sensitivity of f to $\mathbf{X}$ may be estimated by using only evaluations of $\mathcal{L}$ and $\nabla\mathcal{L}$ along with MCMC. This framework also admits additional useful information as by-products of estimating (5.3). More details are given in Section 5.3.

5.3 Computing sensitivities

This section presents the main algorithmic developments of this chapter. The proposed method may be partitioned into three stages:

(i) preprocessing where information is collected about the loss function,

(ii) sampling where samples are drawn from the probability measure (5.2),

(iii) post-processing where sensitivities, as well as additional information, are computed.

In the preprocessing stage, characteristic values for the parameters and the loss function are found. These characteristic values are used to determine the temperature and regularization factor in the PDF ϕ (5.2).

In the sampling stage, the temperature and regularization factor are determined. Subsequently an MCMC sampler is run to collect samples from (5.2).

In the post-processing stage, sensitivities are computed by evaluating the gradient of the loss function at the samples drawn in the sampling stage ii. In addition, the parameter dependencies and the robustness of the sensitivities with respect to perturbations in the temperature are extracted from the existing samples and gradient evaluations. These two pieces of information are by-products of computing sensitivities and require no additional computation. The robustness of the sensitivities has similarities to the robustness analysis in Chapter 3; the results of this chapter are specific to the form of the proposed PDF.

These three stages are described in Subsections 5.3.1, 5.3.2, and 5.3.3, respectively. The method is summarized as a whole in Subsection 5.3.4.

5.3.1 Preprocessing stage

Characteristic magnitudes for $\mathbf{X}$ and $\mathcal{L}$ are needed to determine the regularization factor and temperature. To this end, two auxiliary computations are introduced as a preprocessing step.

The first auxiliary computation runs an optimization routine to minimize $\mathcal{L}$; the choice of optimizer is not essential here. Let $\mathbf{x}^\star$ be the minimizing parameter vector. It is acceptable if $\mathbf{x}^\star$ is not the global minimizer of $\mathcal{L}$ as long as it is sufficiently close to capture characteristic magnitudes of $\mathcal{L}$ in regions of good fit.

The second auxiliary computation uses $\mathbf{x}^\star$ to determine the range of loss function values that the MCMC sampler should explore. Let $c > 0$, and let $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$ be a random vector defined by

$$Z_k \sim \mathcal{U}[(1-c)x_k^\star, (1+c)x_k^\star], \tag{5.4}$$

where all the Z_k's are independent of one another and $\mathcal{U}$ denotes the uniform distribution. Hence, $\mathbf{Z}$ represents uniform uncertainty of $c\%$ about $\mathbf{x}^\star$.

Determining c is an application-dependent problem. In fact, its determination is the only portion of the proposed method that cannot be automated. To choose c it is suggested to,

1. fix a value for c,
2. sample from $\mathbf{Z}$,
3. asses the quality of f's predictions using the sample.

Repeating this sample and assessment process for various values of c allow the user to determine a c that yields reasonable predictions. This step is highly subjective and application dependent; however, it is a very natural means of inserting user specification. One simple way to do this is visualizing the model output for each sample and increasing c until the outputs become unrealistic.

Taking large values for c will result in the PDF ϕ giving significant probability to regions of the parameter space yielding poor fits, and hence sensitivity indices that are not useful. Taking small values for c will result in the PDF ϕ giving significant probability to regions of the parameter space near local minima, thus making the sensitivity indices local. Since the choice of c is strongly user dependent, the robustness of the sensitivity indices with respect to perturbations in c is highly relevant; this is indirectly addressed by Theorem 11.

Once c is specified, then a threshold M, which is used to compute the regularization factor and temperature (see Sub-subsection 5.3.2.1 and Sub-subsection 5.3.2.2), may be easily computed via Monte Carlo integration. Define the threshold

$$M = \mathbb{E}(\mathcal{L}(\mathbf{Z})). \tag{5.5}$$

Note that the expectation in (5.5) is computed with respect to the independent uniform measure; all other expectations in the chapter are computed with respect to the PDF ϕ (5.2).

5.3.2 Sampling stage

An MCMC method is used to sample from ϕ (5.2) through evaluations of the unnormalized density η (5.1). Then the $\mathfrak{S}_k$'s may be computed through evaluations of $\nabla\mathcal{L}$ at the sample points. Many MCMC methods may be used to sample ϕ; see, for example [33, 35, 47, 99, 106, 116].

Determining which MCMC method to use and when it has converged may be challenging. Convergence diagnostics [14, 21] have been developed which may identify when the chain has not

converged; however, they all have limitations and cannot ensure convergence [34]. In Section 5.4, adaptive MCMC [116] is used with the convergence diagnostic from [13].

Assuming that an MCMC sampler is specified, Sub-subsections 5.3.2.1 and 5.3.2.2 focus on determining the temperature and regularization factors, respectively.

5.3.2.1 Determining the regularization factor

To determine the regularization factor λ, consider the function

$$\mathcal{L}_\lambda(\mathbf{x}) = \mathcal{L}(\mathbf{x}) + \lambda||\mathbf{x}||_2^2.$$

The PDF ϕ gives greatest probability to regions where $\mathcal{L}_\lambda$ is small. If $\lambda||\mathbf{x}||_2^2$ is small relative to $\mathcal{L}(\mathbf{x})$, then the local minima of $\mathcal{L}_\lambda$ are near the local minima of $\mathcal{L}$. Ideally $\lambda = 0$ will be used, but in some cases this results in ϕ being too heavy tailed. Instead $\lambda||\mathbf{x}||_2^2 \approx \nu\mathcal{L}_\lambda(\mathbf{x})$ may be taken for some $\nu \in (0, 1)$; that is, the regularization term contributes ν percent of the value of $\mathcal{L}_\lambda$. Setting $\lambda||\mathbf{x}||_2^2 = \nu\mathcal{L}_\lambda(\mathbf{x})$ and replacing $\mathcal{L}(\mathbf{x})$ and $\mathbf{x}$ with M and $\mathbf{x}^\star$, gives

$$\lambda = \frac{\nu M}{(1-\nu)||\mathbf{x}^\star||_2^2}. \tag{5.6}$$

In practice, it is suggested to begin with $\nu = 0$. If the MCMC sampler yields heavy-tailed distributions that converge slowly, then ν may be increased to aid the convergence. This case is illustrated in Section 5.4.

5.3.2.2 Determining the temperature

To determine the temperature δ, define

$$M_\lambda = M + \lambda||\mathbf{x}^\star||_2^2.$$

It is desirable to find δ so that $\mathcal{L}_\lambda(\mathbf{x}) \leq M_\lambda$ with probability α; $\alpha = .99$ is suggested to mitigate wasted computation in regions where $\mathbf{x}$ yields a poor fit. Let $A = \{\mathbf{x} \in \mathcal{X}_r | \mathcal{L}_\lambda(\mathbf{x}) \leq M_\lambda\}$. Note that A is a Lebesgue measurable set since $\mathcal{L}_\lambda$ is continuous and $\mathcal{X}_r$ is Lebesgue measurable. Define the function $\Delta : [\delta_{\min}, \infty) \to [0, 1]$ by

$$\Delta(\delta) = \int_A \phi(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{5.7}$$

Then $\Delta(\delta)$ gives the probability that $\mathcal{L}_\lambda(\mathbf{x}) \leq M_\lambda$. The optimal temperature δ is the solution of $\Delta(\delta) = \alpha$. Four results are given below showing that Δ possesses advantageous properties making the nonlinear equation $\Delta(\delta) = \alpha$ easily solvable.

Theorem 8. *If* $\int_{\mathcal{X}_r} \mathcal{L}_\lambda(\mathbf{x}) e^{-\delta_{min}\mathcal{L}_\lambda(\mathbf{x})} \mathrm{d}\mathbf{x} < \infty$*, then* Δ *is differentiable on* (δ_{min}, ∞) *with*

$$\Delta'(\delta) = (-1 + \Delta(\delta)) \int_A \mathcal{L}_\lambda(\mathbf{x})\phi(\mathbf{x})\mathrm{d}\mathbf{x} + \Delta(\delta) \int_{\mathcal{X}_r \setminus A} \mathcal{L}_\lambda(\mathbf{x})\phi(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{5.8}$$

Proof. Let $U(\delta) = e^{-\delta(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})}$ and $V(\delta) = \int_{\mathcal{X}_r} e^{-\delta(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})}\mathrm{d}\mathbf{x}$. Then

$$\Delta(\delta) = \frac{1}{V(\delta)} \int_A U(\delta)\mathrm{d}\mathbf{x}\,.$$

Using Theorem 6.28 in [65] with $(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})e^{-\delta_{\min}(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})}$ dominating $U'(\delta)$, gives that $\int_A U(\delta)\mathrm{d}\mathbf{x}$ and V are differentiable with

$$\begin{aligned} \frac{d}{d\delta} \int_A U(\delta)\mathrm{d}\mathbf{x} &= -\int_A (\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})U(\delta)\mathrm{d}\mathbf{x}\,, \\ V'(\delta) &= -\int_{\mathcal{X}_r} (\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})e^{-\delta(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})}\mathrm{d}\mathbf{x}\,. \end{aligned}$$

Δ is differentiable since $V(\delta) > 0$, $\forall \delta > \delta_{\min}$, and applying the quotient rule for derivatives gives

$$\Delta'(\delta) = \frac{\frac{\mathrm{d}}{\mathrm{d}\delta}\int_A U(\delta)\mathrm{d}\mathbf{x}}{V(\delta)} - \frac{V'(\delta)}{V(\delta)} \frac{\int_A U(\delta)\mathrm{d}\mathbf{x}}{V(\delta)}\,.$$

Simple manipulations yields

$$\Delta'(\delta) = (-1 + \Delta(\delta)) \int_A (\mathcal{L}_\lambda(\mathbf{x}) - L_{\min}) \frac{U(\delta)}{V(\delta)}\mathrm{d}\mathbf{x} + \Delta(\delta) \int_{\mathcal{X}_r \setminus A} (\mathcal{L}_\lambda(\mathbf{x}) - L_{\min}) \frac{U(\delta)}{V(\delta)}\mathrm{d}\mathbf{x}\,.$$

Writing $\frac{U(\delta)}{V(\delta)} = \phi(\mathbf{x})$ and using the linearity of the integral completes the proof. □

Theorem 9. *If* $\int_{\mathcal{X}_r} \mathcal{L}_\lambda(\mathbf{x}) e^{-\delta_{min}\mathcal{L}_\lambda(\mathbf{x})} \mathrm{d}\mathbf{x} < \infty$*, then* Δ *is a strictly increasing function on* (δ_{min}, ∞)*.*

Proof. The proof of Theorem 8 gives

$$\Delta'(\delta) = (-1 + \Delta(\delta)) \int_A (\mathcal{L}_\lambda(\mathbf{x}) - L_{\min}) \frac{U(\delta)}{V(\delta)}\mathrm{d}\mathbf{x} + \Delta(\delta) \int_{\mathcal{X}_r \setminus A} (\mathcal{L}_\lambda(\mathbf{x}) - L_{\min}) \frac{U(\delta)}{V(\delta)}\mathrm{d}\mathbf{x}.$$

Since

$$\begin{aligned}
&\mathbf{x} \in A \implies \mathcal{L}_\lambda(\mathbf{x}) \leq M_\lambda, \\
&\mathbf{x} \in \mathcal{X}_r \setminus A \implies \mathcal{L}_\lambda(\mathbf{x}) > M_\lambda, \\
&\Delta(\delta) \in [0,1], \\
&\int_{\mathcal{X}_r \setminus A} \phi(\mathbf{x}) \mathrm{d}\mathbf{x} = 1 - \Delta(\delta),
\end{aligned}$$

so

$$\begin{aligned}
\Delta'(\delta) &> (-1 + \Delta(\delta)) \int_A (M_\lambda - L_{\min}) \frac{U(\delta)}{V(\delta)} \mathrm{d}\mathbf{x} + \Delta(\delta) \int_{\mathcal{X}_r \setminus A} (M_\lambda - L_{\min}) \frac{U(\delta)}{V(\delta)} \mathrm{d}\mathbf{x} \\
&= (-1 + \Delta(\delta))(M_\lambda - L_{\min})\Delta(\delta) + \Delta(\delta)(M_\lambda - L_{\min})(1 - \Delta(\delta)) \\
&= 0.
\end{aligned}$$

□

Theorems 8 and 9 yield desirable properties of Δ. The assumption that $\mathcal{L}_\lambda(\mathbf{x}) e^{-\delta_{\min} \mathcal{L}_\lambda(\mathbf{x})}$ is integrable is necessary for $\Delta'(\mathbf{x})$ to be well defined. Note that this assumption follows from Assumption 1 when $\mathcal{X}_r$ is bounded. Theorem 10 and Corollary 2 below give existence and uniqueness, respectively, for the solution of $\Delta(\delta) = \alpha$ under mild assumptions.

Theorem 10. *If $\mathcal{X}_r$ is a bounded set and $\exists \mathbf{x}' \in \mathcal{X}_r$ such that $\mathcal{L}_\lambda(\mathbf{x}') < M_\lambda$, then $\forall \alpha \in (0,1)$ $\exists \delta > \delta_{min}$ such that $\Delta(\delta) > \alpha$.*

Proof. Let $\alpha \in (0,1)$. Define

$$\begin{aligned}
\tilde{M}_\lambda &= \frac{\mathcal{L}(\mathbf{x}') + M_\lambda}{2} < M_\lambda, \\
\tilde{A} &= \{\mathbf{x} \in \mathcal{X}_r | \mathcal{L}_\lambda(\mathbf{x}) \leq \tilde{M}_\lambda\}.
\end{aligned}$$

By Assumption 1, $\mathcal{L}_\lambda$ is continuous so $Vol(\tilde{A}) > 0$. Then $\exists \delta > \delta_{\min}$ such that

$$\frac{Vol(\mathcal{X}_r \setminus A)}{(1-\alpha) Vol(\tilde{A})} < e^{(M_\lambda - \tilde{M}_\lambda)\delta}.$$

Since

$$\frac{\int_A e^{-\delta(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})} \mathrm{d}\mathbf{x}}{\int_{\mathcal{X}_r} e^{-\delta(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})} \mathrm{d}\mathbf{x}} > \alpha \iff (1-\alpha) \int_A e^{-\delta(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})} \mathrm{d}\mathbf{x} > \alpha \int_{\mathcal{X}_r \setminus A} e^{-\delta(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})} \mathrm{d}\mathbf{x},$$

it is enough to show

$$\int_{\mathcal{X}_r \setminus A} e^{-\delta(\mathcal{L}_\lambda(\mathbf{x})-L_{\min})} \mathrm{d}\mathbf{x} < (1-\alpha) \int_A e^{-\delta(\mathcal{L}_\lambda(\mathbf{x})-L_{\min})} \mathrm{d}\mathbf{x}.$$

If $\mathbf{x} \in \mathcal{X}_r \setminus A$, then

$$\mathcal{L}_\lambda(\mathbf{x}) > M_\lambda \implies e^{-\delta(\mathcal{L}_\lambda(\mathbf{x})-L_{\min})} \leq e^{-\delta(M_\lambda - L_{\min})};$$

hence it is enough to show that

$$e^{-\delta(M_\lambda - L_{\min})} Vol(\mathcal{X}_r \setminus A) < (1-\alpha) \int_A e^{-\delta(\mathcal{L}_\lambda(\mathbf{x})-L_{\min})} \mathrm{d}\mathbf{x}.$$

Since the exponential is nonnegative and $\alpha \in (0,1)$, it is equivalent to show

$$\frac{Vol(\mathcal{X}_r \setminus A)}{1-\alpha} < \int_{\tilde{A}} e^{-\delta(\mathcal{L}_\lambda(\mathbf{x})-M_\lambda)} \mathrm{d}\mathbf{x}.$$

If $\mathbf{x} \in \tilde{A}$, then

$$\mathcal{L}_\lambda(\mathbf{x}) \leq \tilde{M}_\lambda \implies \int_{\tilde{A}} e^{-\delta(\mathcal{L}_\lambda(\mathbf{x})-M_\lambda)} \mathrm{d}\mathbf{x} \geq \int_{\tilde{A}} e^{-\delta(\tilde{M}_\lambda - M_\lambda)} \mathbf{x} = e^{-\delta(\tilde{M}_\lambda - M_\lambda)} Vol(\tilde{A}).$$

But

$$\frac{Vol(\mathcal{X}_r \setminus A)}{(1-\alpha) Vol(\tilde{A})} < e^{(M_\lambda - \tilde{M}_\lambda)\delta}$$

by the construction of δ. □

Corollary 2. *If $\alpha \in (\Delta(\delta_{min}), 1)$, $\mathcal{X}_r$ is a bounded set and there exists $\mathbf{x}' \in \mathcal{X}_r$ such that $\mathcal{L}_\lambda(\mathbf{x}') < M_\lambda$, then $\Delta(\delta) = \alpha$ admits a unique solution.*

Proof. Since $\mathcal{X}_r$ is bounded, $e^{-\delta \mathcal{L}_\lambda(\mathbf{x})} \in L^1(\mathcal{X}_r)\ \forall \delta \in \mathbb{R}$. Mimicking the argument of Theorem 8, Δ is differentiable and hence continuous at $\delta_{\min}$. Since $\alpha \in (0,1)$, Theorem 10 gives $\exists \delta > \delta_{\min}$ such that $\Delta(\delta_{\min}) < \alpha < \Delta(\delta)$. Then existence holds by the intermediate value theorem. Uniqueness follows from Theorem 9.

□

The assumption that $\mathcal{X}_r$ is bounded is reasonable in most applications; $\mathcal{X}$ may be unbounded, but $\mathcal{X}_r$ is restricted to physically relevant solutions so it is typically bounded. The assumption that $\mathcal{L}_\lambda(\mathbf{x}') < M_\lambda$ means that M_λ is not chosen as the global minimum, which should always hold in practice. The assumption that $\alpha \in (\Delta(\delta_{\min}), 1)$ is necessary for existence. Typically

$\Delta(\delta_{\min})$ is much less than 1, while α is chosen close to 1. The assumptions Theorem 10 and Corollary 2 hold in most applications

In summary, under mild assumptions, $\Delta(\delta) = \alpha$ is a scalar nonlinear equation admitting a unique solution and Δ possesses nice properties (monotonicity and differentiability). Further, $\Delta(\delta)$ and $\Delta'(\delta)$ may be approximated simultaneously by running MCMC. The challenge is that evaluating $\Delta(\delta)$ and $\Delta'(\delta)$ with high precision requires running a long MCMC chain. In fact, $\Delta'(\delta)$ is significantly more challenging to evaluate than $\Delta(\delta)$. For this reason, using a derivative-free nonlinear solver is suggested, it will still be efficient since Δ is a well-behaved function. In the spirit of inexact Newton methods [62], shorter chains may be run for the early iterations solving $\Delta(\delta) = \alpha$ and the precision increased near the solution. In practice, relatively few evaluations of Δ are needed because of its properties, shown above.

As previously highlighted, the PDF (5.2) corresponds to a Bayesian posterior PDF when $\mathcal{L}$ is a negative log likelihood and $\delta = 1$. If $\delta < 1$, the GSA approach uses a "flatter" PDF than the Bayesian posterior. The determination of δ incorporates information from the parameter optimization procedure to ensure that the sensitivity analysis searches the parameter space in which the optimization routine traverses.

5.3.3 post-processing stage

Having attained samples of $\mathbf{X}$, the sensitivities (5.3) may be estimated by evaluating $\nabla\mathcal{L}$ at the sample points and forming the Monte Carlo estimator for the expectations in (5.3). In addition to computing these sensitivities, two other useful pieces of information may be extracted, namely, the robustness of the sensitivities with respect to perturbations in the temperature and the parameter dependencies. These are described in Sub-subsections 5.3.3.1 and 5.3.3.2, respectively.

