



Technische Universität München
TUM School of Engineering and Design

Uncertainty quantification and separation with high-dimensional engineering models

Maximilian Ehre

Vollständiger Abdruck der von der TUM School of Engineering and Design zur Erlangung
eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr.-Ing. Kai-Uwe Bletzinger

Prüfer der Dissertation:

1. Prof. Dr. sc. tech. Daniel Straub
2. Prof. Faidon-Stelios Koutsourelakis, Ph.D.
3. Prof. Dr. Bruno Sudret

Die Dissertation wurde am 19.07.2021 bei der Technischen Universität München eingereicht
und durch die TUM School of Engineering and Design am 05.02.2022 angenommen.

Contents

I	Synopsis	x
1	Introduction	1
1.1	Motivation & context	1
1.2	Thesis organization & contributions	3
2	Uncertainty Quantification & Separation	5
2.1	Elements of probability theory	5
2.1.1	Probabilistic framework, conditional probability and independence	5
2.1.2	Random variables, distributions, moments	6
2.1.3	Random vectors, joint, marginal and conditional distributions	8
2.1.4	Random functions	10
2.2	Uncertainty Propagation	13
2.2.1	Setup	13
2.2.2	Second moment analysis	14
2.2.3	Monte Carlo	15
2.2.4	Surrogate Modelling	16
2.3	Reliability Analysis	30
2.3.1	Introduction	30
2.3.2	The reliability problem	31
2.3.3	Sampling methods	31
2.3.4	Surrogate-assisted sampling methods	36
2.4	Sensitivity Analysis	39
2.4.1	Sensitivity analysis of model output	39
2.4.2	Variance-based SAMO with surrogate models	40
2.4.3	Reliability-oriented sensitivity analysis	43
2.5	Uncertainty Separation	44
2.5.1	Introduction to Bayesian Analysis	44
2.5.2	Bayesian Uncertainty Separation	46
3	Concluding Remarks & Outlook	51
3.1	High-dimensional surrogate modelling for UQ	51
3.1.1	Summary	51
3.1.2	Outlook	52
3.2	Uncertainty separation	52
3.2.1	Summary	52
3.2.2	Outlook	53

II	Published Papers	68
4	PLS-based adaptation for efficient PCE representation in high dimensions	69
4.1	Introduction	70
4.2	PCE representations	72
4.2.1	Transformed PCE basis	74
4.3	Partial least squares	74
4.3.1	Properties of the PLS matrices	76
4.4	PLS-based PCE	78
4.4.1	Linear PLS-based transformation	78
4.4.2	PCE-driven PLS-based transformation	79
4.5	Examples	82
4.5.1	Linear elastic bar	83
4.5.2	Nonlinear oscillator	88
4.5.3	Steel plate	91
4.6	Concluding remarks	96
4.6.1	Discussion on computational complexity	96
4.6.2	Outlook	96
5	Global sensitivity analysis in high dimensions with PLS-PCE	101
5.1	Introduction	102
5.2	Partial least squares and polynomial chaos expansions	103
5.2.1	Polynomial Chaos Expansion	103
5.2.2	Basis adaptation	104
5.2.3	Partial least squares-based PCE	105
5.3	Global sensitivity analysis	107
5.3.1	Variance-based sensitivity analysis	107
5.3.2	PCE-based sensitivity analysis	107
5.4	Global sensitivity analysis with PLS-PCE	108
5.4.1	Computation in the asymptotic limit $N \rightarrow \infty$	108
5.4.2	Corrections for small samples sizes	110
5.4.3	A comment on variance-based sensitivity analysis with the standard basis adaptation format	112
5.5	Numerical experiments	112
5.5.1	Elastic truss	113
5.5.2	Steel plate	118
5.6	Conclusion	123
5.7	Acknowledgment	126
6	Sequential, active learning of low-dimensional model representations for reliability analysis	130
6.1	Introduction and previous work	131
6.2	Reliability analysis	133
6.3	Sequential importance sampling for rare event estimation	134
6.4	Partial least squares-based polynomial chaos expansions	136
6.4.1	Polynomial Chaos Expansions	136
6.4.2	Basis adaptation via partial least squares	138
6.4.3	PLS-PCE-R	139

6.4.4	PLS-PCE-W	139
6.5	Learning PLS-PCE models in each SIS level	140
6.5.1	The sequential subspace importance sampler	140
6.5.2	Active learning of low-dimensional model representations	143
6.6	Numerical experiments	145
6.6.1	Error measures	147
6.6.2	Low- and medium-dimensional examples	147
6.6.3	High-dimensional example: Steel plate	150
6.7	Concluding remarks	153
6.8	Acknowledgment	154
7	Conditional reliability analysis in high dimensions based on controlled mixture importance sampling and information reuse	160
7.1	Introduction	161
7.2	Conventional reliability analysis	162
7.2.1	Problem Statement	162
7.2.2	Standard MC	163
7.2.3	Importance sampling	163
7.2.4	The iCE method (iCE)	164
7.3	Conditional reliability analysis	167
7.3.1	Problem Statement	167
7.3.2	Connection to RBDO	168
7.3.3	Source selection	169
7.3.4	Information reuse for iCE: mixture-based and controlled importance sampling	171
7.3.5	Preconditioning iCE	173
7.3.6	The computational procedure	173
7.4	Numerical experiments	173
7.4.1	Parameter Study: sequential processing chain	173
7.4.2	Case Study: monopile foundation in plastically behaving soil	179
7.5	Conclusion	184
7.6	Acknowledgment	185
8	A framework for global reliability sensitivity analysis in the presence of multi-uncertainty	189
8.1	Introduction	190
8.2	Global Reliability Sensitivity Analysis	192
8.2.1	Variance-Based Sensitivity Analysis	193
8.2.2	The proposed reliability sensitivity indices	194
8.2.3	Monte-Carlo estimators	195
8.3	Polynomial Basis Surrogate Modelling	196
8.3.1	Polynomial Chaos Expansions	197
8.3.2	Canonical Decomposition	197
8.3.3	Surrogate-Based sensitivity indices	198
8.4	Conditional Surrogate-Based Reliability Sensitivities	199
8.4.1	Level 1	200
8.4.2	Level 2	201
8.4.3	The Framework	202
8.4.4	Computational cost	202

8.5	Numerical examples	205
8.5.1	Elastic truss	205
8.5.2	Monopile Foundation	208
8.6	Concluding Remarks	214
8.6.1	Discussion	215
8.6.2	Outlook	215
8.A	Connecting the proposed Sobol' index to its indicator function-based counterpart . .	216

Abstract

Uncertainty quantification plays a central role in modern engineering practice. It supports judging the adequacy of modelling choices, the description of naturally random phenomena and — ultimately — making optimal decisions in an imperfect, partially known world. Uncertainty quantification in engineering is carried out with a computational model that describes the physical processes underlying an engineering application. This thesis addresses two major challenges within model-based uncertainty quantification.

1. In the face of ever-increasing disposable computational resources, more complex models are employed both in the description of the engineering application as well as the description of uncertainty. This frequently leads to increased numbers of uncertain parameters, which in turn implies high-dimensional models. A common example is in modelling spatially or temporally distributed uncertainty with random fields and processes. Such high-dimensional models require carefully designed approaches to achieve computationally affordable and accurate uncertainty quantification. This need is addressed by introducing a surrogate modelling approach that relies on dimensionality reduction and is able to significantly alleviate the computational demands associated with high-dimensional uncertainty quantification, especially when the engineering model is computationally expensive. Based on this surrogate, approaches for uncertainty propagation, sensitivity and reliability analysis are developed in several original contributions.

2. Within uncertainty quantification, different types of uncertainty can be discerned. This distinction is based on whether uncertainty arises due to an inherently random phenomenon — known as *aleatory uncertainty* — or whether it stems from a lack of knowledge that can be potentially alleviated with more information — known as *epistemic uncertainty*. The separate treatment of different types of uncertainty can be of paramount importance, yet is currently often neglected in practice. This thesis presents a framework for the separate treatment of two generic classes of uncertainty. Original methods for the efficient computation of quantities of interest related to uncertainty separation are proposed. In particular, this entails efficient methods to compute the probability of failure conditional on one of the two uncertainty classes as well as a novel reliability sensitivity measure arising from uncertainty separation.

Zusammenfassung

Die Quantifizierung von Unsicherheiten ist von zentraler Bedeutung für das Ingenieurwesen. Sie dient der Bewertung verschiedener Modelle und der diesen Modellen zugrunde liegenden Annahmen, der Beschreibung von zufälligen Prozessen und in letzter Konsequenz auch der optimalen Entscheidungsfindung im Angesicht unvollständiger Information. Die Quantifizierung von Unsicherheiten in der Ingenieurspraxis beruht üblicherweise auf einem Computermodell, welches die dem untersuchten System zugrunde liegenden physikalischen Prozesse beschreibt. Die vorliegende Arbeit befasst sich mit zwei Problemstellungen innerhalb der modellbasierten Quantifizierung von Unsicherheiten:

1. Die Komplexität und Detailtreue der verwendeten Modelle wächst gemeinsam mit den verfügbaren Rechenkapazitäten stetig. Dies gilt für die Computermodelle, welche die Ingenieursanwendung selbst abbilden wie auch die probabilistischen Modelle, welche zur Beschreibung von Unsicherheit herangezogen werden. Oftmals schlägt sich diese Entwicklung in einer wachsenden Zahl von modellierten Unsicherheiten im probabilistischen Modell nieder, was zu sogenannten hochdimensionalen Modellen führt. Ein verbreitetes Beispiel findet sich in der Modellierung zeitlich und räumlich verteilter Unsicherheit mittels durch hochdimensionale Zufallsvektoren diskretisierter Zufallsfelder und -prozesse. In der vorliegenden Arbeit entwickeln wir ersatzmodellbasierte Methoden zur Analyse solch hochdimensionaler Probleme, welche die benötigte Rechenkapazität signifikant reduzieren, insbesondere wenn das Computermodell rechenintensiv ist. Basierend auf diesem Ersatzmodell entwickeln wir in mehreren Publikationen Methoden zur Propagation von Unsicherheiten wie auch für die Sensitivitäts- und Zuverlässigkeitsanalyse von Computermodellen.

2. Innerhalb der Quantifizierung von Unsicherheiten lassen sich verschiedenen Typen von Unsicherheit unterscheiden: Unsicherheiten, welche inhärent zufällige Prozesse beschreiben, werden als *aleatorische Unsicherheiten* bezeichnet und unterschieden von den *epistemischen Unsicherheiten*, die von unvollständiger Information herrühren. Die getrennte Behandlung dieser beiden Unsicherheiten ist oft von entscheidender Wichtigkeit, wird in der Praxis gegenwärtig allerdings häufig ignoriert. In der vorliegenden Arbeit entwickeln wir ein allgemeines Konzept zur getrennten Behandlung zweier verschiedener Typen von Unsicherheiten. Wir entwickeln neuartige Methoden zur effizienten Berechnung verschiedener Interessengrößen, die sich aus der expliziten Trennung von Unsicherheiten ergeben, so zum Beispiel die Versagenswahrscheinlichkeit bedingt auf einen der beiden Unsicherheitstypen sowie neuartige Zuverlässigkeitssensitivitätsmaße, deren Definition auf der Trennung von Unsicherheiten beruht.

Acknowledgements

The past four years have been a joyful and memorable time to me and doing my PhD in the Engineering Risk Analysis group is a decision I value highly to this date.

I want to thank my supervisors Daniel and Iason for trusting me with this project and supporting me unconditionally throughout the process. Iason, in particular, I have to thank for countless hours of fruitful discussion, constant creative input and his yoga lessons somewhere on the spectrum between soothing and tormenting (more towards the former, though).

I am thankful to Karen Willcox and Bruno Sudret for having me as an academic guest in their respective research groups and the fruitful collaborations that have emerged from these visits.

To my former and current colleagues Felipe, Elizabeth, Max, Sebastian and Antonis, I am very grateful for keeping me company, for our discussions both of professional and rather unprofessional nature as well as myriads of cups of coffee and pints of beer shared over these conversations.

Finally I am grateful to my family for nothing in particular since just having them is the greatest of pleasures. To be more specific, though, I have Lilli to thank for taking care of Eleni for days and weeks and months on end while I was trying to finalize this document and I have my parents to thank for nagging me on a constant basis over when I was going to submit it and I have my sister Hannah to thank for not doing just that.

PART I

SYNOPSIS

1.1 Motivation & context

Uncertainties are ubiquitous in engineering practice. They must be acknowledged and accounted for at various stages of the engineering process ranging from system analysis and design to assessment, retrofitting and decommission decisions. The ever-increasing capabilities of modern computers have rendered model-based simulation and computer-aided engineering in general a sharp tool in the engineer's shed. While increasingly fast computers have largely unlocked the world of deterministic model-based engineering, quantifying uncertainties requires switching to a non-deterministic paradigm, which is usually associated with significantly increased computational analysis expenses. In response to the need for computationally efficient methods in this vein, the relatively young field of uncertainty quantification (UQ) has formed at the intersection of applied and statistical mathematics with applied sciences as, e.g., engineering. UQ is broadly concerned with the model-based systematic and quantitative treatment of uncertainties. The model contains a description of the application, in which uncertainties shall be quantified and typically comes in the form of a computer code that maps inputs to outputs. This model is usually of deterministic nature, which is to say the model output is deterministic conditional on the model input. We refer to this as the computational model. The description and modelling of uncertainty is separated from the computational model and formulated in terms of the model inputs: Any uncertainty about a given system will be formulated as an input to the computational model. It is common practice to resort to probability theory [114, 191] for representing uncertainty. Other representations of uncertainty can be summarized under the term 'imprecise probability' [231], all of which share the common denominator of expressing belief in a manner that is in conflict with the axioms of probability theory (see Section 2.1) in one way or another. Examples are interval probabilities [9, 19, 151], fuzzy numbers [77, 82, 108, 207], p-boxes [69, 200] and random sets [45, 110].

The first step of any UQ analysis is the elicitation and characterization of uncertainty associated with the system. Physical sources of uncertainty in engineering and mechanical systems include randomness in geometric and material properties as well as applied loads. Several other sources, such as uncertainties about the constitutive laws describing the physics underlying the system's behaviour, the discretization of these laws as well as boundary and initial conditions imposed on the system, are related to the computational model itself. The inference of probabilistic models

for input uncertainties often pertains to selecting a suitable distribution model and estimating that distribution's parameters based on measurement data. A detailed account of model selection and parameter estimation is given in [83] and [21], respectively.

A word of warning is in order about the potential dependency of input uncertainties. Failure to recognize and appropriately model such dependencies can lead to a vast misrepresentation of stochastic properties of the model response and consequentially to spurious and misleading UQ results. A second prominent pitfall associated with the uncertain model inputs is their cardinality: if the number of inputs grows beyond a certain limit, both conceptual and computational challenges arise. Computational challenges emerge through the oftentimes exponential scaling of computational cost with the input dimension (take numerical integration as an example) for which the term *curse of dimensionality* was coined in [20]. Conceptual challenges often stem from the counter-intuitive properties of the high-dimensional spaces, in which these inputs reside. Such properties may, however, lend themselves to an advantageous exploitation of some form, which is appropriately referred to as the *blessing of dimensionality* [105]. These phenomena are usually based on *concentration-of-measure* phenomena in high dimensions. Additionally, many high-dimensional problems faced in engineering applications possess a low-rank structure that can be exploited to efficiently compute solutions in a lower-dimensional setting. There is generally no single number of inputs that separates high-dimensional problems from low-dimensional problems as the point at which said challenges arise largely depend on the very algorithm and application at hand. In the context of UQ, this boundary is often quoted as lying somewhere between 10 and 100 dimensions.

The second step of UQ analyses consists in the propagation of uncertainty across the computational model. This propagation can be oriented in one of two ways: if uncertainty is propagated from the uncertain inputs to the response, we speak of *forward* UQ. Vice versa, if the model is used to propagate uncertainty observed in the response or a quantity of interest derived thereof to characterize statistical properties of the inputs, we speak of *inverse* UQ. Examples of inverse UQ techniques are Bayesian model inversion (see Section 2.5) and — to a certain degree — probabilistic sensitivity analysis (see Section 2.4). Depending on the analysis goal, different sub-disciplines of forward UQ can be discerned: If the goal resides in the probabilistic characterization of the model response (another term for the output of the computational model), we speak of uncertainty propagation (see Section 2.2). A second discipline is reliability analysis, the goal of which consists in computing the probability of a previously specified (typically rare) event associated with system failure (see Section 2.3). Sensitivity analysis is concerned with assessing the sensitivity of the response (or quantities of interest derived thereof) with respect to the model inputs in a non-deterministic setting (see Section 2.4) and as such can be grouped into either of both UQ categories (forward and inverse). Other disciplines — featured less prominently in this work yet present — include decision analysis that aims at formalizing optimal decision making under uncertainty based on utility theory and stochastic optimization, where either the optimization objective or constraints contain randomness.

In literature, uncertainty is frequently classified according to whether it represents an inherently random process (e.g., predicting the outcome of a dice throw) or a lack of knowledge (e.g., having to guess the outcome of a dice already thrown that rolled around a corner in the process such that we cannot see it) [66, 70, 85, 161]. The former is often referred to as *aleatory* uncertainty and the latter is termed *epistemic* uncertainty. A similar terminology that reveals why such a distinction may be useful discriminates between *irreducible* (corresponds to aleatory) and *reducible* (corresponds to epistemic) uncertainty, where reducibility means the possibility to reduce uncertainty by gather-

ing more information. While probability theory is the unanimous choice for modelling irreducible uncertainty, the question of how to model epistemic uncertainty is subject to an ongoing debate with different viewpoints including subjective Bayesian probability [161, 49] as well as any of the non-probabilistic approaches mentioned above [207, 70, 85, 151].

1.2 Thesis organization & contributions

Chapter 2 offers a discussion of the theoretical and methodological foundations underlying this thesis. As such, Chapter 2 contains no original research but provides an overview of the state of the art along with a brief discussion of selected well-known methods. In Section 2.1, the necessary elements of probability theory and uncertainty modelling are introduced. Sections 2.2 through 2.4 treat the disciplines of uncertainty propagation, reliability analysis and sensitivity analysis, respectively. Section 2.5 details a brief discussion of Bayesian inference along with an account of uncertainty separation under the Bayesian paradigm. Figure 1.1 illustrates the general uncertainty quantification and separation framework that integrates the various forward and inverse UQ disciplines.

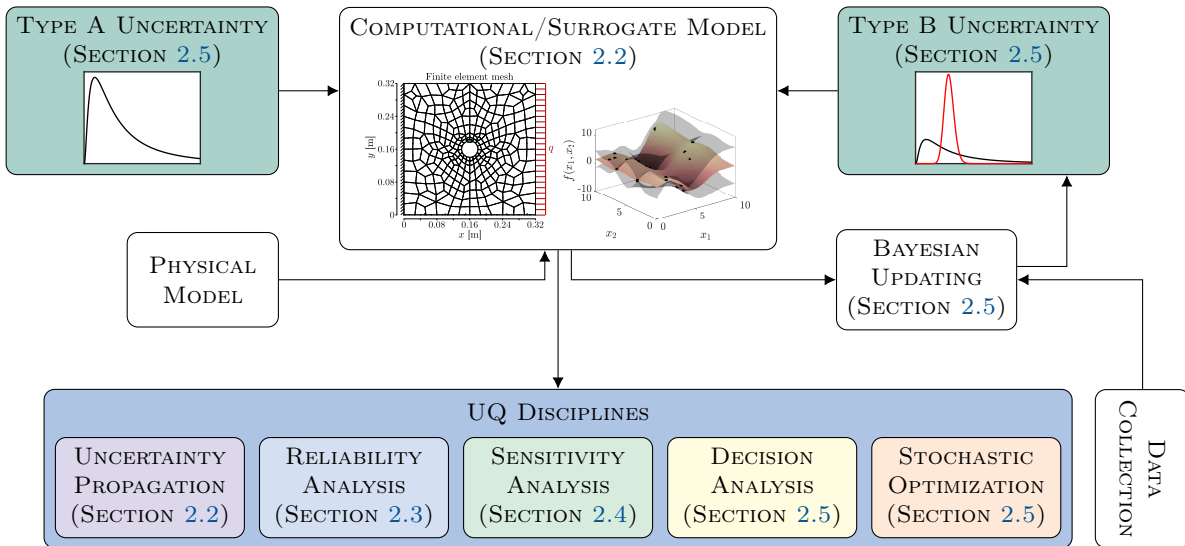


Figure 1.1: The combined uncertainty quantification and separation framework.

In part II, the publications in which this thesis is grounded follow as Chapters 4 to 8.

In Chapters 4 to 6, we focus on developing surrogate-based methods for forward UQ (uncertainty propagation, sensitivity analysis and reliability analysis, respectively) in the context of high-dimensional uncertainty. Following this, in Chapters 7 and 8, we develop a general framework for uncertainty separation and devise methods for reliability sensitivity and conditional reliability analysis. The corresponding original publications are listed in the following:

- Publication 1 (original publication [174]; Chapter 4) introduces PLS-PCE, a novel polynomial chaos expansion (PCE) with built-in dimensionality reduction via a subspace regression tech-

nique termed partial least squares (PLS) that paves the way for high-dimensional surrogate modelling using PCEs.

- Publication 2 (original publication [61]; Chapter 5) builds on publication 1 and derives a backtransformation from any PLS-PCE to standard PCE format. In this way, postprocessing PLS-PCEs is rendered a straightforward and efficient procedure and quantities of interest like global variance-based sensitivity measures are readily computed by analyzing the PLS-PCE model coefficients.
- Publication 3 (original publication [62]; Chapter 6) combines PLS-PCE models with a sequential importance sampling scheme for rare event probability estimation. The efficiency of sequential importance sampling is demonstrably improved by several orders of magnitude as the importance densities are reconstructed using locally and adaptively learned PLS-PCE models.
- Publication 4 (original publication [63]; Chapter 7) proposes an efficient approach to solving a sequence of parametrized reliability problems based on an information reuse strategy and controlled and mixture importance sampling [164]. Such sequences arise in the context of uncertainty separation, in particular in conditional reliability analysis, where the probability of failure is conditional on one of the two separated groups of uncertainties.
- Publication 5 (original publication [60]; Chapter 8) builds on this conditional probability of failure to define a new reliability sensitivity index and proposes a bi-level surrogate modelling approach to efficiently compute the new indices.

Uncertainty Quantification & Separation

2.1 Elements of probability theory

2.1.1 Probabilistic framework, conditional probability and independence

A probability space consists of the triplet $(\Omega, \mathcal{F}, \mathbb{P})$ (the presentation given here follows [114, 191]). Ω is the set of all possible outcomes ω called the sample space. $\mathcal{F} \subseteq 2^\Omega$ is a σ -field on Ω . That is, \mathcal{F} is a set of sets containing subsets of Ω (at least the empty set \emptyset and Ω) and is closed under the set operations of union, complementation and intersection. The most commonly used σ -field is the Borel σ -field of Ω , $\mathcal{B}(\Omega)$, as whenever Ω is a topological space, $\mathcal{B}(\Omega)$ is the smallest possible σ -field that guarantees measurability. Together, Ω and \mathcal{F} constitute a measurable space on which $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ is a probability measure. This means that $\mathbb{P}(E) \in \mathbb{R}_{\geq 0}$, $E \in \mathcal{F}$ (Kolmogorov's first axiom), $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$ (Kolmogorov's second axiom) and $\forall i \neq j : A_i \cap A_j = \emptyset \Rightarrow \mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ (σ -additivity; Kolmogorov's third axiom).

For two events $A, B \in \mathcal{F}$, the probability of A given that B has occurred (that implies $\mathbb{P}(B) > 0$) is called the probability of A *conditional on* B . The notation for such a conditional probability is $\mathbb{P}(A|B)$ and its definition reads

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad (2.1)$$

where $\mathbb{P}(A \cap B)$ is the joint probability of A and B and $\mathbb{P}(A)$, $\mathbb{P}(B)$ are the marginal (or unconditional) probabilities of A and B , respectively. $\mathbb{P}(\cdot|B)$ obeys Kolmogorov's axioms and thus also forms a probability measure on (Ω, \mathcal{F}) . If the probability of A is unaffected by whether B occurs or not, A and B are said to be independent. This is the case, if $\mathbb{P}(A|B) = \mathbb{P}(A)$ or equivalently — by inserting this in Eq. (2.1) — $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. There are at least two notable consequences of Eq. (2.1). The first is the *partition theorem* or *law of total probability*, which states that for an event $A \in \mathcal{F}$ and a partition of Ω $\{B_i \in \mathcal{F}\}_{i=1}^n$, such that $B_1 \cup B_2 \cup \dots \cup B_n = \Omega$ and $B_i \cap B_j = \emptyset$ whenever $i \neq j$, the marginal probability of A can be computed as

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A|B_i)\mathbb{P}(B_i). \quad (2.2)$$

The second is *Bayes' law*, which reverses the conditioning and for $1 \leq k \leq n$ states

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{i=1}^n \mathbb{P}(A|B_i)\mathbb{P}(B_i)}. \quad (2.3)$$

2.1.2 Random variables, distributions, moments

A random variable is a measurable map from $(\Omega, \mathcal{F}, \mathbb{P})$ to some other measurable space. Given the scope of this thesis, it is sufficient to consider real-valued random variables and vectors. Let X be a real-valued random variable, then

$$X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \quad \text{such that} \quad \{\omega \in \Omega : X(\omega) \in B\}, \quad B \in \mathcal{B}(\mathbb{R}). \quad (2.4)$$

That is to say, the measurable function X maps to the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and for each measurable event $B \in \mathcal{B}(\mathbb{R})$, its pre-image under X has to belong to \mathcal{F} . \mathbb{P} , being a probability measure on (Ω, \mathcal{F}) , induces a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, namely $\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(X^{-1}(B))$, $B \in \mathcal{B}(\mathbb{R})$. The Borel measure on the real line is the Lebesgue measure on \mathbb{R} , λ , which assigns to each interval its length. The random variable X may be completely characterized by its cumulative distribution function (CDF) $F_X(x)$, which is defined as

$$F_X(x) = \mathbb{P}(\omega \in \Omega : X(\omega) \leq x) = \mathbb{P}_X((-\infty, x]). \quad (2.5)$$

If not indicated otherwise, upper-case letters represent random variables and corresponding lower-cases indicate *sample space variables*, i.e., deterministic and specific outcomes of the random variable. The CDF has the following properties: $\lim_{x \rightarrow -\infty} F_X(x) = 0$ and $\lim_{x \rightarrow \infty} F_X(x) = 1$. Moreover, $F_X(x)$ is non-decreasing and right-continuous, i.e., $\lim_{x \searrow x_0} F_X(x) = F_X(x_0)$. The expected value of X is defined as

$$\mu_X = \mathbb{E}[X] = \int_{\Omega} X d\mathbb{P} = \int_{\Omega} X(\omega) \mathbb{P}(d\omega) = \int_{\mathbb{R}} x \mathbb{P}_X(dx) = \int_{\mathbb{R}} x d\mathbb{P}_X \quad (2.6)$$

assuming its existence, i.e., $\mathbb{E}[|X|] < \infty$. The integral over ω requires abstract integration and measure theory and is beyond the scope of this thesis. However, for real-valued random variables, the expectation can be cast in terms of a standard integral and the measure of X as on the right-hand side of this equation. Eq. (2.6) indicates that the expectation operator is linear, i.e., $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$. The random variable X is said to be absolutely continuous if whenever $\lambda(B) = 0$, then also $\mathbb{P}_X(B) = 0$ for any $B \in \mathcal{B}(\mathbb{R})$ (this is often written as $\mathbb{P}_X \ll \lambda$ and reads ‘ λ dominates \mathbb{P}_X ’) and we can write [191, Chapter 12]

$$\mathbb{P}_X(B) = \int_B f_X(x) \lambda(dx). \quad (2.7)$$

f_X is a non-negative function called the probability density function (PDF) of X and can be used to completely define an absolutely continuous random variable in place of F_X . In fact, f_X and F_X are closely related. To see this, choose $B = (-\infty, x]$ in Eq. (2.7) such that

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(x') \lambda(dx') = \int_{-\infty}^x f_X(x') dx', \quad (2.8)$$

in which the rightmost equality holds if and only if f_X is Riemann-integrable. It follows that $f_X(x) = dF_X(x)/dx$ wherever the derivative exists. $f_X(x)$ is a non-negative function and since $\lim_{x \rightarrow \infty} F_X(x) = 1$, it is

$$\int_{\mathbb{R}} f_X(x) dx = 1. \quad (2.9)$$

There exists a second frequently used flavor of real-valued random variables that is not absolutely continuous, namely discrete random variables. The equivalent to a PDF in the context of discrete random variables is the probability mass function. However, since discrete random variables are of no further interest to this work, we do not introduce them formally. A useful extension to Eq. (2.6) is the *change of variables theorem* [191, Chapter 6]. It states that, for any measurable, Riemann-integrable function $g : \mathbb{R} \rightarrow \mathbb{R}$, the expectation of g is

$$\mathbb{E}[g(X)] = \int_{\Omega} g(X(\omega)) \mathbb{P}(d\omega) = \int_{\mathbb{R}} g(x) d\mathbb{P}_X = \int_{\mathbb{R}} g(x) f_X(x) dx. \quad (2.10)$$

That is, instead of computing the density of the random variable $g(X)$, f_g , we may use f_X to compute the expectation of $g(X)$. In particular, setting $g(X) = (X - \mathbb{E}[X])^m$ with $m \in \mathbb{N}_{>0}$ yields the m -th central moment of X and with $m = 2$ the definition of *variance* (i.e. the second central moment of X) emerges as

$$\sigma_X^2 = \mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{\mathbb{R}} (x - \mathbb{E}[X])^2 f_X(x) dx, \quad (2.11)$$

if the integral exists. In general, the family of p -integrable real-valued random variables, i.e., $\mathbb{E}[X^p] \leq \infty$, on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ constitute a *Lebesgue space* $L_p^p(\Omega, \mathbb{R})$ together with the norm $\|X\|_p = \mathbb{E}[X^p]^{1/p}$. If $p = 2$, we simply write $\|\cdot\|$ (as we typically work with square-integrable random variables). σ_X is called the standard deviation of X . Properties of the variance operator following from Eq. (2.11) are i.) $\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ and ii.) $\mathbb{V}[aX + b] = a^2 \mathbb{V}[X]$.

In engineering, it is common to define random variables in terms of their CDF or PDF rather than constructing a probability space for a given experiment. However, an underlying probability space still exists and the simplest way of constructing it is by the identity map $X(\omega) = \omega$, such that $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$ and $\mathbb{P} = \mathbb{P}_X$. This is sometimes referred to as the *canonical probability space* [5, Chapter 4.6]. In this work, probability spaces are of canonical type unless explicitly defined otherwise.

An isoprobabilistic transformation from one absolutely continuous real-valued random variable X to another Y , where both X and Y are defined on the same probability space, is defined as the map $T : \mathbb{R} \rightarrow \mathbb{R}$ between the images of X and Y such that $\mathbb{P}(X(\omega)) = \mathbb{P}(Y(\omega)) \forall \omega \in \mathbb{R}$. If the CDFs $F_X(x)$ and $F_Y(y)$ are known this can be expressed as $F_X(x) = F_Y(y)$ and if F_Y is invertible we have $Y = T(X)$ with $T = F_Y^{-1} \circ F_X$. For a sample x_i drawn from the density f_X , we write $x_i \sim f_X$. T can be used to transform samples drawn from f_X into samples drawn from f_Y , $y_i \sim f_Y$. In particular, if X has the uniform distribution on $[0, 1]$, we have $T = F_Y^{-1}$. It is common practice to generate uniformly distributed random samples with a random number generator (see [124, Chapter 8] for an introduction to and overview over common random number generators) and use F_Y^{-1} to transform these samples into samples from the desired distribution. If F_Y^{-1} is not available, alternative methods to sample from f_Y include rejection sampling [159] and Markov Chain Monte Carlo (MCMC) methods [149, 150].

2.1.3 Random vectors, joint, marginal and conditional distributions

A d -dimensional real-valued random vector \mathbf{X} on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is defined as the measurable function $\mathbf{X} : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Note, that boldface letters distinguish vectors and matrices from scalar quantities here as well as in the remainder of this work. The induced measure in the image space of \mathbf{X} , $\mathbb{P}_{\mathbf{X}}$, is the Lebesgue measure on \mathbb{R}^d that assigns to each d -dimensional rectangle its volume and in analogy to Eq. (2.5), the CDF reads

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}) &= \mathbb{P}(X_1 \leq x_1 \cap X_2 \leq x_2 \cap \cdots \cap X_d \leq x_d) \\ &= \mathbb{P}_{\mathbf{X}}(X_1 \in (-\infty, x_1] \cap X_2 \in (-\infty, x_2] \cap \cdots \cap X_d \in (-\infty, x_d]). \end{aligned} \quad (2.12)$$

The construction of a canonical probability space underlying \mathbf{X} if \mathbf{X} is defined in terms of $F_{\mathbf{X}}(\mathbf{x})$ is in analogy to the univariate case discussed in the previous section: $\mathbf{X}(\omega) = \omega$, $\Omega = \mathbb{R}^d$, $\mathcal{F} = \mathcal{B}(\mathbb{R}^d)$ and $\mathbb{P} = \mathbb{P}_{\mathbf{X}}$. Further, if \mathbf{X} is absolutely continuous, its joint PDF $f_{\mathbf{X}}$ is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^d F_{\mathbf{X}}(\mathbf{x})}{\partial x_1 \cdots \partial x_d}. \quad (2.13)$$

Just as its univariate counterpart, $f_{\mathbf{X}}(\mathbf{x})$ is a non-negative function which integrates to 1 over its definition space \mathbb{R}^d . The continuous analogue of marginal probability in the fundamental probability space of Subsection 2.1.1 is the marginal probability distribution. The marginal PDF of an m -variate subset of \mathbf{X} is obtained by integrating the joint PDF $f_{\mathbf{X}}$ over all coordinates of \mathbf{X} that are not present in the subset. Let $\mathbf{I} \subset \{1, \dots, d\}$ be an index set with m entries and let its complement with respect to the set $\{1, \dots, d\}$ be $\mathbf{I}^c = \mathbf{I} \setminus \{1, \dots, d\}$. Then, the marginal PDF of $\mathbf{X}_{\mathbf{I}} = [X_{I_1}, X_{I_2}, \dots, X_{I_m}]$ reads

$$f_{\mathbf{X}_{\mathbf{I}}}(\mathbf{x}_{\mathbf{I}}) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{\mathbf{X}}(\mathbf{x}) dx_{I_1^c} \cdots dx_{I_{d-m}^c}. \quad (2.14)$$

Likewise, the continuous analogue of conditional probability is the conditional probability distribution. Introducing a second index set \mathbf{J} such that $\mathbf{I} \cap \mathbf{J} = \emptyset$ lets us define the PDF of $\mathbf{X}_{\mathbf{I}}$ conditional on $\mathbf{X}_{\mathbf{J}} = \mathbf{x}_{\mathbf{J}}$ in analogy to Eq. (2.1) (with $B = \{\mathbf{X}_{\mathbf{J}} = \mathbf{x}_{\mathbf{J}}\}$, other events may be conditioned upon as well):

$$f_{\mathbf{X}_{\mathbf{I}}|\mathbf{X}_{\mathbf{J}}}(\mathbf{x}_{\mathbf{I}}|\mathbf{X}_{\mathbf{J}} = \mathbf{x}_{\mathbf{J}}) = \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{X}_{\mathbf{J}}}(\mathbf{x}_{\mathbf{J}})}. \quad (2.15)$$

Two absolutely continuous real-valued random variables X_i and X_j are independent if their joint PDF factorizes as $f_{X_i, X_j}(x_i, x_j) = f_{X_i}(x_i)f_{X_j}(x_j)$. Likewise, if the coordinates of \mathbf{X} are pairwise independent, i.e., if X_i and X_j are independent whenever $i \neq j$, then

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^d f_{X_i}(x_i). \quad (2.16)$$

The expectation of \mathbf{X} extends in the obvious way from the univariate case as does the change of variables theorem, i.e.,

$$\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}[\mathbf{X}] = \int_{\mathbb{R}^d} \mathbf{x} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = [\mathbb{E}[X_1], \dots, \mathbb{E}[X_d]]^{\top} = [\mu_{X_1}, \dots, \mu_{X_d}]^{\top}, \quad (2.17)$$

where the integration is performed element-wise and

$$\mathbb{E}[g(\mathbf{X})] = \int_{\mathbb{R}^d} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \quad (2.18)$$

The covariance of two elements of \mathbf{X} , X_i and X_j , is

$$\begin{aligned}\text{COV}[X_i, X_j] &= \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} (x_i - \mathbb{E}[X_i])(x_j - \mathbb{E}[X_j]) f_{X_i X_j}(x_i, x_j) dx_i dx_j.\end{aligned}\quad (2.19)$$

The Pearson correlation coefficient of X_i and X_j is defined as $\rho_{ij} = \text{COV}[X_i, X_j] / \sqrt{\mathbb{V}[X_i]\mathbb{V}[X_j]}$. $\text{COV}[X_i, X_j]$ such that $-1 \leq \rho_{ij} \leq 1$ always since $\text{COV}[A, B] \leq \sqrt{\text{COV}[A, A]\text{COV}[B, B]} = \sqrt{\mathbb{V}[A]\mathbb{V}[B]}$ by the Cauchy-Schwarz-inequality. Assembling all covariances of the elements of \mathbf{X} in a matrix yields the covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{d \times d}$ with $\Sigma_{ij} = \text{COV}[X_i, X_j]$, $i, j = 1, \dots, d$. Analogously, the correlation matrix $\mathbf{R} \in \mathbb{R}^{d \times d}$ can be defined. Both $\mathbf{\Sigma}$ and \mathbf{R} are at least positive-semidefinite matrices, i.e., they have a non-negative spectrum. Covariance and correlation are measures of linear dependence. If two random variables X_i, X_j are independent they are also uncorrelated such that $\rho_{ij} = \text{COV}[X_i, X_j] = 0$ (the reverse is not generally true). Further properties following from Eq. (2.19) are: i.) $\text{COV}[X_i, X_j] = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]$, ii.) $\text{COV}[X_i, X_i] = \mathbb{V}[X_i]$, iii.) $\text{COV}[X_i, X_j] = \text{COV}[X_j, X_i]$, iv.) $\text{COV}[aX_i + bX_j, Z] = a\text{COV}[X_i, Z] + b\text{COV}[X_j, Z]$, v.) $\mathbb{V}[X_i + X_j] = \mathbb{V}[X_i] + \mathbb{V}[X_j] + 2\text{COV}[X_i, X_j]$.

\mathbf{X} is referred to as Gaussian random vector if its joint distribution is multivariate Gaussian. The Gaussian joint PDF is defined as

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{\Sigma}) = \frac{1}{\sqrt{(2\pi)^n \det \mathbf{\Sigma}}} \exp \left\{ -\frac{1}{2} \left[(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \right\} \quad (2.20)$$

with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$. Based on Eq. (2.20), multivariate Gaussian random vectors are completely defined by their mean vector $\boldsymbol{\mu}_{\mathbf{X}}$ and covariance matrix $\mathbf{\Sigma}_{\mathbf{X}}$. Moreover, linear transformations of Gaussian random vectors are Gaussian. If $\boldsymbol{\mu}_{\mathbf{X}} = \mathbf{0}_d$ and $\mathbf{\Sigma}_{\mathbf{X}} = \mathbf{I}_{d \times d}$ (with $\mathbf{0}_d$ a d -dimensional vector of zeros and $\mathbf{I}_{d \times d}$ the d -dimensional identity matrix), \mathbf{X} is said to be *standard-normal*. We write $\varphi_d(\cdot)$ for the d -dimensional standard-normal PDF and $\Phi_d(\cdot)$ for the corresponding CDF. If the elements of \mathbf{X} are uncorrelated they are also independent (this is not generally true but holds for Gaussian random vectors). In general, $\mathbf{\Sigma}_{\mathbf{X}}$ is a real, symmetric matrix and therefore diagonalizable. In particular, it can be written in terms of an orthogonal matrix \mathbf{A} and a diagonal matrix \mathbf{D} as

$$\mathbf{\Sigma}_{\mathbf{X}} = \mathbf{A} \mathbf{D} \mathbf{A}^T. \quad (2.21)$$

The columns of $\mathbf{A} = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ form an orthogonal basis of \mathbb{R}^d and together with the diagonal entries of \mathbf{D} , $\{\lambda_i\}_{i=1}^d$, form solutions to the eigenvalue problem

$$\mathbf{\Sigma}_{\mathbf{X}} \mathbf{v} = \lambda \mathbf{v}. \quad (2.22)$$

The $\{\mathbf{v}_i\}_{i=1}^d$ are called the eigenvectors of $\mathbf{\Sigma}_{\mathbf{X}}$ and the $\{\lambda_i\}_{i=1}^d$ are the corresponding non-negative eigenvalues. By diagonalization of $\mathbf{\Sigma}_{\mathbf{X}}$, any dependent Gaussian random vector \mathbf{X} can be transformed into an independent one \mathbf{X}' , where $\mathbf{X}' = \mathbf{A}^T (\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})$. Vice versa, a correlated Gaussian random vector can be generated from a standard-normal random vector \mathbf{U} by

$$\mathbf{X} = \boldsymbol{\mu}_{\mathbf{X}} + \mathbf{A} \mathbf{D}^{1/2} \mathbf{U} = \boldsymbol{\mu}_{\mathbf{X}} + \sum_{i=1}^d \sqrt{\lambda_i} \mathbf{v}_i U_i. \quad (2.23)$$

Alternatively, the factorization $\mathbf{\Sigma}_{\mathbf{X}} = \mathbf{B} \mathbf{B}^T$ may be used to obtain a standard-normal random vector as $\mathbf{U} = \mathbf{B}^{-1} (\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})$. If \mathbf{B} is a lower triangular matrix, this factorization is the Cholesky

decomposition of $\Sigma_{\mathbf{X}}$ and with $\mathbf{B} = \mathbf{A}\mathbf{D}^{1/2}$, the eigendecomposition of $\Sigma_{\mathbf{X}}$ is recovered.

In the context of this work and UQ subdisciplines such as reliability analysis, it is common practice to formulate problems on a canonical probability space with standard-normal measure. Thus, the transformation of input random vectors \mathbf{X} with arbitrary joint distribution into the standard-normal random vector $\mathbf{U} \sim \varphi_d(\mathbf{u})$ is an important task. If either the joint CDF $F_{\mathbf{X}}$ or a series of conditional CDFs as follows are known, a possible transformation $T : \mathbf{X} \rightarrow \mathbf{U}$ is the Rosenblatt transformation [190] for which

$$T = [\Phi^{-1} \circ F_{X_1}(x_1), \Phi^{-1} \circ F_{X_2|X_1}(x_2|x_1), \dots, \Phi^{-1} \circ F_{X_d|X_1 \dots X_{d-1}}(x_d|x_1, \dots, x_{d-1})]^T. \quad (2.24)$$

While such complete dependence information is rare in practice, knowledge of the marginal distributions of the coordinates of \mathbf{X} and the covariance structure is more common. Based on these two ingredients, the coordinates can be transformed marginally to an independent Gaussian random vector that is in turn transformed to a correlated Gaussian random vector. This transformation is known as the Nataf transformation [157, 134] and reads

$$T = \mathbf{B}^{-1}[\Phi^{-1} \circ F_{X_1}(x_1), \Phi^{-1} \circ F_{X_2}(x_2), \dots, \Phi^{-1} \circ F_{X_d}(x_d)]. \quad (2.25)$$

\mathbf{B} is the lower triangular matrix resulting from the Cholesky decomposition of $\Sigma_{\mathbf{U}}$. The entries of $\Sigma_{\mathbf{U}}$ in turn are computed based on the prescribed $\Sigma_{\mathbf{X}}$ through an integral equation. A third possibility of finding T if partial dependency information is available that is not necessarily limited to marginal distributions and the covariance structure of \mathbf{X} is presented by copula theory. A copula $C : [0, 1]^d \rightarrow [0, 1]$ is a d -variate CDF with uniform marginal distributions $F_{X_1}(x_1), \dots, F_{X_d}(x_d)$ such that

$$F_{\mathbf{X}}(\mathbf{x}) = C(F_{X_1}(x_1), \dots, F_{X_d}(x_d)) = C(x_1, \dots, x_d). \quad (2.26)$$

According to *Sklar's theorem* [209, without proof], [158, Chapter 2] any multivariate joint CDF may be represented by its marginal CDFs and a copula C that contains all the dependency information about the coordinates of \mathbf{X} . Further, if the marginal distributions are continuous, then C is unique. There exist various copula types that model different features such as tail dependence (Gumbel copula [158, Chapter 4]) or conditional dependence amongst groups of variables using graphical models (Vine copulas [18, 223]). The Nataf model has been shown to be equivalent to a Gaussian copula [123]. A detailed discussion of various copula types and inference of copulas is found in [158].

2.1.4 Random functions

Some definitions

Random phenomena need not be scalar quantities but can depend on an independent parameter such as time (random process) or location (random field). We may view the corresponding mathematical object as a real function-valued random variable, sometimes also referred to as random function, $H : (\Omega \times \mathcal{S}, \mathcal{F} \times \mathcal{B}(\mathcal{S})) \rightarrow \mathbb{R}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the index space $\mathcal{S} \subseteq \mathbb{R}^N$ with $\mathcal{B}(\mathcal{S})$ the Borel σ -algebra on \mathcal{S} [80, Chapter 3]. Since \mathbf{s} is a continuous index, $H(\mathbf{s})$ can be viewed as an infinite collection of random variables $\{\mathbf{s} \in \mathcal{S} : H(\mathbf{s}, \omega)\}$ that can be partially characterized in terms of its finite-dimensional distributions. The n^{th} -order CDF of $H(\mathbf{s}, \omega)$ is defined as the joint CDF of the n random variables $\{H(\mathbf{s}_1, \omega), H(\mathbf{s}_2, \omega), \dots, H(\mathbf{s}_n, \omega)\}$,

$$F_{H(\mathbf{s}_1) \dots H(\mathbf{s}_n)}(h_1, \dots, h_n; \mathbf{s}_1, \dots, \mathbf{s}_n) = \mathbb{P}[H(\mathbf{s}_1, \omega) \leq h_1 \cap \dots \cap H(\mathbf{s}_n, \omega) \leq h_n]. \quad (2.27)$$

The expectation and variance of $H(\mathbf{s}, \omega)$, $\mu_H(\mathbf{s})$ and $\sigma_H^2(\mathbf{s})$ are defined via the 1st-order or marginal CDF of H , $F_{H(\mathbf{s})}(h) = F(\mathbf{s}, h)$. If H is an absolutely continuous random function such that its PDF $f(\mathbf{s}, h)$ exists wherever $F(\mathbf{s}, h)$ is differentiable and is computed according to Eq. (2.13), then

$$\begin{aligned}\mu_H(\mathbf{s}) &= \mathbb{E}[H(\mathbf{s}, \omega)] = \int_{\mathbb{R}} hf(\mathbf{s}, h)dh, \\ \sigma_H^2(\mathbf{s}) &= \mathbb{V}[H(\mathbf{s}, \omega)] = \int_{\mathbb{R}} (h - \mu_H(\mathbf{s}))^2 f(\mathbf{s}, h)dh\end{aligned}\tag{2.28}$$

assuming the integrals exist. In analogy to Eq. (2.19), the auto-covariance function Γ_H and auto-correlation function ρ_H of H are based on its 2nd-order distribution and are defined as

$$\begin{aligned}\Gamma_H(\mathbf{s}_1, \mathbf{s}_2) &= \mathbb{E}[(H(\mathbf{s}_1, \omega) - \mu_H(\mathbf{s}_1))(H(\mathbf{s}_2, \omega) - \mu_H(\mathbf{s}_2))], \\ \rho_H(\mathbf{s}_1, \mathbf{s}_2) &= \frac{\Gamma_H(\mathbf{s}_1, \mathbf{s}_2)}{\sigma_H(\mathbf{s}_1)\sigma_H(\mathbf{s}_2)}.\end{aligned}\tag{2.29}$$

Γ_H and ρ_H are bounded (by the Cauchy-Schwarz-inequality), symmetric (to the line $\mathbf{s}_1 = \mathbf{s}_2$) and positive semi-definite functions.

$H(\mathbf{s}, \omega)$ is said to be *stationary* if all of its finite-dimensional distributions are translation invariant [80, Chapter 3], i.e.,

$$F_{H(\mathbf{s}_1)\dots H(\mathbf{s}_n)}(h_1, \dots, h_n; \mathbf{s}_1, \dots, \mathbf{s}_n) = F_{H(\mathbf{s}_1+\boldsymbol{\xi})\dots H(\mathbf{s}_n+\boldsymbol{\xi})}(h_1, \dots, h_n; \mathbf{s}_1 + \boldsymbol{\xi}, \dots, \mathbf{s}_n + \boldsymbol{\xi}).\tag{2.30}$$

In practice, random functions are described by 1st- and 2nd-order distributions only due to the impracticality or impossibility of obtaining higher-order distributions. $H(\mathbf{s}, \omega)$ is said to be *weakly stationary* if its 1st- and 2nd-order distributions are translation-invariant [80, Chapter 3], i.e.,

$$F_{H(\mathbf{s}_1)H(\mathbf{s}_2)}(h_1, h_2; \mathbf{s}_1, \mathbf{s}_2) = F_{H(\mathbf{s}_1+\boldsymbol{\xi})H(\mathbf{s}_2+\boldsymbol{\xi})}(h_1, h_2; \mathbf{s}_1 + \boldsymbol{\xi}, \mathbf{s}_2 + \boldsymbol{\xi}).\tag{2.31}$$

In either case, $\mu_H(\mathbf{s})$ is constant and the second-moment functions $\Gamma_H(\mathbf{s}_1, \mathbf{s}_2)$ and $\rho_H(\mathbf{s}_1, \mathbf{s}_2)$ are merely functions of the separation vector $\boldsymbol{\delta} = \mathbf{s}_1 - \mathbf{s}_2$ [3, Chapter 2]. (Weakly) stationary random functions whose second-moment functions depend equally on each coordinate of $\boldsymbol{\delta}$ are referred to as *isotropic*. This is, e.g., the case, when $\Gamma_H(\mathbf{s}_1, \mathbf{s}_2)$ and $\rho_H(\mathbf{s}_1, \mathbf{s}_2)$ depend on the ℓ_p -norm of $\boldsymbol{\delta}$, $\|\boldsymbol{\delta}\|_p$.

Typical isotropic correlation models are the exponential model $\rho(\boldsymbol{\delta}) = \exp\{-\|\boldsymbol{\delta}\|_1/l\}$, the squared-exponential model $\rho(\boldsymbol{\delta}) = \exp\{-(\|\boldsymbol{\delta}\|/l)^2/2\}$ and the Matérn model [187, Chapter 4]

$$\rho(\boldsymbol{\delta}; \nu, r) = (2^{1-\nu})/\Gamma(\nu)(\sqrt{2\nu}/r\|\boldsymbol{\delta}\|)^\nu K_\nu(\sqrt{2\nu}/r\|\boldsymbol{\delta}\|).\tag{2.32}$$

$\Gamma(\cdot)$ is the Gamma-function and $K_\nu(\cdot)$ is the modified Bessel function of the second kind [2, Chapter 9]. The Matérn model includes both the exponential and the squared-exponential model as special cases with $\nu = 0.5$ and $\nu \rightarrow \infty$, respectively as well as $r = l$. In all three models, l is a scale parameter that controls the rate of decay of the autocorrelation with the separation distance. The major difference between the models are the varying degrees of differentiability they impose upon H in the mean-square sense [3, Chapter 2]. This can be roughly understood as an indication as to how smooth sample trajectories of $H(\mathbf{s}, \omega)$ will be. In particular, if $H(\mathbf{s}, \omega)$ is Gaussian, the Matérn model is $\lfloor \nu \rfloor$ times mean-square differentiable which implies that the exponential model is non-differentiable and the squared-exponential model is infinitely differentiable in the mean-square sense.

Gaussian random functions [1, 187] are random functions whose every finite-dimensional distribution is a Gaussian distribution. In particular, the finite-dimensional distribution of order n is n -variate Gaussian such that the corresponding PDF of the random vector $\mathbf{H} = [H(\mathbf{s}_1, \omega), \dots, H(\mathbf{s}_n, \omega)]^T$ is given by Eq. (2.20) with mean vector $\boldsymbol{\mu}_H = [\mu_H(\mathbf{s}_1), \dots, \mu_H(\mathbf{s}_n)]$ and covariance matrix $\boldsymbol{\Sigma}_H = [\Gamma_H(\mathbf{s}_i, \mathbf{s}_j)]_{n \times n}$. Since $H(\mathbf{s}, \omega)$ is completely specified by $\boldsymbol{\mu}_H$ and $\boldsymbol{\Sigma}_H$, any weakly stationary Gaussian random function is stationary.

Translation functions are an important class of non-Gaussian random functions that arise through marginally transforming Gaussian random functions. That is, a translation function H with non-Gaussian marginal distribution F_H is obtained as $H(\mathbf{s}, \omega) = F_H^{-1} \circ \Phi(U(\mathbf{s}, \omega))$ with $U(\mathbf{s}, \omega)$ a standard-normal random function, i.e., $U(\omega)|\mathbf{s} \sim \mathcal{N}(u|0, 1)$. The covariance structure of $H(\mathbf{s}, \omega)$, $\Gamma_H(\mathbf{s}_1, \mathbf{s}_2)$, implicitly depends on $\Gamma_U(\mathbf{s}_1, \mathbf{s}_2)$ and can be obtained using, e.g., the Nataf model.

Discretization

Due to the impossibility of treating an infinite collection of random variables computationally, random functions are discretized to make them accessible to computation. There exists a variety of discretization methods all of which aim at representing $H(\mathbf{s}, \omega)$ as accurately as possible using a finite number of random variables. These methods are usually divided in point discretization methods, averaging methods and series expansion methods. In point discretization methods, a set of random variables are used to represent the values of $H(\mathbf{s}, \omega)$ at discrete coordinates in the index space \mathcal{S} and are taken in a linear combination with a set of deterministic functions to represent $H(\mathbf{s}, \omega)$ on the entire space \mathcal{S} . Particular methods then typically differ in the choice of these deterministic functions, where examples include the midpoint method ([50]; piecewise constant), the shape function method ([135]; piecewise linear) as well as the optimal linear estimation method ([112]; optimal variance representation). Averaging methods represent the random function as weighted averages of H taken over discrete subdomains of \mathcal{S} and include the spatial averaging approach [228] and the weighted integral method [46, 47]. Finally, series expansion methods also represent $H(\mathbf{s}, \omega)$ as a linear combination of random variables and deterministic functions, where, however, the random variables do not necessarily correspond to local values of $H(\mathbf{s}, \omega)$. Examples include the expansion-optimal linear estimation method [112], the spectral representation method [206] and the Karhunen-Loève-expansion (KLE) [136]. In the following, we briefly describe the midpoint method and the KLE due to their relevance to several example applications showcased in the original publications in Part II.

The midpoint method belongs to the group of point discretization methods and is perhaps the simplest approach to discretizing a random function. The idea is to ensure the m^{th} -order distribution of the approximated random function $\hat{H}(\mathbf{s}, \omega)$ at m discrete points in $\{\mathbf{s}_i \in \mathcal{S}\}_{i=1}^m$ (the midpoints) equals that of $H(\mathbf{s}, \omega)$. The midpoints are typically chosen as the barycenters of a polygonal partitioning of \mathcal{S} and the value $\hat{H}(\mathbf{s}, \omega)$ takes at any midpoint is assumed constant over the corresponding polygon. If $H(\mathbf{s}, \omega)$ is Gaussian, we can generate samples from its m^{th} -order distribution through Eq. (2.23). Likewise, if $H(\mathbf{s}, \omega)$ is a translation function, the Nataf transform is available to approximately sample from $\hat{H}(\mathbf{s}, \omega)$.

The KLE can be viewed as a continuous extension of Eq. (2.23) for $H(\mathbf{s}, \omega)$ with arbitrary finite-dimensional distributions (not necessarily Gaussian) and continuous covariance function $\Gamma_H(\mathbf{s}_1, \mathbf{s}_2)$

that — according to *Mercer's theorem* — admits a spectral decomposition such that the KLE reads

$$H(\mathbf{s}, \omega) = \boldsymbol{\mu}_{\mathbf{X}} + \sum_{i=1}^{\infty} \sqrt{\lambda_i} a_i(\mathbf{s}) \xi_i(\omega) \quad \text{with} \quad \int_{\mathcal{S}} \Gamma_{\mathbf{H}}(\mathbf{s}, \mathbf{s}') a_i(\mathbf{s}') d\mathbf{s}' = \lambda_i a_i(\mathbf{s}). \quad (2.33)$$

The eigenfunctions $\{a_i(\mathbf{s})\}_{i=1}^{\infty} : \mathcal{S} \rightarrow L^2(\mathcal{S})$ form a complete orthogonal basis (COB) of $L^2(\mathcal{S})$. The eigenvalues $\{\lambda_i\}_{i=1}^{\infty}$ are ordered in the sense that $\lambda_i \leq \lambda_j \Leftrightarrow i \leq j$ and have the property $\lim_{i \rightarrow \infty} \lambda_i = 0$. Due to the orthogonality of the $\{a_i(\mathbf{s})\}_{i=1}^{\infty}$, the variance of the random function representation (2.33) reads $\mathbb{V}[H(\mathbf{s}, \omega)] = \sum_{i=1}^{\infty} \lambda_i a_i^2(\mathbf{s})$ meaning that each eigenvalue represents the variance contributed by its associated eigenfunction. Thus, the truncated KLE

$$\widehat{H}_m(\mathbf{s}, \omega) = \boldsymbol{\mu}_{\mathbf{X}} + \sum_{i=1}^m \sqrt{\lambda_i} a_i(\mathbf{s}) \xi_i(\omega) \quad (2.34)$$

is optimal in the sense that it minimizes the mean squared error $\int_{\mathcal{S}} \mathbb{E}[(H(\mathbf{s}, \omega) - \widehat{H}_m(\mathbf{s}, \omega))^2] d\mathbf{s} = \mathbb{E}[\|H(\mathbf{s}, \omega) - \widehat{H}_m(\mathbf{s}, \omega)\|^2] = \sum_{i=m+1}^{\infty} \lambda_i$. In general, the $\{\xi_i\}_{i=1}^m$ are obtained by projecting $H(\mathbf{s}, \omega)$ on the eigenbasis $\{a_i\}_{i=1}^m$ as

$$\xi_i = \frac{1}{\sqrt{\lambda_i}} \int_{\mathcal{S}} [H(\mathbf{s}, \omega) - \boldsymbol{\mu}_H(\mathbf{s})] a_i(\mathbf{s}) d\mathbf{s} \quad (2.35)$$

and are always centered and uncorrelated, i.e. $\mathbb{E}[\xi_i] = 0$ and $\mathbb{E}[\xi_i \xi_j] = \delta_{ij}$. For Gaussian random functions $H(\mathbf{s}, \omega)$, the latter property implies pairwise independence amongst the coordinates of $\boldsymbol{\xi}$. In fact, $\boldsymbol{\xi}$ is a (standard-normal) Gaussian random vector as the integral in Eq. (2.35) can be written as a weighted Riemann sum of Gaussian random variables, which is Gaussian again. Finding the joint distribution of $\boldsymbol{\xi}$ in the general case of non-Gaussian random functions is a non-trivial undertaking. However, translation functions can be generated in the familiar manner by generating a KLE for a Gaussian random function and then transforming it. The integral eigenvalue equation on the right of Eq. (2.33) is a Fredholm equation of the second kind and its solution presents the main computational task when discretizing Gaussian or translation random functions with a KLE. Several numerical solution techniques are discussed in [24].

2.2 Uncertainty Propagation

2.2.1 Setup

Throughout this work, the computational model $\mathcal{Y} : \mathbf{X} \rightarrow Y$ is assumed to be a scalar- and real-valued mapping from the canonical probability space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P}_{\mathbf{X}})$ to the probability space of the model response $Y \in (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_Y)$. In practice, the input random vector \mathbf{X} will always be defined in terms of its joint PDF $f_{\mathbf{X}}(\mathbf{x})$ or its joint CDF $F_{\mathbf{X}}(\mathbf{x})$. We assume that $\mathcal{Y} \in L^2_{f_{\mathbf{X}}}(\mathbb{R}^d, \mathbb{R}) = \{u : \mathbb{R}^d \rightarrow \mathbb{R} \text{ such that } \int_{\mathbb{R}^d} |u(\mathbf{x})|^2 f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \leq \infty\}$. This means that the model response is square-integrable with respect to $f_{\mathbf{X}}$: $\mathbb{E}_{f_{\mathbf{X}}} [|\mathcal{Y}(\mathbf{X})|^2] \leq \infty$. Note, that this setting may be extended to the vector-valued case of \mathcal{Y} with model responses $\mathbf{Y} \in \mathbb{R}^{d_Y}$ by applying the concepts introduced hereafter to each coordinate of \mathcal{Y} , $\{Y_i = \mathcal{Y}_i(\mathbf{X})\}_{i=1}^{d_Y}$. Moreover, \mathcal{Y} is assumed to be a *black box*, meaning the sole information we can retrieve from \mathcal{Y} are point-wise evaluations. In particular, no intermediate results between input and output can be probed and no additional quantities of interest

such as derivatives can be computed alongside the model output (however, numerical differentiation of the model using the black box is possible, of course).

The central objective of uncertainty propagation lies in characterizing the uncertainty in the model response Y . The PDF of the model response, f_Y provides an exhaustive description of said uncertainty. However, estimates of f_Y are rarely available in satisfactory quality due to the large amount of information required to construct them. The target is thus often replaced with partial knowledge of the distribution, e.g., as encoded by a finite number of its moments.

2.2.2 Second moment analysis

The goal here is to compute the first two moments of Y , $\mu_Y = \mathbb{E}_{f_{\mathbf{X}}}[\mathcal{Y}(\mathbf{X})]$ and $\sigma_Y^2 = \mathbb{E}_{f_{\mathbf{X}}}[(\mathcal{Y}(\mathbf{X}) - \mu_Y)^2]$. For the special case of a linear model,

$$Y = \mathbf{a}^T \mathbf{X} + b \quad \text{with } \mathbf{a} \in \mathbb{R}^d, b \in \mathbb{R}, \quad (2.36)$$

by linearity of expectation and the definition of covariance in Eq. (2.19), we have

$$\begin{aligned} \mu_Y &= \mathbf{a}^T \boldsymbol{\mu}_{\mathbf{X}} + b \\ \sigma_Y^2 &= \mathbf{a}^T \boldsymbol{\Sigma}_{\mathbf{X}} \mathbf{a} \end{aligned} \quad (2.37)$$

with $\boldsymbol{\mu}_{\mathbf{X}}$ the mean vector of \mathbf{X} and $\boldsymbol{\Sigma}_{\mathbf{X}}$ its covariance matrix.