5.3.3.1 Robustness with respect to the temperature

As a result of the uncertainty in the determination of δ (computation of $\mathbf{x}^\star$, choice of c, estimation of M, solution of $\Delta(\delta) = \alpha$), the robustness of the sensitivities with respect to δ are analyzed. Consider the functions

$$\begin{aligned} F_k : (\delta_{\min}, \infty) &\to \mathbb{R}, \\ \delta &\mapsto \left(\int_{\mathcal{X}_r} |x_k| \left(\frac{e^{-\delta\mathcal{L}_\lambda(\mathbf{x})}}{\int_{\mathcal{X}_r} e^{-\delta\mathcal{L}_\lambda(\tilde{\mathbf{x}})} d\tilde{\mathbf{x}}} \right) \mathrm{d}\mathbf{x} \right) \left(\int_{\mathcal{X}_r} \left| \frac{\partial\mathcal{L}}{\partial x_k}(\mathbf{x}) \right| \left(\frac{e^{-\delta\mathcal{L}_\lambda(\mathbf{x})}}{\int_{\mathcal{X}_r} e^{-\delta\mathcal{L}_\lambda(\tilde{\mathbf{x}})} d\tilde{\mathbf{x}}} \right) \mathrm{d}\mathbf{x} \right) \end{aligned}$$

$k = 1, 2, \ldots, p$; clearly $F_k(\delta) = \mathfrak{S}_k$. Theorem 11 gives the derivative of the sensitivity index with respect to the temperature δ, namely, $F_k'(\delta)$.

Theorem 11. *If* $\int_{\mathcal{X}_r} \mathcal{L}_\lambda(\mathbf{x}) e^{-\delta_{min}\mathcal{L}_\lambda(\mathbf{x})}\mathrm{d}\mathbf{x}$, $\int_{\mathcal{X}_r} \mathcal{L}_\lambda(\mathbf{x})|x_k| e^{-\delta_{min}\mathcal{L}_\lambda(\mathbf{x})}\mathrm{d}\mathbf{x}$, *and* $\int_{\mathcal{X}_r} \mathcal{L}_\lambda(\mathbf{x})|\frac{\partial \mathcal{L}}{\partial x_k}(\mathbf{x})| e^{-\delta_{min}\mathcal{L}_\lambda(\mathbf{x})}\mathrm{d}\mathbf{x}$ *exist and are finite, then* F_k *is differentiable with*

$$F_k'(\delta) = -\operatorname{Cov}(|X_k|, \mathcal{L}_\lambda(\mathbf{X}))\mathbb{E}\left(\left|\frac{\partial \mathcal{L}}{\partial x_k}(\mathbf{X})\right|\right) - \mathbb{E}\left(|X_k|\right)\operatorname{Cov}\left(\left|\frac{\partial \mathcal{L}}{\partial x_k}(\mathbf{X})\right|, \mathcal{L}_\lambda(\mathbf{X})\right), \tag{5.9}$$

where $\operatorname{Cov}(\cdot,\cdot)$ *is the covariance operator.*

Proof. Let

$$U(\delta) = \int_{\mathcal{X}_r} |x_k| e^{-\delta(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})}\mathrm{d}\mathbf{x}$$

$$V(\delta) = \int_{\mathcal{X}_r} \left|\frac{\partial \mathcal{L}}{\partial x_k}(\mathbf{x})\right| e^{-\delta(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})}\mathrm{d}\mathbf{x}$$

and

$$W(\delta) = \int_{\mathcal{X}_r} e^{-\delta(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})}\mathrm{d}\mathbf{x}.$$

Then

$$F_k(\delta) = \frac{U(\delta)}{W(\delta)}\frac{V(\delta)}{W(\delta)}.$$

Note that $W(\delta) > 0\ \forall \delta > 0$, so it is enough to show that U, V, and W are differentiable. Theorem 6.28 in [65] gives the result using

$$|x_k|\left(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min}\right)e^{-\delta_{\min}(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})},$$

$$\left|\frac{\partial \mathcal{L}}{\partial x_k}(\mathbf{x})\right|\left(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min}\right)e^{-\delta_{\min}(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})},$$

and

$$\left(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min}\right)e^{-\delta_{\min}(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})}$$

to dominate the derivatives of the integrands of U, V, and W, respectively. Applying Theorem 6.28 in [65] to $U(\delta)$, $V(\delta)$, and $W(\delta)$ yields

$$U'(\delta) = -\int_{\mathcal{X}_r} |x_k|\left(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min}\right)e^{-\delta(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})}\mathrm{d}\mathbf{x}$$

$$V'(\delta) = -\int_{\mathcal{X}_r} \left| \frac{\partial \mathcal{L}}{\partial x_k}(\mathbf{x}) \right| (\mathcal{L}_\lambda(\mathbf{x}) - L_{\min}) e^{-\delta(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})} \mathrm{d}\mathbf{x}$$

and

$$W'(\delta) = -\int_{\mathcal{X}_r} (\mathcal{L}_\lambda(\mathbf{x}) - L_{\min}) e^{-\delta(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})} \mathrm{d}\mathbf{x}.$$

An application of the product and quotient rules to F_k yields

$$F_k'(\delta) = \left(\frac{U'(\delta)}{W(\delta)} - \frac{W'(\delta)}{W(\delta)} \frac{U(\delta)}{W(\delta)} \right) \frac{V(\delta)}{W(\delta)} + \left(\frac{V'(\delta)}{W(\delta)} - \frac{W'(\delta)}{W(\delta)} \frac{V(\delta)}{W(\delta)} \right) \frac{U(\delta)}{W(\delta)}.$$

Basic algebra along with the fact that $\frac{e^{-\delta(\mathcal{L}_\lambda(\mathbf{x}) - L_{\min})}}{W(\delta)} = \phi(\mathbf{x})$ yields the result. □

Theorem 11 allows $F_k'(\delta)$ to be computed from the samples and function evaluations used to compute $\mathfrak{S}_k$. For small h, $F_k(\delta + \delta h) \approx F_k(\delta) + h\delta F_k'(\delta)$ so the robustness of $\mathfrak{S}_k$ may be estimated without any further computational expense.

Since the magnitude of $\mathfrak{S}_k$ may depend on δ, it is useful to normalize for each h when assessing robustness. Define

$$\hat{F}_k(\delta + \delta h) = \frac{F_k(\delta) + h\delta F_k'(\delta)}{\sum_{j=1}^{p} (F_j(\delta) + h\delta F_j'(\delta))}, \quad k = 1, 2, \ldots, p, \tag{5.10}$$

which may be plotted for $h \in (-h_{max}, h_{max})$ to assess robustness. Since this is only a local estimate it is suggested to take $h_{max} = \frac{1}{10}$, reflecting a 10% uncertainty about δ.

5.3.3.2 Extracting parameter dependencies

Parameters are typically dependent, and the dependency information is a valuable complement to the sensitivity indices. For instance, if f is sensitive to two parameters that are highly dependent, then it may be possible to remove one of them from f since the other may compensate. In addition, the dependencies may reveal parameter misspecifications in f.

The strength and nature of the dependencies in $\mathbf{X}$ are typically not known a priori. Correlation coefficients may be computed from the MCMC samples and returned as a by-product of computing sensitivity indices. The Pearson correlation coefficient is commonly used to measure dependencies from sampled data. The remainder of the chapter will focus on correlations measured through the Pearson correlation coefficient. Other measures of dependency may be interchanged within the framework as well.

5.3.4 Summary of the method

This subsection summarizes the proposed method. The method is divided into three algorithms, one for each stage described in Section 5.3.

Algorithm 3 performs the auxiliary computations of Subsection 5.3.1. Note that determining c in Line 2 is the only application-specific portion of the proposed method; user discernment is necessary to choose c.

Algorithm 4 requires the user to specify the parameter ν from Sub-subsection 5.3.2.1, the parameter α from Sub-subsection 5.3.2.2, and the number of MCMC samples N. It is suggested to start with $\nu = 0$ and rerunning Algorithm 4 with a larger ν if the convergence results indicate that the PDF is heavy tailed. Hence ν may be viewed as a computed quantity rather than one specified by the user. As mentioned in Sub-subsection 5.3.2.2, $\alpha = .99$ is suggested. It may be considered fixed and the user only needs to change it if they have a specific purpose which requires giving more weight to "poor" parameter choices. The choice of N may be difficult; however, more samples may be appended after an initial run so N can be adapted without any wasted computation.

Algorithm 5 is a simple post-processing of the MCMC samples to compute sensitivity indices, robustness estimates, and parameter dependencies. One may also perform convergence diagnostics on the MCMC estimators of $\mathbb{E}\left(\left|\frac{\partial \mathcal{L}}{\partial x_k}(\mathbf{X})\right|\right)$, $k = 1, 2, \ldots, p$, along with Algorithm 5.

Algorithm 3 Auxiliary Computation

1: compute $\mathbf{x}^\star = \arg\min \mathcal{L}(\mathbf{x})$ via some optimization routine
2: determine $c > 0$ through visualization of model outputs, see (5.4)
3: estimate M via Monte Carlo integration, see (5.5)

Algorithm 4 Sampling

1: **function** (ν, N, α)
2: compute λ using (5.6)
3: solve $\Delta(\delta) = \alpha$, see (5.7)
4: run MCMC sampler to draw N samples from ϕ (5.2)
5: store MCMC samples in a matrix S
6: test convergence of the sampler
7: **end function**

Algorithm 5 Sensitivities, Perturbations, and Correlations

1: evaluate $\nabla\mathcal{L}$ at samples from S
2: estimate $\mathfrak{S}_k$, $k = 1, 2, \ldots, p$
3: estimate $F_k'(\delta)$, $k = 1, 2, \ldots, p$
4: compute empirical correlation matrices from S

5.4 Numerical results

In this section the proposed method is applied to two problems. The first is a synthetic test problem meant to illustrate the methodological details described in Section 5.3. Second is the motivating application where f is a space-time hierarchical Gaussian process used for wind speed prediction [4].

5.4.1 Synthetic test problem

This synthetic problem illustrates the proposed method of GSA and its properties on a simple example with least squares estimation. The difficulty of MCMC sampling with heavy tailed distributions, and the effect of the regularization factor (5.6), is demonstrated.

Mimicking characteristics of the motivating application, consider a space-time process governed by the function

$$f(s, t; \mathbf{X}) = S(s)T(t), \tag{5.11}$$

where

$$S(s) = \alpha_0 + \alpha_1 s + \alpha_2 s^2$$

and

$$T(t) = \beta_0 + \beta_1 e^{-\gamma t} \cos\left(\frac{2\pi}{100} t\right) + \beta_2 \sin\left(\frac{2\pi}{100} t\right) + \beta_3 \frac{1}{1 + e^{-.1(t-50)}}$$

with $s \in [0, 1]$, $t \in [0, 100]$, and

$$\begin{aligned} \mathbf{X} &= (\beta_0, \beta_1, \beta_2, \beta_3, \gamma, \alpha_0, \alpha_1, \alpha_2) \\ &= (2, 10, 3, .01, .01, 1, .01, 1). \end{aligned} \tag{5.12}$$

Drawing 15^2 samples from (5.11) on a uniform grid of $[0, 1] \times [0, 100]$ gives the data

$$\{(s_i, t_i, f(s_i, t_i))\}_{i=1}^{225}.$$

A model $\hat{f}$ parameterized in the same form as (5.11) is proposed, but the parameters are assumed to be unknown. They are determined by minimizing the least squares loss function

$$\mathcal{L}(\mathbf{x}) = \frac{1}{225}\sum_{i=1}^{225}(f(s_i,t_i) - \hat{f}(s_i,t_i))^2.$$

Analytic solutions for the sensitivities are intractable; however, the results may be validated by comparing them with knowledge of the true model that generated the data. In particular, the relative importance of the parameters is clear by examining (5.11) and (5.12). It is expected that β_1 will be the most important parameter and β_3 and α_1 to be the least important parameters.

The proposed method is used with $N = 10^5$, $\alpha = .99$, $c = .1$, and $h_{max} = .1$. Five independent chains are generated from overdispersed initial iterates using adaptive MCMC [116]. When $\nu = \lambda = 0$, the MCMC sampler fails to converge because the tail of ϕ is too heavy. To illustrate this, Figure 5.1 shows the iteration history for the parameter β_1 in each of the five chains after a burn-in period is discarded. The two leftmost frames indicate that ϕ is heavy tailed; the other three chains never reach the tail. A heavy-tailed PDF such as this requires extensive sampling, which makes the reliable computation of sensitivity indices intractable. Therefore, regularization is used to alleviate this problem by increasing ν and monitoring the sampler's convergence; $\nu = .2$ is found to yield converged chains with $N = 10^5$ samples. The chains are deemed convergent by using the potential scale reduction factor (PSRF) [13] as well as visualizing the iteration histories and histograms from each of the five chains.

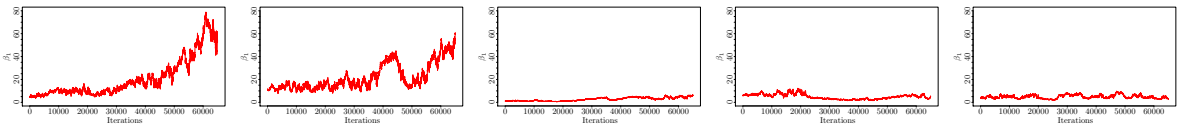

Figure 5.1 Iteration history for parameter β_1. Each frame corresponds to an independent chain.

Plotting the iteration history of $\mathcal{L}_\lambda$ indicates that a burn-in of 3.5×10^4 is sufficient. Then the remaining samples from the five chains are pooled together so that sensitivities and correlations may be computed from them. Figure 5.2 shows the sensitivity indices and Pearson correlation matrix computed from the pooled samples. These results are consistent with expectations, β_1 is seen as the most important parameter and α_1 as the least important. Two primarily blocks are seen in the correlation plot representing the set of spatial variables and the set of temporal variables. Negative correlations are observed on the off diagonal blocks since the spatial and temporal variables are multiplied by one another and hence are inversely related.

Figure 5.3 displays (5.10) plotted for $h \in (-\frac{1}{10}, \frac{1}{10})$, $k = 1, 2, \ldots, 8$. The horizontal lines

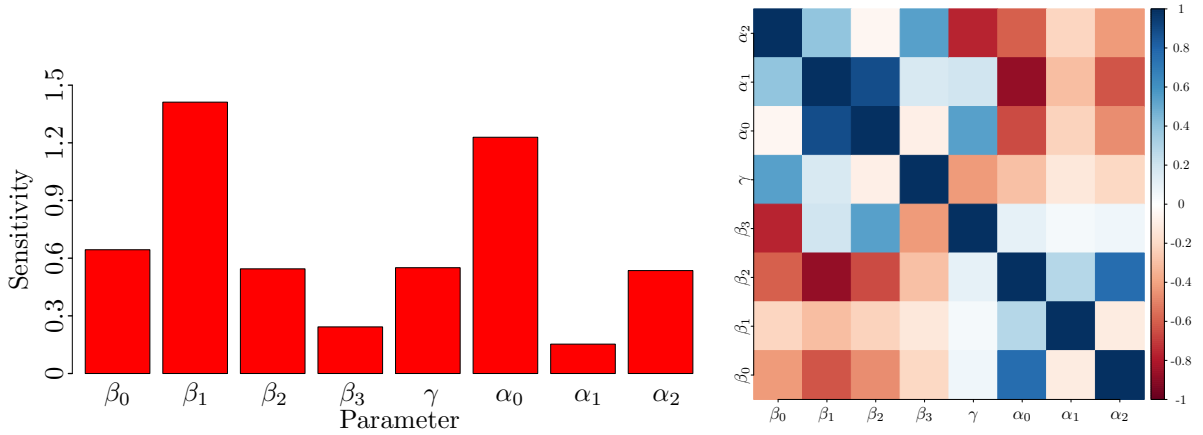

Figure 5.2 Sensitivity indices (left) and Pearson correlation coefficients of the parameters (right) for the synthetic test problem.

indicate that errors in determining δ are immaterial since the analysis would be unchanged by perturbing δ.

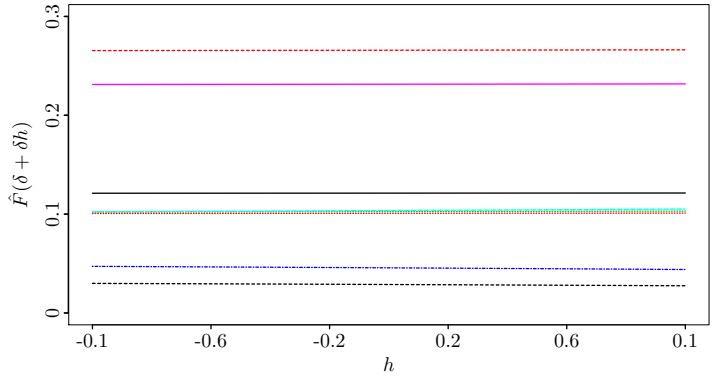

Figure 5.3 Sensitivity index perturbations for the synthetic test problem. Each line corresponds to a given parameter.

5.4.2 Analysis of a space-time Gaussian process

In this section, the proposed method is applied to analyze the motivating statistical model [4]. The model aims at forecasting wind speed by fusing two heterogeneous datasets: numerical weather prediction (NWP) model outputs and physical observations. Let Y_{NWP} denote the output of the NWP model and Y_{Obs} denote the observed measurements. They are modeled as a

bivariate space-time Gaussian process specified in terms of mean and covariance structures as follows,

$$\begin{pmatrix} Y_{\text{Obs}} \\ Y_{\text{NWP}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_{\text{Obs}}(\mathbf{X}) \\ \mu_{\text{NWP}}(\mathbf{X}) \end{pmatrix}, \begin{pmatrix} \Sigma_{\text{Obs}}(\mathbf{X}) & \Sigma_{\text{Obs,NWP}}(\mathbf{X}) \\ \Sigma^T_{\text{Obs,NWP}}(\mathbf{X}) & \Sigma_{\text{NWP}}(\mathbf{X}) \end{pmatrix} \right), \tag{5.13}$$

where $\mathbf{X}$ is the set of parameters that describe the shapes of the means and covariances. The model is expressed in a hierarchical conditional manner to avoid the specification of the full joint covariance in (5.13), indeed the mean and covariance of the distributions $(Y_{\text{Obs}}|Y_{\text{NWP}}) \sim \mathcal{N}\Big(\mu_{Obs|NWP}, \Sigma_{Obs|NWP}\Big)$ and $Y_{\text{NWP}} \sim \mathcal{N}\Big(\mu_{NWP}, \Sigma_{NWP}\Big)$ are specified in time, geographical coordinates, and parameters from the numerical model (the land-use parameter). More precisely,

$$\mu_{NWP}(t,s) = (\alpha_0 \text{LU(s)} + \alpha_1 \text{Lat(s)} + \alpha_2 \text{Long(s)})\, g(t), \tag{5.14}$$

where s is a spatial location, t is time, $g(t)$ represents a sum of temporal harmonics with daily, half-daily and 8hr-periodicities, LU(s) is a categorical variable that represents the land-use associated with location s, Lat and Long are the latitude and longitude coordinates.

$$\Sigma_{NWP}(*, s_i; *, s_j) = \text{Cov}(Y_{\text{NWP}}(*, s_i), Y_{\text{NWP}}(*, s_j)) = (\Psi(s_i)\Gamma_0\Psi(s_j)^T) + \delta_{i-j}\Gamma_{LU(s_i)}, \tag{5.15}$$

Γ_0, $(\Gamma_{LU(s_i)})_{i=1..I}$ are temporal squared exponential covariances expressed as

$$\Gamma_*(t_k, t_l) = \sigma_* \exp(-\rho_*(|t_k - t_l|)^2) + \delta_{k-l}\gamma_*,$$

where δ_{k-l} is the Kronecker delta, σ, ρ, and γ are parameters to be estimated. $(\Gamma_{LU(s_i)})_{i=1..I}$ are land-use specific terms, and Ψ is linear in the latitude and longitude coordinates and quadratic in time. The parameters α_0, α_1, α_2, along with the parameters of $g(t)$, $(\Gamma_{s_i})_{i=0..I}$ and Ψ, are estimated by maximum likelihood. Denote the collection of all these parameters by $\mathbf{X}_{NWP}$.