If \mathcal{M} is nonlinear, we may develop \mathcal{M} in a Taylor series around $\boldsymbol{\mu}_{\mathbf{X}}$ to obtain arbitrary-order approximations to μ_Y and σ_Y . The m -th order approximations of μ_Y and σ_Y require model derivatives of order m and \mathbf{X} -statistics of order m and m^2 , respectively. Estimated quantities are distinguished from their true counterparts with a $\hat{\cdot}$. The second-order approximation of \mathcal{Y} around $\boldsymbol{\mu}_{\mathbf{X}}$ reads

$$\begin{aligned} \hat{\mathcal{Y}}(\mathbf{X}) &= \mathcal{Y}(\boldsymbol{\mu}_{\mathbf{X}}) + \sum_{i=1}^d (X_i - \mu_{X_i}) \frac{\partial \mathcal{Y}}{\partial x_i} \Big|_{\boldsymbol{\mu}_{\mathbf{X}}} + \frac{1}{2} \sum_{j=1}^d \sum_{i=1}^d (X_i - \mu_{X_i})(X_j - \mu_{X_j}) \frac{\partial^2 \mathcal{Y}}{\partial x_i \partial x_j} \Big|_{\boldsymbol{\mu}_{\mathbf{X}}} \\ &= \mathcal{Y}(\boldsymbol{\mu}_{\mathbf{X}}) + [\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}]^T \nabla_{\mathbf{x}} \mathcal{Y}(\boldsymbol{\mu}_{\mathbf{X}}) + \frac{1}{2} [\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}]^T \mathbf{H}_{\mathbf{x}} \mathcal{Y}(\boldsymbol{\mu}_{\mathbf{X}}) [\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}]. \end{aligned} \quad (2.38)$$

$\nabla_{\mathbf{x}} \mathcal{Y} = [\partial \mathcal{Y} / \partial x_1, \dots, \partial \mathcal{Y} / \partial x_d]^T \in \mathbb{R}^d$ is a d -dimensional column vector that represents the gradient of \mathcal{Y} with respect to the coordinates of \mathbf{x} and $\mathbf{H}_{\mathbf{x}} \mathcal{Y} \in \mathbb{R}^{d \times d}$ is the Hessian matrix of second derivatives of \mathcal{Y} with respect to \mathbf{x} with entries $[\mathbf{H}_{\mathbf{x}} \mathcal{Y}]_{ij} = \partial^2 \mathcal{Y} / (\partial x_i \partial x_j)$. Therefore, if second-order statistics of \mathbf{X} and second-order derivatives of \mathcal{Y} are available, the second-order approximation of μ_Y and first-order approximation of σ_Y^2 can be computed as

$$\begin{aligned} \hat{\mu}_Y &= \mathcal{Y}(\boldsymbol{\mu}_{\mathbf{X}}) + \frac{1}{2} \sum_{j=1}^d \sum_{i=1}^d \text{COV}[X_i, X_j] \frac{\partial^2 \mathcal{Y}}{\partial x_i \partial x_j} \Big|_{\boldsymbol{\mu}_{\mathbf{X}}} = \mathcal{Y}(\boldsymbol{\mu}_{\mathbf{X}}) + \frac{1}{2} \|\mathbf{H}_{\mathbf{x}} \mathcal{Y}(\boldsymbol{\mu}_{\mathbf{X}}) \circ \boldsymbol{\Sigma}_{\mathbf{X}}\|_F^2, \\ \hat{\sigma}_Y^2 &= \mathcal{Y}(\boldsymbol{\mu}_{\mathbf{X}}) + \sum_{j=1}^d \sum_{i=1}^d \text{COV}[X_i, X_j] \frac{\partial \mathcal{Y}}{\partial x_i} \Big|_{\boldsymbol{\mu}_{\mathbf{X}}} \frac{\partial \mathcal{Y}}{\partial x_j} \Big|_{\boldsymbol{\mu}_{\mathbf{X}}} = [\nabla_{\mathbf{x}} \mathcal{Y}(\boldsymbol{\mu}_{\mathbf{X}})]^T \boldsymbol{\Sigma}_{\mathbf{X}} \nabla_{\mathbf{x}} \mathcal{Y}(\boldsymbol{\mu}_{\mathbf{X}}). \end{aligned} \quad (2.39)$$

$\mathbf{A} \circ \mathbf{B}$ denotes the Hadamard (i.e., element-wise) product of two matrices \mathbf{A} and \mathbf{B} of equal size and $\|\mathbf{A}\|_F = (\sum_i \sum_j A_{ij})^{1/2}$ is the Frobenius norm of \mathbf{A} .

Computational models of engineering problems frequently do not possess a closed form representation but come as a set of discretized differential equations which represent conservation laws and other first principles. Given the black-box nature of \mathcal{Y} , the required derivatives in Eq. (2.39) must be obtained using numerical differentiation. Alternatively, if \mathcal{Y} is not of black-box type, the equations underlying the model may be manipulated such as to yield explicit expressions for the sought response derivatives. This is known as the *probabilistic perturbation method* [106].

2.2.3 Monte Carlo

Monte Carlo simulation (MC) is an umbrella term for stochastic simulation methods that are based on repeated experiments with artificially generated random samples [196]. MC can be used to characterize the response Y through a finite number of samples from the PDF of y , f_Y . These are obtained by generating a set of n random samples of \mathbf{X} , $\{\mathbf{x}_i\}_{i=1}^n$, and computing the corresponding values of Y , $\{y_i = \mathcal{Y}(\mathbf{x}_i)\}_{i=1}^n$. Moments of Y can be estimated from the sample $\{y_i = \mathcal{Y}(\mathbf{x}_i)\}$. In particular, standard estimators for the mean μ_Y and variance σ_Y^2 of Y read

$$\hat{\mu}_Y = \frac{1}{n} \sum_{i=1}^m y_i, \quad \hat{\sigma}_Y^2 = \frac{1}{n-1} \sum_{i=1}^m (y_i - \hat{\mu}_Y)^2, \quad y_i \stackrel{i.i.d.}{\sim} f_Y \quad (2.40)$$

$y_i \stackrel{i.i.d.}{\sim} f_Y$ indicates that the samples $\{y_i = \mathcal{Y}(\mathbf{x}_i)\}_{i=1}^n$ are independent and identically distributed according to f_Y . These estimators are unbiased, which means $\mathbb{E}[\hat{\mu}_Y] = \mu_Y$ and $\mathbb{E}[\hat{\sigma}_Y^2] = \sigma_Y^2$, and converge to the true mean μ_Y and variance σ_Y^2 asymptotically as $n \rightarrow \infty$ due to the law of large numbers. If the samples are i.i.d., the estimators are random variables of their own and the variance of $\hat{\mu}_Y$ is $\sigma_{\hat{\mu}_Y}^2 = \sigma_Y^2/n$. Estimation errors are typically quantified with their coefficient of variation (CoV)

$$\delta_{\hat{\mu}} = \frac{\sigma_{\hat{\mu}}}{\hat{\mu}} = \frac{\sigma}{\mu\sqrt{n}} \approx \frac{\hat{\sigma}}{\hat{\mu}\sqrt{n}}. \quad (2.41)$$

Eq. (2.41) reveals both the relatively slow convergence of MC ($\propto \sqrt{n}$) as well as the remarkable property of dimension independence.

Often, we are not only interested in estimating moments of Y from samples but also its PDF $f_Y(y)$. The empirical PDF is given as a sum of Dirac delta functions $f_{\text{emp}}(y) = 1/n \sum_{i=1}^n \delta(y - y_i)$, where δ is defined as

$$\delta(y) = \begin{cases} 0, & y \neq 0, \\ 1, & y = 0 \end{cases}, \quad \text{with } \int_{-\infty}^{\infty} \delta(y) dy = 1. \quad (2.42)$$

The most common technique to obtain a smooth PDF estimate \hat{f}_Y is kernel density estimation. The kernel density estimator is given by the convolution of $f_{\text{emp}}(y)$ with a kernel K and reads

$$\hat{f}_Y(y) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right). \quad (2.43)$$

Next to the kernel, the bandwidth h has to be chosen such as to do justice to the data (see Fig. 2.1). If chosen too large, the resulting density estimate is oversmoothed, which leads to information loss; if chosen too small, the resulting $\hat{f}_Y(y)$ will exhibit mostly features of K and contain spurious artifacts

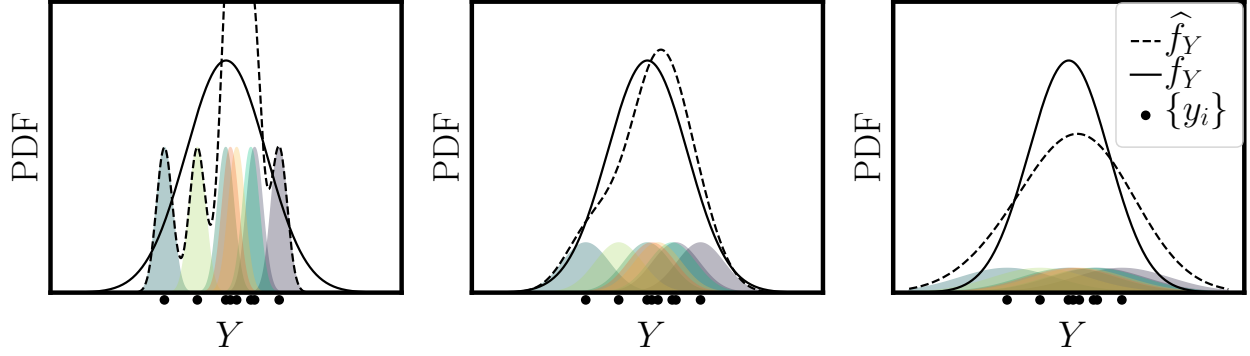


Figure 2.1: Illustration of undersmoothed (left, bandwidth $h \leq h_{\text{opt}}$), optimal (center, bandwidth $h_{\text{opt}} = 1.06\widehat{\sigma}_Y n^{-1/5}$) and oversmoothed (right, bandwidth $h \geq h_{\text{opt}}$) kernel density estimators of a Gaussian distribution with Gaussian kernel K . Colored lines represent the weighted kernel contributions $\{1/(nh)K((y_i - y)/h)\}_{i=1}^n$. The dashed line emerges as the sum of the colored lines.

due to undersmoothing/overfitting. Popular choices of K and h are the Gaussian kernel $\varphi(\cdot)$ and $h = 1.06\widehat{\sigma}_Y n^{-1/5}$ [208, Chapter 3]. If computationally feasible, Monte Carlo is usually the simplest and one of the most reliable approaches to estimate moments or parameters in general. However, when \mathcal{Y} is expensive to evaluate, the number of samples required by MC to achieve a prescribed accuracy is often unaffordable. In such cases, a possible remedy is to construct cheap-to-evaluate surrogate models of \mathcal{Y} , which we discuss in the following section.

2.2.4 Surrogate Modelling

Introduction

Surrogate models are used in uncertainty quantification to alleviate the computational burden of repeated model evaluations if the model is expensive. This is frequently the case in engineering, where — as mentioned above — models often compute solutions of differential equation systems representing first principles such as conservation of mass and energy. Surrogate models are functions $\widehat{\mathcal{Y}}(\mathbf{X}; \boldsymbol{\theta})$ that are parameterized by $\boldsymbol{\theta}$ and considerably cheaper to evaluate than \mathcal{Y} . The goal is to replace the expensive model \mathcal{Y} with the cheap surrogate model $\widehat{\mathcal{Y}}$. For this replacement to be sensible, the parameters are computed based on a set of evaluations of the original model referred to as the design of experiments (DoE) or experimental design, $\mathcal{E} = \{\mathbf{X}, \mathbf{Y}\}_{i=1}^n$ with $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n \in \mathbb{R}^{n \times d}$ and $\mathbf{Y} = \{y_i = \mathcal{Y}(\mathbf{x}_i)\}_{i=1}^n \in \mathbb{R}^n$ such that $\widehat{\mathcal{Y}}$ becomes as similar to \mathcal{Y} as possible in some sense. In what sense is made precise by the loss function $\mathcal{L}(\mathcal{Y}(\mathbf{X}), \widehat{\mathcal{Y}}(\mathbf{X}, \boldsymbol{\theta}))$ that provides a measure of dissimilarity between $\widehat{\mathcal{Y}}$ and \mathcal{Y} for any viable choice of parameters $\boldsymbol{\theta}$, such that the goal is to minimize \mathcal{L} over \mathcal{E} . Without loss of generality, \mathbf{X} is assumed standardized in the remainder of this chapter. That means a linear transformation $\{X_i = (\tilde{X}_i - \mu_{\tilde{X},i})/\sigma_{\tilde{X},i}\}_{i=1}^d$ is applied to the non-standardized input matrix $\tilde{\mathbf{X}}$, where $\boldsymbol{\mu}_{\mathbf{X}}$ and $\boldsymbol{\sigma}_{\mathbf{X}}$ are the columnwise mean and standard deviation of $\tilde{\mathbf{X}}$. Standardization serves to reduce scaling influence amongst different variables in regression methods.

Within the field of *statistical learning theory* [83], surrogate modelling for UQ as introduced above is best described as *supervised, model-based or predictive learning*. Most surrogates are designed to

solve regression problems meaning that the output of \mathcal{Y} is of quantitative nature rather than categorical. Conversely, if the output is categorical, the underlying problem is referred to as classification problem. Sometimes, classification methods are useful in UQ, e.g., in solving reliability and rare event estimation problems [94, 32] (see Chapter 2.3). In high-dimensional applications, surrogate modelling is combined with/preceded by a dimensionality reduction step for which both supervised and unsupervised methods may be deployed. Unsupervised methods are geared towards detecting structures in the \mathbf{X} -data of the DoE only, meaning these methods do not receive output information (also termed *labeled* data in the context of statistical learning theory). If, however, as in the case of a surrogate modelling setting, label information is available, supervised dimensionality reduction techniques usually have an advantage over their unsupervised pendants. Nonparametric methods make use of the DoE to compute predictions immediately rather than identifying parameters $\boldsymbol{\theta}$ first and using the resulting model for prediction. An example is the kernel density estimator in Eq. (2.43).

Different surrogate models arise by choosing different functional forms for $\widehat{\mathcal{Y}}$ as well as different loss functions \mathcal{L} . The black box assumption for \mathcal{Y} dictates that any and all surrogate modelling approaches discussed in the following and used in Part II are so-called *non-intrusive* approaches. An incomplete list of non-intrusive modelling approaches includes linear [83, Section 3.2] and nonlinear [204] as well as generalized [146] and additive models [83, Chapter 9], spectral methods, Gaussian process [187] and support vector regression [55, 210] as well as neural networks [76, 95, 27]. In the following, several regression methods as well as spectral methods such as polynomial chaos expansions (PCE) and spectral low-rank approximations (LRA) are discussed in more detail due to their relevance in Part II. Moreover, an account of dimensionality-reducing regression techniques is given introducing principal component regression (PCR) and linear as well as nonlinear partial least squares (PLS).

Error measures, model selection & validation

When using surrogate models in UQ, the interest is in accurate predictions for future unobserved data. Thus, quantifying the prediction error of $\widehat{\mathcal{Y}}$ away from the training set is key in assessing the surrogate model quality. Upon determining an estimate of the model parameters $\boldsymbol{\theta}$ based on a fixed DoE \mathcal{E} , $\widehat{\boldsymbol{\theta}}$, taking the expectation of $\mathcal{L}(\mathcal{Y}(\mathbf{X}), \widehat{\mathcal{Y}}(\mathbf{X}, \widehat{\boldsymbol{\theta}}))$ with respect to the input PDF $f_{\mathbf{X}}$ yields the *generalization error* or *expected risk* $\text{Err}_{\mathcal{L}}$. As $\widehat{\boldsymbol{\theta}}$ is fixed, we drop it and write

$$\text{Err}_{\mathcal{L}} = \int_{\mathbb{R}} \mathcal{L}(\mathcal{Y}(\mathbf{X}), \widehat{\mathcal{Y}}(\mathbf{X})) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \mathbb{E}_{f_{\mathbf{X}}} \left[\mathcal{L}(\mathcal{Y}(\mathbf{X}), \widehat{\mathcal{Y}}(\mathbf{X})) \right]. \quad (2.44)$$

Evaluating $\text{Err}_{\mathcal{L}}$ requires point-wise knowledge of \mathcal{Y} , which defeats the purpose of constructing a surrogate model $\widehat{\mathcal{Y}}$. An approximation of $\text{Err}_{\mathcal{L}}$ is given by the *empirical error* $\widehat{\text{Err}}_{\mathcal{L}}$, which amounts to evaluating Eq. (2.44) over \mathcal{E} :

$$\widehat{\text{Err}}_{\mathcal{L}} = \frac{1}{n} \sum_{i=1}^n \mathcal{L}(Y_i, \widehat{\mathcal{Y}}(\mathbf{x}_i)). \quad (2.45)$$

Minimizing $\widehat{\text{Err}}_{\mathcal{L}}$ is one way of computing the parameter estimate $\widehat{\boldsymbol{\theta}}$. In regression problems, the square loss function $\mathcal{L} = (\mathcal{Y}(\mathbf{X}) - \widehat{\mathcal{Y}}(\mathbf{X}, \boldsymbol{\theta}))^2$ is a standard choice and is employed throughout this work. We write lower-case *err* for *relative* error estimates. A natural normalization for relative square-loss errors is the variance of Y and the corresponding relative generalization and empirical

errors read

$$\text{err}_{\ell_2} = \frac{\mathbb{E}_{f_{\mathbf{X}}} \left[\left(\mathcal{Y}(\mathbf{X}) - \widehat{\mathcal{Y}}(\mathbf{X}) \right)^2 \right]}{\mathbb{V}_{f_{\mathbf{X}}} [\mathcal{Y}(\mathbf{X})]} \quad \text{and} \quad \widehat{\text{err}}_{\ell_2} = \frac{\frac{1}{n} \sum_{i=1}^n \left(Y_i - \widehat{\mathcal{Y}}(\mathbf{x}_i) \right)^2}{\frac{1}{n-1} \sum_{i=1}^n \left(Y_i - \frac{1}{n} \sum_{i=1}^n Y_i \right)^2}. \quad (2.46)$$

err_{ℓ_2} (and Err_{ℓ_2}) is also referred to as (*relative*) *mean-squared error* (RMSE). They express the accuracy of $\widehat{\mathcal{Y}}$ conditional on a fixed DoE \mathcal{E} (and corresponding $\widehat{\boldsymbol{\theta}}$ given a deterministic estimation procedure). In order to describe the sampling properties of $\widehat{\mathcal{Y}}$ under random DoEs, we may define the pointwise MSE at any input $\mathbf{x} \in \mathbb{R}$ and average over the DoE distribution $f_{\mathcal{E}}$ (or equivalently – but less likely to be accessible – the sampling distribution of $\widehat{\boldsymbol{\theta}}$) to obtain the pointwise MSE

$$\text{Err}_{\ell_2}^{\mathcal{E}}(\mathbf{x}) = \mathbb{E}_{f_{\widehat{\boldsymbol{\theta}}}} \left[\left(\mathcal{Y}(\mathbf{x}) - \widehat{\mathcal{Y}}(\mathbf{x}, \widehat{\boldsymbol{\theta}}) \right)^2 \right] = \mathbb{E}_{f_{\mathcal{E}}} \left[\left(\mathcal{Y}(\mathbf{x}) - \widehat{\mathcal{Y}}(\mathbf{x}, \mathcal{E}) \right)^2 \right]. \quad (2.47)$$

$\text{Err}_{\ell_2}^{\mathcal{E}}$ admits a decomposition in two parts representing systematic and random errors as follows

$$\begin{aligned} \text{Err}_{\ell_2}^{\mathcal{E}}(\mathbf{x}) &= \mathbb{E}_{f_{\mathcal{E}}} \left[\left(\mathcal{Y}(\mathbf{x}) - \widehat{\mathcal{Y}}(\mathbf{x}, \mathcal{E}) \right)^2 \right] \\ &= \left(\mathbb{E}_{f_{\mathcal{E}}} \left[\mathcal{Y}(\mathbf{x}) - \widehat{\mathcal{Y}}(\mathbf{x}, \mathcal{E}) \right] \right)^2 + \mathbb{V}_{f_{\mathcal{E}}} \left[\mathcal{Y}(\mathbf{x}) - \widehat{\mathcal{Y}}(\mathbf{x}, \mathcal{E}) \right] \\ &= \underbrace{\left(\mathcal{Y}(\mathbf{x}) - \mathbb{E}_{f_{\mathcal{E}}} \left[\widehat{\mathcal{Y}}(\mathbf{x}, \mathcal{E}) \right] \right)^2}_{\text{Bias}^2(\mathbf{x})} + \underbrace{\mathbb{V}_{f_{\mathcal{E}}} \left[\widehat{\mathcal{Y}}(\mathbf{x}, \mathcal{E}) \right]}_{\text{Variance}(\mathbf{x})}. \end{aligned} \quad (2.48)$$

The bias can be understood as measuring how far off the surrogate model is on average when reconstructed with infinitely many randomly redrawn DoEs of size n , \mathcal{E} . Thus, the bias accounts for systematic errors, e.g., in the model form, that are independent of the random draw of the DoE. The variance part of the MSE on the other hand represents the variability of $\widehat{\mathcal{Y}}$ due to the random character of the DoE and the fact that it is of limited size n and vanishes in theory as $n \rightarrow \infty$.

Being a function of the DoE only, the empirical error is not informative regarding how well $\widehat{\mathcal{Y}}$ generalizes/extrapolates away from the data. In particular, the relative empirical error and the related coefficient of determination $R^2 = 1 - \widehat{\text{err}}_{\ell_2}$ can always be improved by adding more parameters to \mathcal{Y} . This means that choosing the complexity of the model (e.g., the number of parameters in $\boldsymbol{\theta}$) based on $\widehat{\text{err}}_{\ell_2}$ is likely to result in overly complex models that grant too much importance to the DoE and generalize poorly away from it. This is referred to as *overfitting*. One way of resolving this issue is to select the model and its complexity based on modified versions of err_{ℓ_2} that penalize model complexity (see, e.g., [36] for a correction to the empirical error in linear regression problems). Another is to consider loss functions that impose structure such as sparsity on the model (these approaches are discussed in more detail in Subsection 2.2.4). A less heuristic way of performing model selection is given by *k-fold cross-validation*, where the DoE is divided in k equal-sized partitions that each contain n/k data points. Each partition is then removed from the DoE once and a corresponding model is constructed without this partition present in the DoE leading to a set of k models $\{\widehat{\mathcal{Y}}_i\}_{i=1}^k$. The empirical error of each model is then evaluated over its corresponding left out data partition and thereby measures generalization ability. The relative cross-validation error is computed as the average of these k errors normalized with the response variance:

$$\widehat{\text{err}}_{\text{CV},k} = \frac{\frac{1}{k} \sum_{i=1}^k \widehat{\text{Err}}_{\ell_2,i}}{\mathbb{V}[\mathbf{Y}]} \quad \text{with} \quad \widehat{\text{Err}}_{\ell_2,k} = \frac{1}{n} \sum_{i=1}^n \left(Y_i - \widehat{\mathcal{Y}}(\mathbf{x}_i) \right)^2. \quad (2.49)$$

If $k = n$, n models are constructed and each one is constructed using all points in the DoE except one. Therefore, n -fold cross validation is also known as *leave-one-out cross validation* (LOO-CV).

Linear regression

For each point in \mathcal{E} , (\mathbf{x}_i, y_i) (corresponding to the i -th row of \mathbf{X} and \mathbf{Y} , respectively), the linear regression model assumes a linear relationship contaminated with an error term ϵ_i :

$$y_i = \widehat{\mathcal{Y}}(\mathbf{x}_i) + \epsilon_i = a_0 + a_1 x_{i1} + \cdots + a_d x_{id} + \epsilon_i, \quad i = 1, \dots, n. \quad (2.50)$$

In the context of regression, \mathbf{X}/Y are often referred to as *regressors/regressees*, *independent/dependent variables* or *explaining/explained variables*. Several assumptions are made for the linear regression model:

- A1 (*strict exogeneity of the error terms*) The error terms have zero mean conditional on all observations of the regressors \mathbf{X} , i.e., $\mathbb{E}[\epsilon|\mathbf{X}] = \mathbf{0}_n$ (which implies $\mathbb{E}[\epsilon] = \mathbf{0}_n$).
- A2 (*spherical error model*) The error terms are independent of one another conditional on all observations of the regressors and their variance is constantly σ^2 , i.e., $\mathbb{E}[\epsilon\epsilon^T|\mathbf{X}] = \sigma^2\mathbf{I}_{n \times n}$ (which implies $\mathbb{E}[\epsilon\epsilon^T] = \sigma^2\mathbf{I}_{n \times n}$ as well)
- A3 (*no multicollinearity*) \mathbf{X} has full rank. This last property is required by the solution of the least-squares problem Eq. (2.53).

Sometimes, the assumption of Gaussian errors is added, which allows for maximum-likelihood and Bayesian estimation of \mathbf{a} and yields finite-sample properties of the least squares estimator such as unbiasedness and the covariance structure of the estimated parameters $\widehat{\mathbf{a}}$.

The linear model can include nonlinear functions $\psi(\cdot)$ of one, several or all independent variables as long as the model remains linear in \mathbf{a} . These functions are called features and enhance the flexibility of the linear model. The *design matrix* $\Psi \in \mathbb{R}^{n \times P}$ is the matrix representing P features evaluated at all n DoE-points, i.e., $\Psi_{ij} = \psi_j(\mathbf{x}_i)$. The constant term a_0 can be collected in Ψ by choosing $\psi_1(\mathbf{X}) = 1$, such that the model in matrix notation and with the Gaussianity assumption for ϵ reads

$$\mathbf{Y} = \Psi\mathbf{a} + \epsilon, \quad \text{with} \quad \epsilon|\Psi \sim \mathcal{N}(\mathbf{e}|\mathbf{0}_n, \sigma^2\mathbf{I}_{n \times n}). \quad (2.51)$$

The parameters of this model are $\theta = [\mathbf{a}^T, \sigma^2]^T$. In many presentations of linear regression, the conditioning of ϵ on the design matrix is dropped as Ψ is assumed fixed and deterministic. To maintain that the design matrix in general may be random but the properties of OLS discussed in this paragraph do not account for said randomness, we carry the conditioning along in the notation. The primary goal is to find the coefficient vector $\mathbf{a} \in \mathbb{R}^P$. This may be achieved by minimizing the squared distance between \mathcal{Y} and $\widehat{\mathcal{Y}}$ over \mathcal{E} , which is known as the (ordinary) least squares method (OLS) and implies a quadratic loss function \mathcal{L} , which is also referred to as the *residual sum of squares* (RSS) in the context of regression. The minimization reads

$$\widehat{\mathbf{a}} = \arg \min_{\mathbf{a} \in \mathbb{R}^{P+1}} \text{RSS}(\mathbf{a}) = \arg \min_{\mathbf{a} \in \mathbb{R}^{P+1}} \sum_{i=1}^n (Y_i - \widehat{Y}_i(\mathbf{a}))^2 = \arg \min_{\mathbf{a} \in \mathbb{R}^{P+1}} (\mathbf{Y} - \Psi\mathbf{a})^T (\mathbf{Y} - \Psi\mathbf{a}). \quad (2.52)$$

Eq. (2.52) is a convex and continuous optimization problem and setting the derivative of the RSS with respect to \mathbf{a} to 0 yields the *normal equations*

$$\hat{\mathbf{a}} = (\mathbf{\Psi}^T \mathbf{\Psi})^{-1} (\mathbf{\Psi}^T \mathbf{Y}). \quad (2.53)$$

An unbiased estimate of the error variance σ^2 reads

$$\hat{\sigma}^2 = \frac{1}{n - P} (\mathbf{Y} - \mathbf{\Psi} \hat{\mathbf{a}})^T (\mathbf{Y} - \mathbf{\Psi} \hat{\mathbf{a}}). \quad (2.54)$$

The number $n - P$ is referred to as the *degrees of freedom* of the regression model and can be understood as the number of unconstrained residuals (P residuals are constrained by Eq. (2.53), which limits its variability [84]). An estimator with few degrees of freedom is prone to overfitting. In the limit $n = P$, the problem is ill-posed and Eq. (2.54) is not defined. Strategies for treating the $n \leq P$ -case are discussed in Subsection 2.2.4. An alternative derivation of Eq. (2.53) can be obtained by maximizing the log-likelihood of \mathbf{a} given \mathcal{E} [155, Section 7.3]. This means, the least-squares estimator of the linear model coincides with its maximum-likelihood-estimator under the assumed Gaussian and spherical error model. Based on assumptions A1 and A2 above, unbiasedness and the covariance structure of $\hat{\mathbf{a}}$ are readily stated as

$$\begin{aligned} \mathbb{E}[\hat{\mathbf{a}} - \mathbf{a} | \mathbf{\Psi}] &= \mathbb{E}[(\mathbf{\Psi}^T \mathbf{\Psi})^{-1} \mathbf{\Psi}^T \boldsymbol{\epsilon} | \mathbf{\Psi}] \\ &= (\mathbf{\Psi}^T \mathbf{\Psi})^{-1} \mathbf{\Psi}^T \mathbb{E}[\boldsymbol{\epsilon} | \mathbf{\Psi}] \stackrel{A1}{=} \mathbf{0} \\ \mathbb{V}[\hat{\mathbf{a}} | \mathbf{\Psi}] &= \mathbb{V}[\hat{\mathbf{a}} - \mathbf{a} | \mathbf{\Psi}] = \mathbb{V}[(\mathbf{\Psi}^T \mathbf{\Psi})^{-1} \mathbf{\Psi}^T \boldsymbol{\epsilon} | \mathbf{\Psi}] \\ &= (\mathbf{\Psi}^T \mathbf{\Psi})^{-1} \mathbf{\Psi}^T \mathbb{V}[\boldsymbol{\epsilon} | \mathbf{\Psi}] \mathbf{\Psi} (\mathbf{\Psi}^T \mathbf{\Psi})^{-T} \stackrel{A2}{=} \sigma^2 (\mathbf{\Psi}^T \mathbf{\Psi})^{-1}. \end{aligned} \quad (2.55)$$

If $\boldsymbol{\epsilon}$ is assumed Gaussian, $\hat{\mathbf{a}}$ asymptotically has distribution $\hat{\mathbf{a}} \sim \mathcal{N}(\boldsymbol{\alpha} | \mathbf{a}, \sigma^2 (\mathbf{\Psi}^T \mathbf{\Psi})^{-1})$ as $n \rightarrow \infty$. The Gauss theorem states that Eq. (2.53) is the best linear unbiased estimator (BLUE) of \mathbf{a} given the model Eq. (2.51) [203]. If $\boldsymbol{\epsilon}$ is more generally distributed according to a member of the *exponential family* of distributions, the model parameters can be estimated with the maximum likelihood principle as well. Such models are known as generalized linear models (GLM) [146].

Frequently, modellers need to decide which subset amongst all available features to include in a linear regression model. This is a model selection problem and as mentioned above, cross-validation is frequently used to select the subset that minimizes the cross-validation error. In the case of linear regression, the leave-one-out cross validation error is particularly convenient to evaluate [203] and reads

$$\widehat{\text{Err}}_{\text{LOO}} = \sum_{i=1}^n \left(\frac{\mathcal{Y}(\mathbf{x}_i) - \hat{\mathcal{Y}}(\mathbf{x}_i)}{1 - h_i} \right)^2, \quad (2.56)$$

with h_i being the i -th main diagonal element of the matrix $\mathbf{\Psi} (\mathbf{\Psi}^T \mathbf{\Psi})^{-1} \mathbf{\Psi}^T$. This means the LOO-CV may be computed without constructing n different models, which renders it an attractive tool for model selection in linear regression problems.

Regularization, ridge regression & subset selection

Eq. (2.53) explains the necessity of assumption A3 above: if $\mathbf{\Psi}$ is of full rank, $(\mathbf{\Psi}^T \mathbf{\Psi})$ is invertible and the least squares solution exists. In the context of regression, the problem of rank-deficient

Ψ is often referred to as *multicollinearity*. Even if the features (i.e., the columns of Ψ) are not linearly dependent, they may be strongly correlated, which renders the computation of the inverse numerically unstable in many instances. A remedy is provided by regularization of the loss function such that

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a} \in \mathbb{R}^{P+1}} (\mathbf{Y} - \Psi \mathbf{a})^T (\mathbf{Y} - \Psi \mathbf{a}) + \lambda \mathcal{R}(\mathbf{a}). \quad (2.57)$$

λ controls the amount of regularization and is usually identified with cross-validation and $\mathcal{R}(\mathbf{a})$ is a penalty term. Choosing the penalty term as $\mathcal{R}(\mathbf{a}) = \|\mathbf{a}\|^2 = \mathbf{a}^T \mathbf{a}$ penalizes the complexity of the parameter vector. This penalty is known as *Tikhonov regularization* and the associated regression technique is sometimes called *ridge regression*. Eq. (2.57) is a convex and continuous optimization problem as well and attains its global minimum at

$$\hat{\mathbf{a}} = (\Psi^T \Psi + \lambda \mathbf{I}_{P \times P})^{-1} (\Psi^T \mathbf{Y}). \quad (2.58)$$

As the introduced penalty term leads to the addition of λ to all main diagonal entries of $(\Psi^T \Psi)$, it has a stabilizing effect on the computation of the inverse matrix. In other words, the regression problem is well-posed again. The result of Eq. (2.58) can also be obtained in a Bayesian setting as the maximum a-posteriori point (MAP) by reusing the Gaussian OLS-likelihood and placing a suitable prior on \mathbf{a} . In particular, $\mathbf{a} \sim \mathcal{N}(\cdot | \mathbf{0}, \tau^2 \mathbf{I})$, and known σ^2 then yield the MAP Eq. (2.58) with $\lambda = \sigma^2 / \tau^2$ [155, Section 7.5].