The conditional distribution is expressed through its mean and covariance:

$$\mu_{Obs|NWP}(t,s) = \mu(t,s) + (\Lambda Y_{\text{NWP}})(t,s)\,,$$

where $\mu(t,s)$ is written similarly to $\mu_{NWP}(t,s)$ as a product of temporal harmonics and a linear combination of the coordinates latitude and longitude. Λ is a projection matrix specified in time, latitude, longitude and the land-use parameter. The covariance $\Sigma_{Obs|NWP}$ is parametrized with a similar shape to (5.15) with a different set of parameters. Parameters of these functions are denoted as $\mathbf{X}_{Obs|NWP}$.

The model is fit by maximum likelihood estimation on the two datasets with respect to the parameters $\mathbf{X} = (\mathbf{X}_{NWP}, \mathbf{X}_{Obs|NWP})$. The negative log likelihood of the model can be

decomposed as

$$\mathcal{L}(\mathbf{x}) = \mathcal{L}_{\text{NWP}}(\mathbf{x}_{NWP}) + \mathcal{L}_{\text{Obs|NWP}}(\mathbf{x}_{Obs|NWP}), \tag{5.16}$$

where $\mathcal{L}_{\text{NWP}}(\mathbf{x}_{NWP})$ and $\mathcal{L}_{\text{Obs|NWP}}(\mathbf{x}_{Obs|NWP})$ are the negative log likelihoods for the marginal distribution of Y_{NWP} and the conditional distribution $Y_{\text{Obs}}|Y_{\text{NWP}}$, respectively. Since the model decomposes in this way, analysis of the parameters in Y_{NWP} and $Y_{\text{Obs}}|Y_{\text{NWP}}$ are considered separately. The dataset consists of 27 days of measurements from August 2012; details may be found in [4]. The parameter sensitivity during the first 13 days for Y_{NWP} and $Y_{\text{Obs}}|Y_{\text{NWP}}$ are analyzed in Sub-subsection 5.4.2.1 and Sub-subsection 5.4.2.2, respectively. Inferences are drawn from this analysis and validated using the later 14 days of data in Sub-subsection 5.4.2.3.

5.4.2.1 Parameter sensitivity analysis for Y_{NWP}

In this sub-subsection, the proposed method is applied to determine the sensitivity of the marginal model for Y_{NWP} to its 41 parameters during a 13-day period. The sensitivities being computed are with respect to the parameters in $\mathbf{X}_{NWP}$, but for notational simplicity denote them by $\mathbf{X}$ in this sub-subsection.

In order to determine $\mathbf{x}^\star$ (Line 1 of Algorithm 3), the L-BFGS-B algorithm is used to minimize $\mathcal{L}_{\text{NWP}}$. Visualizing the model predictions for various choices of c yields $c = .35$ (Line 2 of Algorithm 3). Then M (5.5) is estimated with $\ell = 5000$ Monte Carlo samples (Line 3 of Algorithm 3). It returns an estimate $M = 4160$ with standard deviation 2; hence ℓ is considered to be sufficiently large. These steps complete the preprocessing stage by providing characteristic values for the parameters $\mathbf{X}$ and the loss function $\mathcal{L}$.

The PDF ϕ is found to be heavy tailed, so $\nu = .1$ is chosen to reduce this effect. Then the equation $\Delta(\delta) = \alpha$ is solved with $\alpha = .99$ by evaluating Δ and manually updating δ. The solution $\delta = .07$ is obtained. This converged in very few iterations because of the nice properties of the equation $\Delta(\delta) = \alpha$. Any other derivative-free nonlinear solver may be used; however, manual tuning is preferable in many cases because of the simplicity of the equation and the stochasticity of the function evaluations. Having determined ν and δ, the PDF ϕ is now well defined and samples may be drawn from it.

Adaptive MCMC [116] is used with a desired acceptance rate of .15. Five chains of length $N = 4 \times 10^5$ each are generated independently from overdispersed initial iterates, and the first 10^5 iterates are discarded as burn-in. The PSRF convergence diagnostic from [13] is used on X_k and $\frac{\partial \mathcal{L}}{\partial x_k}(\mathbf{X})$, $k = 1, 2, \ldots, p$, to assess convergence of each of the $2p$ estimates. The PSRFs for all parameters lie in the intervals $(1, 1.025)$ and $(1, 1.048)$, respectively. Other visual diagnostics are applied as well, along with comparing sensitivity indices from each of the chains. The sensitivity estimation appears to converge.

Figure 5.4 displays the sensitivity indices estimated from each of the chains. The five different colors represent the five different chains; their comparability demonstrates that MCMC estimation errors are negligible. The intercept terms in the mean and the covariance kernel parameters are the most influential. The terms parameterizing Ψ are less influential, particularly the quadratic temporal terms.

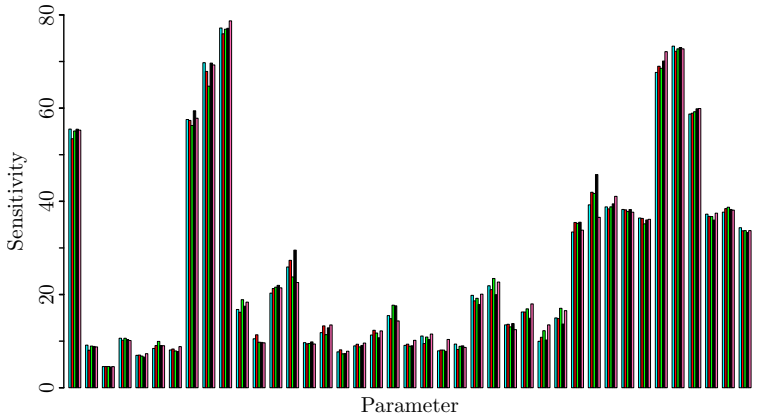

Figure 5.4 Sensitivity indices for Y_{NWP}. The five colors represent the sensitivity indices computed from each of the five chains.

As discussed in Sub-subsection 5.3.3.1, the robustness of the sensitivities with respect to errors in δ may be estimated as a by-product of computing sensitivities. Figure 5.5 displays (5.10) plotted for $h \in (-1/10, 1/10)$, $k = 1, 2, \ldots, 41$. Most of the curves are nearly horizontal, and those that not horizontal display small variation that does not change the resulting inference. Thus the sensitivities are robust with respect to δ, and hence any errors made when determining δ are negligible.

As mentioned in Sub-subsection 5.3.3.2, parameter correlation information is a useful complement to sensitivity indices. Figure 5.6 displays the empirical Pearson correlation matrix computed from the 1.5×10^6 MCMC samples retained after removing burn-in and pooling the chains. Strong positive correlations are observed between the three land-use dependent spatial intercepts in the mean. Strong negative correlations are observed between the temporal range ρ and the nugget term γ parameterizing the land-use specific covariance kernels $\Gamma_{LU(s_i)}$. This correlation is expected since the nugget term represents the variance of the signal that is not explained by the exponential part.

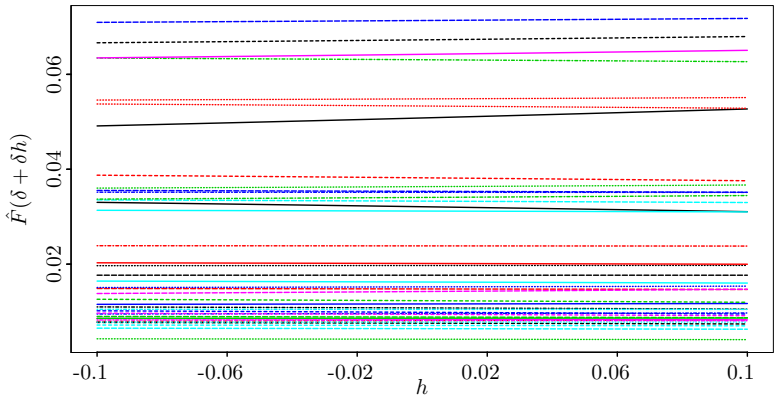

Figure 5.5 Sensitivity index perturbations for Y_{NWP}. Each line corresponds to a given parameter.

5.4.2.2 Parameter sensitivity analysis for $Y_{\text{Obs}}|Y_{\text{NWP}}$

In this sub-subsection, the proposed method is applied to determine the sensitivity of the model for $Y_{\text{Obs}}|Y_{\text{NWP}}$ to its 54 parameters during the same 13-day period used in Sub-subsection 5.4.2.1. The sensitivities being computed are with respect to the parameters in $\mathbf{X}_{Obs|NWP}$, but for notational simplicity denote them by $\mathbf{X}$ in this sub-subsection.

In a similar fashion to Sub-subsection 5.4.2.1, the L-BFGS-B algorithm is used to determine $\mathbf{x}^\star$ (Line 1 of Algorithm 3). Visualizing the model predictions for various choices of c yields $c = .05$ (Line 2 of Algorithm 3). Then M (5.5) is estimated with $\ell = 5000$ Monte Carlo samples (Line 3 of Algorithm 3). It returns an estimate $M = 2973$ with standard deviation 2; hence ℓ is considered to be sufficiently large. These steps complete the preprocessing stage by providing characteristic values for the parameters $\mathbf{X}$ and the loss function $\mathcal{L}$. One may note that c and M are significantly smaller for $Y_{\text{Obs}}|Y_{\text{NWP}}$ than for Y_{NWP}. Their difference is unsurprising since Y_{NWP} and $Y_{\text{Obs}}|Y_{\text{NWP}}$ model two different processes.

The PDF ϕ is found to be heavy tailed, so $\nu = .15$ is chosen to reduce the tail of ϕ. Analogously to Sub-subsection 5.4.2.1, $\Delta(\delta) = \alpha = .99$ is solved yielding $\delta = .06$ (Line 3 of Algorithm 4).

Adaptive MCMC [116] is used with a desired acceptance rate of .15. Five chains of length $N = 4 \times 10^5$ each are generated independently from overdispersed initial iterates, and the first 10^5 iterates are discarded as burn-in. The convergence diagnostics discussed in Sub-subsection 5.4.2.1 are used. The PSRFs for all parameters lie in the intervals $(1, 1.191)$ and $(1, 1.036)$, respectively. A few sensitivity indices have not fully converged; however, the remaining uncertainty in their estimation is sufficiently small for the purposes of this sensitivity analysis. These uncertain sensitivities are among the largest in magnitude. Since the goal is encouraging model parsimony, and derivative based approaches are known to have limitations in ordering the most influential

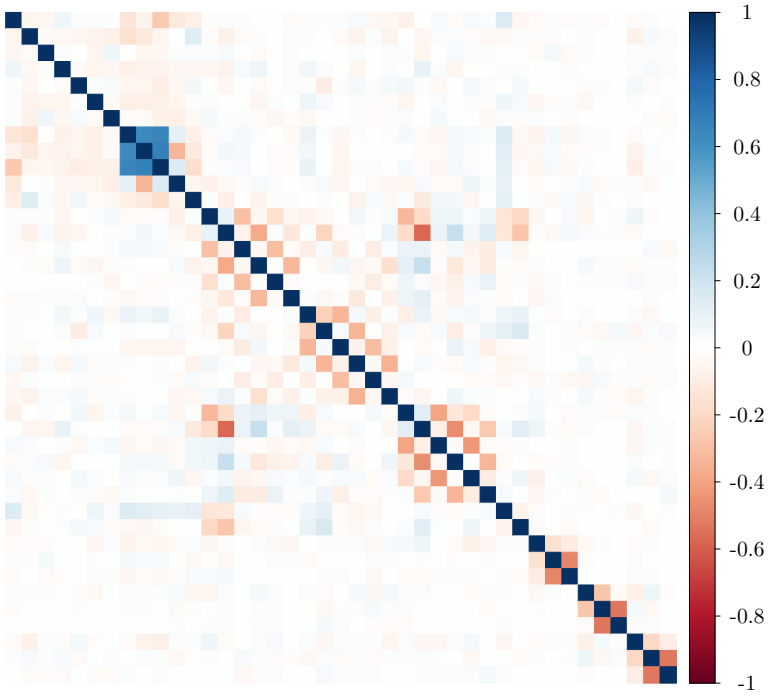

Figure 5.6 Pearson correlation coefficients for the parameters of Y_{NWP}.

parameters, then the remaining uncertainty is of secondary importance.

Figure 5.7 displays the sensitivity indices estimated from each of the chains. The five different colors represent the five different chains. The sensitivities with greatest uncertainties are demonstrated by the differences in their estimated values in each chain; however, these discrepancies are sufficiently small that they do not alter the resulting inference. The longitudinal terms in the mean and Ψ are observed to have little influence since their sensitivity indices are nearly zero. The most influential parameters are the spatial weights in the matrix Λ which acts on Y_{NWP}.

Figure 5.8 displays (5.10) plotted for $h \in (-\delta/10, \delta/10)$, $k = 1, 2, \ldots, 54$. Most of the curves are nearly horizontal, and those that are not horizontal display small variation that does not change the resulting inference.

Figure 5.9 displays the empirical Pearson correlation matrix computed from the 1.5×10^6 MCMC samples retained after burn-in and pooling the chains. Strong correlations are observed between the spatial weights in the matrix Λ. Similar to Y_{NWP}, strong negative correlations are also observed between the temporal range ρ and the nugget term γ of the land-use specific covariance kernels.

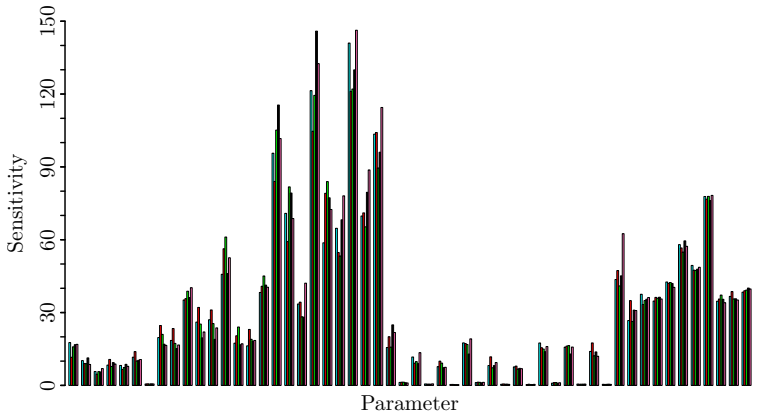

Figure 5.7 Sensitivity indices for $Y_{\text{Obs}}|Y_{\text{NWP}}$. The five colors represent the sensitivity indices computed from each of the five chains.

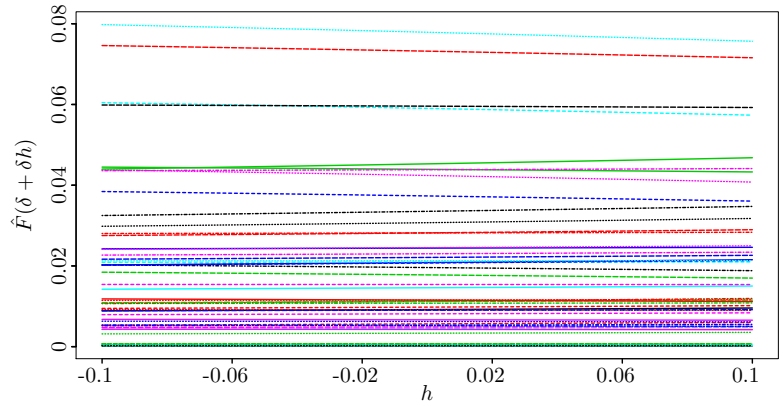

Figure 5.8 Sensitivity index perturbations for $Y_{\text{Obs}}|Y_{\text{NWP}}$. Each line corresponds to a given parameter.

5.4.2.3 Inference and validation of results

In some cases with mathematical models one may set a threshold and fix all parameters whose sensitivity is below the threshold. This approach is not suitable for statistical models because the parameters must be understood in light of their contribution to the model structure and their correlation with other parameters. Rather, the sensitivity indices and correlation structures should be used to re-parameterize the statistical model in a simpler way. For instance, if a collection of spatially dependent parameters are all found to be unimportant then the user may consider replacing them by a single parameter which is not spatially dependent.

Using the results of Sub-subsection 5.4.2.1, and considerations of the model structure, it is determined that the Y_{NWP} model is insensitive to the temporal quadratic terms in the parameterization of Ψ. Similarly, coupling the results of Sub-subsection 5.4.2.2, and the model

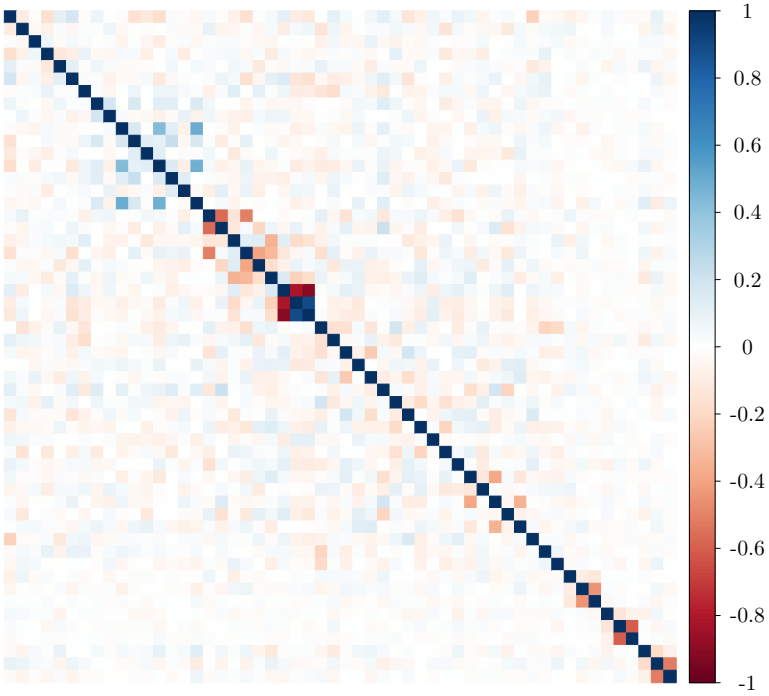

Figure 5.9 Pearson correlation coefficients for the parameters of $Y_{\text{Obs}}|Y_{\text{NWP}}$.

structure, it is determined that the $Y_{\text{Obs}}|Y_{\text{NWP}}$ model is insensitive to several of the longitude terms. Specifically, the longitude term in the parameterization of the mean and the nine longitude terms in the parameterization of Ψ. This conclusion confirms what one would expect from the physics considerations. The flow is predominantly east-west and thus the north-south correlation is relatively weaker. The east-west information is likely to be well captured by Y_{NWP} and hence is not needed in $Y_{\text{Obs}}|Y_{\text{NWP}}$.

The 16 insensitive parameters are removed from the model, yielding a more parsimonious model, call it the reduced model. To validate inferences, the original model and the reduced model are used to predict on the other 14 days of data not use in the sensitivity analysis. Leave-one-out cross-validation is used to fit each of the models and assess their predictive capabilities.

For each of the 14 days, 1,000 scenarios are simulated and two metrics are used to quantify the predictive capacity of the full and reduced models, namely, the energy score and the root mean square error. The energy score [38, 94], a measure of distance between distributions, is computed for the joint distribution of all spatial locations at each day; hence, 14 energy scores are computed. The root mean square error is computed as the square root of the time average

squared error at each spatial location; hence, there are 11 root mean square errors. Figure 5.10 displays the energy scores on the left and root mean square errors on the right. The reduced model has slightly smaller (and hence better) energy scores and root mean square errors in the majority of the cases. The sum of energy scores for full model and reduced model is 121.1 and 120.7, respectively. The sum of root mean squared errors for the full model and reduced model is 11.4 and 11.3, respectively.

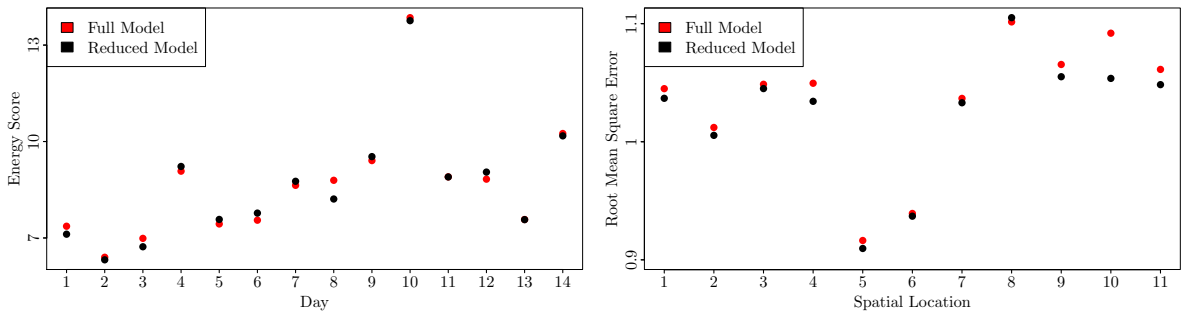

Figure 5.10 Left: energy scores for each of the 14 days being predicted; right: root mean square error for each of the 11 spatial locations being predicted. The full model is red and the reduced model is black.

To further illustrate the difference between the full and reduced models, the simulated scenarios are displayed for a typical case. Specifically, the spatial location with median root mean square error and six days with median energy score is chosen to plot the 1,000 scenarios along with the observed data. Figure 5.11 displays the results with the full model on the left and reduced model on the right.

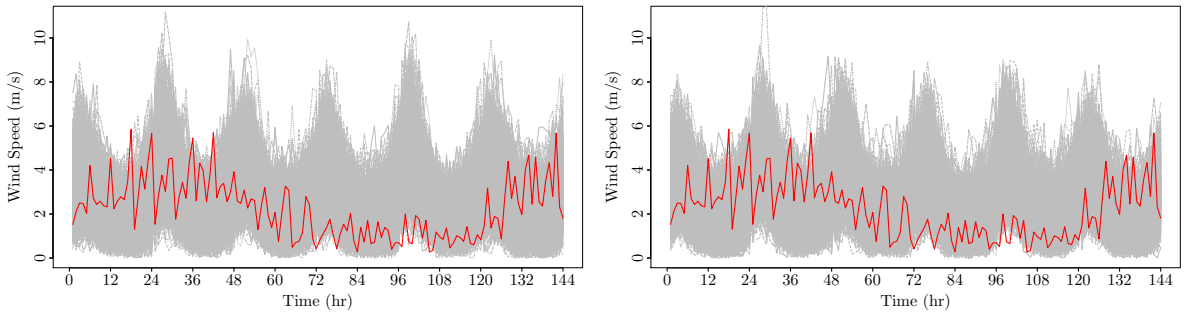

Figure 5.11 Predictions using the full model (left) and reduced model (right) for 6 days at a fixed spatial location. The red curve is the observed wind speed and the grey curves are 1000 simulations generated from each model.

The sensitivity analysis enabled the parameterization of the model to be simplified from having 95 parameters to 79. The reduced model has equal or better predictive capability and is simpler to fit and analyze. Further, the reduced model typically has fewer outlying scenarios, as evidenced in Figure 5.11.

For this application a strong insensitivity of $Y_{\text{Obs}}|Y_{\text{NWP}}$ to some of its terms contributing longitudinal information is observed. This is likely because the Y_{NWP} model captures longitudinal information well and hence the Gaussian process does not need to fit terms contributing longitudinal information. Thus, inferences may also be made on the data being input to the statistical model through the parameter sensitivities.

These inferences and simplifications are useful for multiple reasons. First, long term weather prediction is difficult so the parameters must be optimized frequently to accommodate changing weather patterns. Hence a large optimization problem must be solved frequently, reducing the number of parameters allows for faster and more robust optimization. Second, by removing unimportant parameters the model is more robust and can be more easily integrated into a larger workflow, namely power scheduling. Third, the underlying system may be better understood through these inferences, for instance, the unimportance of longitudinal information.

5.5 Conclusion

A new method for GSA of statistical model parameters was introduced. It addresses the challenges and exploits the problem structure specific to parameterized statistical models. The method is nearly fully automated; one step depends on the user's discretion, but this level of user specification is likely necessary for any alternative method. The proposed method also admits robustness analysis at no additional computational cost, thus yielding sensitivities accompanied with certificates of confidence in them.

The method was motivated by, and applied to, a Gaussian process model aimed at wind simulation. Sensitivities were computed and the model parameterization simplified by removing 17% of the model parameters. This simpler model was validated and shown to provide equal or superior predictive capability compared with the original model.

The proposed method has two primary limitations. First, it relies heavily on Markov Chain Monte Carlo sampling for which convergence diagnostics are notoriously challenging. Second, regularization may be needed to eliminate heavy-tailed distributions. Determining the regularization constant is simple in principle but may require drawing many samples to resolve. However, these limitations are classical and have been observed in various applications previously.

CHAPTER

6

DERIVATIVE-BASED GSA FOR PDE-CONSTRAINED OPTIMIZATION

6.1 Introduction

This chapter considers GSA for the solution of PDE-constrained optimization problems with respect to uncertain parameters (which are not optimization variables). Throughout this thesis, the function f has been a mapping from parameters to a QoI or model output; in this chapter f is a mapping from uncertain parameters to the solution of a PDE-constrained optimization problem. Hence evaluating $f(\mathbf{X})$ is computationally intensive, and in general, $f(\mathbf{X})$ will assume values in a function space. Building off the work of Brandes and Griesse [11], a GSA paradigm and computational framework is developed. A derivative-based approach to GSA is adopted in order to utilize the work of [11] which provides a foundation for efficiently computing derivatives of the optimal solution with respect to the uncertain parameters, i.e. local sensitivities. This chapter extends their work in three ways, (i) an algorithmic extension, (ii) a scalable C++ implementation, and (iii) a systematic framework for GSA in PDE-constrained optimization. The framework is illustrated by applying it to a multi-physics nonlinear system. The related works of [15, 39, 40, 41, 42, 43, 44] consider various aspects of local sensitivity analysis for PDE-constrained optimization problems, in particular, focusing on conditions for differentiability and the computation of derivational derivatives.

6.2 Overview of sensitivity analysis for PDE-constrained optimization

This section overviews the work of Brandes and Griesse [11] to provide necessary background. Throughout the chapter the optimization variables will be referred to as the control, though the results of this chapter apply to design and deterministic inverse problems as well. The uncertain parameters will be denoted by $\mathbf{x}$ (or $\mathbf{X}$ when treated as random variables) and called the uncertain parameters or simply parameters (instead of input variables or variables as in some chapters of this thesis). Local sensitivity analysis is studied first to develop a computational framework, it is extended to GSA later in the chapter.

6.2.1 Control problem formulation

Consider the PDE-constrained control problem

$$\begin{aligned} &\min_{y,u} J(y,u) \\ \text{s.t. } &e(y,u,\mathbf{x}) = 0 \\ &y \in Y, u \in U \end{aligned} \tag{6.1}$$

where Y is the state space, U is the control space, $\mathbf{x} \in \mathcal{X}$ are fixed parameters, $J : Y \times U \to \mathbb{R}$ is an objective function, and $e : Y \times U \times \mathcal{X} \to Z^\star$ is the weak form of a PDE; $Z^\star$ denotes the dual of Z. The PDE represented by e may be time dependent or stationary. It is assumed that Y and Z are reflexive Banach spaces, and U and $\mathcal{X}$ are Hilbert spaces. The stronger assumptions on U and $\mathcal{X}$ are to enable subsequent analysis with the Singular Value Decomposition (SVD). The spaces $Y, U, \mathcal{X}$, and Z may be finite or infinite dimensional. In particular, the parameter space $\mathcal{X}$ may represent a finite number of parameters or a functional representation (for instance, a spatially dependent parameter), in which case $\mathcal{X}$ is infinite dimensional. In contrast to the previous chapters when GSA focused on functions whose domain was $\mathbb{R}^p$ (or a subset of it), this chapters permits a more general parameter space.

The goal is to develop a computational framework which may by used for time dependent, multi-physics, and/or multi-scale nonlinear problems in two or three spatial dimensions, with infinite (or large finite) dimensional parameter uncertainty. A first step towards this end is taken in this chapter through developing a framework which is applied to a multi-physics nonlinear steady state problem in two spatial dimensions with spatially dependent parameters. Extensions of the framework are highlighted.

The objective is to analyze the sensitivity of the optimal control solution with respect to changes in the parameters $\mathbf{x}$. To study (6.1) as an unconstrained optimization problem, define

the Lagrangian $\mathcal{L} : Y \times U \times Z \times \mathcal{X} \to \mathbb{R}$ as

$$\mathcal{L}(y, u, \lambda, \mathbf{x}) = J(y, u, \mathbf{x}) + \langle \lambda, e(y, u, \mathbf{x}) \rangle, \tag{6.2}$$

where $\lambda \in Z$ is the Lagrange multiplier, in the context of control it is called the adjoint state. Throughout the chapter, the subscripts J_* and e_* will denote the Fréchet derivative of J and e with respect to $*$, respectively. After discretizing, J_* and e_* will denote the gradient of J and Jacobian of e, respectively.

The following assumptions are made for the remainder of the chapter:

1. J and e are twice continuously differentiable with respect to $(y, u, \mathbf{x})$.
2. $\mathbf{x}_0 \in \mathcal{X}$ are nominal parameters and (y_0, u_0) is a local minimum of (6.1), with corresponding adjoint state λ_0, when $\mathbf{x} = \mathbf{x}_0$.
3. $e_y(y_0, u_0, \mathbf{x}_0)$ is surjective and the adjoint state λ_0 is unique.
4. There exists $\rho > 0$ such that

$$\mathcal{L}_{(y,u),(y,u)}(y_0, u_0, \lambda_0, \mathbf{x}_0)((y, u), (y, u)) \geq \rho ||(y, u)||_{Y \times U}$$

   for all (y, u) in the null space of $e_{(y,u)}(y_0, u_0, \mathbf{x}_0)$.

These assumptions imply that

$$\mathcal{L}_y(y_0, u_0, \lambda_0, \mathbf{x}_0) = f_y(y_0, u_0) + \langle \lambda, e_y(y_0, u_0, \mathbf{x}_0) \rangle = 0 \tag{6.3}$$

$$\mathcal{L}_u(y_0, u_0, \lambda_0, \mathbf{x}_0) = f_u(y_0, u_0) + \langle \lambda, e_u(y_0, u_0, \mathbf{x}_0) \rangle = 0 \tag{6.4}$$

$$\mathcal{L}_\lambda(y_0, u_0, \lambda_0, \mathbf{x}_0) = e(y_0, u_0, \mathbf{x}_0) = 0 \tag{6.5}$$

which are the well known necessary conditions for optimality.

6.2.2 Sensitivity formulation

The following result of [11] provides the foundation for local sensitivity analysis.

Theorem 12. *There are neighborhoods $\mathcal{N}(\mathbf{x}_0) \subset \mathcal{X}$ of $\mathbf{x}_0$ and $\mathcal{N}(y_0, u_0, \lambda_0) \subset Y \times U \times Z$ of (y_0, u_0, λ_0) and a continuously differentiable function $f : \mathcal{N}(\mathbf{x}_0) \to \mathcal{N}(y_0, u_0, \lambda_0)$ such that for all $\mathbf{x} \in \mathcal{N}(\mathbf{x}_0)$, $f(\mathbf{x})$ is the unique element of $\mathcal{N}(y_0, u_0, \lambda_0)$ which satisfies*

$$\nabla \mathcal{L}(f(\mathbf{x}), \mathbf{x}) = 0,$$

i.e. is the unique critical point of (6.1) *in* $\mathcal{N}(y_0, u_0, \lambda_0)$*. Further, the Fréchet derivative of* f *at* $\mathbf{x}_0$ *is given by*

$$f'(\mathbf{x}_0) = \mathcal{K}^{-1}\mathcal{B}$$

where

$$\mathcal{K} = \begin{pmatrix} \mathcal{L}_{yy} & \mathcal{L}_{yu} & e_y^\star \\ \mathcal{L}_{uy} & \mathcal{L}_{uu} & e_u^\star \\ e_y & e_u & 0 \end{pmatrix},$$

$$\mathcal{B} = -\begin{pmatrix} \mathcal{L}_{y\mathbf{x}} \\ \mathcal{L}_{u\mathbf{x}} \\ \mathcal{L}_{\lambda\mathbf{x}} \end{pmatrix},$$

with the second derivatives of $\mathcal{L}$ *evaluated at* $(y_0, u_0, \lambda_0, \mathbf{x}_0)$*;* $\star$ *denotes the adjoint of an operator. The operator* $\mathcal{K}$ *is called the Karush-Kuhn-Tucker (KKT) operator.*

Theorem 12 gives a computable expression for the change in the optimal state, control, adjoint triple when the parameters $\mathbf{x}_0 \in \mathcal{X}$ are perturbed. Notice that this is a local result in the sense that the optimization problem (6.1) may have multiple local minima but f depends upon the specific local minimum (y_0, u_0, λ_0). Restricting to one local minimum ensures the existence and well-poseness of f and its derivative.

In what follows, the arguments of $\mathcal{L}$ and its derivatives are omitted implying that they are evaluated at $(y_0, u_0, \lambda_0, \mathbf{x}_0)$. From Theorem 12, the directional derivative of optimal state, control, adjoint triple in the direction $\mathbf{x}$ is given by the solution, (y, u, λ), of the linear system

$$\mathcal{K}\begin{pmatrix} y \\ u \\ \lambda \end{pmatrix} = \mathcal{B}\mathbf{x}. \tag{6.6}$$

To determine the sensitivity of the optimal controller to changes in $\mathbf{x}$, define the projection operator $\Pi : Y \times U \times Z \to U$ by

$$\Pi\begin{pmatrix} y \\ u \\ \lambda \end{pmatrix} = u.$$

Then the Fréchet derivative of the optimal control solution with respect to the parameters $\mathbf{x}$

may be expressed as the linear operator $\mathcal{D} : \mathcal{X} \to U$,

$$\mathcal{D} = \Pi \mathcal{K}^{-1} \mathcal{B}, \tag{6.7}$$

which, thanks to Theorem 12, is well defined in a neighborhood of $\mathbf{x}_0$. Note that the sensitivity of the state, or a function of the state and control, may be considered by using a different Π. This chapter focuses on the sensitivity of the optimal control strategy.

Assume that $\mathcal{D}$ is a compact operator, see [11] more further discussion of its compactness. Since $\mathcal{X}$ and U are Hilbert spaces, the sensitivity of the control solution u_0 to changes in $\mathbf{x}_0$ may be analyzed through the SVD of $\mathcal{D}$. Theorem 13 below recalls fundamental results for the SVD of compact linear operators on Hilbert spaces.

Theorem 13. *There exists $\sigma_k \geq 0$, $\mathbf{x}_k \in \mathcal{X}$, $u_k \in U$, $k = 1, 2, \ldots$, such that*

- $\mathcal{D}\mathbf{x} = \sum_{k=1}^{\infty} \sigma_k(\mathbf{x}, \mathbf{x}_k) u_k \quad \forall \mathbf{x} \in \mathcal{X}$
- $\sigma_k \geq \sigma_{k+1}$ $\forall k$
- $\{\mathbf{x}_k\}_{k=1}^{\infty}$ *are orthonormal in* $\mathcal{X}$
- $\{u_k\}_{k=1}^{\infty}$ *are orthonormal in* U
- $\mathcal{D}\mathbf{x}_k = \sigma_k u_k$, *k=1,2,...*

Since $\mathbf{x}_k$ and u_k are normalized, the singular value σ_k measures the magnitude of the change in control solution when $\mathbf{x}_0$ is perturbed in the direction $\mathbf{x}_k$. The ordering of the singular values enables an identification of important directions in parameter space. In practice, the K leading singular triples $(\sigma_k, \mathbf{x}_k, u_k)$, $k = 1, 2, \ldots, K$, are computed, where K is typically small (for instance, $K = 4$ is sufficient for the application in Section 6.9), but the choice of K dependents on the spectrum of $\mathcal{D}$. Problems where $\mathcal{D}$ has a low rank structure are of particular interest, which corresponds to when the controller is only sensitivity to a small subset of the uncertain parameters. Section 6.8 elaborates on how these singular triples are used to infer the influence of the parameters.

6.2.3 Discretization

Let $Y_h \subseteq Y, U_h \subseteq U, Z_h \subseteq Z, \mathcal{X}_h \subseteq \mathcal{X}$ be finite dimensional subspaces inheriting norms from the larger spaces. In practice they typically arise from a discretization of the PDE, for instance, a finite element discretization. Let $\mathbf{x}_0 \in \mathcal{X}_h$ be nominal parameters, and $(y_0, u_0, \lambda_0) \in Y_h \times U_h \times Z_h$ be a local solution of the discretization of (6.1) with nominal parameters $\mathbf{x}_0$. Assume that (y_0, u_0, λ_0) satisfies Assumptions 3 and 4.

Let $\{\phi_1, \phi_2, \ldots, \phi_p\}$ be a basis for $\mathcal{X}_h$ and $\{\psi_1, \psi_2, \ldots, \psi_n\}$ be a basis for U_h. In practice, the implementation of $\Pi, \mathcal{K}$, and $\mathcal{B}$ will act on coordinate representations of functions. Specifically, $\mathbf{x} \in \mathcal{X}_h$ is represented by $\mathbf{x}_c \in \mathbb{R}^p$ implying that

$$\mathbf{x} = \sum_{i=1}^{p} (\mathbf{x}_c)_i \phi_i,$$

and similarly with coordinate representation for U_h. Define the coordinate transformation operators

$$E_{\mathcal{X}_h} : \mathbb{R}^p \to \mathcal{X}_h, \quad C_{\mathcal{X}_h} : \mathcal{X}_h \to \mathbb{R}^p, \quad E_{U_h} : \mathbb{R}^n \to U_h, \quad \text{and} \quad C_{U_h} : U_h \to \mathbb{R}^n$$

which map between the coordinates in Euclidean space and the corresponding function in the discretized function space ($\mathcal{X}_h$ or U_h).

Computational routines used in practice will represent the action of $\mathcal{D}$ on functions by applying the matrix

$$D = C_{U_h} \mathcal{D} E_{\mathcal{X}_h} = C_{U_h} \Pi \mathcal{K}^{-1} \mathcal{B} E_{\mathcal{X}_h} \in \mathbb{R}^{n \times p} \tag{6.8}$$

to the coordinate representation of the function.

Applying a standard SVD routine to the matrix D is not appropriate because the inner products are taken in Euclidean space rather than in $\mathcal{X}_h$ and U_h. As observed in [11], an SVD routine using Euclidean inner products should be applied to the matrix

$$R_{U_h} D R_{\mathcal{X}_h}^{-1}, \tag{6.9}$$

where $R_{\mathcal{X}_h}$ and R_{U_h} are the Cholesky factors of the symmetric positive definite mass matrices $M_{\mathcal{X}_h} = R_{\mathcal{X}_h}^T R_{\mathcal{X}_h} \in \mathbb{R}^{p \times p}$ and $M_{U_h} = R_{U_h}^T R_{U_h} \in \mathbb{R}^{n \times n}$ defined by

$$(M_{\mathcal{X}_h})_{i,j} = (\phi_i, \phi_j)_{\mathcal{X}} \qquad (M_{U_h})_{i,j} = (\psi_i, \psi_j)_U.$$

Computing the SVD of (6.9) is problematic for large scale applications because the Cholesky factors $R_{\mathcal{X}_h}$ and R_{U_h} are dense, whereas the mass matrices $M_{\mathcal{X}_h}$ and M_{U_h} are sparse. To alleviate this challenge, [11] computes the SVD of (6.9) by reformulating it as a non-symmetric eigenvalue problem which does not require the Cholesky factors, and solves the eigenvalue problem with a iterative non-symmetric solver. A new formulation will be given in Section 6.4 which has advantageous properties to facilitate scalable computation.

6.2.4 An illustrative example

Computing the SVD of (6.7) is computationally intensive because it requires repeatedly solving KKT systems of the form (6.6). A natural alternative is to fix the controller to the optimal u_0 and analyze the sensitivity of the state (or objective function) to changes in the parameters; this eliminates the KKT solve thus reducing the computational complexity significantly. This approach suffers two drawbacks:

- The optimal controller will change when the parameters change, so fixing the controller may lead to sensitivity which occurs because of a poor control strategy.
- There may be parameters for which the state is sensitive but the optimal control is not, or vice versa. Since the control strategy is our end goal, fixing the controller does not give the sensitivities we need. Uncertainty in the state should be assessed through forward uncertainty quantification once a robust control strategy has been determined.