By choosing $\mathcal{R}(\mathbf{a}) = \|\mathbf{a}\|_1 = \sum_{i=1}^P |a_i|$, it is possible to penalize the number of non-zero parameters, which leads to *subset selection* amongst the available regressors and produces sparse solutions. This regularization scheme is also known as *Least Absolute Shrinkage and Selection Operator* (LASSO) [220] or — in a slight variation of formulation — *Basis Pursuit* [38]. The associated optimization problem is non-continuous and thus does not possess a simple closed-form solution like ordinary or ridge regression. Instead, algorithms such as least-angle regression (LARS) [59], coordinate descent and proximal gradients [155, Section 13.4] may be used to solve for the parameter vector.

Subset selection can be achieved with various penalty types, e.g., by choosing an ℓ_0 -regularization term $\mathcal{R}(\mathbf{a}) = \|\mathbf{a}\|_0 = \sum_{i=1}^P \mathbf{I}(a_i = 0)$, which however results in a non-convex optimization problem. Algorithms for solving such problems proceed by adding regressors to the solution in a stepwise manner and decide which regressors to add in each step based on heuristic principles. Popular examples are forward-backward selection [244], matching pursuit [141] and its orthogonal variants [180, 43, 224, 53]. Similar to OLS and ridge regression, subset selection can be achieved by stating the regression problem in a Bayesian setting and choosing an appropriate (that is, sparsity-inducing) prior — such as the Laplace distribution — for the model parameters. A Bayesian framework for sparse linear models was proposed in [222] and the idea went on to produce a plethora of descendants in the realm of sparse Bayesian learning ever since [103, 11, 225]. For a detailed overview, the reader is referred to [155, Section 13.7].

Both subset selection and ridge regression can be viewed as trading bias for variance in Eq. (2.48) to decrease the MSE compared to OLS. The estimator variance reduces as less parameters have to be estimated based on a fixed DoE size, whereas the bias increases as a consequence of artificially forcing otherwise small parameters to zero (subset selection) or at least shrinking them below their OLS-value (ridge regression). Subset selection also renders the model more interpretable compared to OLS through reducing the number of explaining variables retained in the model.

Nonlinear regression

The nonlinear regression model no longer assumes linearity in \mathbf{a} so that any nonlinear function of \mathbf{X} and \mathbf{a} may be postulated:

$$\mathbf{Y} = h(\mathbf{X}, \mathbf{a}) + \boldsymbol{\epsilon}, \quad \text{with} \quad \boldsymbol{\epsilon}|\mathbf{X} \sim \mathcal{N}(\mathbf{e}|\mathbf{0}_n, \sigma^2 \mathbf{I}_{n \times n}). \quad (2.59)$$

In special cases, this model can be linearized by a suitable transformation of \mathbf{X} and/or \mathbf{Y} . Linearization, however, bears implications concerning the error model: if, for example, in the original regression model, the error is assumed additive spherical Gaussian, then a log-transformation of Y leads to multiplicative errors with log-normal distribution. Whether this is appropriate has to be checked carefully. Even if no such transformation can be applied, Eq. (2.59) may be solved using least squares and an iterative procedure known as Gauss-Newton algorithm. Upon choosing an initial value $\hat{\mathbf{a}}_k \in \mathbb{R}^P$ at $k = 0$, expanding $h(\mathbf{X}, \mathbf{a})$ in a first-order Taylor series about $\hat{\mathbf{a}}_k$ yields

$$h(\mathbf{X}, \mathbf{a}) \approx \tilde{h}(\mathbf{X}, \mathbf{a}) = h(\mathbf{X}, \hat{\mathbf{a}}_k) + \nabla_{\mathbf{a}} h(\mathbf{X}, \hat{\mathbf{a}}_k) \underbrace{(\mathbf{a} - \hat{\mathbf{a}}_k)}_{\Delta \mathbf{a}}. \quad (2.60)$$

$\nabla_{\mathbf{a}} h(\mathbf{X}, \hat{\mathbf{a}}_k)$ is short notation for $\nabla_{\mathbf{a}} h(\mathbf{X}, \mathbf{a})|_{\mathbf{a}=\hat{\mathbf{a}}_k} \in \mathbb{R}^{n \times P}$. \tilde{h} represents the tangent plane to the true nonlinear model parameter surface at $\hat{\mathbf{a}}_k$. Minimizing the distance between \mathbf{Y} and \tilde{h} yields the best possible \mathbf{a} on this plane, which is used as \mathbf{a}_{k+1} and identified through the following updating rule:

$$\begin{aligned} \hat{\mathbf{a}}_{k+1} &= \hat{\mathbf{a}}_k + \widehat{\Delta \mathbf{a}}_{\text{opt}} \\ &= \hat{\mathbf{a}}_k + \arg \min_{\Delta \mathbf{a} \in \mathbb{R}^P} (\mathbf{Y} - h(\mathbf{X}, \hat{\mathbf{a}}_k) - \nabla_{\mathbf{a}} h(\mathbf{X}, \hat{\mathbf{a}}_k) \Delta \mathbf{a})^T (\mathbf{Y} - h(\mathbf{X}, \hat{\mathbf{a}}_k) - \nabla_{\mathbf{a}} h(\mathbf{X}, \hat{\mathbf{a}}_k) \Delta \mathbf{a}) \\ &= \hat{\mathbf{a}}_k + ([\nabla_{\mathbf{a}} h(\mathbf{X}, \hat{\mathbf{a}}_k)]^T \nabla_{\mathbf{a}} h(\mathbf{X}, \hat{\mathbf{a}}_k))^{-1} [\nabla_{\mathbf{a}} h(\mathbf{X}, \hat{\mathbf{a}}_k)]^T (\mathbf{Y} - h(\mathbf{X}, \hat{\mathbf{a}}_k)). \end{aligned} \quad (2.61)$$

The second equality amounts to selecting the optimal update $\widehat{\Delta \mathbf{a}}_{\text{opt}}$ as the one that minimizes the distance between \mathbf{Y} and \tilde{h} . The last equality makes use of the known linear least squares solution, Eq. (2.53). There exist several extensions to the Gauss-Markov algorithm that improve both efficiency and stability, e.g., the Levenberg-Marquardt-algorithm that uses the ℓ_2 -regularized least squares solution in Eq. (2.53) rather than OLS to compute the update.

An asymptotic estimate of the covariance structure of the final parameter estimate $\hat{\mathbf{a}}$ follows from invoking the linear approximation Eq. (2.60) once more. Upon convergence, $\hat{\mathbf{a}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y}$, where $\mathbf{A} = \nabla_{\mathbf{a}} h(\mathbf{X}, \hat{\mathbf{a}}_k)$. Based on this, it can be shown that $\hat{\mathbf{a}} - \mathbf{a} \sim \mathcal{N}(\boldsymbol{\alpha}|\mathbf{0}, \mathbb{V}[\hat{\mathbf{a}} - \mathbf{a}|\mathbf{X}])$ for large n [204, Theorem 2.1]. The covariance matrix of $\hat{\mathbf{a}}$ is then given by

$$\begin{aligned} \mathbb{V}[\hat{\mathbf{a}}|\mathbf{X}] &= \mathbb{V}[\hat{\mathbf{a}} - \mathbf{a}|\mathbf{X}] \approx \mathbb{V}[(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \boldsymbol{\epsilon}|\mathbf{X}] \\ &= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbb{V}[\boldsymbol{\epsilon}|\mathbf{X}] \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-T} \stackrel{A2}{=} \sigma^2 (\mathbf{A}^T \mathbf{A})^{-1}. \end{aligned} \quad (2.62)$$

Principal component regression and partial least squares

We briefly return to the linear least squares problem to review a collection of subspace methods that are relevant for the construction of dimensionality-reducing surrogate models in Part II. The

idea is to solve the linear regression problem in a linear subspace of \mathbb{R}^d as defined by $\mathbf{W} \in \mathbb{R}^{d \times m}$ such that $\mathbf{Z} = \mathbf{W}^T \mathbf{X} \in \mathbb{R}^m$ and $\mathbf{W}^T \mathbf{W} = \mathbf{I}_{m \times m}$ with $m < p$. Such an approach is useful when the regression model has only a few degrees of freedom in the original space or even $d > n$. It also serves as an alternative to regularized regression approaches in that it can handle multicollinearity in \mathbf{X} and potentially reduce $\mathbb{V}[\mathbf{a}]$. It is straightforward to show that the matrix maximizing $\mathbb{V}[\mathbf{Z}]$ and minimizing the reconstruction error $\|\mathbf{X} - \mathbf{W}\mathbf{W}^T \mathbf{X}\|_F$ in Frobenius-norm (as defined below Eq. (2.39)) is given by the m eigenvectors of the empirical covariance of \mathbf{X} , $\widehat{\Sigma}_{\mathbf{X}}$, associated with the m largest eigenvalues of that matrix [155, Section 12.2]. Assuming \mathbf{X} was centered such that it has column mean equal 0, the full eigendecomposition of $\widehat{\Sigma}_{\mathbf{X}}$ reads

$$\widehat{\Sigma}_{\mathbf{X}} = \frac{1}{n} \mathbf{X}^T \mathbf{X} = \mathbf{V} \mathbf{D} \mathbf{V}^T \quad \text{with } \mathbf{V} = [\mathbf{v}_1 \dots, \mathbf{v}_d] \in \mathbb{R}^{d \times d} \quad \text{and } \mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_d) \in \mathbb{R}^{d \times d}, \quad (2.63)$$

where the eigenpairs $\{\lambda_i, \mathbf{v}_i\}_{i=1}^d$ are solutions of $\widehat{\Sigma}_{\mathbf{X}} \mathbf{v}_i = \lambda_i \mathbf{v}_i$. Thus, setting $\mathbf{W} = [\mathbf{v}_1 \dots, \mathbf{v}_m] = \mathbf{V}_m$, the empirical covariance of \mathbf{Z} is $\widehat{\Sigma}_{\mathbf{Z}} = \mathbf{V}_m^T \widehat{\Sigma}_{\mathbf{X}} \mathbf{V}_m = \text{diag}(\lambda_1, \dots, \lambda_m)$. That is, the eigenvalues equal the variance represented by their associated principal component. For the sake of stability, computing the PCR subspace \mathbf{V}_m via the eigendecomposition in Eq. (2.63) is typically replaced with computing the *singular value decomposition* (SVD) of \mathbf{X} (in particular, the *thin* SVD [75, Section 2.4]). The SVD is a matrix factorization that can be understood as generalizing the eigendecomposition to non-square matrices [75, Section 2.4]. For any real-valued matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$, the SVD consists of the orthogonal matrices $\mathbf{U} \in \mathbb{R}^{n \times n}$ and $\tilde{\mathbf{V}} \in \mathbb{R}^{d \times d}$ as well as the rectangular diagonal matrix $\mathbf{S} \in \mathbb{R}^{n \times d}$ with the singular values of \mathbf{X} , $\{s_i\}_{i=1}^{\min(n,d)}$ on the main diagonal in decreasing order. The SVD reads

$$\mathbf{X} = \mathbf{U} \mathbf{S} \tilde{\mathbf{V}}^T = \sum_{i=1}^{\min(n,d)} s_i \mathbf{u}_i \mathbf{v}_i^T \quad \text{with } \mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_d] \quad \text{and } \tilde{\mathbf{V}} = [\mathbf{v}_1, \dots, \mathbf{v}_d]. \quad (2.64)$$

The second expression in Eq. (2.64) may be understood as representing \mathbf{X} with a linear combination of rank-1 matrices. By $\widehat{\Sigma}_{\mathbf{X}} = \mathbf{X}^T \mathbf{X} / n = \tilde{\mathbf{V}} \mathbf{S}^2 \tilde{\mathbf{V}}^T / n$, we see that $\tilde{\mathbf{V}}$ equals \mathbf{V} such that the column space defined by $\tilde{\mathbf{V}}$ equals the eigenspace of $\widehat{\Sigma}_{\mathbf{X}}$ and $\lambda_i = s_i^2 / n$. Computing the regression coefficients in the PCA subspace amounts to solving m univariate linear regression problems due to the orthogonality of the principal components. Choosing $m = d$ recovers the OLS solution. PCR can also be applied to feature vectors, in which case $m < P$ directions in feature space are identified.

While PCR can be effective in stabilizing and improving the OLS procedure, it is suboptimal for surrogate modelling as the DoE output data \mathbf{Y} lies waste and the linear subspace is computed solely based on the input data covariance. Supervised methods like *canonical correlation analysis* (CCA) [92], *sliced inverse regression* [129] and *partial least squares* (PLS) [237, 239] offer a remedy by using \mathbf{Y} in the construction of a relevant subspace.

In the following, we discuss PLS1, a PLS variant for scalar-valued output data. The extension to vector-valued output is straight-forward and known as PLS2. Just as PCA, PLS identifies a set of principal components that however maximizes the squared empirical cross-covariance between input and output $(\mathbf{X}^T \mathbf{Y} / n)^2$ (the quadratic form prevents positively and negatively correlated entries from cancelling one another) rather than the autocovariance of the input (as is done in PCA). The direction maximizing this quantity is the first *weight*

$$\mathbf{w}_1 = \frac{\mathbf{X}^T \mathbf{Y}}{\|\mathbf{X}^T \mathbf{Y}\|}. \quad (2.65)$$

The input data projected along the new direction is referred to as the first *score* $\mathbf{t}_1 = \mathbf{X}\mathbf{w}_1 \in \mathbb{R}^n$. The rank-1 reconstruction of \mathbf{X} , $\widehat{\mathbf{X}}_1 = \mathbf{t}_1\mathbf{p}_1^T$ is obtained by solving the linear regression problem to find the first *load* $\mathbf{p}_1 = \mathbf{X}^T\mathbf{t}_1/\mathbf{t}_1^T\mathbf{t}_1 \in \mathbb{R}^d$. In *linear* PLS, the *inner mapping* between the univariate projected input \mathbf{t}_1 and \mathbf{Y} is taken to be a linear regression model $\mathbf{Y} = b_1\mathbf{t}_1 + \boldsymbol{\epsilon}$, whereby $b_1 = \mathbf{t}_1^T\mathbf{Y}/\mathbf{t}_1^T\mathbf{t}_1$. The rank-1 approximation of \mathbf{Y} is thus $\widehat{\mathbf{Y}}_1 = b_1\mathbf{t}_1$. This completes the first step of linear PLS. The following step proceeds identically on the residual input and output matrices $\mathbf{E}_2 = \mathbf{X} - \widehat{\mathbf{X}}_1$ and $\mathbf{F}_2 = \mathbf{Y} - \widehat{\mathbf{Y}}_1$ to extract the second principal component and so forth (compare Fig. 2.2). Various termination criteria have been proposed to identify the optimal number of components in the PLS subspace, e.g., based on cross-validation or simply by tracking residual output variability (as in Fig. 2.2). Upon termination with k components, the weights, loads and scores are gathered in matrices $\mathbf{W} \in \mathbb{R}^{d \times k}$, $\mathbf{T} \in \mathbb{R}^{n \times k}$ and $\mathbf{P} \in \mathbb{R}^{d \times k}$ to compute $\mathbf{R} = \mathbf{W}(\mathbf{P}^{-T}\mathbf{W})^{-1} \in \mathbb{R}^{d \times k}$, where \mathbf{R} maps the input data to the PLS scores, i.e., $\mathbf{T} = \mathbf{R}\mathbf{X}$. The procedure is visualized in Fig. 2.2. To overcome the limitation of linear inner mappings in linear PLS models, any residual \mathbf{F}_k may be

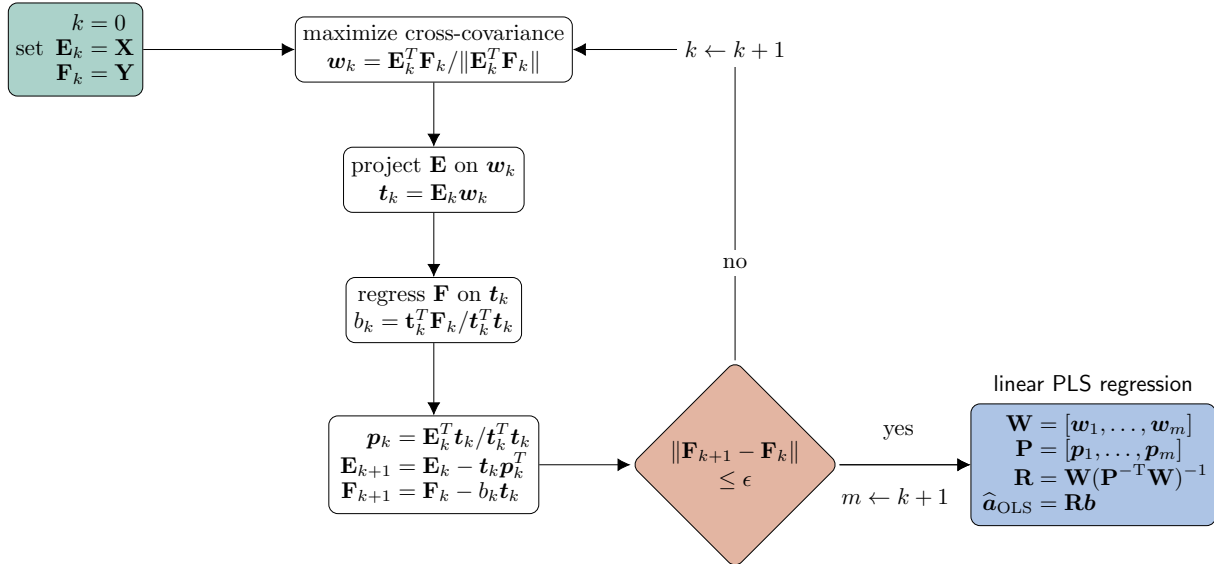


Figure 2.2: Flow chart of linear partial least squares regression with univariate output (PLS1).

assumed as an arbitrary nonlinear function of \mathbf{t}_k parametrized by \mathbf{b}_k along the k -th direction [238]. This implies solving a nonlinear regression problem for the model

$$\mathbf{F}_k = h(\mathbf{E}_k\mathbf{w}_k; \mathbf{b}_k) + \boldsymbol{\epsilon}, \quad \text{with} \quad \boldsymbol{\epsilon}|\mathbf{X} \sim \mathcal{N}(\mathbf{e}|\mathbf{0}_n, \sigma^2\mathbf{I}_{n \times n}). \quad (2.66)$$

[13] presents a modified version of the approach put forward in [238] that is reminiscent of the Gauss-Markov algorithm: in the first step, the nonlinear inner mapping is fitted to obtain \mathbf{b}_k (an initial guess of \mathbf{t}_k is computed with linear PLS) and then a first-order approximation as in Eq. (2.61) is used to compute a weight update $\Delta\mathbf{w}_k$. This process is repeated until a stationary solution for the set of coefficients $\{\mathbf{w}_k, \mathbf{b}_k\}$ is achieved (according to an upper bound on the norm of $\Delta\mathbf{w}_k$, see also Fig. 2.3). Several other nonlinear PLS methods have been proposed, a comprehensive overview of which (including kernel PLS) is provided in [192]. In publication 1 (Chapter 4), we have proposed a nonlinear PLS technique to construct PCEs for high-dimensional surrogate modelling problems. PCEs belong to the class of spectral expansions that are discussed in the next section.

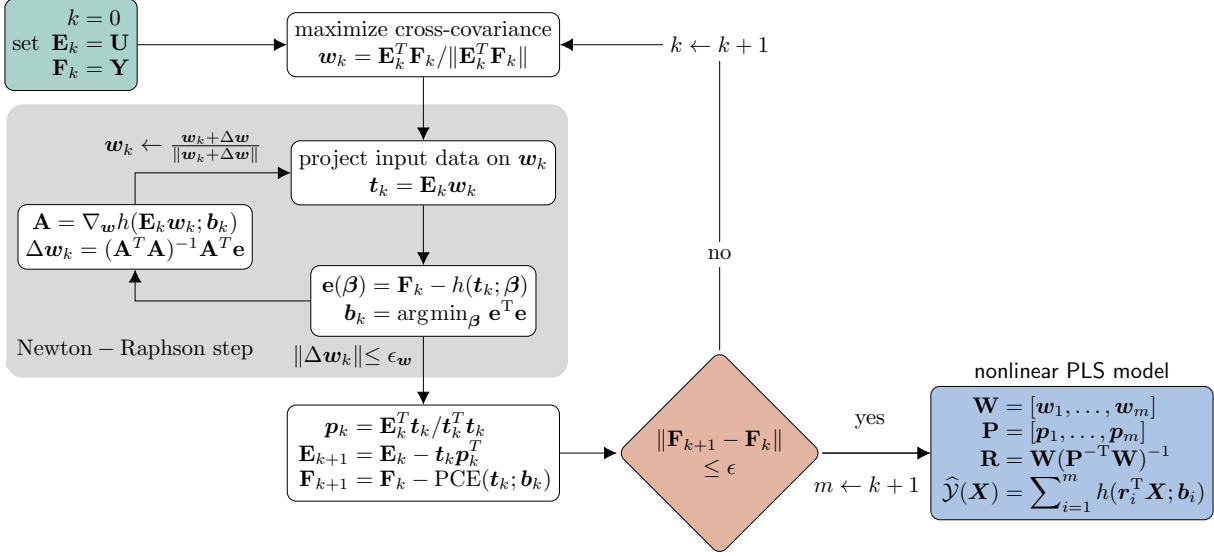


Figure 2.3: Flow chart of nonlinear partial least squares regression.

Stochastic Spectral Methods

Stochastic spectral methods are based on representing \mathcal{Y} on a complete orthonormal basis (CONB) of the weighted Lebesgue space $L^2_{f_{\mathbf{X}}}(\mathbb{R}^d, \mathbb{R})$ (a Lebesgue space with $p = 2$ is a *Hilbert space*). This approach was first introduced in [236] for Gaussian product measures as *homogeneous chaos* and repopularized in the context of *stochastic finite element methods* (SFEM) in [73]. The latter approach uses polynomial chaos expansions as ansatz functions for a Galerkin procedure in order to solve stochastic boundary value problems. This is an example of an intrusive approach that requires manipulating the governing equations describing the problem and is not compatible with the black box limitation assumed in this work. Other examples of intrusive spectral stochastic methods are [122, 12]. The vast majority of Hilbert spaces arising in the description of physical systems are *separable* and any Hilbert spaces used in the following discussion are assumed to be separable. The inner product associated with $L^2_{f_{\mathbf{X}}}(\mathbb{R}^d, \mathbb{R})$ is defined as

$$\langle u, v \rangle_{f_{\mathbf{X}}} = \int_{\mathbb{R}^d} u(\mathbf{x})v(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} = \mathbb{E}_{f_{\mathbf{X}}}[u(\mathbf{X})v(\mathbf{X})]. \quad (2.67)$$

By definition, an orthonormal basis of $L^2_{f_{\mathbf{X}}}(\mathbb{R}^d, \mathbb{R})$, $\{h_i(\mathbf{x})\}_{i \in \mathbb{N}}$, satisfies

$$\langle h_i, h_j \rangle_{f_{\mathbf{X}}} = \delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = 1 \end{cases}. \quad (2.68)$$

Then, since \mathcal{Y} belongs to the space spanned by $\{h_i(\mathbf{x})\}_{i \in \mathbb{N}}$, we may express it as a linear combination of the basis functions:

$$\mathcal{Y}(\mathbf{X}) = \sum_{i=0}^{\infty} a_i h_i(\mathbf{X}). \quad (2.69)$$

Due to the orthonormality property of the $\{h_i(\mathbf{x})\}_{i \in \mathbb{N}}$, the coefficients $\{a_i\}_{i \in \mathbb{N}}$ can be evaluated by projecting \mathcal{Y} on the corresponding basis elements:

$$\langle \mathcal{Y}, h_j \rangle_{f_{\mathbf{X}}} = \left\langle \sum_{i=0}^{\infty} a_i h_i, h_j \right\rangle_{f_{\mathbf{X}}} = \sum_{i=0}^{\infty} a_i \langle h_i, h_j \rangle_{f_{\mathbf{X}}} = \sum_{i=0}^{\infty} a_i \delta_{ij} = a_j. \quad (2.70)$$

If the coordinates of \mathbf{X} are pairwise independent, then $f_{\mathbf{X}}$ is a product measure as in Eq. (2.16). Consequently, $L_{f_{\mathbf{X}}}^2(\mathbb{R}^d, \mathbb{R})$ is isomorphic to a tensor product of the set of univariate Hilbert spaces weighted with the marginal densities $\{f_{X_k}(x_k)\}_{k=1}^d$,

$$L_{f_{X_1}}^2(\mathbb{R}, \mathbb{R}) \otimes L_{f_{X_2}}^2(\mathbb{R}, \mathbb{R}) \otimes \cdots \otimes L_{f_{X_d}}^2(\mathbb{R}, \mathbb{R}). \quad (2.71)$$

Let $\{\tilde{h}_j^{(k)}(x_k)\}_{j=0}^{p_k}$ be a CONB of $L_{f_{X_k}}^2(\mathbb{R}, \mathbb{R})$. A CONB of $L_{f_{\mathbf{X}}}^2(\mathbb{R}^d, \mathbb{R})$ can then be constructed as

$$\{h_i(\mathbf{x})\}_{i \in \mathbb{N}} = \left\{ \bigotimes_{k=1}^d \tilde{h}_{\mathcal{I}_k}^{(k)}(x_k) \right\}_{\mathcal{I} \in \mathbb{N}^d}. \quad (2.72)$$

By using copulas as defined in Eq. (2.26), an isoprobabilistic transformation $T : \mathbf{X} \rightarrow \tilde{\mathbf{X}}$ may be used to transform any dependent random vector $\tilde{\mathbf{X}}$ into an independent one \mathbf{X} such that $\mathcal{Y} \circ T$ is an element of Eq. (2.71) and the above basis construction can be applied. In an alternative approach, [214] constructs a CONB for $L_{f_{\tilde{\mathbf{X}}}}^2(\mathbb{R}^d, \mathbb{R})$ by assuming that $f_{\tilde{X}_k}$ is known and $L_{f_{\tilde{X}_k}}^2(\mathbb{R}, \mathbb{R})$ admits the CONB $\{\tilde{h}_j^{(k)}(x_k)\}_{j=0}^{p_k}$ such that

$$\{h_i(\tilde{\mathbf{x}})\}_{i \in \mathbb{N}} = \left\{ \frac{\bigotimes_{k=1}^d \sqrt{f_{\tilde{X}_k}}(\tilde{x}_k) \tilde{h}_{\mathcal{I}_k}^{(k)}(\tilde{x}_k)}{\sqrt{f_{\tilde{\mathbf{X}}}(\tilde{\mathbf{x}})}} \right\}_{\mathcal{I} \in \mathbb{N}^d}. \quad (2.73)$$

By using the copula PDF $c(\mathbf{u}) = \partial^d C(\mathbf{u}) / (\partial u_1 \cdots \partial u_d)$, this reduces to

$$\{h_i(\tilde{\mathbf{x}})\}_{i \in \mathbb{N}} = \left\{ \frac{\bigotimes_{k=1}^d \tilde{h}_{\mathcal{I}_k}^{(j)}(\tilde{x}_k)}{\sqrt{c(F_{\tilde{X}_1}(\tilde{x}_1), \dots, F_{\tilde{X}_d}(\tilde{x}_d))}} \right\}_{\mathcal{I} \in \mathbb{N}^d}. \quad (2.74)$$

Thus, going forward, we consider independent \mathbf{X} only as either of the two described methods may be used to construct stochastic spectral expansions of dependent random vectors.

Truncating representation Eq. (2.69) after P terms gives

$$\hat{\mathcal{Y}}_P(\mathbf{X}) = \sum_{i=0}^{P-1} a_i h_i(\mathbf{X}). \quad (2.75)$$

This approximation is guaranteed to converge to $\mathcal{Y}(\mathbf{X})$ in mean-square as $p \rightarrow \infty$ [183].

Polynomial chaos expansions

Polynomial chaos expansions are spectral representations that arise by restricting the space of admissible basis functions to orthonormal polynomials. Such an orthonormal polynomial basis of $L^2_{f_{\mathbf{X}}}(\mathbb{R}^d, \mathbb{R})$ can be constructed, e.g., with a Gram-Schmidt-orthonormalization procedure [188] for arbitrary $f_{\mathbf{X}}$. [242] proposed the *generalized polynomial chaos expansion* by exploiting the fact that different known families of orthogonal polynomials can be normalized to form CONBs of Hilbert spaces weighted with PDFs of different standard distribution types (see Tab. 2.1 for a few important examples).

Upon identification of the univariate bases associated with X_k , $\{\psi_i^{(k)}(\mathbf{X})\}_{i \in \mathbb{N}}$ and assuming \mathbf{X} is an independent random vector (or has been transformed into one prior to the PCE construction), the PCE reads

$$\mathcal{Y}_P(\mathbf{X}) = \sum_{\mathcal{I} \in \mathbb{N}^d} a_{\mathcal{I}} \Psi_{\mathcal{I}}(\mathbf{X}) \quad \text{with} \quad \Psi_{\mathcal{I}}(\mathbf{X}) = \prod_{j=1}^d \psi_{\mathcal{I}_j}^{(j)}(X_j). \quad (2.76)$$

An intuitive way of truncating Eq. (2.76) is limiting the admissible total polynomial degree $p = \|\mathbf{p}\|_1 = \sum_{k=1}^d \mathcal{I}_k$, $j = 1, \dots, P$. The resulting set of multivariate polynomial basis functions has cardinality

$$P = \sum_{q=0}^p \binom{q+d-1}{q} = \binom{d+p}{d} \quad (2.77)$$

and the corresponding approximation reads

$$\hat{\mathcal{Y}}_p(\mathbf{X}) = \sum_{\|\mathcal{I}\|_1 \leq p} a_{\mathcal{I}} \Psi_{\mathcal{I}}(\mathbf{X}). \quad (2.78)$$

We can summarize all P possible combinations of univariate basis elements in a matrix $\boldsymbol{\alpha} \in \mathbb{N}_{\geq 0}^{P \times d}$, such that the polynomial orders of the univariate basis elements are stored columnwise and the P different multivariate basis elements are stored rowwise. With this definition, Eq. (2.78) becomes

$$\hat{\mathcal{Y}}_P(\mathbf{X}) = \sum_{i=0}^{P-1} a_i \prod_{k=1}^d \psi_{\alpha_{ik}}^{(k)}(X_k). \quad (2.79)$$

Finding the set of all possible combinations of univariate polynomials up to total polynomial order p is a combinatorial problem for which an efficient solution is presented in [219, Chapter 3].

Computing the coefficients $\{a_i\}_{i=0}^{P-1}$ by projection as in Eq. (2.70) can be accomplished using either the integral or the expectation formulation in Eq. (2.67). The former leads to *quadrature methods*, in which the coefficients are identified using multi-dimensional numerical integration schemes [211, 109, 12, 241]. The latter gives rise to *simulation methods* in which the expectation in Eq. (2.67) is approximated with either crude MC sampling [74] or more elaborate approaches such as latin hypercube sampling [147, 165] or quasi-random sequence sampling [160, 34, 109].

Eq. (2.79) may be understood as a linear regression model with the multivariate basis elements representing features. Thus, the PCE coefficients may be identified by solving a linear regression problem, the solution of which is given by Eq. (2.53) with design matrix $\boldsymbol{\Psi} = [\Psi_{\boldsymbol{\alpha}_1}(\mathbf{X}), \dots, \Psi_{\boldsymbol{\alpha}_P}(\mathbf{X})]$. This

solution is asymptotically equivalent to the projection solution [22] as $n \rightarrow \infty$.