The following example illustrates differences in these two approaches.

Example 6.2.1. *Consider the optimization problem*

$$\begin{aligned} \min_{y,u} \quad & J(y,u) \\ s.t. \quad & y = \frac{1}{1+e^{-x_1 u}} + x_2 \end{aligned} \tag{6.10}$$

where

$$J(y,u) = (y-2)^2 + \frac{.001}{2}u^2,$$

and $\mathbf{x} = (x_1, x_2)$ *are uncertain parameters.*

Let $y_{opt}(\mathbf{x}), u_{opt}(\mathbf{x})$*, as a function of* $\mathbf{x}$*, be the optimal solution of* (6.10)*. Then*

$$\frac{\partial u_{opt}}{\partial x_1}(0.5, 0.5) = 9.99 \qquad \textit{and} \qquad \frac{\partial u_{opt}}{\partial x_2}(0.5, 0.5) = 3.12.$$

If instead, (6.10) *is solved with* $\mathbf{x} = (0.5, 0.5)$ *and the derivative of*

$$g(x_1, x_2) = \left(\frac{1}{1+e^{-x_1 u_{opt}(0.5,0.5)}} + x_2 - 2 \right)^2 + \frac{.001}{2} u_{opt}(0.5, 0.5)^2$$

is computed with respect to (x_1, x_2)*, then*

$$\frac{\partial g}{\partial x_1}(0.5, 0.5) = 0.135 \qquad \textit{and} \qquad \frac{\partial g}{\partial x_2}(0.5, 0.5) = 1.03.$$

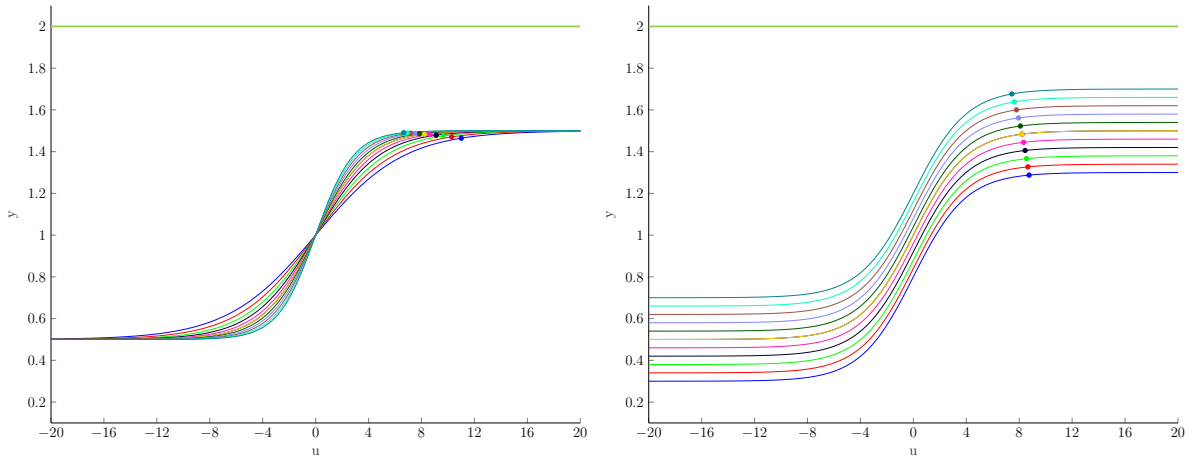

Figure 6.1 Plot of the constraint y in (6.10) as a function of u for different values of $\mathbf{x}$. Left: varying x_1 from 0.3 to 0.7 with $x_2 = 0.5$ fixed; right: x_2 varying from 0.3 to 0.7 with $x_1 = 0.5$ fixed. Each curve corresponds to a different $\mathbf{x}$ and each dot corresponds to the solution of (6.10) for that given $\mathbf{x}$. The green horizontal line is the target state $y = 2$.

Hence computing the sensitivity of the control solution to $\mathbf{x}$ *gives a different conclusion than the sensitivity of the objective function (at the optimal control) to* $\mathbf{x}$*. Figure 6.1 gives some intuition for this result. The constraint is plotted with* y *as a function of* u*, i.e. the state as a function of the control. Curves are plotted for different values of* $\mathbf{x}$ *and the solution of* (6.10) *for each fixed* $\mathbf{x}$ *is given by the dot on the curve. The left (right) panel has* $x_2 = 0.5$ *(*$x_1 = 0.5$*) fixed and varies* x_1 *(*x_2*) from* 0.3 *to* 0.7*. This demonstrates that as* x_1 *varies, left panel, the control solution* u *varies significantly while the state* y *is kept nearly constant; whereas as* x_2 *varies, right panel, the control solution* u *is nearly constant while the state* y *varies significantly.*

6.3 Overview of extensions and contributions

The approach of [11] presented in Section 6.2 is local sensitivity analysis in the sense that the results are only valid in a neighborhood of the nominal parameters $\mathbf{x}_0$. A natural extension for GSA is to compute local sensitivities, as described in Section 6.2, at a variety of samples from $\mathcal{X}_h$, i.e. different $\mathbf{x}_0$'s. To this end, a mathematical and computational framework is developed to compute the local sensitivities quickly, in particular, levering parallelism whenever possible.

The contributions of this chapter to the efficient and scalable computation of local sensitivities are twofold:

1. developing a software infrastructure, within the Trilinos framework, to leverage state of the art (matrix free) PDE-constrained optimization software with parallel linear algebra constructs within,

2. reformulating the non-symmetric eigenvalue problem of [11] as a symmetric generalized eigenvalue problem and utilizing a randomized eigenvalue solver which affords an embarrassingly parallel distribution of the KKT solves.

Having developed the capability to compute local sensitivities quickly, the work of [11] is extended to be global. Specifically, local sensitivities are computed at various samples in $\mathcal{X}$ (not global in the optimization variables); however, computational cost limits a full exploration of the parameter space. The formulation of global sensitivities is rigorously developed and its benefits and limitations are highlighted. A systematic approach is presented to compute, visualize, interpret, and make inferences with the sensitivities. The proposed framework is demonstrated on a optimal control problem constrained by the nonlinear multi-physics Boussinesq flow equations with 153 uncertain parameters.

6.4 Proposed reformulation

In this section the eigenvalue problem originally posed in [11] is reformulated. For the remainder of the chapter, analysis will be performed using coordinate representations. For notational simplicity, $\mathbf{x}$ and u are used to denote the coordinate representations for parameters and controllers. The singular values and singular vectors of (6.9) correspond to the positive eigenvalues and eigenvectors of the Jordan-Wielandt matrix

$$W = \begin{pmatrix} 0 & R_{U_h} D R_{\mathcal{X}_h}^{-1} \\ (R_{\mathcal{X}_h}^{-1})^T D^T (R_{U_h})^T & 0 \end{pmatrix}.$$

This gives the symmetric eigenvalue problem

$$W \begin{pmatrix} u \\ \mathbf{x} \end{pmatrix} = \alpha \begin{pmatrix} u \\ \mathbf{x} \end{pmatrix} \tag{6.11}$$

for which the largest eigenvalues are to be computed.

To avoid computing the Cholesky factors of the mass matrices, define

$$S = \begin{pmatrix} R_{U_h}^{-1} & 0 \\ 0 & R_{\mathcal{X}_h}^{-1} \end{pmatrix}$$

and

$$\begin{pmatrix} \tilde{u} \\ \tilde{\mathbf{x}} \end{pmatrix} = S \begin{pmatrix} u \\ \mathbf{x} \end{pmatrix}.$$

Then (6.11) may be reformulated as

$$(S^{-1})^T W S^{-1} \begin{pmatrix} \tilde{u} \\ \tilde{\mathbf{x}} \end{pmatrix} = \alpha (S^{-1})^T S^{-1} \begin{pmatrix} \tilde{u} \\ \tilde{\mathbf{x}} \end{pmatrix}. \tag{6.12}$$

Letting

$$A = (S^{-1})^T J S^{-1} = \begin{pmatrix} 0 & M_{U_h} D \\ D^T M_{U_h} & 0 \end{pmatrix}$$

and

$$B = (S^{-1})^T S^{-1} = \begin{pmatrix} M_{U_h} & 0 \\ 0 & M_{\mathcal{X}_h} \end{pmatrix},$$

(6.12) may be rewritten as $A\mathbf{v} = \alpha B\mathbf{v}$. Since A is symmetric and B is symmetric positive definite, a symmetric generalized eigenvalue problem must be solved. Unlike the eigenvalue problem posed in [11], this formulation benefits from symmetry [27]. It is proposed to solve the generalized eigenvalue problem with a randomized algorithm, whose advantages will be elaborated on in Section 6.5.

Assume that the K largest eigenvalues $\alpha_k > 0$ and corresponding eigenvectors $(\tilde{u}_k, \tilde{\mathbf{x}_k})$, $k = 1, 2, \ldots, K$, for $A\mathbf{v} = \alpha B\mathbf{v}$, have been computed. The corresponding singular values, right singular vectors, and left singular vectors of (6.9) are α_k, $R_{\mathcal{X}_h}\tilde{\mathbf{x}}_k$, and $R_{U_h}\tilde{u}_k$, $k = 1, 2, \ldots, K$, respectively. It appears that the mass matrix Cholesky factor is needed to compute these singular vectors; however, as shown in [11], the coordinate representation of the right and left singular vectors of (6.7) (using P_h and U_h inner products) are given by

$$\frac{\tilde{p}_k}{\tilde{p}_k^T M_{P_h} \tilde{p}_k} \qquad \text{and} \qquad \frac{\tilde{u}_k}{\tilde{u}_k^T M_{U_h} \tilde{u}_k},$$

$k = 1, 2, \ldots, K$, respectively. Hence only the dominant eigenvalues and eigenvectors of $A\mathbf{v} = \alpha B\mathbf{v}$ are computed, and post-processed by multiplying with the mass matrices.

6.5 Overview of randomized linear algebra

This section presents necessary background on randomized linear algebra and eigenvalue computation, and its advantages in the context of the proposed sensitivity analysis. The articles [48, 101] are utilized for this review and an algorithm from [101] is presented.

Randomized linear algebra has recently emerged as a powerful tool in scientific computation. Randomized methods have been developed for many of the classical problems such as solving

linear systems of equations, computing eigenvalues, and computing SVD's. The utility of randomized methods is that they permit a reordering of the computation which may better exploit computing architectures. Most traditional algorithms are inherently serial, for instance, constructing Krylov subspaces require sequential matrix-vector products since each vector is formed using the previous ones. In contrast, randomized methods may require a comparable number of matrix-vector products which can be computed in parallel. They also enable distributed memory computation with minimal communication overhead. These features motivate the use of randomized methods in this work.

A randomized method seeks to project a large matrix onto a low dimensional subspace through random sampling of its range space, then performs the desired computation in the low dimensional subspace. Consider a generic matrix A. One may sample independent random vectors $\{\mathbf{v}_i\}_{i=1}^{\ell}$ and compute $A\mathbf{v}_i$ for $i = 1, 2, \ldots, \ell$. This computation is easily parallelized because the $\mathbf{v}_i$'s are independent. For modest values of ℓ, the vectors $\{A\mathbf{v}_i\}_{i=1}^{\ell}$ will approximate the eigenspace corresponding to the dominant eigenvalues; the reader is referred to [48] for a more complete theoretical justification. Approximating the eigenspace with $\text{span}\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_\ell\}$ may not be as accurate as using a ℓ dimensional Krylov subspace, but can be sufficiently accurate to be useful when high precision is not required. Computation may be performed on a smaller matrix formed from $\{A\mathbf{v}_i\}_{i=1}^{\ell}$ at a cost which is negligible compared to the cost of applying A. In this way, the computational bottleneck of sequentially applying A (as is typically required in traditional algorithms) is now embarrassingly parallel.

As described in Section 6.4, it is proposed to compute an SVD by reformulating it as a symmetric generalized eigenvalue problem. An algorithm from [101] is given below to provide the necessary background for this approach. Consider the generalized symmetric eigenvalue problem

$$A\mathbf{v} = \alpha B\mathbf{v} \tag{6.13}$$

where $A \in \mathbb{R}^{M\times M}$ is a symmetric matrix and $B \in \mathbb{R}^{M\times M}$ is a symmetric positive definite matrix, $M = p + n$ is the sum of dimensions of $\mathcal{X}_h$ and U_h. It is known that the eigenvalues are real and the eigenvectors are orthogonal with respect to the B inner product, $\langle \mathbf{v}, \mathbf{w}\rangle_B = \mathbf{v}^T B\mathbf{w}$.

Algorithm 6 presents the "two-pass algorithm" from [101]. It inputs the desired number of eigenvalues $2K$ (the eigenvalues of (6.12) come in positive/negative pairs so $2K$ eigenvalues are needed to determine the K largest eigenvalues) and an oversampling factor L, and returns an approximation to the $2K$ eigenvalues of greatest magnitude and their corresponding eigenvectors. The oversampling factor L is a user defined parameter which scales the cost versus accuracy. Typically $L < 20$ (or even $L < 10$) is sufficient, see [48, 101] and references therein.

In Line 1 of Algorithm 6, one typically samples $\mathbf{v}_i \in \mathbb{R}^M$ by drawing each entry of $\mathbf{v}_i$ independently from a standard normal distribution. Line 2 is embarrassingly parallel since each

$\mathbf{v}_i$ is independent. The resulting vectors will define the subspaces over which the eigenvectors may be computed. In Line 3, the vectors $\mathbf{y}_i$, $i = 1, 2, \ldots, 2K + L$ are collected together to form a matrix $Y \in \mathbb{R}^{M\times(2K+L)}$ and the QR decomposition of Y is computed with respect to B inner products, Algorithm 7 below details this procedure. The columns of the matrix $Q \in \mathbb{R}^{M\times(2K+L)}$ are a B orthogonal basis for $\text{span}\{B^{-1}A\mathbf{v}_1, B^{-1}A\mathbf{v}_2, \ldots, B^{-1}A\mathbf{v}_{2K+L}\}$. The matrix A is applied to each column of Q in Line 4, this is also embarrassingly parallel. After computing $AQ \in \mathbb{R}^{M\times(2K+L)}$, Q^T is applied on the left to form $T \in \mathbb{R}^{(2K+L)\times(2K+L)}$, a small square matrix. This may be viewed as a projection of A using B inner products. The eigenvalue decomposition of T is computed in Line 5, this is computationally light since T is small. The eigenvalues of T approximate the dominant eigenvalues of (6.13). To approximate the eigenvectors of (6.13), Q is applied to the eigenvectors $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{2K}\}$ of T in Line 6. Then $Q\mathbf{w}_i$, $i = 1, 2, \ldots, 2K$ are B orthogonal vectors which approximate the eigenvectors of (6.13).

In summary, the computational cost of the algorithm is dominated by Lines 2 and 4 where $B^{-1}A$ and A are applied; however, computing these $2K + L$ matrix vector products is embarrassingly parallel which enables exploitation of computational resources. In fact, since $2K + L$ typically is not too large (for instance, less than 20), the functions which define A and B^{-1} may also be parallelized, giving two levels of parallelism in the algorithm.

Algorithm 6 Randomized Two-Pass Generalized Eigenvalue Algorithm

Input: desired number of eigenvalues $2K$, oversampling factor L
1: sample independent random vectors $\mathbf{v}_i \in \mathbb{R}^M$, $i = 1, 2, \ldots, 2K + L$
2: compute $\mathbf{y}_i = B^{-1}A\mathbf{v}_i$, $i = 1, 2, \ldots, 2K + L$
3: compute QR factorization $Y = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{2K+L}) = QR$ where $Q^TBQ = I$
4: form $T = Q^TAQ$
5: compute eigenvalue decomposition $T = VEV^T$, $V = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{2K+L})$
6: form $Q\mathbf{w}_i$, $i = 1, 2, \ldots, 2K$
Return: approximate eigenvalues from E and eigenvectors $Q\mathbf{w}_i$, $i = 1, 2, \ldots, 2K$

There are multiple ways to compute a QR factorization with respect to B inner products, see [101] for more details. Algorithm 7 below is adopted from [101] and used in the implementation of Line 3 in Algorithm 6. Notice that the operations in Algorithm 7 are all standard and can be efficiently executed using existing linear algebra libraries. Since the computational cost of applying B is small relative to applying A, and $2K + L$ is relatively small, the cost of using Algorithm 7 in Line 3 of Algorithm 6 is negligible.

The benefits of using a randomized eigenvalue solver in the context of the current sensitivity analysis are threefold:

Algorithm 7 QR Factorization with B Inner Products

Input: matrix Y, symmetric positive definite matrix B
1: compute a QR decomposition $Y = ZS$ where $Z^T Z = I$, S upper triangular
2: form $C = Z^T BZ$
3: compute a Cholesky factorization $C = R_C^T R_C$
4: form $Q = ZR_C^{-1}$
5: form $R = R_C S$
Return: Q,R, such that $Y = QR$ and $Q^T BQ = I$

1. The bottleneck in the eigenvalue computation is applying A, which is embarrassingly parallel with the randomized solver. This is particularly useful for large scale problems where the parallel efficiency of the underlying linear algebra constructs decreases with a large number of processors. Better parallel efficiency may be attained by distributing the matrix vector products.

2. The randomized solver is particular well suited for problems with a low rank structure. A low rank structure corresponds to having greatest sensitivity in a small subset of the high dimensional parameter space. Such low rank structures in the parameter space are commonly observed in practice.

3. Low precision may be permitted when computing sensitivities since they typically are not used for additional computation downstream. The oversampling factor L provides an easy mechanism to control the computational cost.

6.6 Software implementation

This section provides an overview of the implementation and highlights its important features. The software is implemented in the Rapid Optimization Library (ROL) [67] of Trilinos [54], a collection of C++ libraries for scientific computation. The implementation is matrix free, has two levels of parallelism (three if multiple samples are taken in parameter space, see Section 6.8), and may be easily adapted to variety of applications. Its parallelism and matrix free design make it scalable.

Algorithm 6 provides an outline of the implementation. Lines 1, 3, 5, and 6 in Algorithm 6 are easily executed through calls to standard libraries and applying B to vectors, so it is sufficient to focus on applying A, B, and B^{-1}. Since the action of B^{-1} on a vector is easily computed with iterative linear solvers, it is sufficient to focus on applying A and B, which requires the ability to apply $M_{\mathcal{X}_h}$, M_{U_h}, Π, $\mathcal{K}^{-1}$, and $\mathcal{B}$. The subsections below highlight the implementation of the action of each of these operators.

The optimization problem (6.1) is solved using the PDE-constrained optimization tools in ROL. It provides an interface to evaluate the objective function and constraint, and the action of their derivatives. The ROL implementation also provides access to the entries of $M_{\mathcal{X}_h}$ and M_{U_h} through calls to the underlying finite element library. These tools are leveraged to efficiently implement the action of the operators needed for sensitivity analysis.

6.6.1 Applying $M_{\mathcal{X}_h}$ and M_{U_h}

The entries of $M_{\mathcal{X}_h}$ and M_{U_h} are easily accessible through the underlying finite element library. Their action on vectors is implemented in an efficient matrix free, and parallelizable, manner.

6.6.2 Applying Π

The operator Π is a projection which inputs the state, control, adjoint triple, and return the control. Hence its implementation is trivial. The action of its adjoint, which is needed in A, is also trivial to implement.

6.6.3 Applying $\mathcal{K}^{-1}$

Applying $\mathcal{K}^{-1}$ is the most computationally intensive portion of the sensitivity computation. Its action requires the solution of a large linear system whose dimension is the sum of degrees of freedom from the state, control, and adjoint variables. It may be rewritten as the minimization of a quadratic objective subject to a linear constraint [11]. All operators in the objective and constraint are composed from the objective Hessian and constraint Jacobian of (6.1), which may be accessed through ROL. The quadratic linear minimization problem is solved using ROL and leverages the existing implementation used when solving (6.1).

6.6.4 Applying $\mathcal{B}$

The operator $\mathcal{B}$ is the term in the sensitivity computation which encodes the parameter uncertainty. The action of $\mathcal{B}$ on $\mathbf{v}$ is the directional derivative of the Lagrangian's (6.2) gradient in the direction $\mathbf{v}$. ROL provides access to the Lagrangian's gradient and $\mathcal{B}\mathbf{v}$ is approximated with finite differences. In many applications, the parameters $\mathbf{x}$ appears linearly as they may be coefficients multiplying a state or differential operator in the PDE. In this case, the finite difference directional derivative is exact. Future developments may include using algorithmic differentiation [44] to compute the parameter derivatives efficiently when the dependence is nonlinear.

The adjoint of $\mathcal{B}$ must also be applied when applying A. In the implementation, this is done using the property that the i^{th} entry of a matrix vector product $C^T\mathbf{v}$ is given by $\mathbf{v}^T C\mathbf{s}_i$, where

$\mathbf{s}_i$ is a vector whose has 1 in its i^{th} entry and 0 elsewhere. Hence the action of the adjoint of $\mathcal{B}$ in the discretized coordinate representation may be computed by applying $\mathcal{B}$ p times (p is the dimension of $\mathcal{X}_h$). Since the computational cost of applying $\mathcal{K}^{-1}$ is significantly more than applying $\mathcal{B}$, the computational cost of this approach is negligible. Future developments may include using algorithmic differentiation to efficiently compute the action of the adjoint $\mathcal{B}$.

6.7 Formulation of global sensitivity indices

The previous sections developed a computationally efficient and scalable algorithm (and implementation) to perform local sensitivity analysis with fixed parameters $\mathbf{x}_0 \in \mathcal{X}_h$. The goal of this section is to formally define global sensitivity indices which do not depend on a specific $\mathbf{x}_0$, but rather a probability distribution for the parameters that reflects their uncertainty. This section highlights challenges, identifies assumptions, and formally defines global sensitivity indices. Computational complexity typically prohibits accurate estimation of these global sensitivity indices. Section 6.8 builds from this formulation to provide guidelines for computing, visualizing, and interpreting sensitivities in practice.

The ideas of derivative-based global sensitivity measures are mimicked for this extension. There are two important, and related, points to consider when defining global sensitivity indices:

- The term global sensitivity index is referring to being global in the parameter space, not in the optimization variables (y, u). In general, only local optimally of the optimization problem (6.1) may be ensured, but analysis may be performed at different $\mathbf{x}_0$'s thus making the analysis global in $\mathcal{X}_h$.

- For a fixed $\mathbf{x}_0$, Theorem 12 only ensures that the Fréchet derivative is well defined for a given local minimum (y_0, u_0). For this reason, it is necessary to account for different local minima associated with each $\mathbf{x}_0$.

To this end, two random vectors/fields are introduced.

Let $\mathbf{X}$ be a random vector of length p (the dimension of $\mathcal{X}_h$) encoding the uncertainty in the parameters. The user has freedom to define $\mathbf{X}$ to capture whatever uncertainty or statistical characterization is known in the application.

As highlighted above, the presence of multiple local minima poses an additional challenge beyond what is typically considered in global sensitivity analysis. To formally defined sensitivity indices, let $\mathbf{Z}$ be a random vector whose length is $\dim(Y_h) + \dim(U_h)$, we will use realizations of $\mathbf{Z}$ as the initial iterates when solving (6.1). Assume that a deterministic optimization algorithm is used to solve (6.1) so that each realization of $\mathbf{Z}$ is uniquely mapped to local minimum (y_0, u_0). This ensures that the global sensitivity index defined below is well-posed. If the user wishes to

have zero as the initial iterate, then $\mathbf{Z}$ will equal zero with probability 1. Otherwise, the user may permit a random initial iterate in which case there may be different local minima found for different realizations of $\mathbf{Z}$.

For a given realization $\mathbf{x}_0$ of $\mathbf{X}$ and z_0 of $\mathbf{Z}$, Theorem 12 gives a unique Fréchet derivative whose SVD may be computed. Let $(\sigma_k(\mathbf{X},\mathbf{Z}), \mathbf{x}_k(\mathbf{X},\mathbf{Z}), u_k(\mathbf{X},\mathbf{Z}))$, $k = 1, 2, \ldots, K$ be the K leading random singular triplets that come from computing the SVD of (6.8) for each realization of $\mathbf{X}, \mathbf{Z}$. Then define the global sensitivity index for the i^{th} parameter as

$$\mathcal{S}_i = \mathbb{E}_{\mathbf{X},\mathbf{Z}}\left[\sum_{k=1}^{K} \sigma_k(\mathbf{X},\mathbf{Z})\left|(M_{\mathcal{X}_h}\mathbf{x}_k(\mathbf{X},\mathbf{Z}))_i\right|\right] \tag{6.14}$$

for $i = 1, 2, \ldots, p$. The term $\left|(M_{\mathcal{X}_h}\mathbf{x}_k(\mathbf{X},\mathbf{Z}))_i\right|$ is the absolute value of the i^{th} entry of the random vector $M_{\mathcal{X}_h}\mathbf{x}_k(\mathbf{X},\mathbf{Z})$; the absolute value is taken since the singular vector is only unique up to a sign. The singular vector is multiplied by $M_{\mathcal{X}_h}$ to properly map the coordinate sensitivity to the sensitivity of the basis function in $\mathcal{X}_h$. This definition of sensitivity has a similar intuition as the activity scores proposed in [20].

The sensitivity index $\mathcal{S}_i$ may be related to the solution of the optimization problem (6.1) as follows. Let $g : \mathbb{R}^p \times \mathbb{R}^{\dim(Y_h)+\dim(U_h)} \to U_h$ be defined by $g(\mathbf{x},\mathbf{z}) = \Pi f(E_{\mathcal{X}_h}\mathbf{x})$, where $\mathbf{z}$ is an initial iterate which is assumed to uniquely define the solution of (6.1) for the given $\mathbf{x}$. Theorem 14 bounds the expected difference in control solutions in terms of sensitivity indices.

Theorem 14. *Let $i \in \{1, 2, \ldots, p\}$ and assume that*

- *the Hessian of g with respect to $\mathbf{x}$, $g_{\mathbf{x},\mathbf{x}}$, exists,*
- *$\exists C > 0$ such that $||g_{\mathbf{x},\mathbf{x}}((\mathbf{x},\mathbf{z}),(\mathbf{x},\mathbf{z}))||_U \leq C \ \forall (\mathbf{x},\mathbf{z}) \in \mathbb{R}^p \times \mathbb{R}^{dim(Y_h)+dim(U_h)}$,*
- *$\exists \epsilon_i > 0$ such that*

$$\sum_{k=K+1}^{\min\{dim(U_h),p\}} \sigma_k(\mathbf{X},\mathbf{Z})|(\phi_i, E_{\mathcal{X}_h}\mathbf{x}_k(\mathbf{X},\mathbf{Z}))_{\mathcal{X}}| \leq \epsilon_i \text{ almost everywhere.}$$

Then

$$\mathbb{E}_{\mathbf{X},\mathbf{Z}}\left[||g(\mathbf{X}+\delta\mathbf{c}_i,\mathbf{Z}) - g(\mathbf{X},\mathbf{Z})||_U\right] \leq |\delta|\mathcal{S}_i + |\delta|\epsilon_i + \frac{C}{2}\delta^2,$$

where $\mathbf{c}_i \in \mathbb{R}^p$ equals 1 in its i^{th} entry and 0 elsewhere.

Proof. Let $(\mathbf{x},\mathbf{z})$ be a fixed realization of $(\mathbf{X},\mathbf{Z})$. Taylor's theorem implies that

$$||g(\mathbf{x}+\delta\mathbf{c}_i,\mathbf{z}) - g(\mathbf{x},\mathbf{z})||_U \leq ||\mathcal{D}E_{\mathcal{X}_h}\delta\mathbf{c}_i||_U + \frac{C}{2}\delta^2.$$

Leveraging the Singular Value Decomposition of $\mathcal{D}$ and coordinate transformations yields

$$||g(\mathbf{x}+\delta\mathbf{c}_i,\mathbf{z})-g(\mathbf{x},\mathbf{z})||_U \leq \sum_{k=1}^{\min\{\dim(U_h),p\}} \sigma_k|\delta||(\phi_i, E_{\mathcal{X}_h}\mathbf{x}_k)_{\mathcal{X}}|||u_k||_U + \frac{C}{2}\delta^2$$
$$\leq |\delta|\sum_{k=1}^{K}\sigma_k|(M_{\mathcal{X}_h}\mathbf{x}_k)_i| + |\delta|\epsilon_i + \frac{C}{2}\delta^2,$$

where $(\sigma_k, \mathbf{x}_k, u_k)$, $k = 1, 2, \ldots, \min\{\dim(U_h), p\}$ are the singular triplets of $\mathcal{D}$. Taking the expectation over all $(\mathbf{x}, \mathbf{z})$ completes the proof. □

The constant C in Theorem 14 corresponds to the nonlinearity of the operator g, if g is approximately linear then C will be small. The constant ϵ_i corresponds to the SVD truncation error, which is small. Theorem 14 implies that the average change in the control solution when the nominal parameters are perturbed in the i^{th} coordinate direction is bounded by the sensitivity index for the i^{th} parameter, the SVD truncation error, and a measure of the nonlinearity of g.

6.8 Computation, visualization, and interpretation of sensitivities

Estimating (6.14) accurately through sampling based approaches may require extensive computational effort. However, useful information may be obtained by computing local sensitivities at a sparse number of samples from the parameter space. This section details practical considerations computing, visualizing, and interpreting samples of local sensitivities. The proposed approach is demonstrated in Section 6.9.

Assume that control solutions and local sensitivities have been computed at N (chosen based upon the user's computational budget) different realizations of $\mathbf{X}, \mathbf{Z}$, denote them as $\overline{\mathbf{x}}^j, \overline{z}^j$, $j = 1, 2, \ldots, N$. For each $j = 1, 2, \ldots, N$, this gives,

- the state y^j_{opt} and control u^j_{opt} solution of (6.1) with initial iterate $\overline{z}^j$ when $\mathbf{x} = \overline{\mathbf{x}}^j$,
- the K leading singular triples of (6.8), $(\sigma^j_k, \mathbf{x}^j_k, u^j_k)$, $k = 1, 2, \ldots, K$.

The spectrum of (6.8) summarizes its low rank structure (or lack thereof). The singular values σ^j_k, $k = 1, 2, \ldots, K$, $j = 1, 2, \ldots, N$, may be visualized in a scatter plot which readily reveals the structure of the leading singular values for different samples in parameter space. Ideally a decay in the singular values will be observed for each fixed j indicating a low rank structure. If such a low rank structure exists, the K singular values and singular vectors may be used to analyze the sensitivities.

Define the local sensitivity index as

$$\mathcal{S}_i^j = \sum_{k=1}^{K} \sigma_k |(M_{\mathcal{X}_h} \mathbf{x}_k^j)_i| \tag{6.15}$$

for $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, N$. The scalar $\mathcal{S}_i^j$ measures the influence of parameter i, i.e. the i^{th} component of the discretized parameter vector, for the j^{th} parameter sample. Then $\mathcal{S}_i^j$, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, N$, may be easily visualized in a scatter plot which reveals the relative influence of the various parameters and the variability of their influence over the parameter samples $\overline{\mathbf{x}}^j$, $j = 1, 2, \ldots, N$.

A user may consider computing the average of $\mathcal{S}_i^j$, $j = 1, 2, \ldots, N$, which estimates (6.14), $i = 1, 2, \ldots, p$. Rather than adopting this approach, the user should consider the empirical distribution of $\mathcal{S}_i^j$, or at least its mean and standard deviation; this is inline with the classical Morris method for screening [57, 86]. The reason for this suggestion is twofold. First, computational cost prohibits taking enough samples to reliably estimate the local sensitivities averaged over parameter space, i.e. the average of $\mathcal{S}_i^j$, $j = 1, 2, \ldots, N$, as $N \to \infty$. The variability in the samples indicates indicates how the sensitivities for different parameters may be compared, and, as in the Morris method, gives a measure of the nonlinearity of f. Second, $\mathcal{S}_i^j$ depends upon which local minimum, (y_{opt}^j, u_{opt}^j) of (6.1), the optimization routine determines. The challenges of global optimization are well known and for this reason it is acknowledged that this as a undesirable property of $\mathcal{S}_i^j$ which cannot be alleviated. Nonetheless, considering the empirical distribution of $\mathcal{S}_i^j$ allows the user to observe variability that may arise from sensitivities computed at different local minima. Since variability in $\mathcal{S}_i^j$ is contributed by the parameter sampling and the nominal local minimum, both of which are meaningful, the empirical distribution of $\mathcal{S}_i^j$ gives a practical strategy to visualize and interpret these measures of sensitivity.

Along with the spectrum and parameter sensitivity information above, the resulting changes in the control strategy when the parameters are perturbed may be visualized as well. In particular, u_k^j is the change in the control strategy if the parameter vector $\overline{\mathbf{x}}^j$ is perturbed in the direction $\mathbf{x}_k^j$. The samples u_k^j may be visualized by overlaying them on a plot, or plotting statistical quantities computed from the sample $\{u_k^j\}_{j=1}^N$. In both cases, it indicates which part of the control strategy will change as the uncertain parameters change.

The sensitivity information discussed above indicates which parameters the control strategy is most sensitive too, and how it will change if they are perturbed. The control problem (6.1) is solved at each sample as part of the sensitivity computation. The variability of the state and control strategy over the samples may be observed as a by-product of the computation and visualized by overlaying them on a plot, or plotting statistical quantities computed from the samples.

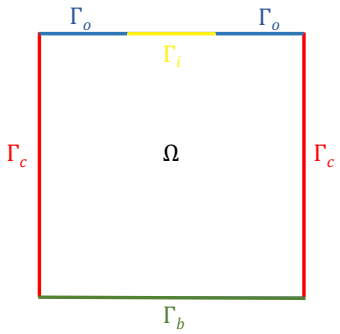

Figure 6.2 Computation domain and boundaries for (6.16).

6.9 Numerical results

In this section the proposed sensitivity analysis is demonstrated on the steady state nonlinear multi-physics Boussinesq flow equations in two spatial dimensions. This arises from an application in high pressure chemical vapor deposition (CVD) reactors, see [60]; the control problem is implemented in the ROL PDE-OPT examples [67]. Reactant gases are injected in the top of a reactor and flow downwards to create a epitaxial film on the bottom. Vorticities created by buoyancy-driven convection inhibit some gases from reaching the bottom of the reactor. Thermal fluxes are controlled on the side walls of the reactor in order to minimize the vorticity. Consider the control problem,

$$\min_{y,u} \frac{1}{2}\int_\Omega (\nabla \times v) d\Omega + \frac{\gamma}{2}\int_{\Gamma_c} u^2 d\Gamma \tag{6.16}$$

$s.t.$

$$\begin{aligned}
&-\epsilon(\mathbf{x})\nabla^2 v + (v\cdot\nabla)v + \nabla q + \eta(\mathbf{x})Tg = 0 && \text{in } \Omega\\
&\nabla\cdot v = 0 && \text{in } \Omega\\
&-\kappa(\mathbf{x})\Delta T + v\cdot\nabla T = 0 && \text{in } \Omega\\
&T = 0 \quad \text{and} \quad v = v_i && \text{on } \Gamma_i\\
&\kappa(\mathbf{x})\frac{\partial T}{\partial n} = 0 \quad \text{and} \quad v = v_o && \text{on } \Gamma_o\\
&T = T_b(\mathbf{x}) \quad \text{and} \quad v = 0 && \text{on } \Gamma_b\\
&\kappa(\mathbf{x})\frac{\partial T}{\partial n} + \nu(\mathbf{x})(u-T) = 0 \quad \text{and} \quad v = 0 && \text{on } \Gamma_c
\end{aligned}$$

where $\Omega = (0,1)\times(0,1)$, $\Gamma_i = [1/3, 2/3]\times\{1\}$, $\Gamma_o = [0,1/3]\times\{1\}\cup[2/3,1]\times\{1\}$, $\Gamma_b = [0,1]\times\{0\}$, and $\Gamma_c = \{0,1\}\times[0,1]$. Figure 6.2 depicts the domain and boundaries. The state y consists of

horizontal and vertical velocities $v = (v_1, v_2)$, the pressure q, and the temperature T; the control u is a function defined on the left and right boundaries of Ω, their union is denoted Γ_c. The deterministic inflow and outflow conditions v_i and v_o are given by

$$v_i(s) = -4\left(s - \frac{1}{3}\right)\left(\frac{2}{3} - s\right)$$

and

$$v_o(s) = \begin{cases} 2\left(\frac{1}{3} - s\right)s & \text{if } s \in \left[0, \frac{1}{3}\right] \\ 2\left(s - \frac{2}{3}\right)(1 - s) & \text{if } s \in \left[\frac{2}{3}, 1\right] \end{cases}.$$

Uncertainties enter the the model through the finite dimensional vector $\mathbf{x} \in \mathbb{R}^p$ where $p = 2m_b + 2m_\ell + 2m_r + 3$. In particular, the boundary term T_b is defined through the sum

$$T_b(s, \mathbf{x}) = 1 + 0.2\sum_{k=1}^{m_b}\left(\mathbf{x}_k\frac{\sin(\pi k s)}{k} + \mathbf{x}_{m_b+k}\frac{\cos(\pi k s)}{k}\right). \tag{6.17}$$

In addition, the boundary term $\nu(s, \mathbf{x})$, a function defined on Γ_c, is expressed as a sum in the form of (6.17) with $2m_\ell$ parameters in the sum defining the left boundary term and $2m_r$ parameters in the sum defining the right boundary term. Along with these $2m_b + 2m_\ell + 2m_r$ parameters which define T_b and ν, there are three parameters, $\mathbf{x}_{p-2}, \mathbf{x}_{p-1}, \mathbf{x}_p$, which appear in the uncertain scalar quantities ϵ, η, and κ, as

$$\epsilon = \frac{1}{Re} = \frac{1 + 0.05\mathbf{x}_{p-2}}{100}, \qquad \kappa = \frac{1}{RePa} = \frac{1 + 0.05\mathbf{x}_{p-1}}{72}$$
$$\text{and } \eta = \frac{Ge}{Re^2} = 1 + 0.05\mathbf{x}_p,$$

where Re, Ge, and Pa are the Reynolds number, Grashof number, and Prandtl number respectively. Each $\mathbf{x}_k$, $k = 1, 2, \ldots, p$, is assumed to be independent and uniformly distributed on $[-1, 1]$. The initial iterate is taken to be zero for all samples; similar results were found using a random initial iterate.

The PDE is discretized with finite elements on a 99x99 rectangular mesh. The velocity and pressure are represented with the Q2-Q1 Taylor-Hood finite element pair and the temperature is represented with the Q2 finite element. Taking $m_b = m_\ell = m_r = 25$ yields a total of $p = 153$ uncertain parameters. The deterministic control problem is solved using the full space composite step algorithm in ROL with the control penalty $\gamma = 0.01$. Figure 6.3 displays the uncontrolled (left) and controlled (right) velocity field with parameters $x_k = 0$, $k = 1, 2, \ldots, 153$. The undesired vorticities are observed in the uncontrolled velocity field and are reduced by the control strategy.