Based on the orthonormality of the polynomial basis, mean and variance of $\hat{\mathcal{Y}}$ are readily computed as

$$\mathbb{E}_{f_{\mathbf{X}}}[\hat{\mathcal{Y}}_p(\mathbf{X})] = a_0 \quad \text{and} \quad \mathbb{V}_{f_{\mathbf{X}}}[\hat{\mathcal{Y}}_p(\mathbf{X})] = \sum_{0 < \|\mathcal{I}\|_1 \leq p} a_{\mathcal{I}}^2 = \sum_{i=1}^{P-1} a_i^2. \quad (2.80)$$

Moreover, global variance-based sensitivities of $\hat{\mathcal{Y}}_p(\mathbf{X})$ with respect to \mathbf{X} can be computed based on the PCE coefficients directly [217], which is detailed in Subsection 2.4.2 as well as publications 2 & 5 (Chapters 5 and 8).

The basis cardinality P given the total polynomial order p grows factorially with d and quickly surpasses feasibility bounds for many applications due to several informational and computational bottlenecks. Most prominently, for accurate estimation, the DoE size n has to grow relative to the number of explanatory variables in the regression problem, somewhere between $n = 2P$ [217, 10] and $n = \mathcal{O}(P^2)$ [40, 156] depending on the complexity of the problem. This is an informational bottleneck. The number of required true model evaluations quickly defeats the purpose of constructing a surrogate model (that purpose being to evaluate the original model as scarcely as possible). Subset selection as discussed in Subsection 2.2.4 is well-suited to tackle this particular problem and has been successfully used to construct sparse PCEs in many variations. A comprehensive overview is provided in [138]. Beyond this issue, the computational bottleneck of constructing and storing the multi-index set α additionally limits the application of PCEs in high-dimensional problems. Sorting the P multi-indices will require an effort of $2^{\mathcal{O}(P)}$ operations. The disadvantageous scaling of P with d is often relaxed through restricting the number of basis elements heuristically, e.g., by computing the total degree with a q -norm with $q < 1$ (this is referred to as hyperbolic truncation) or by only allowing a small fraction of the d coordinates to possess non-constant contributions to a multivariate basis element at a time (this is referred to as interaction order). In our experience, retaining computability of PCEs for models with $d \geq 100$ requires sacrificing expressivity by considerably restricting the basis in one way or another.

A fundamentally different approach is proposed in [221] under the name *basis adaptation*, in which the rotational symmetry of Gaussian PDFs is exploited to project high-dimensional Gaussian input random vectors to low-dimensional (one-dimensional in the extreme case) linear subspaces. These linear combinations of the original inputs will have a Gaussian measure again and thus lend themselves to a (low-dimensional) PC representation using Hermite polynomials. Thus, if a ‘good’ subspace can be identified, the dimensional limitations of PCEs may be overcome elegantly. In publication 1 (Chapter 4), we demonstrate how to use nonlinear PLS (see Subsection 2.2.4) to identify such a subspace along with an optimal subspace PCE.

Table 2.1: Some standard distributions with PDFs f_X and normalized orthogonal polynomial family spanning $L^2_{f_X}$. The gamma function is defined as $\Gamma(s) = \int_0^\infty u^{s-1} \exp\{-u\} du$. The normalization constant of the Jacobi polynomials is $c(r, s, n) = \sqrt{\frac{(2n+r+s+1)\Gamma(r+n+1)\Gamma(r+1)\Gamma(s+1)}{n!\Gamma(r+s+n+1)\Gamma(n+s+1)\Gamma(r+s+2)}}$.

disitribution	support	PDF $f_X(x)$	family	$\psi_n(x)$
Normal(0, 1)	$(-\infty, \infty)$	$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	Hermite	$\sqrt{n!} \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{(-1)^k}{k!(n-2k)!} \frac{x^{n-2k}}{2^k}$
Uniform(-1, 1)	$[-1, 1]$	$f_X(x) = \frac{1}{2}$	Legendre	$\sqrt{2n-1} \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{(-1)^k (2n-2k)! x^{n-2k}}{(n-k)!(n-2k)!k!2^n}$
Gamma(1, $\alpha + 1$)	$[0, \infty)$	$\frac{1}{\Gamma(\alpha+1)} x^\alpha e^{-x}$	Laguerre(α)	$\sqrt{\frac{k!\Gamma(\alpha+1)}{\Gamma(n+\alpha+1)}} \sum_{k=0}^n (-1)^k \binom{n+\alpha}{n-k} \frac{x^k}{k!}$
Beta($r + 1, s + 1$)	$[-1, 1]$	$\frac{\Gamma(r+s+2)}{\Gamma(r+1)\Gamma(s+1)} \frac{(x+1)^r (1-x)^s}{2^{r+s+1}}$	Jacobi(r, s)	$c(r, s, n) \sum_{k=0}^n \binom{n}{k} \frac{\Gamma(r+s+n+k+1)}{\Gamma(r+k+1)} \left(\frac{x-1}{2}\right)^k$

Low-rank apximations

Canonical decompositions (also termed *CANDECOMP*, *PARAFRAC* or *Canonical Polyadics*) represent a generalization of the SVD to tensors of dimension ≥ 3 and have first been introduced in [86]. In the same way the SVD decomposes a matrix in a weighted sum of outer products of two vectors in Eq. (2.64), canonical decompositions represent a d -way tensor as the weighted sum of outer products of d vectors. The tensor rank r is defined as the number of components required in this decomposition to exactly recover the original tensor [113]. Thus, each component corresponds to a rank 1-tensor. The canonical decomposition is the simplest of several low-rank tensor approximation formats but also places the strongest assumptions on the low-rank data structure in the full tensor. Other formats include the *tensor train* [163], the Tucker format [226] and the closely related higher-order SVD [44] as well as the hierarchical Tucker format [78]. Comprehensive reviews of low-rank tensor decomposition and approximation techniques are given in [113, 79].

By replacing the discrete d -way tensor with a d -variate continuous function that is approximated by a linear combination of rank 1-functions, a continuous canonical decomposition emerges that is referred to as *low-rank approximation* (LRA) [54, 39] in the context of spectral stochastic methods. Due to the product structure of the underlying $L^2_{f_X}(\mathbb{R}^d, \mathbb{R})$, a spectral decomposition of \mathcal{Y} is achieved by selecting the univariate orthonormal polynomial bases introduced for PCEs to build up the rank-1 functions. The resulting format is illustrated in Fig. 2.4 reads

$$\hat{\mathcal{Y}}(\mathbf{X}) = \sum_{i=1}^r a_i \underbrace{\prod_{k=1}^d \sum_{j=0}^{m_k} z_{ijk} \psi_j^{(k)}(X_k)}_{f_i^{(k)}(X_k)}, \quad (2.81)$$

where m_k is the maximally considered polynomial order in each coordinate of \mathbf{X} and thus the coefficient tensor \mathbf{z} has a total of $r \sum_{k=1}^d (m_k + 1)$ entries. The format of Eq. (2.81) can be transformed into a PCE by collecting duplicate basis functions $\{\Psi_i\}_{i=1}^{P-1}$ that arise by taking the $f_i^{(k)}(X_k)$ in a product and summing up the corresponding coefficients. Solving for all coefficients (\mathbf{a} and \mathbf{z}) at once is an ill-posed problem. A greedy approach is devised in [39], which is inspired by the greedy construction of a canonical decomposition in the discrete setting [113] and leads to a series of well-posed optimization problems. In particular, rank 1-functions are added to the model one

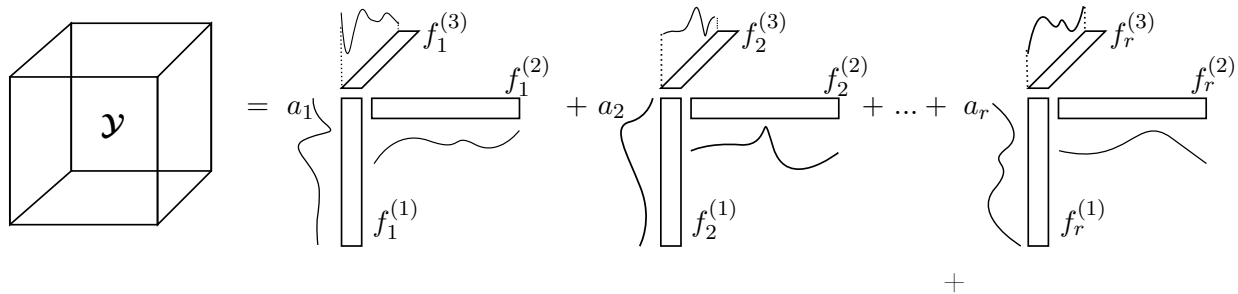


Figure 2.4: Visualization of the LRA format for a three-dimensional model \mathcal{Y} .

by one, where each new rank 1-function is fitted on the residual between the current and the true model (based on \mathcal{E}). The coefficients of the i -th single rank 1-function $\mathbf{z}_{i:}$ are computed with an alternating least squares approach. The idea is to fix all coefficients but those associated with the k -th direction and solve an ordinary least squares problem for $\mathbf{z}_{ik:}$. This procedure is repeated over all directions until the RSS (see Eq. (2.52)) rate of change drops below a prescribed threshold. This completes the *correction step*, the first of two steps that are repeated alternately. The correction step is followed by the *updating step*, in which the outer coefficients \mathbf{b} of all rank 1-functions added to the model so far are refit with a least squares procedure conditional on \mathbf{z} . The optimal stopping rank may be identified with k -fold cross validation (in the leave-one-out case Eq. (2.56) can be used).

The combination of greedy construction, ALS and linear growth of $\text{card}(\mathbf{z})$ with d make the LRA better suited to high-dimensional problems compared to (sparse) PCEs. However, while PCEs asymptotically ($P \rightarrow \infty$) converge to the true model in mean-square, there is no such guarantee for LRAs and the greedy construction is suboptimal compared to a direct approach with fixed rank r [39].

In the context of this thesis, on the one hand LRAs are used to compute surrogate-based reliability sensitivity measures within the framework introduced in publication 5 (Section 2.4). On the other hand, in publication 1 (Chapter 4), LRAs serve as a reference method for constructing high-dimensional surrogate models.

2.3 Reliability Analysis

2.3.1 Introduction

Reliability analysis is concerned with estimating the probability of failure of a model, where failure is defined in terms of unacceptable model outcomes, i.e., a (typically unlikely) subset of all possible model outcomes. Estimating a probability of failure is usually a *rare event probability estimation* [193] problem, which is a statistical discipline in its own right due to the difficulties associated with estimating small probabilities (say, of order $\leq \mathcal{O}(10^{-3})$). The reliability of a system is a key quantity in many decision problems either via constraints arising from modern structural design codes (design decision) [148] or as the target quantity driving a decision (risk-based decision analysis) [65]. A variety of approaches gathered under the terms *structural reliability methods* (SRM), *rare event simulation methods* and *rare event probability estimation methods* have been proposed to estimate the probability of failure associated with a computational model \mathcal{Y} and a given *limit-state function*

(LSF) g . The LSF is a function that defines which model responses are deemed unacceptable.

Reliability problems may be very challenging tasks due to a number of reasons: the typically small target probability (probability of failure) renders crude MC approaches inefficient due to the large required number of samples, while large numbers of random inputs present challenges for many approaches that suffer from the *curse of dimensionality* (e.g., numerical integration). Additional complexity may be due to strong non-linearities in both the underlying computational model \mathcal{Y} and the LSF g as well as through the consideration of time-dependent reliability problems (see [52, Chapter 15] for an introduction and further references) and system reliability problems [52, Chapter 14] & [124, Chapter 9], in which the interaction of several components and their reliability has to be considered. Time-dependent and system reliability problems are, however, beyond the scope of this thesis.

In the following, after a formal statement of the reliability problem, we summarize the state of the art of both sampling-based and surrogate-assisted rare event simulation methods and give an account of a few select methods in view of their relevance in Part II of this thesis.

2.3.2 The reliability problem

By convention, the LSF $g(\mathbf{X})$ describes unacceptable realizations in the input as $g(\mathbf{X}) \leq 0 \Leftrightarrow$ Failure. The associated failure event reads $F = \{\mathbf{x} \in \mathbb{R}^d : g(\mathbf{x}) \leq 0\}$. g often explicitly depends on the response of the computational model \mathcal{Y} describing the engineering system. A typical example is the exceedance of a threshold, for example if the response of \mathcal{Y} represents stresses or deformations. The corresponding LSF reads $g(\mathbf{X}) = \text{threshold} - \mathcal{Y}(\mathbf{X})$. We consider \mathbf{X} on the canonical probability space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P})$. The *probability of failure* is expressed as

$$p = \mathbf{P}(F) = \int_{\mathbb{R}^d} \mathbb{I}[g(\mathbf{x}) \leq 0] f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \mathbb{E}_{f_{\mathbf{X}}} [\mathbb{I}(g(\mathbf{X}) \leq 0)]. \quad (2.82)$$

Therein, $\mathbb{I}[\cdot]$ is an indicator function assuming the value 1 whenever its argument resolves as true and 0 otherwise. The above reliability problem can be stated in standard-normal space using an isoprobabilistic transformation $T : \mathbf{X} \rightarrow \mathbf{U}$, e.g., as in Eqs. (2.24) to (2.26). With the transformed LSF $G = g \circ T^{-1}$, the probability of failure reads

$$p = \int_{\mathbb{R}^d} \mathbb{I}[G(\mathbf{u}) \leq 0] \varphi_d(\mathbf{u}) d\mathbf{u} = \mathbb{E}_{\varphi_d} [\mathbb{I}(G(\mathbf{U}) \leq 0)], \quad (2.83)$$

where φ_d denotes the d -dimensional independent standard-normal PDF (stated here again for convenience). The failure event expressed in the probability space of \mathbf{U} reads $F = \{\mathbf{u} \in \mathbb{R}^d : G(\mathbf{u}) \leq 0\}$.

2.3.3 Sampling methods

SRM can be classified into approximation-based methods (e.g., the first- and second-order reliability method as discussed in [184, 48]) that approximate the LSF and compute the failure probability associated with the approximation, and sampling-based methods that aim at reducing the variance of the crude MC estimator. Examples include importance sampling (IS) [201, 33, 64, 6], sequential

importance sampling (SIS) [15, 176, 175], subset simulation (SUS) [7], cross-entropy-based importance sampling (CE-IS) [118, 121, 234, 196, 175], line-sampling [88, 117, 177] as well as multi-level [227, 230] and multi-fidelity [181] MC methods. In this thesis, simulation methods such as crude MC, IS, SIS, SUS and CE-IS are used and modified and are thus detailed in the following.

The crude MC estimate of Eq. (2.83) can be constructed using n independent and identically distributed (i.i.d.) samples $\{\mathbf{u}^k\}_{k=1}^n$ as

$$\hat{p}_{\text{MC}} = \frac{1}{n} \sum_{k=1}^n \mathbb{I}[G(\mathbf{u}^k) \leq 0], \quad \mathbf{u}^k \stackrel{i.i.d.}{\sim} \varphi_d. \quad (2.84)$$

The crude MC estimator is *consistent*, i.e., it converges to p *in probability* as $n \rightarrow \infty$ (in fact, by virtue of the strong law of large numbers, it converges *almost surely*, which is stronger than convergence in probability) and unbiased, i.e., $\mathbb{E}[\hat{p}_{\text{MC}}] = p$. The estimator's variance and CoV are

$$\mathbb{V}[\hat{p}_{\text{MC}}] = \frac{p(1-p)}{n} \quad \text{and} \quad \delta_{\text{MC}} = \frac{\sqrt{\mathbb{V}[\hat{p}_{\text{MC}}]}}{\mathbb{E}[\hat{p}_{\text{MC}}]} = \sqrt{\frac{1-p}{np}}. \quad (2.85)$$

The number of samples required to compute \hat{p}_{MC} at a prescribed CoV δ_0 reads

$$n_0 = \frac{1-p}{\delta_0^2 p} \stackrel{p \ll 1}{\approx} \frac{1}{\delta_0^2 p}. \quad (2.86)$$

This is the reason crude MC is inefficient for rare event probability estimation as, by definition, $p \ll 1$ and thus n_0 becomes large. An intuitive explanation of this circumstance is offered in Fig. 2.5: samples drawn from f_X have probability p of being failure samples. Since p is small, a large number of samples from f_X is necessary to obtain non-zero contributions in Eq. (2.84).¹

Instead of fixing δ_0 , we may as well fix n_0 and ask how the estimator variance may be reduced while keeping this fixed computational budget. Such methods are referred to as *variance reduction methods* and span a large variety of approaches. A detailed discussion of several variance reduction methods is provided in [166, Chapters 7 - 9]. Arguably one of the most widespread variance reduction techniques is IS including numerous variations on the basic theme [166, Chapter 8]. The principal idea is to express Eq. (2.83) as

$$p = \int_{\mathbb{R}^d} \mathbb{I}(G(\mathbf{u}) \leq 0) \underbrace{\frac{\varphi_d(\mathbf{u})}{h(\mathbf{u})}}_{\omega(\mathbf{u})} h(\mathbf{u}) \, d\mathbf{u} = \mathbb{E}_h [\mathbb{I}(G(\mathbf{U}) \leq 0)\omega(\mathbf{U})]. \quad (2.87)$$

¹Of course, if there is no failure sample amongst a large number n of MC samples, this *does* contain some information on the rarity of the failure event. A consistent way to exploit this information is by way of constructing a Bayesian estimate of p with a Binomial likelihood (observing m successful trials out of n total trials with an underlying success probability p and success in this case being failure) [246]. A straight-forward approach is to select a conjugate prior (for a binomial likelihood this will be a Beta-distributed prior) and compute the probability of failure estimate as the posterior mean. If the uniform distribution on $[0, 1]$ is selected as prior (this is a special case of the Beta distribution on $[0, 1]$), the posterior mean reads $\hat{p}_{\text{MC, Bayes}} = (m+1)/(n+2)$, where m is the number of failure samples. If n is large, the estimator CoV can be approximated as $\delta_{\text{MC}} \approx 1/\sqrt{m+1}$. That is, if no failure samples are encountered although n is large, $p \approx (n+2)^{-1}$ with CoV 1. [23] investigates the conservativeness of credible intervals computed based on such posterior distributions and recommends the uniform prior as the one producing the most conservative upper credible bounds on p .

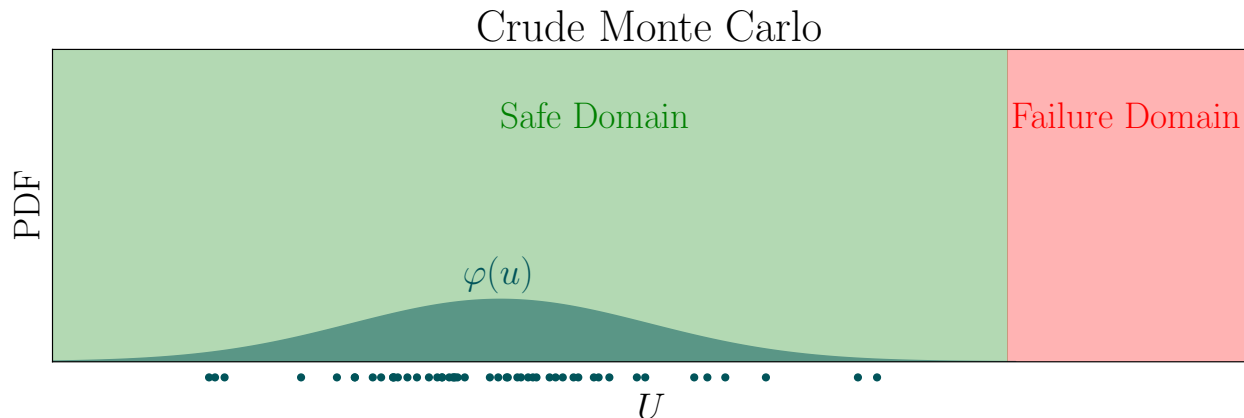


Figure 2.5: Monte Carlo sampling in one-dimensional standard-normal space from the nominal density $\varphi(u)$: the right tail of $\varphi(u)$ decays too fast to offer fair chances of generating failure samples from this density.

Eq. (2.87) is defined if the weight function $\omega(\mathbf{u})$ is defined, which is precisely the case if $\{\mathbf{u} \in \mathbf{F} : h(\mathbf{u}) > 0\}$. The IS estimator then reads

$$\hat{p}_{\text{IS}} = \frac{1}{n} \sum_{k=1}^n \mathbb{I}[G(\mathbf{u}^k) \leq 0] \omega(\mathbf{u}^k), \quad \mathbf{u}^k \stackrel{i.i.d.}{\sim} h. \quad (2.88)$$

It has variance

$$\sigma_{\text{IS}}^2 = \frac{1}{n} \int_{\mathbb{R}^d} (\mathbb{I}(G(\mathbf{u}) \leq 0) \omega(\mathbf{u}) - p)^2 h(\mathbf{u}) d\mathbf{u} = \frac{1}{n} \mathbb{E}_h [(\mathbb{I}(G(\mathbf{U}) \leq 0) \omega(\mathbf{U}) - p)^2] \quad (2.89)$$

and coefficient of variation $\delta_{\text{IS}} = \sigma_{\text{IS}} / \hat{p}_{\text{IS}}$. This CoV can be estimated based on samples from h with

$$\hat{\delta}_{\text{IS}} = \frac{1}{\hat{p}_{\text{IS}}} \sqrt{\frac{1}{n(n-1)} \sum_{k=1}^n (\mathbb{I}[G(\mathbf{u}^k) \leq 0] \omega(\mathbf{u}^k) - \hat{p}_{\text{IS}})^2}, \quad \mathbf{u}^k \stackrel{i.i.d.}{\sim} h. \quad (2.90)$$

\hat{p}_{IS} is a consistent and unbiased estimator of p .

In the context of IS, the original sampling density $\varphi_d(\mathbf{u})$ is referred to as *nominal density* and h is the *biasing density* or simply IS density. The performance of the IS estimator hinges on the selection of h . A standard choice for h when estimating failure probabilities is a Gaussian density centered around the *design point* as determined with FORM [201]. This is the point on the hypersurface $G(\mathbf{U}) = 0$ that is closest to the origin in standard-normal space and thus is the *most probable point of failure* (MPPF). MPPF-based IS approaches usually suffer from poor performance in high-dimensional problems as identifying the MPPF becomes an expensive task. They are moreover not well suited to tackle series system reliability problems or more generally problems with several relevant failure domains that are not captured by a single MPPF. There exists an optimal IS density $h^* = \mathbb{I}[G(\mathbf{u}) \leq 0] \varphi_d(\mathbf{u})$ (see Fig. 2.6) in the sense that one obtains $\sigma_{\text{IS}}^2 = 0$ in Eq. (2.89). However, since p is the normalizing constant of this *optimal IS density* h^* , it is of little relevance in practice. Yet, h^* provides information on the shape of potentially powerful IS densities and can be used to guide the search for a useful h .

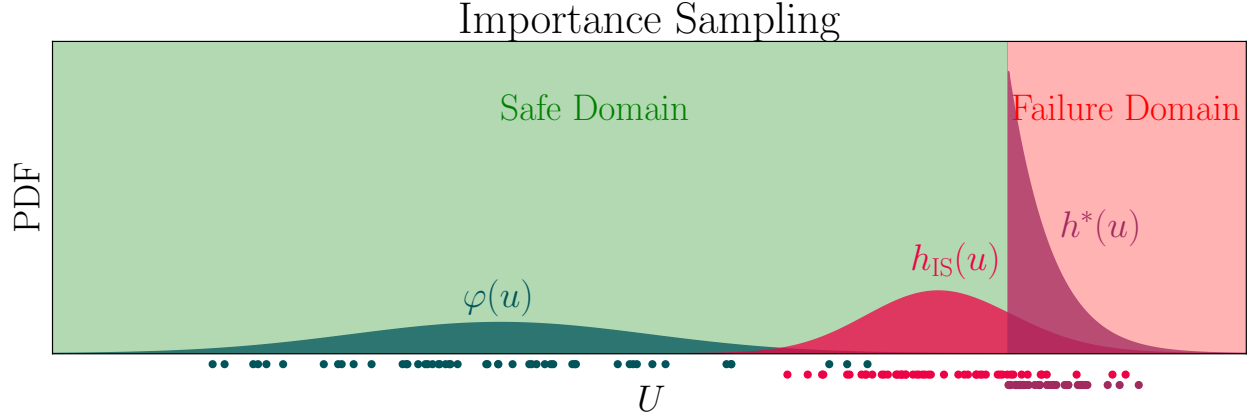


Figure 2.6: Importance sampling visualized in one-dimensional standard-normal space: a Gaussian IS density $h_{\text{IS}}(u)$ along with the optimal IS density $h^*(u)$ and associated samples drawn from each density in matching colors: the optimal IS density is merely the cut-off right tail of $\varphi(u)$ renormalized with p .

Modern rare event simulation methods capitalize on this theoretical result for optimal variance reduction by approaching h^* with a sequence of IS densities. Depending on the method, IS densities are represented with ensembles of samples [176], non-parametric KDE estimates [153, 14] and parametric PDF models [118, 196]. [234] uses a von Mises-Fisher mixture which effectively exploits the *concentration of measure*-phenomenon associated with Gaussian densities in high dimensions. Such measure concentrations are a *blessing of dimensionality* [105], i.e., a collection of geometric properties of high-dimensional spaces facilitating efficient data analysis in high-dimensional problems. In [175], the von Mises-Fisher mixture is augmented with a Nakagami distribution modelling an additional free distribution parameter (namely, the radius of the *important ring* [107]).

Inspired by h^* , two possible choices for the sequence of IS densities are

$$h_i(\mathbf{u}) = \frac{1}{P_i} \mathbb{I}[G(\mathbf{u}) \leq t_i] \varphi_d(\mathbf{u}) \quad (2.91a) \quad \text{and} \quad h_i(\mathbf{u}) = \frac{1}{P_i} \Phi\left(-\frac{G(\mathbf{u})}{t_i}\right) \varphi_d(\mathbf{u}), \quad (2.91b)$$

where $P_i = \int_{\mathbb{R}^d} \mathbb{I}[G(\mathbf{u}) \leq t_i] \varphi_d(\mathbf{u}) d\mathbf{u}$ in Eq. (2.91a) and $P_i = \int_{\mathbb{R}^d} d\mathbf{u}$ in Eq. (2.91b). In both cases, the sequence starts at $i = 0$ with $t_0 = \infty$ and $P_0 = 1$, which recovers the nominal density $\varphi_d(\mathbf{u})$. Then, by choosing $t_{i+1} < t_i$ in each step, both sequences converge to h^* as $\lim_{t_i \rightarrow 0} h_i = h^*$.

When using Eq. (2.91a), the sequence $\{t_i\}_{i=1}^m$ can be understood as defining a nested set of failure events $\{F_i = \{\mathbf{u} : G(\mathbf{u}) \leq t_i\}\}_{i=1}^m$. In each step, t_{i+1} is selected as the ρ -quantile of $h_i(\mathbf{u})$, which is approximated as the ρn -th order statistic of $\{G(\mathbf{u}_i^k)\}_{k=1}^n$, where $\mathbf{u}_i^k \sim h_i$.

In cross-entropy-based IS [194, 196], a parametric density model is used to approximate the h_i of Eq. (2.91a) in each step based on the *elite samples*, i.e., those samples that are contained in F_i . This is done by solving a stochastic optimization problem, namely a cross-entropy minimization between the parametric model and h_i based on the $n\rho$ elite samples. Typical choices are $\rho = 10^{-2} \dots 10^{-1}$. After m steps, h_m will be sufficiently close to h^* and is used to construct an IS estimate of p ac-

ording to Eq. (2.88). Details of this method are described in publication 4 (Chapter 7).

In SUS [7] (see Fig. 2.7), the IS density sequence Eq. (2.91a) is given in terms of samples only. Rather than fitting a new parametric model on the $n\rho$ elite samples of h_i in each step and sampling n new samples from the fitted distribution, $n(1-\rho)$ additional new samples are drawn from h_i with an MCMC sampler using the elite samples as seeds. By design, $\mathbb{P}(F_{i+1}|F_i) \approx \rho$ and thus the ratio of two subsequent normalizing constants is $P_{i+1}/P_i \approx \rho$. To verify this, we note that $\mathbb{P}(F_i) = P_i$ and write

$$\rho \cdot P_i \approx \mathbb{P}(F_{i+1}|F_i)\mathbb{P}(F_i) = \mathbb{P}(F_i|F_{i+1})\mathbb{P}(F_{i+1}) = \mathbb{P}(F_{i+1}) = P_{i+1} \quad \Leftrightarrow \quad \rho \approx \frac{P_{i+1}}{P_i}. \quad (2.92)$$

Typical choices are $\rho = 0.1 \dots 0.3$ [246]. If at least ρn samples lie in F (the original failure domain associated with $t_i = 0$), the algorithm terminates by setting $m = i$, $t_m = 0$. The SUS estimate is then an IS estimate using h_{m-1} as IS density such that

$$\hat{p}_{\text{SUS}} = \frac{1}{n} \sum_{k=1}^n \mathbb{I}[G(\mathbf{u}^k) \leq 0] \frac{P_{m-1}}{\mathbb{I}[G(\mathbf{u}^k) \leq t_{m-1}]} = \frac{\rho^{m-1}}{n} \sum_{k=1}^n \mathbb{I}[G(\mathbf{u}^k) \leq 0], \quad \mathbf{u}^k \stackrel{i.i.d.}{\sim} h_{m-1}. \quad (2.93)$$

This is an interpretation of SUS as a special case of SIS [176]. The original motivation of Eq. (2.93) invokes the chain rule of probability to write $\hat{p}_{\text{SUS}} = \mathbb{P}(\cap_{i=1}^m F_i) = \mathbb{P}(F_1) \prod_{i=1}^{m-1} \mathbb{P}(F_{i+1}|F_i) = \rho^{m-1}/n \sum_{k=1}^n \mathbb{I}[G(\mathbf{u}^k) \leq 0]$ yielding an identical estimator. SUS produces biased estimates of p due to the adaptive estimation of the $\{t_i\}_{i=1}^m$ [30] as well as correlation present in the MCMC samples [7]. Efficient MCMC algorithms for SUS are discussed in [179] and several other enhancements are proposed in [246]. Notably, within both CE-based IS and SUS, a fraction of 70 – 99% of samples

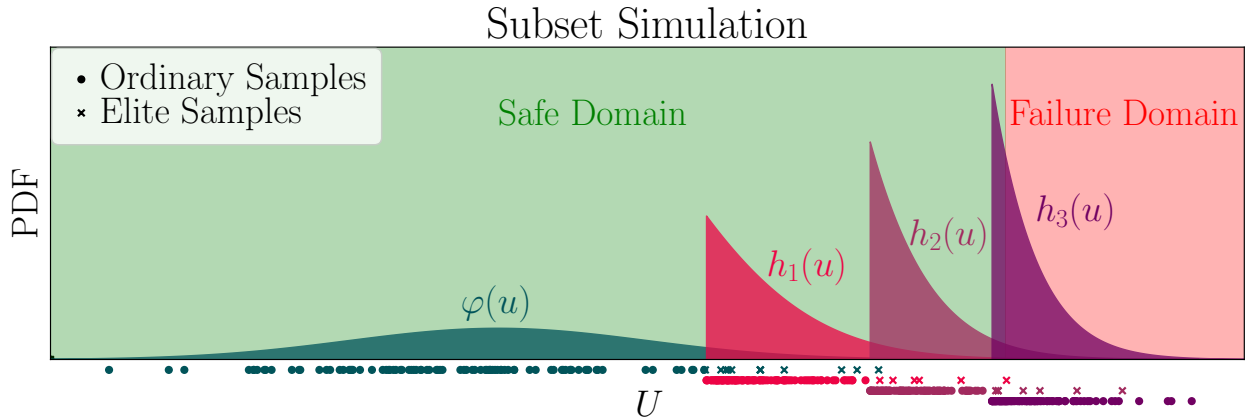


Figure 2.7: Subset simulation visualized in one-dimensional standard-normal space with associated sampling densities and samples (circles) drawn from each density in matching colors: at each step, the $\rho = 10\%$ of samples with the smallest g -values are selected as elite samples (crosses). The associated quantile determines the threshold for the next conditional sampling density from which more samples are generated by simulating Markov chains with the current elite samples as seeds.

lie waste at each level. This shortcoming may be rectified by replacing Eq. (2.91a) with Eq. (2.91b). Then, instead of classifying samples into elite and non-elite samples binarily, each sample is reused in computing the subsequent IS density with an associated weight that expresses ‘eliteness’ on a continuous scale between 0 and 1. [176] introduces this weighting for MCMC-based SIS and [175] does the same for CE-based IS. Both the SIS and the CE-IS version are described in detail in publications 3 (Chapter 6) and 4 (Chapter 7), respectively and are illustrated in Fig. 2.8.