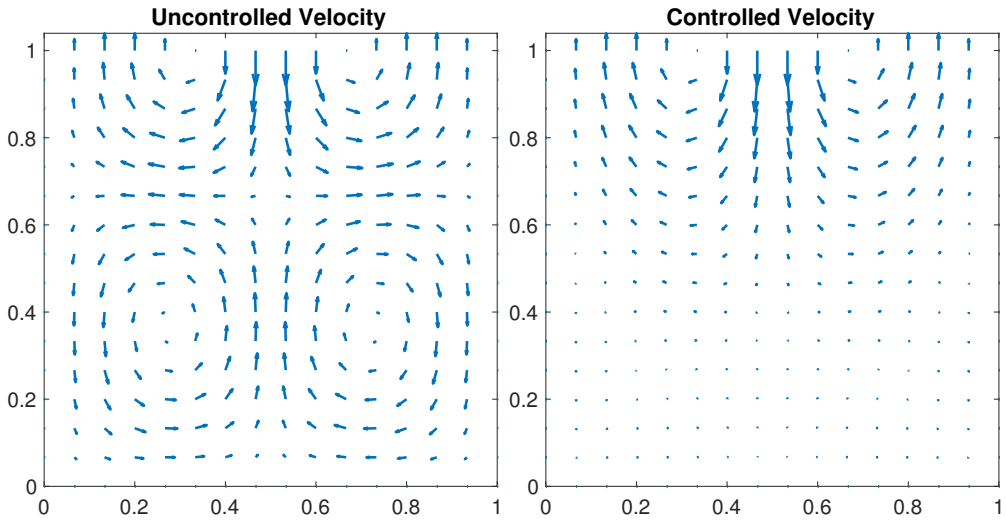

Figure 6.3 Uncontrolled (left) and controlled (right) velocity field.

Local sensitivities are evaluated at $N = 20$ samples from parameter space. Figure 6.4 displays the optimal control solutions for these 20 samples. The left and right panels display the controller on the left and right boundary, respectively. Each curve corresponds to the control solution for a different parameter sample. There is significant variability in the solutions which indicates a strong dependence of the controller on the uncertain parameters. The objective in the sensitivity analysis is to determine which parameters cause the greatest changes in the controller so that uncertainty quantification and robust optimization may focus on them rather than the full set of 153 parameters.

The leading $K = 4$ singular triples of (6.8) are computed and the approach presented in Section 6.8 is followed to analyze them. Since the generalized eigenvalue problem is an augmented system, the leading $2K = 8$ eigenvalues must be computed to extract the leading 4 singular values; an oversampling factor of $L = 8$ is used. Figure 6.5 shows the leading 4 singular values from each of the 20 parameter samples. Each vertical slice in Figure 6.5 gives the 4 singular values for the fixed parameter sample. Observe that there are 2 dominant singular values (thus validating that $K = 4$ is sufficiently large) and that the singular values do not vary significantly over the different parameter samples.

The sensitivity indices (6.15) are computed and plotted in Figure 6.6. There are 20 circles in each vertical slice of Figure 6.6 corresponding to the local sensitivity index for a fixed parameter over the 20 parameter samples. Observe several interesting features:

- The local sensitivity analysis yields similar results for each parameter sample.
- Only around 10% of the uncertain parameters exhibit significant influence on the control

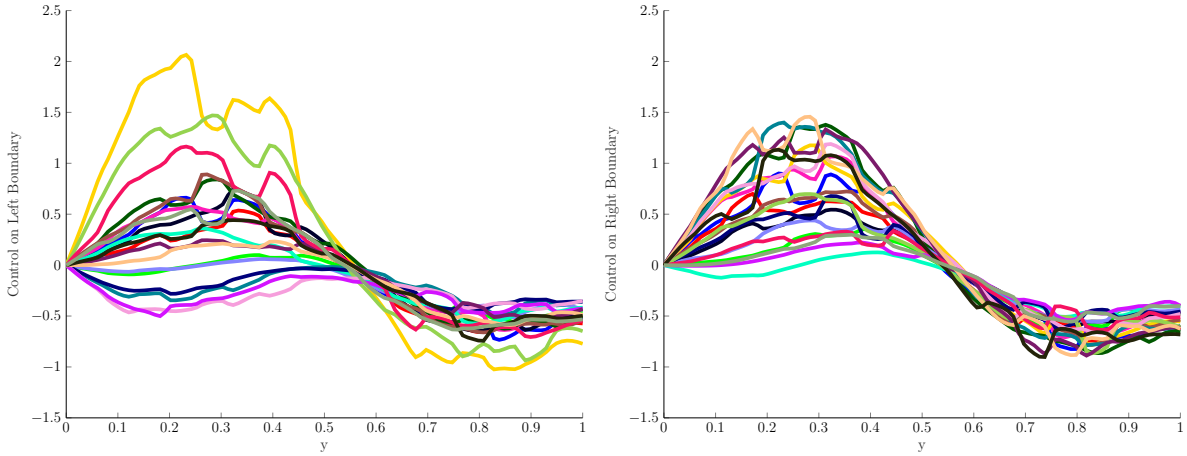

Figure 6.4 Control solutions for (6.16) corresponding to 20 different parameter samples. The left and right panels are the controllers on the left and right boundaries, respectively. Each curve is a control solution for a given parameter sample.

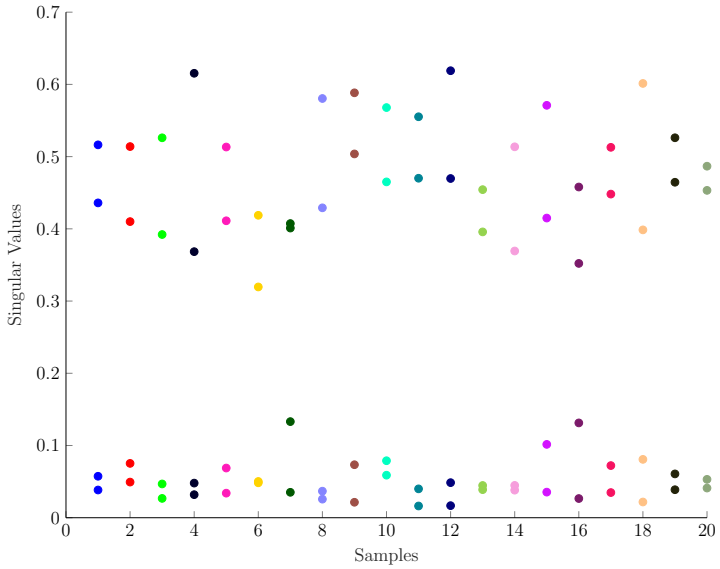

Figure 6.5 Leading 4 singular values of $\mathcal{D}$ at 20 different parameter samples. Each vertical slice corresponds to the leading 4 singular values for a fixed sample.

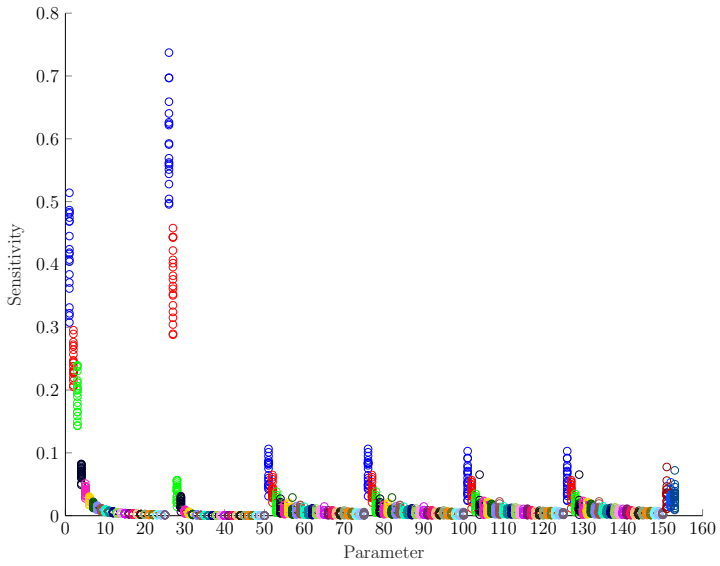

Figure 6.6 Parameter sensitivities (6.15) for the 153 uncertain parameters in (6.16). The 20 circles in each vertical slice indicates the sensitivity index for a fixed parameter as it varies over the 20 parameter samples. The repeating color scheme indicates the grouping of parameters as they correspond to sine and cosine components of each boundary condition.

strategy.

- The bottom boundary condition, T_b, appears to have the greatest influence on the control strategy.

- The cosine components of T_b appear to be more important in the first two frequencies, the sine component appears more important in the third frequency.

To complement these parameter sensitivities, Figure 6.7 displays the singular vectors, u_k, $k = 1, 2$, corresponding to the leading parameter perturbations $\mathbf{x}_k$, $k = 1, 2$. The top left and top right panels of Figure 6.7 show the first singular vector on the left and right boundaries, respectively; the bottom left and bottom right panels are the second singular vector on the left and right boundaries, respectively. The singular vectors u_1 and u_2 may be interpreted as the change in the control strategy if the parameters are perturbed according to the singular vectors $\mathbf{x}_1$ and $\mathbf{x}_2$. Observe that the control strategy will change more near the bottom of the domain, an unsurprising result since the greatest sensitivity is in the bottom boundary condition.

The computation was performed using 20 compute nodes, each containing 16 processors. By taking 20 parameter samples and 16 vectors in the randomized eigenvalue solver, the embarrassingly parallel loop in the eigenvalue solver distributed the computation over all 320 processors thus reducing the overall wall clock time to approximately one matrix vector product

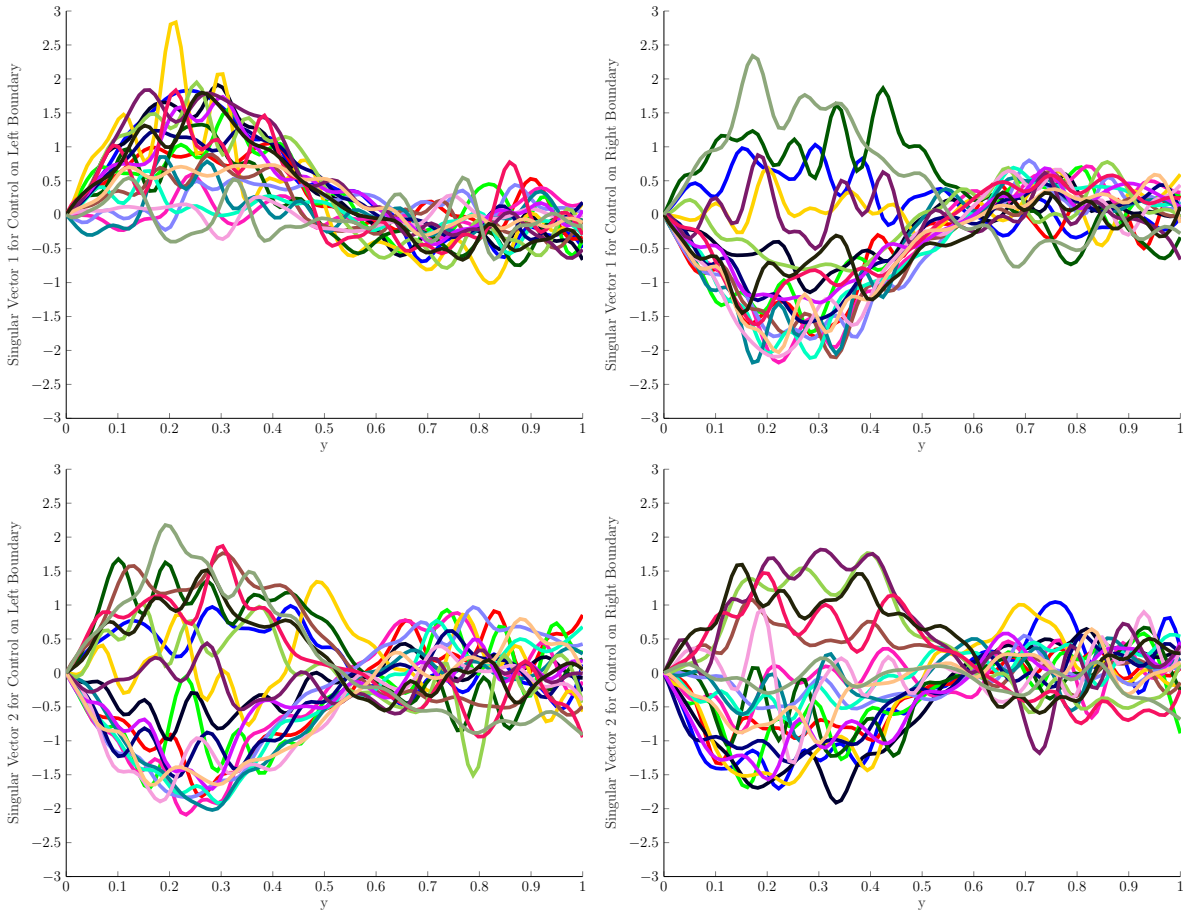

Figure 6.7 First two singular vectors u_k, $k = 1, 2$, of $\mathcal{D}$, i.e. $\mathcal{D}\mathbf{x}_k = \sigma_k u_k$. The top (bottom) row shows the first (second) singular vector on the left and right boundaries, respectively. Each curve corresponds to a different parameter sample.

in Line 2 and one matrix vector product in Line 4 of Algorithm 6. Since the KKT solve (6.6) dominates the computational cost, this yields that the total wall clock time for computing 20 local sensitivities is approximately equal to 4 KKT solves.

6.10 Conclusion

In this chapter the work of [11] has been extended to facilitate GSA for large scale PDE-constrained optimization problems. By coupling a C++ implementation and randomized eigenvalue solver, a scalable software infrastructure has been developed to determine the sensitivity of the optimal control solution to changes in uncertain parameters. The framework is able to exploit low dimensional structure which is commonly found in high dimensional parameter spaces, and is scalable in both the complexity of the parameter space and the underlying PDE itself. Finding this low dimensional structure has many useful applications. For instance, performing uncertainty quantification and robust optimization (the ultimate goal) are challenging in high dimensions. Reducing the number of parameters may enable experimental and computation analysis which would otherwise be infeasible in the high dimensional parameter space. The sensitivities may also direct model development by identifying which model parameters (and hence corresponding physical effects) exhibit the greatest influence on the control strategy.

Computing these sensitivities requires solving the original optimization problem with fixed parameters, and multiple large linear system solves thereafter. The randomized eigenvalue solver allows the for loop over linear system solves to be parallelized thus reducing the wall clock time to approximately four large linear system solves. The primary limitation of the method is that the sensitivities are local in parameter space and hence need to be evaluated at several different parameter samples. Challenges in high dimensional sampling prohibit a complete exploration of parameter space; however, as observed in the numerical results, local sensitivities may be taken at a modest number of sample points and inferences may be drawn if the results do not change significantly between samples. Theoretical development is needed to aid in determining when sparse sampling is sufficient. A possible extension of this work is to perform adaptive sampling in the parameter space and/or utilize a multi-fidelity approach; this is made possible by the efficiency of the local sensitivity computation.

The framework was demonstrated on a steady state nonlinear multi-physics PDE in two spatial dimensions. Since the implementation is matrix free and parallelizable, it may be extended to more complex problems such as having three spatial dimensions, time dependence, and/or multiscale phenomena. The ability to evaluate the gradient and apply the Hessian from the original optimization problem is required; hence, if the optimization problem is computationally tractable then the sensitivity analysis will be as well. The software infrastructure is amenable to these extensions because of the underlying Trillinos constructs.

CHAPTER

7

CONCLUSION

This chapter highlights other extensions of GSA and questions which arise in a similar manner to those presented in this thesis. Existing work by other authors is highlighted and open questions are discussed.

A scalar quantity of interest was assumed in Chapter 1. In most applications, there may be multiple quantities of interest, spatial dependent quantities of interest, or time dependent quantities of interest; [30, 31, 82, 83, 119] address some of these extensions.

In the classical framework of Chapter 1, the model inputs are assumed to be a vector of parameters in $\mathbb{R}^p$. A natural extension is allow functional inputs which may correspond to spatially dependent parameters, for instance. One approach is to represent the functional input with its Karhunen Loeve expansion and truncate so that the function is parameterized by a finite number of parameters. The works of [58, 83] offer other possible extensions.

The assumption of statistically independent input variables has been critical for much of the theoretical developments in GSA. Chapter 2 provides a step toward developing a more robust theory with dependent variables for the Sobol' indices. Results such as Theorem 1 facilitate better understanding of how various GSA methods relate to one another. However, Theorem 1, along with other similar results [90], depend on the assumption of independence. Inequalities relating the different GSA methods with dependent variables are needed to better understand how to use and interpret them. Theoretical properties of derivative-based approaches are particularly interesting because of the computational efficiency they may afford when gradients are computed

through an adjoint equation.

In many cases, such as in Chapter 6, computational cost prohibits drawing enough samples to accurately estimate global sensitivity indices. In [97], the authors seek to address this challenge using a multi-fidelity approach. There is significant opportunity to utilize multi-fidelity methods for GSA when model evaluations are computationally expensive. Screening through sparse sampling is another area of exploration for problems with computationally costly models. This approach is adopted in Chapter 6. Additional research is needed to better understand how to sample and make inferences with a relatively small number of local sensitivities. Multi-fidelity approaches may be beneficial for designing sampling strategies in this case.

One of the most common approaches to GSA with computationally expensive models is computing the Sobol' indices of a surrogate model. There are a plurality of open questions concerning how to construct surrogate models for the purpose of GSA, and how error in the surrogate model propagates through the computation of sensitivity indices. The work of [61] is a first step toward answering these questions. One possible extension of the work in this thesis is applying the framework of Chapter 3 to uncertainties in the model rather than the PDF. This would provide a measure of the robustness of the Sobol' indices with respect to the surrogate model, giving a heuristic to measure how close the Sobol' indices of the surrogate model may be to the Sobol' indices of the true model.

BIBLIOGRAPHY

[1] Alexanderian, A., Winokur, J., Sraj, I., Srinivasan, A., Iskandarani, M., Thacker, W. C. & Knio, O. M. "Global sensitivity analysis in an ocean general circulation model: a sparse spectral projection approach". *Computational Geosciences* **16**.3 (2012), pp. 757–778.

[2] Arampatzis, G., Katsoulakis, M. & Pantazis, Y. "Accelerated sensitivity analysis in high-dimensional stochastic reaction networks". *PLOS ONE* **10** (2015), e0130825–1–e0130825–24.

[3] Beckman, R. J. & McKay, M. D. "Monte Carlo estimation under different distributions using the same simulation". *Technometrics* **29**.2 (1987), pp. 153–160.

[4] Bessac, J., Constantinescu, E. M. & Anitescu, M. "Stochastic simulation of predictive space–time scenarios of wind speed using observations and physical model outputs". *Ann. Applied Statistics* **12**.1 (2018), pp. 432–458.

[5] Bhatia, R. & Davis, C. "A better bound on the variance". *Mathematical Association of America* **107**.4 (2000), pp. 353–357.

[6] Blatman, G. & Sudret, B. "Efficient computation of global sensitivity indices using sparse polynomial chaos expansions". *Reliability Engineering & System Safety* **95**.11 (2010), pp. 1216–1229.

[7] Bolado-Lavin, R. & Badea, A. C. *Review of sensitivity analysis methods and experience for geological disposal of radioactive waste and spent nuclear fuel.* Tech. rep. European Commission, 2009.

[8] Borgonovo, E. "A new uncertainty importance measure". *Reliability Eng. Sys. Safety* **92** (2007), pp. 771–784.

[9] Borgonovo, E. & Iooss, B. "Moment-independent and reliability-based importance measures". *Handbook for Uncertainty Quantification.* Ed. by Ghanem, R., Higdon, D. & Owhadi, H. Springer, 2016, pp. 1265–1287.

[10] Borgonovo, E., Morris, M. D., & Plischke, E. "Functional ANOVA with multiple distributions: Implications for the sensitivity analysis of computer experiments". *SIAM/ASA J. Uncertain. Quantif.* **6**.1 (2018), pp. 397–427.

[11] Brandes, K. & Griesse, R. "Quantitative stability analysis of optimal solutions in PDE-constrained optimization". *Journal of Computational and Applied Mathematics* **206** (2007), pp. 908–926.

[12] Breiman, L., Friedman, J., Olshen, R. & Stone, C. *Classification and regression trees.* Wadsworth Advanced Books and Software, 1984.

[13] Brooks, S. P. & Gelman, A. "General methods for monitoring convergence of iterative simulations". *Journal of Computational and Graphical Statistics* **7**.4 (1998), pp. 434–455.

[14] Brooks, S. P. & Roberts, G. O. "Assessing convergence of Markov Chain Monte Carlo algorithms". *Statistics and Computing* **8** (1997), pp. 319–335.