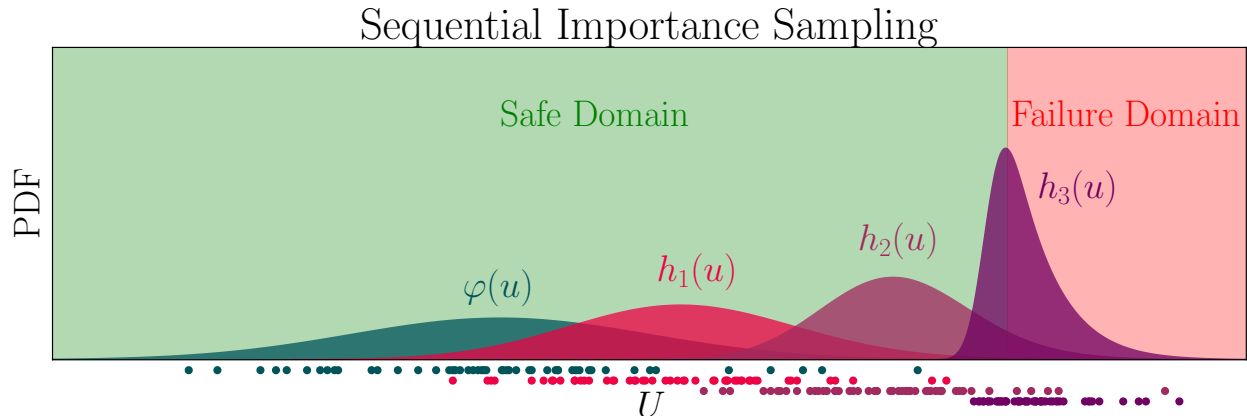


Figure 2.8: Sequential importance sampling visualized in one-dimensional standard-normal space with associated sampling densities and samples drawn from each density in matching colors: step by step the IS densities $\{h_i\}_{i=1}^3$ approach the optimal IS density.

2.3.4 Surrogate-assisted sampling methods

A different approach to solving Eq. (2.82) efficiently replaces the expensive computational model \mathcal{Y} with one or several surrogate models to reduce the cost per sample rather than the overall number of samples required. [67] used a polynomial response surface method in reliability analysis as early as 1989. Improved response surfaces are proposed, e.g., in [81]. A variety of surrogate modelling techniques have been put forward in the meantime. In the context of reliability analysis, approaches such as artificial neural networks (ANN) [171, 95, 202, 170], support vector machines [94, 32, 31], Gaussian process regression [57, 56, 93, 16] and projection to polynomial bases including polynomial chaos expansions (PCE) and low-rank tensor approximations (LRA) [133, 128, 127, 218, 116] have been investigated.

Early surrogate-based approaches to rare event probability estimation rely on the construction of a globally accurate surrogate model based on which the probability of failure is estimated with a large number of (computationally cheap) samples. A drawback of such approaches is that by definition and with high probability, information on failure events is scarce in a DoE that is drawn from the joint input PDF $f_{\mathbf{X}}$ as failure domains are typically located in the tails of $f_{\mathbf{X}}$. To some extent this shortcoming may be rectified by choosing the DoE adaptively to represent the domain of interest better than a random sample from $f_{\mathbf{X}}$. In statistics, this is known as *optimal experimental design* [68] and a recent increase in interest in such concepts in the context of machine learning is linked to the term of *active learning* [205]. Amongst a variety of existing active learning approaches, the version encountered most often in the context of adaptive surrogate construction is the *pool-based* approach (Fig. 2.9): an initial DoE is used to construct the surrogate model. Based on this surrogate model, an *acquisition function* or *learning function* is evaluated at a large set of points randomly sampled from $f_{\mathbf{X}}$. The output of \mathcal{Y} at these points is unknown, thus they are referred to as *unlabeled* data points. These points constitute the *pool* in *pool-based active learning*. In each iteration, those members of the pool that maximise the learning function, are labeled: the original model \mathcal{Y} is evaluated at these points and the new labeled data tuples are added to the DoE. The surrogate model is then reconstructed with the new DoE and the process starts anew until a convergence condition is met. The learning function usually encodes two preferences: 1. Minimizing some form of prediction

uncertainty or error measure of the surrogate model. 2. Balancing exploitation of regions in which the surrogate model is highly uncertain/error-prone with exploring other potentially influential regions of the model input space. In rare event probability estimation, this pertains to both balancing

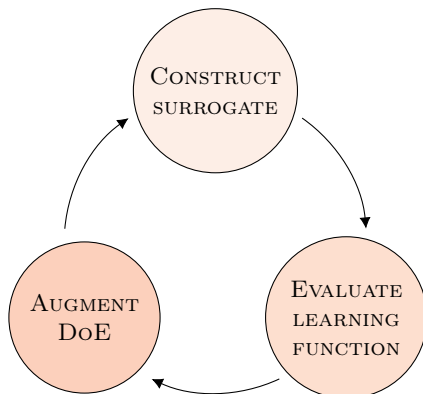


Figure 2.9: Three steps of active learning: 1. Construct the surrogate model $\hat{\mathcal{Y}}$ based on the current DoE. 2. Evaluate the learning function on all unlabeled points in the pool. 3. Evaluate \mathcal{Y} at one/several points maximising the learning function and add the new points to the DoE. Repeat.

proximity of a given point to the failure hypersurface with prediction uncertainty in that point (see Fig. 2.10) as well as balancing the contributions from different disjunct failure subdomains to the overall probability of failure. The latter is often achieved with clustering techniques. For example, [199] identify a prescribed number of clusters in the pool of unlabeled data points and select one point maximizing the learning function in each cluster. In this way several disjunct failure domains are more likely to be represented evenly in the DoE (for example in Fig. 2.10, both the prominent egg-shaped failure domain in the back and the quarter circle in the left corner should be identified with two clusters). The learning function must be cheap to evaluate (as in each iteration it is evaluated at each point in the pool of unlabeled data). The majority of adaptive surrogate-assisted approaches in the context of rare event probability estimation are based on Gaussian process regression [25, 57, 17, 56, 14, 199]. This is mainly due to the fact that Gaussian process models provide a pointwise uncertainty measure (the conditional GP variance) that proves handy in constructing learning functions. In [143], PCEs are paired with active learning for rare event probability estimation. The pointwise uncertainty measure in this approach is obtained with a bootstrap resampling step [58]. Pool-based adaptive surrogate-assisted approaches, however, face similar limitations as crude MC: The smaller the probability of failure, the larger the random sample from $f_{\mathbf{X}}$ has to be in order to obtain a certain number of points in or close to the failure domain. The situation is less sensible for these methods since most of these points will not have to be evaluated with \mathcal{Y} , yet the memory requirements for storing the pool quickly become prohibitive (e.g., if $p = 10^{-6}$ such that at least, say, 10^8 pool members are desirable and the problem input dimension is $d = 100$, the required storage for the resulting 10^{10} double entries equals 80 gigabytes).

Much in the same way sequential sampling techniques solved this issue for crude MC, coupling these sequential sampling techniques with surrogate-based approaches solves the storage problem for pool-based active learning. This usually leads to schemes, in which the surrogate model is reconstructed for maximum local accuracy in each step of the sequential sampling procedure. [170] couples SUS with neural networks — however, without using adaptive DoEs. In [32, 31], SUS is combined with support vector regression and an adaptive approach to the local DoEs. [16] combines SUS with Bayesian GPRs and [243] use a sequential IS approach based on acceptance-rejection-

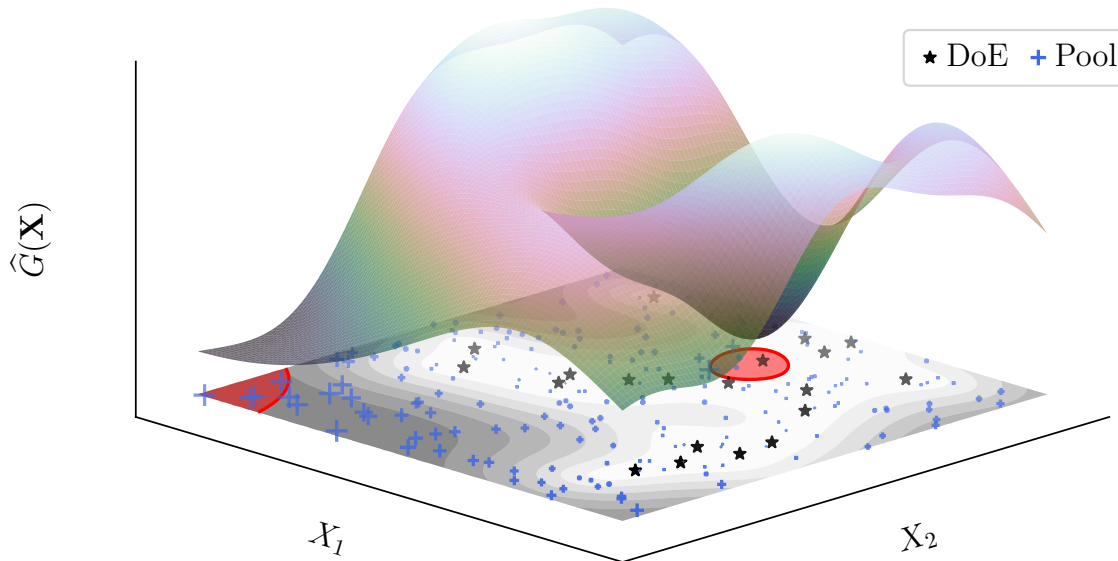


Figure 2.10: Surrogate-based active learning illustrated: the LSF surrogate mean prediction \hat{G} (surface plot) as estimated based on a DoE (black stars) of 20 points is shown along with a measure of prediction uncertainty of \hat{G} , in this case the surrogate standard deviation $\sigma_{\hat{G}}$ (contour plot, dark indicates large $\sigma_{\hat{G}}$ - light indicates low $\sigma_{\hat{G}}$). The selection of a new DoE point from the current pool of unlabeled points (blue crosses) is based on each test points' proximity to the failure domain as predicted by the GPR model (i.e., where $\hat{G} \leq 0$, red surface) on the one hand and the uncertainty of \hat{G} at each of these points on the other hand. The size of the crosses corresponds to the inverse U -function [57] such that the largest cross exposes the combination of model uncertainty and proximity to the failure domain deemed most critical. Therefore, the largest blue cross corresponds to the pool element added to the DoE in this iteration of the active learning algorithm run with the U -function.

sampling along with GPRs. Sequential sampling techniques for rare event probability estimation based on locally reconstructed and/or actively learned surrogate models represent a contemporary, active field of research and giving a full account of recent approaches along these lines is beyond the scope of this work. We notice that many approaches that have been published within the past five years and that go unmentioned here are making promising contributions to this field. Fig. 2.11 depicts a general schematic of this class of methods.

A largely unaddressed issue with such approaches arises with the reintroduction of a problem dimensionality bottleneck through the use of surrogate models. While several sequential sampling techniques such as SUS and SIS with carefully chosen MCMC proposal distributions [176] or parametric IS density models [175] overcome this dependence on the problem dimension, the same cannot be said of surrogate models in many instances. [173] proposes *sequential subspace importance sampling* (SSIS), in which the intermediate IS densities of a sequential IS approach are reconstructed using PLS-PCE models. PLS-PCE models are introduced briefly in Subsection 2.2.4 and in full detail

in publication 1 (Chapter 4) and combine surrogate modelling with dimensionality reduction, which makes them suitable for high-dimensional problems. In publication 3 (Chapter 6), we detail *active sequential subspace importance sampling* (ASSIS), which complements the sequential reconstruction of IS densities in low-dimensional subspaces with an active learning algorithm that optimizes the local DoEs with respect to both the subspace and the surrogate modelling error. Besides PLS, active subspaces [41], sliced inverse regression [129] and autoencoders are some alternative techniques for identifying linear subspaces that have been explored recently to construct surrogate models for rare event simulation [169, 104, 132].

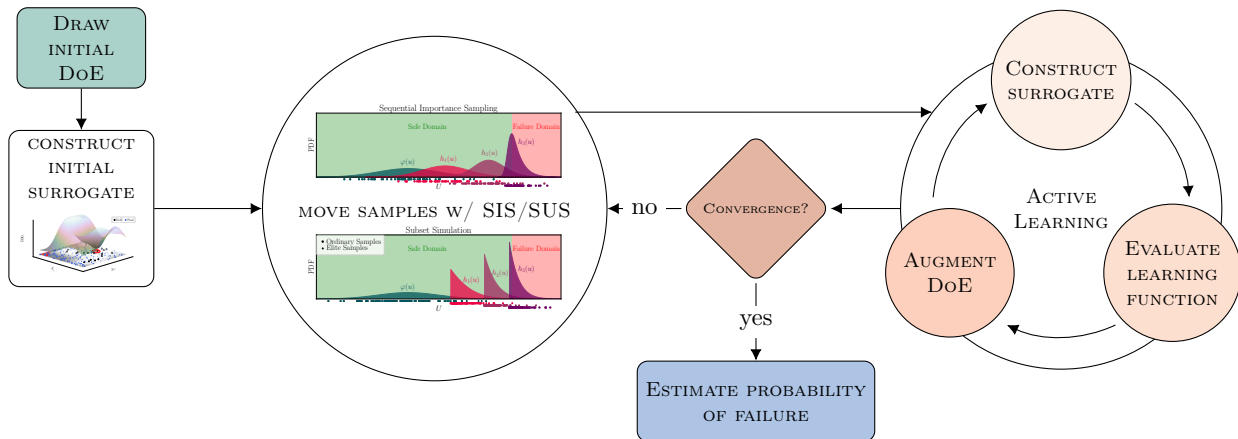


Figure 2.11: A general schematic of sequential sampling schemes for rare event estimation based on actively learned surrogates: after having learned an initial surrogate based on a DoE sampled from $f_{\mathbf{X}}$, an alternating succession of sample propagation and surrogate reconstruction steps follow. Convergence criteria are based on how well the current sample set covers the target failure domain according to the current surrogate model.

2.4 Sensitivity Analysis

2.4.1 Sensitivity analysis of model output

Given a model, sensitivity analysis quantifies how model inputs affect either model output or quantities of interest derived thereof. The classical setting is sensitivity analysis of model output (SAMO) [197], which is discussed in this section followed by a discussion of reliability-oriented sensitivity analysis (ROSA, [35]). In literature, a variety of metrics for quantifying sensitivities with respect to different quantities of interest can be found. Different metrics may be suitable for different purposes depending on what kind of decision shall be made: [198] distinguishes *factor prioritization*, *fixing* and *mapping* as well as *variance cutting* as potential goals of a sensitivity analysis. [29] complements these with the goals of investigating *model structure*, *sign of change* and *stability*. We are interested in producing a ranking of variables based on their importance, which corresponds to factor prioritization and – to a lesser degree – factor fixing. The various metrics proposed in sensitivity analysis literature can be roughly grouped according to ...

- **...scope:** Local sensitivity measures are based on derivatives of the model \mathcal{Y} with respect to the inputs using numerical [98] and automatic differentiation [96] or adjoint solvers [4] and provide local sensitivity measures to small perturbations in the input. [154] proposed averaging these local measures over the input space to obtain a *derivative-based global sensitivity measure* (DGSM).
- **...measure type:** [154] marked the prelude to the invention of various global sensitivity measures. These include but are not limited to other global derivative-based methods [212, 120], variance-based [213, 89, 99], moment-free [28] and more generally f -divergence-based sensitivity measures [229, 185] as well as Shapely effects [167].
- **...variable type:** Some methods hinge on the type of variables with respect to which sensitivities are assessed. These variables can be either uncertain model inputs, i.e., coordinates of \mathbf{X} , or deterministic parameters (e.g., parameters of \mathcal{Y} , G or $f_{\mathbf{X}}$). This differentiation is not so much a factor in SAMO but more commonly encountered in ROSA, in the context of which IS methods can be used to compute sensitivities specifically with respect to parameters of $f_{\mathbf{X}}$ [152] or G [172].
- **...variable dependency:** If sensitivities are computed with respect to random quantities, a distinction between dependent and independent inputs is of paramount importance. While classical variance-based sensitivity analysis requires independent variables by construction (see Subsection 2.4.2), a variety of different extensions to dependent inputs are proposed in [91, 126, 37, 119, 142]. These different approaches may result in different importance rankings for the sample model [235], which suggests a certain ambiguity as to how the results of such an analysis are to be interpreted. The moment-free approach of [28] appears to be more naturally suited to the scenario of dependent variables.
- **...analysis goal:** Finally, a sensitivity is not only computed *with respect to* a set of variables but also *of* a quantity of interest. The most common analysis goal is the output of a computational model. But a variety of other quantities of interest may be considered as well with the probability of failure being a prominent example discussed in the next subsection.

A more comprehensive introduction to sensitivity analysis is given in [197] and [97] is a recent overview of global sensitivity analysis methods for model output.

2.4.2 Variance-based SAMO with surrogate models

Variance-based sensitivity indices are motivated by decomposing the model output variance $\mathbb{V}[\mathcal{Y}(\mathbf{X})]$ into fractions that can be attributed to the coordinates of the model inputs \mathbf{X} . To this end, any square-integrable function $\mathcal{Y} \in L^2_{f_{\mathbf{X}}}(\mathbb{R}^d, \mathbb{R})$ with independent inputs \mathbf{X} (that is, $f_{\mathbf{X}}$ is a product of univariate PDFs¹) can be written as a *Sobol'-Hoeffding-decomposition* [87, 213] or *high-dimensional model representation* (HDMR):

$$\mathcal{Y}(\mathbf{X}) = \mathcal{Y}_0 + \sum_{i=1}^d \mathcal{Y}_i(X_i) + \sum_{i=1}^d \sum_{j=i+1}^d \mathcal{Y}_{ij}(X_i, X_j) + \cdots + \mathcal{Y}_{12\dots d}(\mathbf{X}). \quad (2.94)$$

¹[213] originally assumes marginally uniformly distributed inputs on the d -dimensional unit cube. It is, however, simple to extend the assumptions in [213] to arbitrary product measures of \mathbf{X} [182].

Therein $\mathcal{Y}_0 = \mathbb{E}_{f_{\mathbf{X}}}[\mathcal{Y}(\mathbf{X})]$ is a constant. By defining a multi-index $\mathbf{v} \in \mathcal{P}(\{1, \dots, d\} \setminus \emptyset)$ (this is the power set of $\{1, \dots, d\}$ excluding the empty set \emptyset , which has a total of $2^d - 1$ elements) and $\mathbf{X}_{\mathbf{v}} = \{X_j\}_{j \in \mathbf{v}}$, any remaining summand on the right-hand side of Eq. (2.94) may be written as $\mathcal{Y}_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}})$. If the $\{\mathcal{Y}_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}})\}_{\mathbf{v} \in \mathcal{P}(\{1, \dots, d\}) \setminus \emptyset}$ are pairwise orthogonal, i.e., if $\mathbf{v} \neq \mathbf{w} \Leftrightarrow \mathbb{E}_{f_{\mathbf{X}}}[\mathcal{Y}_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}})\mathcal{Y}_{\mathbf{w}}(\mathbf{X}_{\mathbf{w}})] = 0$, the HDMR in Eq. (2.94) exists uniquely and [213]

$$\mathcal{Y}_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}}) = \mathbb{E}_{f_{\mathbf{X}_{\sim \mathbf{v}}}}[\mathcal{Y}(\mathbf{X})|\mathbf{X}_{\mathbf{v}}] - \sum_{\mathbf{w} \subset \mathbf{v}} \mathbb{E}_{f_{\mathbf{X}_{\sim \mathbf{w}}}}[\mathcal{Y}(\mathbf{X})|\mathbf{X}_{\mathbf{w}}].^2 \quad (2.95)$$

$\mathbf{X}_{\sim \mathbf{v}}$ denotes the complement of $\mathbf{X}_{\mathbf{v}}$ on \mathbf{X} , i.e., all coordinates of \mathbf{X} not contained in $\mathbf{X}_{\mathbf{v}}$. Due to the pairwise orthogonality of the $\{\mathcal{Y}_{\mathbf{v}}\}_{\mathbf{v} \in \mathcal{P}(\{1, \dots, d\})}$, the variance fraction associated with \mathbf{v} is

$$D_{\mathbf{v}} = \mathbb{V}_{f_{\mathbf{X}_{\mathbf{v}}}}[\mathcal{Y}_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}})] = \mathbb{V}_{f_{\mathbf{X}_{\mathbf{v}}}}[\mathbb{E}_{f_{\mathbf{X}_{\sim \mathbf{v}}}}[\mathcal{Y}(\mathbf{X})|\mathbf{X}_{\mathbf{v}}]] - \sum_{\mathbf{w} \subset \mathbf{v}} \mathbb{V}_{f_{\mathbf{X}_{\mathbf{v}}}}[\mathbb{E}_{f_{\mathbf{X}_{\sim \mathbf{w}}}}[\mathcal{Y}(\mathbf{X})|\mathbf{X}_{\mathbf{w}}]]. \quad (2.96)$$

Based on Eq. (2.96), the variance $\mathbb{V}[\mathcal{Y}(\mathbf{X})]$ can be decomposed into $2^d - 1$ variance fractions contributed by the $2^d - 1$ possible combinations of the d inputs. Four types of variance-based sensitivity indices are commonly defined as a consequence of this based on how terms associated with \mathbf{v} are collected (see also Fig. 2.12):

$$\begin{aligned} \text{Sobol' index [213]:} & \quad S_{\mathcal{Y}, \mathbf{v}} = \frac{D_{\mathbf{v}}}{\mathbb{V}[\mathcal{Y}(\mathbf{X})]} \\ \text{Closed Sobol' index [182]:} & \quad S_{\mathcal{Y}, \mathbf{v}}^{\text{clo}} = \frac{\sum_{\mathbf{w} \subseteq \mathbf{v}} D_{\mathbf{w}}}{\mathbb{V}[\mathcal{Y}(\mathbf{X})]} \\ \text{Total Sobol' index [89]:} & \quad S_{\mathcal{Y}, \mathbf{v}}^T = \frac{\sum_{\mathbf{w} \cap \mathbf{v} \neq \emptyset} D_{\mathbf{w}}}{\mathbb{V}[\mathcal{Y}(\mathbf{X})]} \\ \text{Superset index [90]:} & \quad S_{\mathcal{Y}, \mathbf{v}}^S = \frac{\sum_{\mathbf{v} \subseteq \mathbf{w}} D_{\mathbf{w}}}{\mathbb{V}[\mathcal{Y}(\mathbf{X})]} \end{aligned} \quad (2.97)$$

That is, the Sobol' index measures the variance contribution associated precisely with the interaction of the coordinates of $\mathbf{X}_{\mathbf{v}}$. The closed Sobol' index (also *lower index* [168]) is a superset of the Sobol' index and measures all variance contributions by one or several coordinates of $\mathbf{X}_{\mathbf{v}}$. The total Sobol' index (sometimes also *total-effect index* or *upper index* [168]) measures the variance contributions stemming from all interactions of any subset of $\mathbf{X}_{\mathbf{v}}$ with any subset of $\mathbf{X}_{\sim \mathbf{v}}$. Finally, the total Sobol' indices defined in Eq. (2.97) are closed total Sobol' indices, which leaves a fourth type of index to be defined, namely the non-closed total Sobol' index or *superset index* [90]. This index accounts for variance contributions of interactions of the entire input subset $\mathbf{X}_{\mathbf{v}}$ with any subset of $\mathbf{X}_{\sim \mathbf{v}}$. The superset index is thus more restrictive compared to the closed total Sobol' index in the same manner the Sobol' index is more restrictive than the closed Sobol' index. The superset index, however, is stated merely for completeness and bears no practical relevance in this work. For univariate indices, i.e., if $\|\mathbf{v}\|_0 = 1$, the Sobol' index and the closed Sobol' index are identical as are the total Sobol' index and the superset index. An illustration of how Sobol' indices are computed for a two variable model is given in Fig. 2.13. Variance-based sensitivity indices may also be computed based on previously constructed surrogate models such as PCEs [217], canonical LRA [115], spectral tensor-train LRA [26] and GPR models [144, 233]. This computation is particularly simple for PCEs: comparing the PCE format in Eq. (2.79) with the HDMR in Eq. (2.94) reveals that if \mathcal{Y} is a PCE, each $\mathcal{Y}_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}})$

²Due to the independence of the coordinates of \mathbf{X} , $f_{\mathbf{X}_{\sim \mathbf{w}}|\mathbf{X}_{\mathbf{w}}}(\mathbf{x}_{\sim \mathbf{w}}|\mathbf{x}_{\mathbf{w}}) = f_{\mathbf{X}_{\sim \mathbf{w}}}(\mathbf{x}_{\sim \mathbf{w}})$.

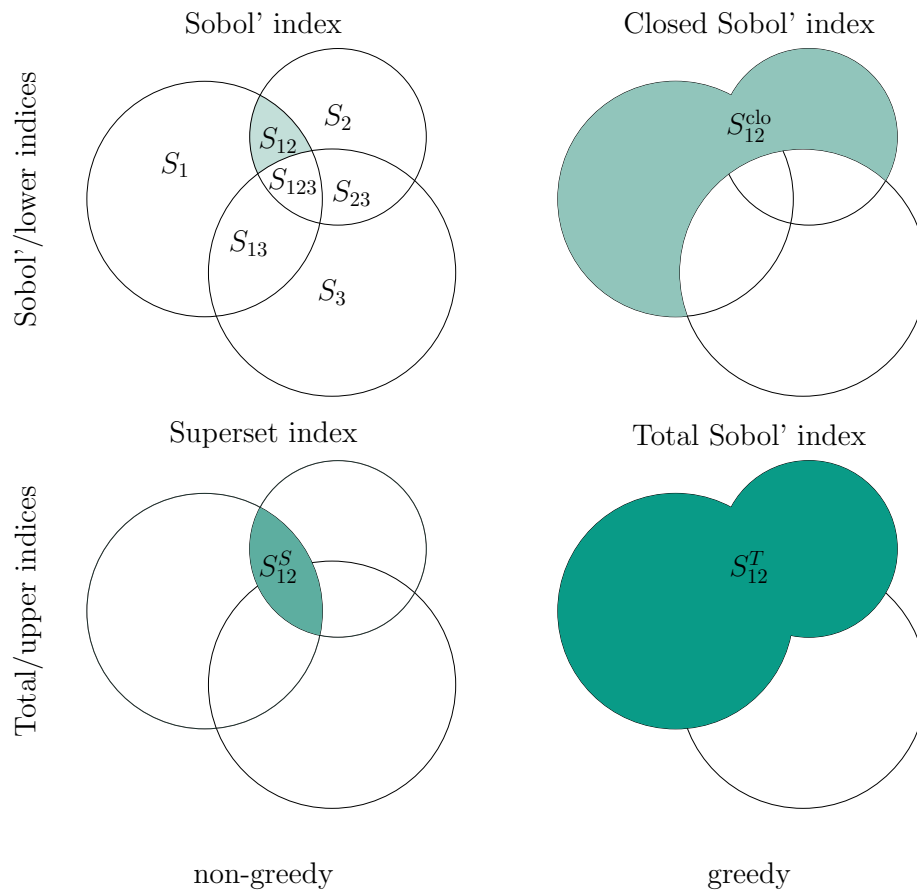


Figure 2.12: Venn diagrams of variance fractions in a three-variable model: the different variance fractions as gathered for the four types of variance-based sensitivity indices of variables 1 and 2 are filled in green.

is simply the sum of basis functions depending on \mathbf{X}_v . By Eq. (2.80), the associated variance D_v is thus equal to the sum of the squared coefficients associated with these basis functions, which yields the Sobol' index of the PCE [217]. In a similar way the other indices can be computed by collecting basis functions and summing the corresponding coefficients squared. A detailed discussion is found in publications 2 & 5 (Chapters 5 and 8).

Publication 2 (Chapter 5) establishes how to compute variance-based sensitivity measures of PLS-PCE models. To this end, we derive a backtransformation that allows to represent PLS-PCEs in the standard PCE format. In this way, upon having transformed the PLS-PCE model, sensitivities can be obtained with the well-established procedures described in the previous paragraph.

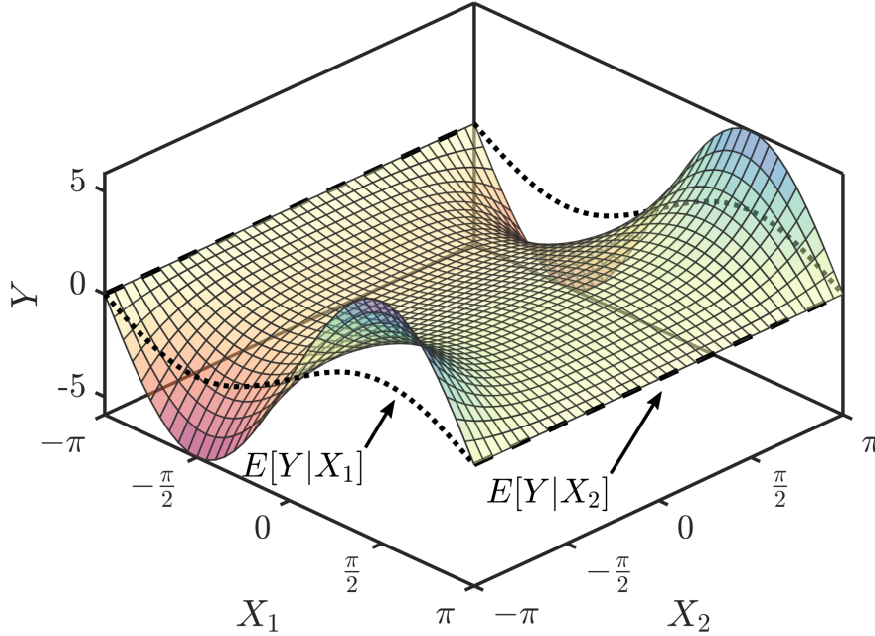


Figure 2.13: Illustration of the meaning of Sobol' indices in a two variable model $Y = \mathcal{Y}(X_1, X_2)$ (surface plot) with $X_1, X_2 \sim \mathcal{U}(-\pi, \pi)$: the conditional expectations (black dotted for X_1 , black dashed for X_2) with respect to either input variable are the result of a 'slicewise' probability integral along the X_2/X_1 -direction at each point along the X_1/X_2 -axis. The Sobol' indices are then proportional to the variances of the black solid lines that is seen to yield 0 for X_2 immediately.

2.4.3 Reliability-oriented sensitivity analysis

For decision-making purposes, the absolute value of the probability of failure of a system is often less important than the sensitivity of this value with respect to input parameters and variables. These can be parameters of the input joint PDF $f_{\mathbf{X}}$, the computational model \mathcal{Y} or the LSF G . Many approaches dedicated to reliability sensitivities are concerned with determining local derivative-based sensitivity measures of the probability of failure with respect to deterministic model parameters. [195] introduced *score functions* to estimate these derivatives with an IS approach. A similar approach is taken in [240] based on adaptive IS. Several other sampling methods have been used in the following to estimate these derivative-based indices, e.g., line sampling [8, 137], SUS [215] and SIS [172]. [232] propose a DGSM of the probability of failure to obtain a global reliability sensitivity measure. In the context of approximative SRM such as FORM and SORM, global reliability sensitivity measures may also be obtained as a byproduct of the MPPF calculation [139, 48]. These so-called α -factors can be interpreted as variance-based sensitivity indices of the FORM- and SORM-based LSF approximations [178]. [111] extend the notion of α -factors by accounting for several design points and regions along the failure hypersurface by means of a Gaussian mixture (GM) model and the definition of a participation factor for the different subdomain contributions of the GM components.

[42] introduced a global sensitivity measure for the probability of failure based on the notion of moment-independent sensitivity put forward in [28] for SAMO. That is, they compute probability-weighted distances (one absolute distance and one quadratic distance) between the unconditional

and conditional probability of failure. Similarly, [125] defines *perturbed-law* indices for ROSA with respect to parameters of $f_{\mathbf{X}}$. These indices measure the relative signed change in probability of failure upon introducing small perturbations in the distribution parameters. [131] computes global, variance-based sensitivities for the indicator function of the LSF and show that these are equivalent to the squared distance measure of [42]. They use a surrogate model to relax the computational cost induced by the sampling-based sensitivity computation approach. Variance-based sensitivity measures for the probability of failure conditioned upon input distribution parameters are investigated in [152, 233]. Efficient computation of variance-based indices of the conditional failure probability as well as perturbed-law indices is available through IS by running what [166, Section 9.14] calls *what-if simulations* based on a set of failure samples obtained with the unconditional/unperturbed input density $f_{\mathbf{X}}$. Recently, [216] proposed to directly measure the sensitivity of a reliability-based engineering decision (assessment and design decisions are considered) with respect to the input variables and parameters based on *value of information*. While often computationally more expensive than other reliability sensitivity methods, this approach yields easily interpretable results and handles variable dependency seamlessly (as do the moment-free approaches of [42, 125]).

In publication 5 (Chapter 8), we introduce a reliability analysis framework that foos on uncertainty separation (detailed in Section 2.5). There, we define sensitivity indices using a similar notion of conditional probability of failure as [42, 152]. We devise a bi-level surrogate modelling strategy to compute the proposed indices of the magnitude of the probability of failure with respect to inputs that may be both uncertain model inputs \mathbf{X} or model/LSF/input distribution parameters.

2.5 Uncertainty Separation

2.5.1 Introduction to Bayesian Analysis

Bayesian analysis is a framework for statistical inference, i.e., for computing probability distributions based on observations. The distinguishing feature of Bayesian methods compared to other statistical inference methods is the utilization of prior belief along with observations. The variables we want to infer are referred to as the unobserved random variables $\mathbf{X} \in \mathbb{R}^d$, whereas the variables on which observations are made, are the observed random variables $\mathbf{Y} \in \mathbb{R}^m$. As usual, the corresponding sample space variables are written as the lower-case letters \mathbf{x} and \mathbf{y} , respectively. The first step in Bayesian inference is the specification of a joint statistical model for \mathbf{x} and \mathbf{y} , $f_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})$ such that conclusions about \mathbf{X} can be drawn based on the observations [71, Section 1.1]. We consider n \mathbf{Y} -observations of equality type and summarize these as data $\mathcal{D} = \{\mathbf{Y} = \mathbf{y}_1, \mathbf{Y} = \mathbf{y}_2, \dots, \mathbf{Y} = \mathbf{y}_n\}$. Drawing conclusions about \mathbf{X} based on \mathcal{D} (i.e., inference) then takes the form of a conditional PDF that follows from *Bayes' law* as

$$f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathcal{D}) = \frac{\overbrace{f_{\mathbf{Y}|\mathbf{X}}(\mathcal{D}|\mathbf{x})}^{\text{Likelihood } L(\mathbf{x})} \overbrace{f_{\mathbf{X}}(\mathbf{x})}^{\text{Prior}}}{\underbrace{f_{\mathbf{Y}}(\mathcal{D})}_{\text{Evidence}}} = \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathcal{D}|\mathbf{x})f_{\mathbf{X}}(\mathbf{x})}{\int f_{\mathbf{Y}|\mathbf{X}}(\mathcal{D}|\mathbf{x})f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}}. \quad (2.98)$$

Following the form of Eq. (2.98), $f_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})$ is usually specified via $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ and $f_{\mathbf{X}}(\mathbf{x})$, the *prior* PDF. The conditional PDF evaluated at \mathcal{D} is referred to as the *likelihood* $L(\mathbf{x}) = f_{\mathbf{Y}|\mathbf{X}}(\mathcal{D}|\mathbf{x})$. The

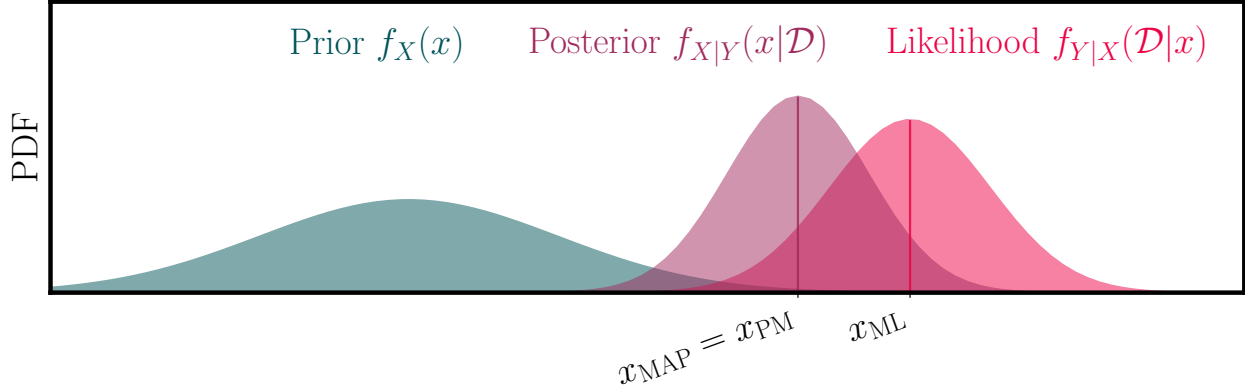


Figure 2.14: Bayesian parameter inference with conjugate priors: Gaussian prior and Gaussian likelihood with known variance σ^2 lead to a Gaussian posterior where the posterior mean (x_{PM}) is a convex combination of prior and data mean. In this case, posterior mean and maximum a-posteriori (x_{MAP}) coincide due to the symmetry of the Gaussian. The influence of the prior compared to the pure maximum likelihood (x_{ML}) estimate is clearly visible.

likelihood establishes a connection between observed and unobserved variables that is required for performing inference. The prior on the other hand is where the abovementioned prior belief enters that sets Bayesian inference apart from, e.g., solely likelihood-based inference. The likelihood quantifies how likely any given value of \mathbf{x} is to generate the observed data \mathcal{D} according to the chosen conditional model $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$, whereas the prior assigns a belief of how likely any value of \mathbf{x} is to occur to begin with. The denominator of Eq. (2.98) represents the *model evidence*, which is a measure for the plausibility of the chosen model $f_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})$. The model evidence is therefore a way of comparing different models (both likelihood and prior models) based on the available data.

Two common settings in which the above general framework finds application in the context of engineering are *Bayesian parameter inference* and *Bayesian model inversion*. In the former case, \mathbf{X} are parameters of the PDF $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ and the goal is to compute the posterior distribution of these distributional parameters after seeing the data on \mathbf{Y} . Upon selecting representative points from $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathcal{D})$, such as the maximum a-posteriori point (MAP) $\mathbf{x}_{\text{MAP}} = \arg \max f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathcal{D})$ or the posterior mean $\mathbf{x}_{\text{PM}} = \int \mathbf{x} f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathcal{D}) d\mathbf{x}$, the updated distribution of \mathbf{Y} is obtained by plugging these points in the conditional model, i.e., as $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}_{\text{MAP}})$ or $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}_{\text{PM}})$ (compare Fig. 2.14). Alternatively, if the uncertainty in the parameters is large, their entire posterior PDF may be used to compute the *posterior predictive* PDF of \mathbf{Y} , $f_{\tilde{\mathbf{Y}}|\mathbf{Y}}(\tilde{\mathbf{y}}|\mathcal{D})$, as

$$f_{\tilde{\mathbf{Y}}|\mathbf{Y}}(\tilde{\mathbf{y}}|\mathcal{D}) = \int f_{\mathbf{Y}|\mathbf{X},\mathbf{Y}}(\tilde{\mathbf{y}}|\mathbf{x}, \mathcal{D}) f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathcal{D}) d\mathbf{x} = \int f_{\mathbf{Y}|\mathbf{X}}(\tilde{\mathbf{y}}|\mathbf{x}) f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathcal{D}) d\mathbf{x}. \quad (2.99)$$

The second equality in Eq. (2.99) holds due to the independence assumption of $\tilde{\mathbf{Y}}$ and \mathcal{D} conditional on \mathbf{X} , i.e., $f_{\tilde{\mathbf{Y}}|\mathbf{X},\mathbf{Y}}(\tilde{\mathbf{y}}|\mathbf{x}, \mathcal{D}) = f_{\tilde{\mathbf{Y}}|\mathbf{X}}(\tilde{\mathbf{y}}|\mathbf{x})$.

In the second setting — Bayesian model inversion — the Bayesian framework is used to infer the input distribution based on observations corresponding to the output of a computational model $\mathcal{Y} : \mathbf{X} \rightarrow \mathbf{Y}$. This is typically achieved by modelling the discrepancy between each observation and the corresponding model output with a probability distribution: $\epsilon_i = \mathbf{y}_i - \mathcal{Y}(\mathbf{x}) \sim f_{\mathbf{E}_i}(\epsilon_i)$. $f_{\mathbf{E}_i}(\epsilon_i)$ is frequently taken as a zero mean Gaussian PDF, i.e., $\mathbf{E}_i \sim \mathcal{N}(\epsilon_i|\mathbf{0}, \Sigma)$, where Σ is a covari-

ance matrix describing possible linear dependencies amongst the observation components. Further, assuming statistically independent measurements $\{\mathbf{y}_i\}_{i=1}^n$, the likelihood reads

$$L(\mathbf{x}) = \prod_{i=1}^n \mathcal{N}(\mathbf{y}_i | \mathcal{Y}(\mathbf{x}), \Sigma). \quad (2.100)$$

As mentioned above, choosing a prior is the subjective part of a Bayesian analysis. The degree of subjectivity may vary, however, as it may be based on personal belief, expert elicitation (which may be equivalent to personal belief if the reader considers themselves an expert in the relevant field), selection rules [162] and/or guiding principles to construct noninformative priors such as Jeffreys's priors [102, 101] or priors satisfying the maximum entropy principle [100]. While many of these principles rest on the idea to minimize the influence the prior exerts on the posterior distribution and thus aim at 'letting the data speak', it is flat/weak priors in particular that can lead to overly confident inference results [72]. A single layer of priors may not do justice to complex models with a large number of unobserved variables [71, Section 2.8]. Instead, hierarchical models may be used, where additional so-called *hyperpriors* are placed on the parameters of the prior distribution of \mathbf{x} (in turn, the parameters of the hyperpriors can be equipped with a distribution again and in this manner hierarchical models with a potentially arbitrary number of layers can be constructed). Moreover, priors are frequently selected functionally, e.g., to achieve closed-form solutions (so-called *conjugate priors* [186, 51]), to provide regularization for ill-posed inference problems [140] or to induce sparse solutions [222].

With the exception of conjugate priors, where for a given likelihood, the prior is chosen such that prior and posterior are of the same distribution type, computing the posterior exactly and in closed form is rarely possible (a few non-conjugate special cases with exact closed-form solutions exist but are not further discussed). Therefore, a compendium of approximate and simulation-based methods for sampling from posterior distributions and computing the model evidence has formed. These approaches include MC-based numerical integration, acceptance-rejection and importance sampling, Laplace approximations, ABC (Approximate Bayesian Computation) methods, variational Bayesian inference and MCMC methods. We refer to [71, Chapters 10 - 13] for a review and a detailed account of original literature on all these methods.

2.5.2 Bayesian Uncertainty Separation

Many QoIs arising in UQ can be cast in terms of an expectation of a function of the computational model \mathcal{Q} , where we restrict ourselves to the case of scalar model output $Y = \mathcal{Y}(\mathbf{X})$. Any such QoI can be written as

$$q = \mathbb{E}[\mathcal{Q}] = \int_{\mathbb{R}^d} \mathcal{Q} \circ \mathcal{Y}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^d} \mathcal{Q} \circ \mathcal{Y} \circ T^{-1}(\mathbf{u}) \varphi(\mathbf{u}) d\mathbf{u}. \quad (2.101)$$

Depending on the UQ discipline, \mathcal{Q} takes different forms:

Uncertainty Propagation $\mathcal{Q} = Y^m$	Sensitivity Analysis $\mathcal{Q} = \mathbb{E}[Y X_v]^2$	Reliability Analysis $\mathcal{Q} = \mathbb{I}[g(Y) \leq 0]$	Decision Analysis $\mathcal{Q} = \mathcal{L}(Y, a)$	Optimization $\mathcal{Q} = \mathcal{P}(Y, \delta)$ $\mathcal{Q} = \mathcal{C}(Y, \delta)$
---	--	--	---	--

In uncertainty propagation, the interest is in moments of the model output (potentially also in its entire distribution), in (variance-based) sensitivity analysis it is in the output variance conditional

on subsets of the model inputs \mathbf{X}_v and in reliability analysis it is the probability of failure. In (stochastic) optimization, the quantity of interest is the output of an uncertain performance function \mathcal{P} (e.g. in robust optimization, a combination of the noisy performance criterion's mean and standard deviation are maximized) or a probabilistic constraint \mathcal{C} (e.g., in reliability-based design optimization, a deterministic objective is maximized under a stochastic constraint, namely on the probability of failure of the system) that depend on the optimization parameters $\boldsymbol{\delta}$. Finally, the decision analysis scenario can be considered generalizing stochastic optimization in the sense that it formulates the search of optima over a set of decisions $\{\mathbf{a}\}$ rather than a set of parameters $\{\boldsymbol{\delta}\}$. Typically, a loss function $\mathcal{L}(Y, a)$ measures the expected loss given any of the available decisions and needs to be minimized.

Frequently, a subset of the model inputs is of particular interest and we would like to investigate how q depends on these inputs. We call them *type B inputs* and write \mathbf{X}_B and all remaining inputs are referred to as the *type A inputs* \mathbf{X}_A . Without loss of generality, one may assume that \mathbf{X} can be ordered such that $\mathbf{X} = [\mathbf{X}_A, \mathbf{X}_B]$. As q is an expectation taken over all inputs, Eq. (2.101) establishes no direct relation between these inputs and the QoI. To obtain such a relation, any $q = \mathbb{E}_{f_{\mathbf{X}_B}}[\mathcal{Q}(\mathbf{X}_B)]$ can be evaluated conditional on the subset of the model inputs we are interested in:

$$Q(\mathbf{X}_B) = \mathbb{E}_{f_{\mathbf{X}_A}}[\mathcal{Q}|\mathbf{X}_B] = \int_{\mathbb{R}^{d_A}} \mathcal{Q} \circ \mathcal{Y}(\mathbf{x}_A, \mathbf{X}_B) f_{\mathbf{X}_A|\mathbf{X}_B}(\mathbf{x}_A|\mathbf{X}_B) d\mathbf{x}_A. \quad (2.102)$$

This is the principle idea of uncertainty separation: as opposed to q , $Q(\mathbf{X}_B)$ is no longer a deterministic quantity but a random variable whose properties may be analyzed in order to characterize how \mathbf{X}_B affects Q . Just as Eq. (2.101), Eq. (2.103) can be cast in standard-normal space by means of the isoprobabilistic transformation $T : \mathbf{X} \rightarrow \mathbf{U}$ as

$$Q(\mathbf{U}_B) = \mathbb{E}_{\varphi_{d_A}}[\mathcal{Q}|\mathbf{U}_B] = \int_{\mathbb{R}^{d_A}} \widehat{\mathcal{Q}} \circ \mathcal{Y} \circ T^{-1}(\mathbf{u}_A, \mathbf{U}_B) \varphi_{d_A}(\mathbf{u}_A) d\mathbf{u}_A, \quad (2.103)$$

where $d_A = \dim(\mathbf{X}_A)$. If T is a bijective map (e.g., if T is given by a Rosenblatt transform or a marginal probability integral transform), the uncertainty separation $\mathbf{X} = [\mathbf{X}_A, \mathbf{X}_B]$ directly translates to standard-normal space $\mathbf{U} = [\mathbf{U}_A, \mathbf{U}_B]$.

A possible motivation for uncertainty separation is to treat non-learnable (type A) and learnable inputs (type B) separately. Learning here refers to the collection of additional data that can subsequently be used to reduce uncertainty in the type B inputs using a Bayesian approach. Depending on the nature of the collected data, this may be done through either parameter inference or model inversion as described in Subsection 2.5.1. The influence of the learnable inputs on the unconditional QoI q can be characterized by first computing and analyzing the conditional QoI. Its variance or interval bounds containing 90%/95%/99% probability mass may be used to communicate the uncertainty in the conditional QoI and thus the potential of reducing uncertainty in type B inputs (see publication 4 in Chapter 7).

In a next step, Q may be subjected to a sensitivity analysis to identify an importance ranking amongst the type B inputs to determine which of the type B inputs should be learned in order to optimally reduce uncertainty in the QoI. While many sensitivity metrics are potentially applicable to obtain such a ranking, the crux resides in approximating $Q(\mathbf{X}_B)$ with sufficient accuracy, which is often too computationally expensive. For example, if $\mathcal{Q} = \mathbb{I}[g(Y) \leq 0]$ (i.e., the reliability setting), each evaluation of Q requires solving a reliability problem over a d_A -dimensional input space, while efficient MC estimates of Sobol' indices call for $n(d_B + 2)$ model evaluations (which in this

case means solving $n(d_B + 2)$ reliability problems) with $n \sim \mathcal{O}(10^2 - 10^3)$ [99]. The conditional reliability analysis method of publication 4 in Chapter 7 can be used to solve this series of $n(d_B + 2)$ reliability problems efficiently. Alternatively, the bi-level surrogate modelling approach mentioned in Section 2.4 and described originally in publication 5 (Chapter 8), can be used to efficiently compute a surrogate of Q , from which variance-based sensitivities of the conditional probability of failure with respect to \mathbf{X}_B may be derived. These sensitivities can be understood as a preposterior (prior to actually collecting more data and reducing uncertainty) heuristic of how much variability there is to reduce in Q by reducing uncertainty in any subset of \mathbf{X}_B (i.e., learning it).

A related, decision-oriented approach to such a preposterior analysis is discussed in [216]. Here, a sensitivity ranking amongst the type B variables is computed for a concrete decision directly. Such a decision is defined as choosing between a set of available options given a loss function $\mathcal{L}(Y, a)$ that associates each option a with a loss or cost. Investigated decision types include whether or not to repair a system (a is binary) and how to design a system (a is continuous). The sensitivity of the decision with respect to a subset of type B inputs is then defined by the expected loss reduction induced by knowing the value of said subset with certainty (this is referred to as partial perfect information). In [216], the conditional probability of failure forms the basis for computing the conditional QoI (which is the loss conditional on various type B input subsets \mathbf{v}) in two versions: once using the optimal a-priori decision without any additional information, a_{opt} (the associated loss being $\mathcal{L}(Y, a_{\text{opt}}|X_{B,\mathbf{v}})$) and once using the optimal decision conditional on knowing $X_{B,\mathbf{v}}$ perfectly, $a_{\text{opt}|X_{B,\mathbf{v}}}$ (the associated loss being $\mathcal{L}(Y, a_{\text{opt}|X_{B,\mathbf{v}}}|X_{B,\mathbf{v}})$). The difference between these two is referred to as the *conditional value of partial perfect information* (CVPPi) and the sensitivity index is obtained by taking the expected value of the CVPPi with respect to $f_{\mathbf{X}_{B,\mathbf{v}}}$, which yields the *expected value of partial perfect information* (EVPPI). [216] refers to this as *decision-theoretic reliability sensitivity* since loss functions based on failure probabilities are considered exclusively. The general concept, however, can be transferred to loss functions depending on other conditional statistics of the model output. Other such statistics may include quantile values or tail-constrained expectations (e.g., in relation to the *Conditional Value at Risk* [189]), moments, et cetera. Again, the decisive computational requirement is the ability to compute $Q(\mathbf{X}_B)$ efficiently and accurately.

Once \mathbf{X}_B is updated with novel data \mathcal{D} on the observable variables (observable variables could be either \mathbf{X}_B itself or the model output Y , in which case the inference mode switches to Bayesian model inversion as discussed in Subsection 2.5.1) in the Bayesian framework, we obtain either a closed-form expression/approximation or samples of the posterior PDF $f_{\mathbf{X}_B|Y}(\mathbf{x}_B|\mathcal{D})$. Once a surrogate model of the conditional QoI $\widehat{Q}(\mathbf{X}_B)$ is established in the course of a preposterior analysis as laid out in the previous paragraphs, q can be updated approximately at no additional cost via

$$\begin{aligned} \widehat{q}_{\text{post}} &= \frac{1}{n} \sum_{k=1}^n \widehat{Q}(\mathbf{x}_B^k), \quad \mathbf{x}_B^k \sim f_{\mathbf{X}_B|Y} \\ &= \widehat{\mathbb{E}}_{f_{\mathbf{X}_B|Y}}[\widehat{Q}(\mathbf{X}_B)|\mathcal{D}] \approx \int_{\mathbb{R}^{d_B}} \widehat{Q}(\mathbf{x}_B) f_{\mathbf{X}_B|Y}(\mathbf{x}_B|\mathcal{D}). \end{aligned} \tag{2.104}$$

The estimate $\widehat{q}_{\text{post}}$ involves two sources of inaccuracy, namely the sampling error and the surrogate approximation error. While sampling errors can be efficiently controlled through the number of posterior samples, the quality of the surrogate approximation depends largely on how different the prior and posterior PDFs $f_{\mathbf{X}_B}(\mathbf{x}_B)$ and $f_{\mathbf{X}_B|Y}(\mathbf{x}_B|\mathcal{D})$ are. As the surrogate model is constructed over a DoE sampled from the prior PDF, it can only be expected to accurately predict Q at posterior samples if the posterior PDF support remains somewhat close to the prior PDF support.

Illustration

In the following, we demonstrate the concept of uncertainty separation in reliability analysis ($q = \mathbb{P}(F)$) of a wind turbine foundation that is also investigated in Publication 4 & 5 (Chapters 7 and 8). The considered monopile foundation is described in Fig. 2.15. Specific values for distributional and computational model parameters are provided in Chapters 7 and 8. In the following we briefly summarize the main features of this model for convenience. The monopile interacts with stiff plastic clay. The clay stiffness s_u along the monopile boundary is modelled by a series of springs. s_u is random and varies with the vertical coordinate z whence it is modelled as a random field. The random field trend in turn is modelled linear in z and s_0 and s_1 represent its intercept and slope, respectively ($\mu_s(z) = s_0 + s_1 z$). s_0 and s_1 are random themselves (hyperparameters) and modelled with lognormal distributions. H is the load acting on the turbine blades and is assumed to follow a Gumbel distribution with parameters a_H and b_H , which are hyperparameters as well and assumed to have log-normal distributions. In the illustrated scenario, all hyperparameters (a_H, b_H, s_0, s_1) as well as the soil stiffness random field itself are considered learnable.

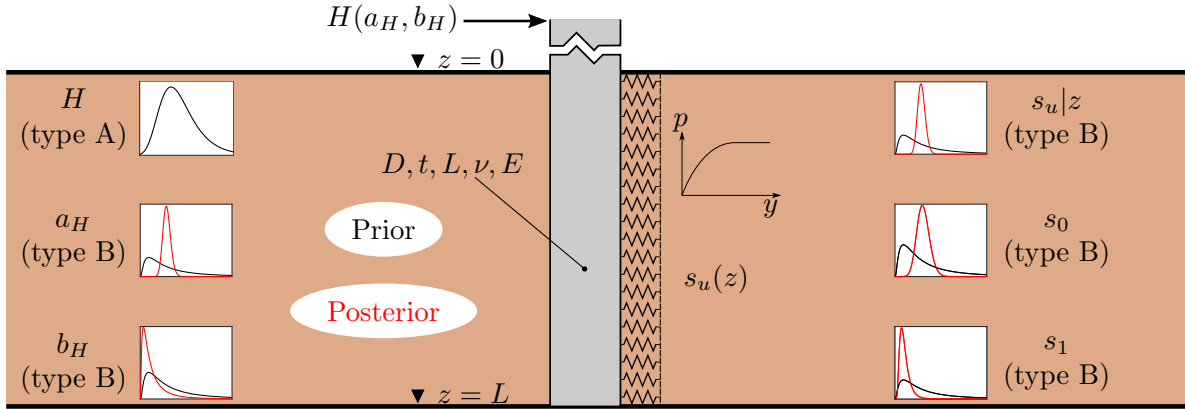


Figure 2.15: \mathcal{Y} is a FE model of the interaction between a wind turbine monopile foundation and stiff plastically behaving soil under wind load H .

The unconditional QoI q in this example is the probability of failure, where failure occurs if the maximum occurring stress in the foundation exceeds a threshold, i.e., $q = \mathbb{P}(F)$. Based on the approaches developed in publications 4 & 5 (Chapters 7 and 8), a DoE for the mapping from $\mathbf{X}_B = [\boldsymbol{\xi}^T, s_0, s_1, a_H, b_H]^T$ to $Q = \mathbb{P}(F|\mathbf{X}_B)$ can be computed based on which a surrogate $\hat{Q}(\mathbf{X}_B)$ can be constructed. The PDF of the conditional probability of failure $f_{\hat{Q}}$ is shown in Fig. 2.16 (top). By integrating $\hat{Q}(\mathbf{X}_B)$ with respect to different subsets of the \mathbf{X}_B vector, the PDF of \hat{Q} conditional on all \mathbf{X}_B coordinates not integrated over emerges. This process is called marginalization. Fig. 2.16 depicts the resulting PDFs of \hat{Q} conditional on various subsets of \mathbf{X}_B obtained through different marginalization steps. Evidently, the PDFs conditional on b_H and – to a lesser extent – a_H exhibit considerably larger dispersion than those conditional on any of the remaining parameters. This implies that b_H and a_H are most influential for the conditional failure probability and should be prioritized when deciding on which of the parameters additional information should be gathered. This is consistent with the surrogate-based reliability sensitivity analysis results obtained for the

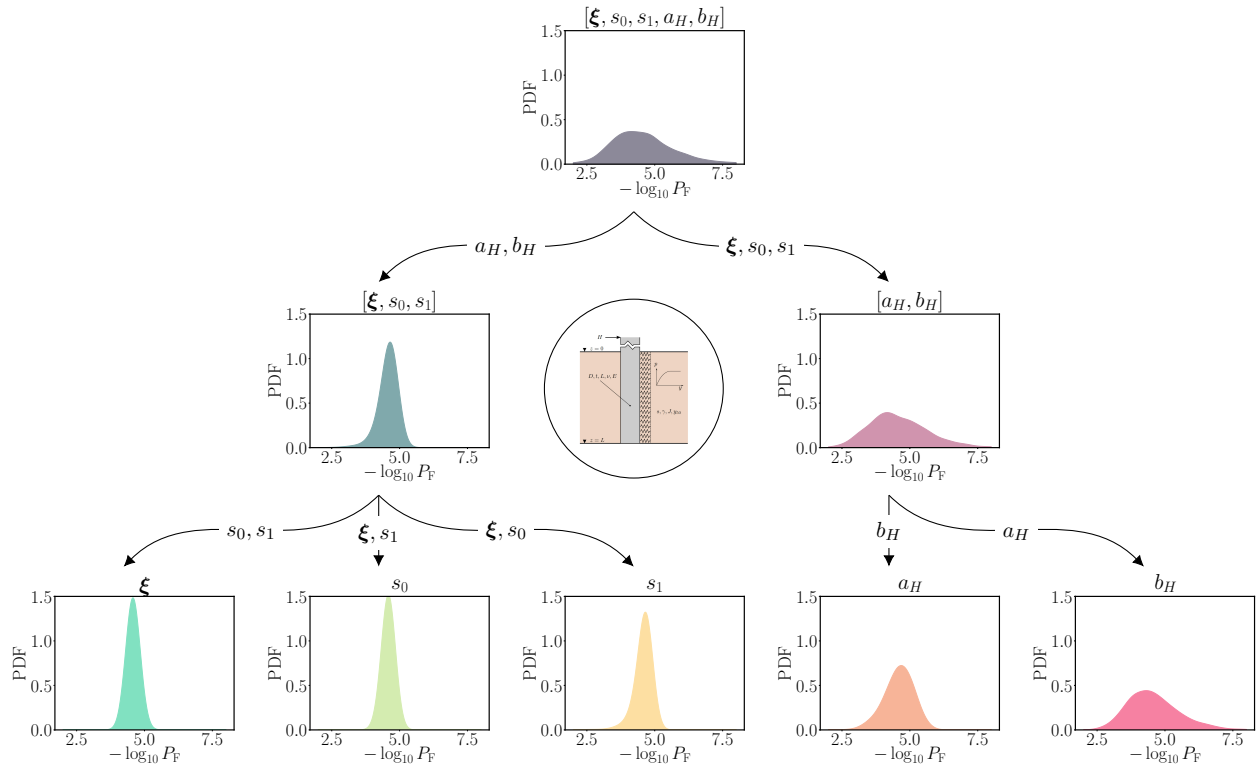


Figure 2.16: Hierarchy of conditional failure magnitude densities: each density is estimated with a KDE using 2000 samples of $-\log_{10} \hat{\mathbb{P}}(\mathbf{F}|\mathbf{X}_B)$ with different subsets \mathbf{X}_B . Starting at the top with a model for $-\log_{10} \hat{\mathbb{P}}(\mathbf{F}|\mathbf{X}_B)$, other conditional models are obtained via marginalization (the marginalized variables are indicated along the edges connecting two conditional models in the graph).

monopile example in publication 5 (Chapter 8) and the conditional densities estimated in publication 4 (Chapter 7).

Concluding Remarks & Outlook

3.1 High-dimensional surrogate modelling for UQ

3.1.1 Summary

In Chapter 4, we addressed the problem of constructing surrogate models for problems with high-dimensional probabilistic input. While most conventional surrogate models suffer from the curse of dimensionality and are thus not fit to tackle such high-dimensional problems, we propose to construct polynomial chaos expansions in low-dimensional linear subspaces of the original high-dimensional input space. This subspace is constructed with a dimensionality-reducing regression technique known as nonlinear partial least squares. The procedure is designed to optimally match the identified subspace with the polynomial chaos expansion constructed therein. Our method offers polynomial computational complexity in the number of inputs d , as opposed to standard polynomial chaos expansions that scale exponentially with d .

In Chapter 5, we showed how to transform any partial least squares-based polynomial chaos expansion as introduced in Chapter 4 into a standard polynomial chaos format. The latter admits a sampling-free and exact way of computing moments of the surrogate output as well as variance-based sensitivity indices thereof. Thus, by means of our transformation, partial least squares-based polynomial chaos expansions inherit these properties, which allows for efficient surrogate-based global sensitivity analysis of high-dimensional models. We provide both a simple asymptotic (large sample) approximation for this backtransformation as well as correction formulae for the small sample case (the sample size referred to here is the size of the design of experiments based on which the surrogate is constructed).

Finally, in Chapter 6, we demonstrate how to use our surrogate model for efficient rare event probability estimation of high-dimensional models. Surrogate models that are constructed with a non-adaptive design of experiments drawn from the input distribution cannot be expected to accurately predict the behaviour of the model in the tails of the model output, which, however, is crucial for accurate rare event probability estimation. We circumvent this issue by combining our surrogate approach with a sequential sampling technique that allows to gradually shift the design of

experiments towards the failure domain. This allows for constructing tail-accurate surrogate models. We introduced an active learning approach that allows for adaptively augmenting the design of experiments in each step of the sequential sampling technique such as to control the error of the partial least squares-based subspace as well as the polynomial chaos expansion constructed therein.