[15] Büskens, C. & Griesse, R. "Parametric sensitivity analysis of perturbed PDE optimal control problems with state and control constraints". *Journal of Optimization Theory and Applications* **131**.1 (2006), pp. 17–35.

[16] Chastaing, G., Gamboa, F. & Prieur, C. "Generalized Hoeffding-Sobol' decomposition for dependent variables-Application to sensitivity analysis". *Electronic Journal of Statistics* (2012).

[17] Chastaing, G., Prieur, C. & Gamboa, F. "Generalized Sobol' sensitivity indices for dependent variables: numerical methods". *Journal of Statistical Computation and Simulation* (2014), pp. 1–28.

[18] Chick, S. E. "Input distribution selection for simulation experiments: Accounting for input uncertainty". *Operations Research* **49**.5 (2001), pp. 744–758.

[19] Constantine, P. *Active Subspaces: Emerging ideas for dimension reduction in parameter studies*. SIAM, 2015.

[20] Constantine, P. G. & Diaz, P. "Global sensitivity metrics from active subspaces". *Reliability Engineering & System Safety* **162** (2017), pp. 1–13.

[21] Cowles, M. K. & Carlin, B. P. "Markov Chain Monte Carlo convergence diagnostics: A comparative review". *Journal of the American Statistical Association* **91**.434 (1996).

[22] Crestaux, T., Maitre, O. L. & Martinez, J.-M. "Polynomial chaos expansion for sensitivity analysis". *Reliability Engineering & System Safety* **94**.7 (2009). Special Issue on Sensitivity Analysis, pp. 1161 –1172.

[23] Da Veiga, S. "Global sensitivity analysis with dependence measures". *Journal of Statistical Computation and Simulation* **85** (2015).

[24] Degasperi, A. & Gilmore, S. "Sensitivity analysis of stochastic models of bistable biochemical reactions". *Formal Methods for Computational Systems Biology*. Springer, 2008, pp. 1–20.

[25] Durrett, R. *Probability: Theory and examples*. 4th ed. Cambridge University Press, 2010.

[26] El Samad, H., Khammash, M., Petzold, L. & Gillespie, D. "Stochastic modelling of gene regulatory networks". *International Journal of Robust and Nonlinear Control* **15**.15 (2005), pp. 691–711.

[27] Fassbender, H. & Kressner, D. "Structured eigenvalue problems". *GAMM-Mitteilungen* **29**.2 (2006), pp. 297–318.

[28] Friedman, J. "Multivariate adaptive regression splines". *Ann. Stat.* **19** (1991), pp. 1–67.

[29] Friedman, J. *Fast MARS*. Tech. rep. 110. Laboratory for Computational Statistics, Department of Statistics, Stanford University, 1993.

[30] Gamboa, F., Janon, A., Klein, T. & Lagnoux, A. "Sensitivity analysis for multidimensional and functional outputs". *Electronic Journal of Statistics* **8** (2014), pp. 575–603.

[31] Gamboa, F., Klein, T. & Lagnoux, A. "Sensitivity analysis based on Cramér–von Mises distance". *SIAM/ASA Journal on Uncertainty Quantification* **6**.2 (2018), pp. 522–548.

[32] Gao, L., Bryan, B. A., Nolan, M., Connor, J. D., Song, X. & Zhao, G. "Robust global sensitivity analysis under deep uncertainty via scenario analysis". *Environmental Modelling & Software Software* **76** (2016), pp. 154–166.

[33] Geyer, C. J. "Importance sampling, simulated tempering, and umbrella sampling". Chapman and Hall/CRC, 2011. Chap. 11, pp. 295–312.

[34] Geyer, C. J. "Introduction to Markov Chain Monte Carlo". Chapman and Hall/CRC, 2011. Chap. 1, pp. 3–48.

[35] Geyer, C. J. & Thompson, E. A. "Annealing Markov Chain Monte Carlo with applications to ancestral inference". *Journal of the American Statistical Association* **90**.431 (1995), pp. 909–920.

[36] Gillespie, D. "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". *J. Comput. Phys.* **22** (1976), pp. 403–434.

[37] Gillespie, D. "Exact stochastic simulation of coupled chemical reactions". *J. Phys. Chem.* **81** (1977), pp. 2340–2361.

[38] Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L. & Johnson, N. A. "Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds". *TEST* **17**.2 (2008), pp. 211–235.

[39] Griesse, R. & Volkwein, S. "Parametric sensitivity analysis for optimal boundary control of a 3D reaction-difusion system". *Nonconvex Optimization and its Applications*. Ed. by Pillo, G. D. & Roma, M. Vol. 83. Springer, Berlin, 2006.

[40] Griesse, R. "Parametric sensitivity analysis in optimal control of a reaction-diffusion system – part II: practical methods and examples". *Optimization Methods and Software* **19**.2 (2004), pp. 217–242.

[41] Griesse, R. "Parametric sensitivity analysis in optimal control of a reaction diffusion System. I. Solution differentiability". *Numerical Functional Analysis and Optimization* **25**.1-2 (2004), pp. 93–117.

[42] Griesse, R. *Stability and sensitivity analysis in optimal control of partial differential equations*. 2007.

[43] Griesse, R. & Vexler, B. "Numerical sensitivity analysis for the quantity of interest in PDE-constrained optimization". *SIAM Journal on Scientific Computing* **29**.1 (2007), pp. 22–48.

[44] Griesse, R. & Walther, A. "Parametric sensitivities for optimal control problems using automatic differentiation". *Optimal Control Applications and Methods* **24** (2003), pp. 297–314.

[45] Groena, E. & Heijungs, R. "Ignoring correlation in uncertainty and sensitivity analysis in life cycle assessment: what is the risk?" *Environmental Impact Assessment Review* **62** (2017), pp. 98–109.

[46] Gunawan, R., Cao, Y., Petzold, L. & Doyle III, F. "Sensitivity analysis of discrete stochastic systems". *Biophys. J.* **88** (2005), pp. 2530–2540.

[47] Haario, H., Laine, M., Mira, A. & Saksman, E. "DRAM: Efficient adaptive MCMC". *Statistics and Computing* **16** (2006), pp. 339–354.

[48] Halko, N., Martinsson, P. G. & Tropp, J. A. "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions". *SIAM Review* **53**.2 (2011), pp. 217–288.

[49] Hall, J. "Uncertainty-based sensitivity indices for imprecise probability distributions". *Reliability Engineering and System Safety* **91** (2006), pp. 1443–1451.

[50] Hart, J., Saunders, C., Novak, V. & Gremaud, P. "Transcranial doppler-based surrogates for cerebral blood flow: A statistical study". *PLoS ONE* **11**.11 (2016).

[51] Hart, J., Alexanderian, A. & Gremaud, P. "Efficient computation of Sobol' indices for stochastic models". *SIAM J. Sci. Comput.* **39**.4 (2017), A1514–A1530.

[52] Hart, J., Bessac, J. & Constantinescu, E. "Global sensitivity analysis for statistical model parameters". *SIAM/ASA J. Uncertain. Quantif.* (2018).

[53] Hart, J. & Gremaud, P. "An approximation theoretic perspective of Sobol' indices with dependent variables". *International Journal for Uncertainty Quantification* **8**.6 (2018), pp. 483–493.

[54] Heroux, M., Bartlett, R., Hoekstra, V. H. R., Hu, J., Kolda, T., Lehoucq, R., Long, K., Pawlowski, R., Phipps, E., Salinger, A., Thornquist, H., Tuminaro, R., Willenbring, J. & Alan Williams ", i.S.N. L. *An overview of Trilinos*. Tech. rep. SAND2003-2927. 2003.

[55] Hooker, G. "Generalized functional ANOVA diagnostics for high- dimensional functions of dependent variables". *Journal of Computational and Graphical Statistics* 16 (2007), pp. 709–732.

[56] Hu, Z., Cao, J. & Hong, L. J. "Robust simulation of global warming policies using the DICE model". *Management Science* **58**.12 (2012), pp. 2190–2206.

[57] Iooss, B. & Lemaître, P. "A review on global analysis methods". *Uncertainty management in simulation-optimization of complex systems*. Ed. by Dellino, G. & Meloni, C. Springer, 2015. Chap. 5, pp. 543–501.

[58] Iooss, B. & Ribatet, M. "Global sensitivity analysis of computer models with functional inputs". *Reliability Engineering & System Safety* **94**.7 (2009), pp. 1194–1204.

[59] Iooss, B. & Saltelli, A. "Introduction to sensitivity analysis". *Handbook for Uncertainty Quantification*. Ed. by Ghanem, R., Higdon, D. & Owhadi, H. Springer, 2016, pp. 1103–1122.

[60] Ito, K. & Ravindran, S. S. "Optimal control of thermally convected fluid flows". *SIAM J. on Scientific Computing* **19** (1998), pp. 1847–1869.

[61] Janon, A., Nodet, M. & Prieur, C. "Uncertainties assessment in global sensitivity indices estimation from metamodels". *Int. J. Uncert. Quant.* **4** (2104), pp. 21–36.

[62] Kelley, C. *Iterative methods for linear and nonlinear equations*. SIAM, 1995.

[63] Kim, D., Debusschere, B. & Najm, H. "Spectral methods for parametric sensitivity in stochastic dynamical systems". *Biophys. J.* **92** (2007), pp. 379–393.

[64] Kindermann, R. & Snell, J. L. *Markov random fields and their applications*. Vol. 1. Contemporary Mathematics. AMS, 1980.

[65] Klenke, A. *Probability Theory: A comprehensive course*. 2nd ed. Universitext. Springer, 2014.

[66] Komorowski, M., Costa, M., Rand, D. & Stumpf, M. "Sensitivity, robustness, and identifiability in stochastic kinetics models". *Proc. Natl. Acad. Sci. USA* **108** (2011), pp. 8645–8650.

[67] Kouri, D. P., Winckel, G. von & Ridzal, D. *ROL: Rapid Optimization Library*.

[68] Kucherenko, S., Tarantola, S. & Annoni, P. "Estimation of global sensitivity indices for models with dependent variables". *Computer Physics Communications* **183** (2012), pp. 937–946.

[69] Kucherenko, S. & Iooss, B. "Handbook of Uncertainty Quantification". Springer, 2016. Chap. Derivative based global sensitivity measures.

[70] Lacirignola, M., Blanc, P., Girard, R., Pérez-López, P. & Blanc, I. "LCA of emerging technologies: addressing high uncertainty on inputs' variability when performing global sensitivity analysis". *Science of the Total Environment* **578** (2017), pp. 268–280.

[71] Lamboni, M., Iooss, B., Popelin, A.-L. & Gamboa, F. "Derivative-based global sensitivity measures: General links with Sobol' indices and numerical tests". *Mathematics and Computers in Simulation* **87** (2013), pp. 45–54.

[72] Le Gratiet, L., Marelli, S. & Sudret, B. "Metamodel-based sensitivity analysis: Polynomial chaos expansions and Gaussian processes". *Handbook on Uncertainty Quantification*. Ed. by Ghanem, R., Higdon, D. & Owhadi, H. Springer, 2016.

[73] Li, G., Rabitz, H., Yelvington, P. E., Oluwole, O. O., Bacon, F., Kolb, C. E. & Schoendorf, J. "Global sensitivity analysis for systems with independent and/or correlated Inputs". *J. Phys. Chem.* **114**.19 (2010), pp. 6022–6032.

[74] Lorenz, E. N. "Deterministic non-periodic flow". *Journal of the Atmospheric Sciences* **20** (1963), pp. 130–141.

[75] Maître, O. L. & Knio, O. "PC analysis of stochastic differential equations driven by Wiener noise". *Reliability Engineering & System Safety* **135** (2015), pp. 107–124.

[76] Maître, O. L., Knio, O. & Moraes, A. "Variance decomposition in stochastic simulators". *J. Chem. Phys.* **142** (2015), pp. 244115–1–244115–13.

[77] Majda, A. & Gershgorin, B. "Quantifying uncertainty in climate change science through empirical information theory". *Proc. Natl. Acad. Sci. USA* **107** (2010), pp. 14958–14963.

[78] Majda, A. & Gershgorin, B. "Improving model fidelity and sensitivity for complex systems through empirical information theory". *Proc. Natl. Acad. Sci. USA* **108** (2011), pp. 10044–10049.

[79] Mara, T. & Tarantola, S. "Variance-based sensitivity analysis of computer models with dependent inputs". *Reliability Eng. Sys. Safety* **107** (2012), pp. 115–121.

[80] Marrel, A., Iooss, B., Da Veiga, S. & Ribatet, M. "Global sensitivity analysis of stochastic computer models with joint metamodels". *Stat. Comput.* **22** (2012), pp. 833–847.

[81] Marrel, A., Iooss, B., Laurent, B. & Roustant, O. "Calculations of Sobol' indices for the Gaussian process metamodel". *Reliability Engineering and System Safety* **94** (2009), pp. 742–751.

[82] Marrel, A., Iooss, B., Van Dorpe, F. & Volkova, E. "Global sensitivity analysis for models with spatially dependent outputs". *Environmetrics* **22** (2011), pp. 383–397.

[83] Marrel, A., Saint-Geours, N. & Lozzo, M. D. "Sensitivity analysis of spatial and/or temporal phenomena". *Handbook for Uncertainty Quantification*. Ed. by Ghanem, R., Higdon, D. & Owhadi, H. Springer, 2016, pp. 1327–1357.

[84] Marzban, C. "Variance-based sensitivity analysis: An illustration on the Lorenz'63 model". *American Meteorological Society* (2013), pp. 4069–4079.

[85] Millner, A., Dietz, S. & Heal, G. "Scientific ambiguity and climate policy". *Environ Resource Econ* **55** (2013), pp. 21–46.

[86] Morris, M. "Factorial sampling plans for preliminary computational experiments". *Technometrics* **33** (1991), pp. 161–174.

[87] Nakayama, M., Goyal, A. & Glynn, P. "Likelihood ratio sensitivity analysis for Markovian models of highly dependable systems". *Stoch. Models* **10** (1994), pp. 701–717.

[88] Narayan, A. & Xiu, D. "Distributional sensitivity for uncertainty quantification". *Commun. Comput. Phys.* **10**.1 (2011), pp. 140–160.

[89] Owen, A. "Better estimation of small Sobol' sensitivity indices". *ACM Trans. Mod. Comput. Simul.* **23** (2013), 11–1:11–17.

[90] Owen, A. "Sobol' indices and Shapley value". *SIAM/ASA J. Uncertain. Quantif.* **2**.1 (2014), pp. 245–251.

[91] Owen, A. & Prieur, C. "On Shapley value for measuring importance of dependent inputs". *SIAM/ASA Journal on Uncertainty Quantification* **5**.1 (2017), pp. 986–1002.

[92] Paleari, L. & Confalonieri, R. "Sensitivity analysis of a sensitivity analysis: We are likely overlooking the impact of distributional assumptions". *Ecological Modelling* **340** (2016), pp. 57–63.

[93] Pianosi, F. & Wagener, T. "A simple and efficient method for global sensitivity analysis based on cumulative distribution functions". *Environmental Modelling & Software* **67** (2015), pp. 1–11.

[94] Pinson, P. & Girard, R. "Evaluating the quality of scenarios of short-term wind power generation". *Applied Energy* **96** (2012), pp. 12–20.

[95] Plyasunov, S. & Arkin, A. "Efficient stochastic sensitivity analysis of discrete event systems". *J. Comput. Phys.* **221** (2007), pp. 724–738.

[96] Prieur, C. & Tarantola, S. "Variance-based sensitivity analysis: Theory and estimation algorithms". *Handbook of Uncertainty Quantification*. Ed. by Ghanem, R., Higdon, D. & Owhadi, H. Springer, 2017.

[97] Qian, E., Peherstorfer, B., O'Malley, D., Vesselinov, V. V. & Willcox, K. "Multifidelity Monte Carlo estimation of variance and sensitivity indices". *SIAM/ASA Journal on Uncertainty Quantification* **6**.2 (2018), pp. 683–706.

[98] Rathinam, M., Sheppard, P. & Khammash, M. "Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks". *J. Chem. Phys* **132** (2010), pp. 034103–1–034103–13.

[99] Rosenthal, J. S. "Optimal proposal distributions and adaptive MCMC". Chapman and Hall/CRC, 2011. Chap. 4, pp. 93–112.

[100] Roustant, O., Barthe, F. & Iooss, B. "Poincaré inequalities on intervals – application to sensitivity analysis". *Electronic Journal of Statistics* **11**.2 (2017), pp. 3081–3119.

[101] Saibaba, A. K., Lee, J. & Kitanidis, P. K. "Randomized algorithms for generalized Hermitian eigenvalue problems with application to computing Karhunen-Loève expansion". *Numerical Linear Algebra with Applications* **23** (2016), pp. 314–339.

[102] Saltelli, A. "Making best use of model evaluations to compute sensitivity indices". *Computer Physics Communications* **145**.2 (2002), pp. 280–297.

[103] Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M. & Tarantola, S. "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index". *Computer Physics Communications* **181** (2010), pp. 259–270.

[104] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. & Tarantola, S. *Global sensitivity analysis: the primer*. Wiley, 2008.

[105] Shapley, L. S. "A value for n-person games". *Contributions to the Theory of Games*. Ed. by Kuhn, H. W. & Tucker, A. W. Vol. 2. Annals of Mathematics Studies. Princeton University Press, 1953, pp. 307–318.

[106] Sherlock C.and Fearnhead, P. & Roberts, G. "The random walk Metropolis: Linking theory and practice through a case study". *Statistical Science* **25**.2 (2010), pp. 172–190.

[107] Sobol', I. & Kucherenko, S. "Derivative based global sensitivity measures and the link with global sensitivity indices". *Math. Comp. Simul.* **79** (2009), pp. 3009–30017.

[108] Sobol', I. & Kucherenko, S. "A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices". *Comput. Phys. Comm.* **181** (2010), pp. 1212–1217.

[109] Sobol', I. "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates". *Mathematics and Computers in Simulation* **55** (2001), pp. 271–280.

[110] Sobol', I., Tarantola, S., .Gatelli, D, Kucherenko, S. & Mauntz, W. "Estimating the approximation error when fixing unessential factors in global sensitivity analysis". *Reliability Engineering & System Safety* **92**.7 (2007), pp. 957–960.

[111] Song, E., Nelson, B. L. & Staum, J. "Shapley effects for global sensitivity analysis: Theory and computation". *SIAM/ASA J. Uncertain. Quantif.* **4** (2016), pp. 1060–1083.

[112] Stone, C. J. "The use of polynomial splines and their tensor products in multivariate function estimation". *The Annals of Statistics* 22 (1994), pp. 118–171.

[113] Storlie, C., Swiler, L., Helton, J. & Sallaberry, C. "Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models". *Reliability Eng. Sys. Safety* **94** (2009), pp. 1735–1763.

[114] Sudret, B. "Global sensitivity analysis using polynomial chaos expansions". *Reliability Engineering & System Safety* **93**.7 (2008), pp. 964 –979.

[115] Tarantola, S. & Mara, T. A. "Variance-based sensitivity indices of computer models with dependent inputs: The Fourier Amplitude Sensitivity Test". *International Journal of Uncertainty Quantication* **7**.6 (2017), pp. 511–523.

[116] Vihola, M. "Robust adaptive Metropolis algorithm with coerced acceptance rate". *Statistics and Computing* **22**.5 (2011), pp. 997–1008.

[117] Vilar, J. M., Kueh, H. Y., Barkai, N. & Leibler, S. "Mechanisms of noise-resistance in genetic oscillators". *Proceedings of the National Academy of Sciences* **99**.9 (2002), pp. 5988–5992.

[118] Wilkinson, D. "Stochastic modelling for quantitative description of heterogeneous biological systems". *Nature* **10** (2009), pp. 122–133.

[119] Xiao, S., Lu, Z. & Xu, L. "Multivariate sensitivity analysis based on the direction of eigen space through principal component analysis". *Reliability Engineering and System Safety* **165** (2017), pp. 1–10.

[120] Xu, C. & Gertner, G. Z. "Uncertainty and sensitivity analysis for models with correlated parameters". *Reliability Eng. Sys. Safety* **93**.10 (2008), pp. 1563–1573.