3.1.2 Outlook

The workhorse of our surrogate modelling approach is a nonlinear partial least squares algorithm. The algorithm is based on a nonlinear regression (Gauss-Newton) approach to iteratively identify subspace directions along with low-dimensional PCEs constructed in said subspace. This is done using a local linearization of the PCE model. In principle, this idea can be adapted to any other type of surrogate model. It is worth noting, however, that the gradients required for the linearization step can be obtained in analytically closed form when paired with polynomial chaos expansions, which makes for a particularly fast and powerful algorithm. We have observed that Gaussian processes tend to destabilize the nonlinear partial least squares algorithm, perhaps due to their increased flexibility compared to polynomial chaos expansions.

As partial least squares naturally extends to multivariate output, where it identifies subspace directions in both the input and output space simultaneously, extending our surrogate approach to problems with both high-dimensional inputs and outputs is straight-forward so that it can be used for sensitivity analysis of multivariate output [130].

Both in the context of uncertainty propagation and reliability analysis, a flexibilization of the dimensionality reduction approach could be achieved by giving up on the premise of a *linear* underlying subspace. This however means sacrificing several convenient properties of the latent variables such that it becomes necessary to think about how consistent surrogate models can be constructed on nonlinear manifolds.

3.2 Uncertainty separation

3.2.1 Summary

In Chapters 7 and 8, the concept of uncertainty separation is introduced with a focus on failure probability targets. The goal is to compute the probability of failure of an engineering system conditional on a subset of uncertain model inputs that are of special interest. Often this heightened interest arises from the fact that these inputs are learnable in the sense that more data can potentially be gathered that would lead to the reduction of uncertainty about these inputs. Computing the conditional probability allows for an in-depth analysis of how each of the learnable inputs influences the reliability of the system and consequently which of the inputs should be learned with priority. The associated computations are, however, expensive as a series of failure probabilities have to be estimated (one for each realization of the conditioned upon input subset at which the conditional failure probability is to be evaluated). This can be understood as a sequence of rare event probability estimation problems that are parametrised by the realizations of the conditioned upon input subset.

To render the solution of such problem series computationally feasible with expensive engineering models, we devised an algorithm that transfers and reuses information in between problems in Chapter 7. In particular, the approach is based on reusing importance sampling densities from other problems in the series in the context of a cross-entropy-based importance sampling method. Combining this idea with mixture and control variates importance sampling [164] yields an efficient way of solving series of hundreds to thousands of reliability problems. A central finding of ours is that the efficiency of the approach depends inversely on the variability of the target conditional probability. In order to alleviate computational burden of uncertainty separation independent of the variability contained in the conditioned upon variable subset, surrogate models may be invoked.

Such a surrogate-based approach is described in Chapter 8. Here, we aim at computing variance-based sensitivities of the conditional probability of failure with respect to each of the conditioned upon (i.e., learnable) model inputs. Based on these metrics, we aim at supporting decisions on which of these inputs additional information should be collected. The computational approach rests on a bi-level surrogate modelling framework, in which the first level emulates the model output and on the second level this first surrogate is used to construct a second surrogate model for the conditional probability of failure in function of the learnable inputs directly. This second surrogate model can then be used to compute the desired sensitivity metrics at negligible computational cost.

3.2.2 Outlook

Our uncertainty separation formulation has several connections to other domains of computational engineering. Foremost, there is a direct relationship with optimization under uncertainty and reliability-based optimization. In these fields, statistics of the model output such as mean, variance and tail probabilities have to be evaluated repeatedly at varying values of design parameters as the optimization progresses. By artificially considering the design parameters a set of conditioned upon random variables, the computational tools developed for uncertainty separation in this thesis are directly applicable for optimization under probabilistic constraints.

A second noteworthy connection exists with stochastic simulators [145, 245]. In essence, stochastic simulators are non-deterministic models in the sense that they may produce different outputs when evaluated at the same input values several times. Thus, only a fraction of the system uncertainty described by a stochastic simulator can be formulated as model input while another fraction is inherent to the model itself (examples are epidemiological or wind turbine performance models). A typical goal is to accurately emulate the model output distribution conditional on the model inputs. By defining the model inputs as a conditioned upon variable set and the model-inherent uncertainty as the remaining set of uncertainties, uncertainty separation can be applied to immediately obtain conditional moments or tail probabilities of such stochastic models. A promising future research direction could therefore consist in computing full conditional output densities based on uncertainty separation.

A third application of uncertainty separation is found in a recent work on decision-oriented reliability sensitivity analysis [216]. There, the value of information of perfectly knowing each of the learnable inputs is assessed in the framework of a particular system reliability-based decision such as whether or not to repair the system or how to choose the design of a system component. The

underlying computations in this approach require knowledge of the conditional probability of failure as defined and computed in Chapters 7 and 8.

Finally, we have addressed the possibility of using the conditional surrogate models introduced in Section 2.5 for performing approximate Bayesian updating of quantities of interest given newly available parameter data. We pointed out that for such an approximation to be satisfactory, the conditional surrogate should perform well where the parameter posterior has considerable probability mass and that this is only the case if parameter prior (from which the DoE of the surrogates is usually sampled) and posterior are sufficiently close to one another. In a future project, the approach could be extended to cases where prior and posterior significantly differ by developing adaptive strategies to train the surrogate on the parameter posterior rather than its prior.

References

- [1] P. Abrahamsen. *A Review of Gaussian Random Fields and Correlation Functions*. Technical Report 917, Norwegian Computing Center, Oslo, Norway., 1997.
- [2] M. Abramowitz and I. A. Stegun. *Handbook of mathematical function: with formulas, graphs and mathematical tables*. New York: Dover Publications, 1965.
- [3] R. Adler. *The Geometry of Random Fields*. Probability and Statistics Series. Wiley, 1981.
- [4] J. S. Arora and E. J. Haug. “Methods of Design Sensitivity Analysis in Structural Optimization”. In: *AIAA Journal* 17.9 (1979), pp. 970–974.
- [5] R. Ash et al. *Probability and Measure Theory*. Elsevier Science, 2000.
- [6] S. Au and J. Beck. “A new adaptive importance sampling scheme for reliability calculations”. In: *Structural Safety* 21.2 (1999), pp. 135–158.
- [7] S.-K. Au and J. L. Beck. “Estimation of small failure probabilities in high dimensions by subset simulation”. In: *Probabilistic Engineering Mechanics* 16.4 (2001), pp. 263–277.
- [8] S. Au. “Reliability-based design sensitivity by efficient simulation”. In: *Computers & structures* 83.14 (2005), pp. 1048–1061.
- [9] J. M. Aughenbaugh and C. J. Paredis. “The value of using imprecise probabilities in engineering design”. In: *Journal of Mechanical Design* 128.4 (2006), pp. 969–979.
- [10] P. C. Austin and E. W. Steyerberg. “The number of subjects per variable required in linear regression analyses”. In: *Journal of Clinical Epidemiology* 68.6 (2015), pp. 627–636.
- [11] S. D. Babacan, R. Molina, and A. K. Katsaggelos. “Bayesian Compressive Sensing Using Laplace Priors”. In: *IEEE Transactions on Image Processing* 19.1 (2010), pp. 53–63.
- [12] I. Babuska, F. Nobile, and R. Tempone. “A Stochastic Collocation Method for Elliptic Partial Differential Equations with Random Input Data”. In: *SIAM J. Numerical Analysis* 45 (2007), pp. 1005–1034.
- [13] G. Baffi, E. Martin, and A. Morris. “Non-linear projection to latent structures revisited (the neural network PLS algorithm)”. In: *Computers & Chemical Engineering* 23.9 (1999), pp. 1293–1307.

- [14] M. Balesdent, J. Morio, and J. Marzat. “Kriging-based adaptive Importance Sampling algorithms for rare event estimation”. In: *Structural Safety* 44.Supplement C (2013), pp. 1–10.
- [15] P. Beaurepaire et al. “Reliability-based optimization using bridge importance sampling”. In: *Probabilistic Engineering Mechanics* 34 (2013), pp. 48–57.
- [16] J. Bect, L. Li, and E. Vazquez. “Bayesian Subset Simulation”. In: *SIAM/ASA Journal on Uncertainty Quantification* 5.1 (2017), pp. 762–786.
- [17] J. Bect et al. “Sequential design of computer experiments for the estimation of a probability of failure”. In: *Statistics and Computing* 22.3 (2012), pp. 773–793.
- [18] T. Bedford and R. M. Cooke. “Vines—a new graphical model for dependent random variables”. In: *Ann. Statist.* 30.4 (Aug. 2002), pp. 1031–1068.
- [19] M. Beer et al. “Reliability analysis with scarce information: Comparing alternative approaches in a geotechnical engineering context”. In: *Structural Safety* 41 (2013), pp. 1–10.
- [20] R. Bellman. *Adaptive Control Processes*. Princeton University Press, 1961.
- [21] J. Benjamin and C. Cornell. *Probability, Statistics, and Decision for Civil Engineers*. Dover Books on Engineering. Dover Publications, 2014.
- [22] M. Berveiller, B. Sudret, and M. Lemaire. “Stochastic finite element: a non intrusive approach by regression”. In: *European Journal of Computational Mechanics* 15.1-3 (2006), pp. 81–92.
- [23] W. Betz, I. Papaioannou, and D. Straub. *Bayesian postprocessing of Monte Carlo simulation in reliability analysis*. Manuscript in preparation. 2021.
- [24] W. Betz, I. Papaioannou, and D. Straub. “Numerical methods for the discretization of random fields by means of the Karhunen-Loève expansion”. In: *Computer Methods in Applied Mechanics and Engineering* 271 (2014), pp. 109–129.
- [25] B. J. Bichon et al. “Efficient Global Reliability Analysis for Nonlinear Implicit Performance Functions”. In: *AIAA Journal* 46.10 (2008), pp. 2459–2468.
- [26] D. Bigoni, A. P. Engsig-Karup, and Y. M. Marzouk. “Spectral Tensor-Train Decomposition”. In: *SIAM Journal on Scientific Computing* 38.4 (2016), A2405–A2439.
- [27] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.
- [28] E. Borgonovo. “A new uncertainty importance measure”. In: *Reliability Engineering & System Safety* 92.6 (2007), pp. 771–784.
- [29] E. Borgonovo and E. Plischke. “Sensitivity analysis: a review of recent advances”. In: *European Journal of Operational Research* 248.3 (2016), pp. 869–887.
- [30] Z. I. Botev and D. P. Kroese. “Efficient Monte Carlo simulation via the generalized splitting method”. In: *Statistics and Computing* 22.1 (2012), pp. 1–16.
- [31] J.-M. Bourinet. “Rare-event probability estimation with adaptive support vector regression surrogates”. In: *Reliability Engineering & System Safety* 150 (2016), pp. 210–221.
- [32] J.-M. Bourinet, F. Deheeger, and M. Lemaire. “Assessing small failure probabilities by combined subset simulation and Support Vector Machines”. In: *Structural Safety* 33.6 (2011), pp. 343–353.
- [33] C. G. Bucher. “Adaptive sampling — an iterative fast Monte Carlo procedure”. In: *Structural Safety* 5.2 (1988), pp. 119–126.

- [34] R. E. Caflisch. “Monte Carlo and quasi-Monte Carlo methods”. In: *Acta Numerica* 7 (1998), pp. 1–49.
- [35] V. Chabridon. “Reliability-oriented sensitivity analysis under probabilistic model uncertainty – Application to aerospace systems”. PhD thesis. Université Clermont Auvergne, Nov. 2018.
- [36] O. Chapelle, V. Vapnik, and Y. Bengio. “Model selection for small sample regression”. In: *Machine Learning* 48.1-3 (2002), pp. 9–23.
- [37] G. Chastaing, F. Gamboa, and C. Prieur. “Generalized Hoeffding-Sobol Decomposition for Dependent Variables -Application to Sensitivity Analysis”. In: *Electronic Journal of Statistics* 6 (Dec. 2011).
- [38] S. S. Chen, D. L. Donoho, and M. A. Saunders. “Atomic Decomposition by Basis Pursuit”. In: *SIAM Journal on Scientific Computing* 20.1 (1998), pp. 33–61.
- [39] M. Chevreuil et al. “A Least-Squares Method for Sparse Low Rank Approximation of Multivariate Functions”. In: *SIAM/ASA Journal on Uncertainty Quantification* 3.1 (2015), pp. 897–921.
- [40] A. Cohen and G. Migliorati. “Optimal weighted least-squares methods”. In: *SMAI Journal of Computational Mathematics* 3 (2017), pp. 181–203.
- [41] P. G. Constantine, E. Dow, and Q. Wang. “Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces”. In: *SIAM Journal on Scientific Computing* 36.4 (2014), A1500–A1524.
- [42] L. Cui, Z. Lü, and X. Zhao. “Moment-independent importance measure of basic random variable and its probability density evolution solution”. In: *Science China Technological Sciences* 53.4 (2010), pp. 1138–1145.
- [43] G. M. Davis, S. G. Mallat, and Z. Zhang. “Adaptive time-frequency decompositions”. In: *Optical Engineering* 33.7 (1994), pp. 2183–2191.
- [44] L. De Lathauwer, B. De Moor, and J. Vandewalle. “A Multilinear Singular Value Decomposition”. In: *SIAM Journal on Matrix Analysis and Applications* 21.4 (2000), pp. 1253–1278.
- [45] A. P. Dempster. “Upper and Lower Probabilities Induced by a Multivalued Mapping”. In: *Ann. Math. Statist.* 38.2 (Apr. 1967), pp. 325–339.
- [46] G. Deodatis. “Weighted Integral Method. I: Stochastic Stiffness Matrix”. In: *Journal of Engineering Mechanics* 117.8 (1991), pp. 1851–1864.
- [47] G. Deodatis and M. Shinozuka. “Weighted Integral Method. II: Response Variability and Reliability”. In: *Journal of Engineering Mechanics* 117.8 (1991), pp. 1865–1877.
- [48] A. Der Kiureghian. “First-and second-order reliability methods”. In: *Engineering Design Reliability Handbook*. Ed. by E. Nikolaidis, D. M. Ghiocel, and S. Singhal. Boca Raton, FL: CRC Press, 2005. Chap. 14.
- [49] A. Der Kiureghian and O. Ditlevsen. “Aleatory or epistemic? Does it matter?” In: *Structural Safety* 31.2 (2009), pp. 105–112.
- [50] A. Der Kiureghian and J.-B. Ke. “The stochastic finite element method in structural reliability”. In: *Probabilistic Engineering Mechanics* 3.2 (1988), pp. 83–91.
- [51] P. Diaconis and D. Ylvisaker. “Conjugate Priors for Exponential Families”. In: *The Annals of Statistics* 7.2 (1979), pp. 269–281.
- [52] O. Ditlevsen and H. O. Madsen. *Structural reliability methods*. John Wiley & Sons Ltd, 1996.

- [53] A. Doostan and H. Owhadi. “A non-adapted sparse approximation of PDEs with stochastic inputs”. In: *Journal of Computational Physics* 230.8 (2011), pp. 3015–3034.
- [54] A. Doostan, A. Validi, and G. Iaccarino. “Non-intrusive low-rank separated approximation of high-dimensional stochastic models”. In: *Computer Methods in Applied Mechanics and Engineering* 263 (2013), pp. 42–55.
- [55] H. Drucker et al. “Support Vector Regression Machines”. In: *Advances in Neural Information Processing Systems*. Ed. by M. C. Mozer, M. Jordan, and T. Petsche. Vol. 9. MIT Press, 1997.
- [56] V. Dubourg, B. Sudret, and F. Deheeger. “Metamodel-based importance sampling for structural reliability analysis”. In: *Probabilistic Engineering Mechanics* 33 (2013), pp. 47–57.
- [57] B. Echard, N. Gayton, and M. Lemaire. “AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation”. In: *Structural Safety* 33.2 (2011), pp. 145–154.
- [58] B. Efron. “Bootstrap methods: Another look at the jackknife”. In: *Annals of Statistics* 7.1 (1979), pp. 1–26.
- [59] B. Efron et al. “Least angle regression (with discussion)”. In: *The Annals of Statistics* 32.2 (2004), pp. 407–499.
- [60] M. Ehre, I. Papaioannou, and D. Straub. “A framework for global reliability sensitivity analysis in the presence of multi-uncertainty”. In: *Reliability Engineering & System Safety* 195 (2020), p. 106726.
- [61] M. Ehre, I. Papaioannou, and D. Straub. “Global sensitivity analysis in high dimensions with PLS-PCE”. In: *Reliability Engineering & System Safety* 198 (2020), p. 106861.
- [62] M. Ehre et al. *Active sequential learning of low-dimensional model representations for reliability analysis*. Submitted to SIAM J. Comp. Sci. 2021.
- [63] M. Ehre et al. “Conditional reliability analysis in high dimensions based on controlled mixture importance sampling and information reuse”. In: *Computer Methods in Applied Mechanics and Engineering* 381 (2021), p. 113826.
- [64] S. Engelund and R. Rackwitz. “A benchmark study on importance sampling techniques in structural reliability”. In: *Structural Safety* 12.4 (1993), pp. 255–276.
- [65] M. Faber. *Statistics and Probability Theory: In Pursuit of Engineering Decision Support*. Topics in Safety, Risk, Reliability and Quality. Springer Netherlands, 2012.
- [66] M. H. Faber. “On the treatment of uncertainties and probabilities in engineering decision analysis”. In: *Journal of Offshore Mechanics and Arctic Engineering* 127.3 (2005), pp. 243–248.
- [67] L. Faravelli. “Response Surface Approach for Reliability Analysis”. In: *Journal of Engineering Mechanics* 115.12 (1989), pp. 2763–2781.
- [68] V. Fedorov. *Theory of Optimal Experiments*. Probability and Mathematical Statistics. Academic Press, Jan. 1972. 306 pp.
- [69] S. Ferson et al. *Constructing probability boxes and Dempster-Shafer structures*. University of North Texas Libraries, Jan. 2003.
- [70] S. Ferson et al. “Summary from the epistemic uncertainty workshop: consensus amid diversity”. In: *Reliability Engineering & System Safety* 85.1 (2004), pp. 355–369.
- [71] A. Gelman et al. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013.

- [72] A. Gelman and Y. Yao. “Holes in Bayesian statistics”. In: *Journal of Physics G: Nuclear and Particle Physics* 48.1 (Dec. 2020), p. 014002.
- [73] R. Ghanem and P. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer-Verlag, Berlin, 1991.
- [74] D. M. Ghiocel and R. G. Ghanem. “Stochastic Finite-Element Analysis of Seismic Soil-Structure Interaction”. In: *Journal of Engineering Mechanics* 128.1 (2002), pp. 66–77.
- [75] G. H. Golub and C. F. van Loan. *Matrix Computations*. Fourth. JHU Press, 2013.
- [76] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016.
- [77] W. Graf and J. U. Sickert. “Time-dependent fuzzy stochastic reliability analysis of structures”. In: *Applied Mechanics and Materials*. Vol. 104. Trans Tech Publ. 2012, pp. 45–54.
- [78] L. Grasedyck. “Hierarchical Singular Value Decomposition of Tensors”. In: *SIAM Journal on Matrix Analysis and Applications* 31.4 (2010), pp. 2029–2054.
- [79] L. Grasedyck, D. Kressner, and C. Tobler. “A literature survey of low-rank tensor approximation techniques”. In: *GAMM-Mitteilungen* 36.1 (2013), pp. 53–78.
- [80] M. Grigoriu. *Stochastic Systems*. Springer Series in Reliability Engineering 978-1-4471-2327-9. Springer, Dec. 2012.
- [81] X. Guan and R. Melchers. “Effect of response surface parameter variation on structural reliability estimates”. In: *Structural Safety* 23.4 (2001), pp. 429–444.
- [82] M. Hanss and S. Turrin. “A fuzzy-based approach to comprehensive modeling and analysis of systems with epistemic uncertainties”. In: *Structural Safety* 32.6 (2010), pp. 433–441.
- [83] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., 2001.
- [84] F. Hayashi. *Econometrics*. Princeton Univ. Press, 2000. XXIII, 683.
- [85] J. Helton and W. Oberkampf. “Alternative representations of epistemic uncertainty”. In: *Reliability Engineering & System Safety* 85.1 (2004). Alternative Representations of Epistemic Uncertainty, pp. 1–10.
- [86] F. L. Hitchcock. “The Expression of a Tensor or a Polyadic as a Sum of Products”. In: *Journal of Mathematics and Physics* 6.1-4 (1927), pp. 164–189.
- [87] W. Hoeffding. “A Class of Statistics with Asymptotically Normal Distribution”. In: *The Annals of Mathematical Statistics* 19.3 (1948), pp. 293–325.
- [88] M. Hohenbichler and R. Rackwitz. “Improvement Of Second-Order Reliability Estimates by Importance Sampling”. In: *Journal of Engineering Mechanics* 114.12 (1988), pp. 2195–2199.
- [89] T. Homma and A. Saltelli. “Importance measures in global sensitivity analysis of nonlinear models”. In: *Reliability Engineering & System Safety* 52.1 (1996), pp. 1–17.
- [90] G. Hooker. “Discovering Additive Structure in Black Box Functions”. In: *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '04. Seattle, WA, USA: Association for Computing Machinery, 2004, pp. 575–580.
- [91] G. Hooker. “Generalized Functional ANOVA Diagnostics for High-Dimensional Functions of Dependent Variables”. In: *Journal of Computational and Graphical Statistics* 16.3 (2007), pp. 709–732.

- [92] H. Hotelling. “Relations between two sets of variates”. In: *Biometrika* 28.3-4 (Dec. 1936), pp. 321–377.
- [93] X. Huang, J. Chen, and H. Zhu. “Assessing small failure probabilities by AK–SS: An active learning method combining Kriging and Subset Simulation”. In: *Structural Safety* 59 (2016), pp. 86–95.
- [94] J. E. Hurtado. “Filtered importance sampling with support vector margin: a powerful method for structural reliability analysis”. In: *Structural Safety* 29.1 (2007), pp. 2–15.
- [95] J. E. Hurtado and D. A. Alvarez. “Neural-network-based reliability analysis: a comparative study”. In: *Computer Methods in Applied Mechanics and Engineering* 191.1 (2001), pp. 113–132.
- [96] D. Hwang, D. W. Byun, and M. Talat Odman. “An automatic differentiation technique for sensitivity analysis of numerical advection schemes in air quality models”. In: *Atmospheric Environment* 31.6 (1997), pp. 879–888.
- [97] B. Iooss and P. Lemaître. “A review on global sensitivity analysis methods”. In: *Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications*. Ed. by C. Meloni and G. Dellino. Springer, 2015.
- [98] J. Iott, R. T. Haftka, and H. M. Adelman. “Selecting step sizes in sensitivity analysis by finite differences”. In: *NASA Technical Memorandum 86382* (Aug. 1985).
- [99] M. J. Jansen. “Analysis of variance designs for model output”. In: *Computer Physics Communications* 117.1 (1999), pp. 35–43.
- [100] E. Jaynes. “On the rationale of maximum-entropy methods”. In: *Proceedings of the IEEE* 70.9 (1982), pp. 939–952.
- [101] E. T. Jaynes. “Prior Probabilities”. In: *IEEE Transactions on Systems Science and Cybernetics* 4.3 (1968), pp. 227–241.
- [102] H. Jeffreys. “An invariant form for the prior probability in estimation problems”. In: *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 186.1007 (1946), pp. 453–461.
- [103] S. Ji, Y. Xue, and L. Carin. “Bayesian Compressive Sensing”. In: *IEEE Transactions on Signal Processing* 56.6 (2008), pp. 2346–2356.
- [104] Z. Jiang and J. Li. “High dimensional structural reliability with dimension reduction”. In: *Structural Safety* 69 (2017), pp. 35–46.
- [105] P. C. Kainen. “Utilizing Geometric Anomalies of High Dimension: When Complexity Makes Computation Easier”. In: *Computer Intensive Methods in Control and Signal Processing: The Curse of Dimensionality*. Ed. by M. Kárný and K. Warwick. Boston, MA: Birkhäuser Boston, 1997, pp. 283–294.
- [106] M. Kaminski. *The Stochastic Perturbation Method for Computational Mechanics*. John Wiley & Sons, Ltd, 2013.
- [107] L. Katafygiotis and K. Zuev. “Geometric insight into the challenges of solving high-dimensional reliability problems”. In: *Probabilistic Engineering Mechanics* 23.2 (2008). 5th International Conference on Computational Stochastic Mechanics, pp. 208–218.
- [108] A. Kaufman and M. M. Gupta. *Introduction to fuzzy arithmetic*. Van Nostrand Reinhold Company New York, USA, 1991.

- [109] A. Keese. “Numerical Solution of Systems with Stochastic Uncertainties: A General Purpose Framework for Stochastic Finite Elements”. PhD thesis. Apr. 2004.
- [110] D. G. Kendall. “Foundations of a theory of random sets”. In: *Stochastic geometry*. Ed. by E. F. Harding and D. G. Kendall. Wiley, 1974, pp. 322–376.
- [111] T. Kim and J. Song. “Generalized Reliability Importance Measure (GRIM) using Gaussian mixture”. In: *Reliability Engineering & System Safety* 173.C (2018), pp. 105–115.
- [112] C.-C. L. A. D. Kiureghian. “Optimal Discretization of Random Fields”. In: *Journal of Engineering Mechanics* 119.6 (1993), pp. 1136–1154.
- [113] T. G. Kolda and B. W. Bader. “Tensor Decompositions and Applications”. In: *SIAM Review* 51.3 (Sept. 2009), pp. 455–500.
- [114] A. Kolmogorov, N. Morrison, and A. Bharucha-Reid. *Foundations of the Theory of Probability*. AMS Chelsea Publishing Series. Chelsea Publishing Company, 1956.
- [115] K. Konakli and B. Sudret. “Global sensitivity analysis using low-rank tensor approximations”. In: *Reliability Engineering & System Safety* 156.Supplement C (2016), pp. 64–83.
- [116] K. Konakli and B. Sudret. “Reliability analysis of high-dimensional models using low-rank tensor approximations”. In: *Probabilistic Engineering Mechanics* 46 (2016), pp. 18–36.
- [117] P. Koutsourelakis, H. Pradlwarter, and G. Schuëller. “Reliability of structures in high dimensions, part I: algorithms and applications”. In: *Probabilistic Engineering Mechanics* 19.4 (2004), pp. 409–417.
- [118] D. P. Kroese, R. Y. Rubinstein, and P. W. Glynn. “Chapter 2 - The Cross-Entropy Method for Estimation”. In: *Handbook of Statistics: Machine Learning: Theory and Applications*. Vol. 31. Handbook of Statistics. Elsevier, 2013, pp. 19–34.
- [119] S. Kucherenko, S. Tarantola, and P. Annoni. “Estimation of global sensitivity indices for models with dependent variables”. In: *Computer Physics Communications* 183.4 (2012), pp. 937–946.
- [120] S. Kucherenko et al. “Monte Carlo evaluation of derivative-based global sensitivity measures”. In: *Reliability Engineering & System Safety* 94.7 (2009). Special Issue on Sensitivity Analysis, pp. 1135–1148.
- [121] N. Kurtz and J. Song. “Cross-entropy-based adaptive importance sampling using Gaussian mixture”. In: *Structural Safety* 42 (2013), pp. 35–44.
- [122] O. Le Matre et al. “A stochastic projection method for fluid flow. I. Basic formulation”. In: *Journal of Computational Physics* 173.2 (2001), pp. 481–511.
- [123] R. Lebrun and A. Dutfoy. “An innovating analysis of the Nataf transformation from the copula viewpoint”. In: *Probabilistic Engineering Mechanics* 24.3 (2009), pp. 312–320.
- [124] M. Lemaire, A. Chateaufneuf, and J.-C. Mitteau. *Structural reliability*. Wiley-ISTE, 2009.
- [125] P. Lemaitre et al. “Density modification-based reliability sensitivity analysis”. In: *Journal of Statistical Computation and Simulation* 85.6 (2015), pp. 1200–1223.
- [126] G. Li et al. “Global Sensitivity Analysis for Systems with Independent and/or Correlated Inputs”. In: *The Journal of Physical Chemistry A* 114.19 (May 2010), pp. 6022–6032.
- [127] J. Li, J. Li, and D. Xiu. “An efficient surrogate-based method for computing rare failure probability”. In: *Journal of Computational Physics* 230.24 (2011), pp. 8683–8697.

- [128] J. Li and D. Xiu. “Evaluation of failure probability via surrogate models”. In: *Journal of Computational Physics* 229.23 (2010), pp. 8966–8980.
- [129] K.-C. Li. “Sliced Inverse Regression for Dimension Reduction”. In: *Journal of the American Statistical Association* 86.414 (1991), pp. 316–327.
- [130] L. Li, I. Papaioannou, and D. Straub. *Partial least squares-based polynomial chaos expansion for global sensitivity analysis of dynamic models in high dimensions*. Manuscript. 2019.
- [131] L. Li et al. “Moment-independent importance measure of basic variable and its state dependent parameter solution”. In: *Structural Safety* 38 (2012), pp. 40–47.
- [132] M. Li and Z. Wang. “Deep learning for high-dimensional reliability analysis”. In: *Mechanical Systems and Signal Processing* 139 (2020), p. 106399.
- [133] R. Li and R. Ghanem. “Adaptive polynomial chaos expansions applied to statistics of extremes in nonlinear random vibration”. In: *Probabilistic Engineering Mechanics* 13.2 (1998), pp. 125–136.
- [134] P.-L. Liu and A. Der Kiureghian. “Multivariate distribution models with prescribed marginals and covariances”. In: *Probabilistic Engineering Mechanics* 1.2 (1986), pp. 105–112.
- [135] W. K. Liu, T. Belytschko, and A. Mani. “Random field finite elements”. In: *International Journal for Numerical Methods in Engineering* 23.10 (1986), pp. 1831–1845.
- [136] M. Loeve. *Probability theory II*. 4th ed. Springer-Verlag New York, 1978.
- [137] Z. Lu et al. “Reliability sensitivity method by line sampling”. In: *Structural Safety* 30.6 (2008), pp. 517–532.
- [138] N. Lüthen, S. Marelli, and B. Sudret. *Sparse Polynomial Chaos Expansions: Literature Survey and Benchmark*. 2021.
- [139] M. Hohenbichler and R. Rackwitz. “Sensitivity and importance measures in structural reliability”. In: *Civil Engineering Systems* 3.4 (1986), pp. 203–209.
- [140] D. J. C. MacKay. “Bayesian interpolation”. In: *Neural computation* 4 (1992), pp. 415–447.
- [141] S. G. Mallat and Zhifeng Zhang. “Matching pursuits with time-frequency dictionaries”. In: *IEEE Transactions on Signal Processing* 41.12 (1993), pp. 3397–3415.
- [142] T. A. Mara and S. Tarantola. “Variance-based sensitivity indices for models with dependent inputs”. In: *Reliability Engineering and System Safety* 107 (2012), pp. 115–121.
- [143] S. Marelli and B. Sudret. “An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis”. In: *Structural Safety* 75 (2018), pp. 67–74.
- [144] A. Marrel et al. “Calculations of Sobol indices for the Gaussian process metamodel”. In: *Reliability Engineering & System Safety* 94.3 (2009), pp. 742–751.
- [145] A. Marrel et al. “Global sensitivity analysis of stochastic computer models with joint metamodels”. In: *Statistics and Computing* 22.3 (2012), pp. 833–847.
- [146] P. McCullagh and J. Nelder. *Generalized Linear Models, Second Edition*. Chapman and Hall/CRC Monographs on Statistics and Applied Probability Series. Chapman & Hall, 1989.
- [147] M. D. McKay, R. J. Beckman, and W. J. Conover. “Comparison of three methods for selecting values of input variables in the analysis of output from a computer code”. In: *Technometrics* 21.2 (1979), pp. 239–245.

- [148] R. E. Melchers. *Structural reliability analysis and prediction*. Civil Engineering Series. John Wiley, 1999.
- [149] N. Metropolis and S. Ulam. “The Monte Carlo method”. In: *Journal of the American statistical association* 44.247 (1949), pp. 335–341.
- [150] N. Metropolis et al. “Equation of State Calculations by Fast Computing Machines”. In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092.
- [151] B. Möller and M. Beer. “Engineering Computation Under Uncertainty - Capabilities of Non-traditional Models”. In: *Comput. Struct.* 86.10 (May 2008), pp. 1024–1041.
- [152] J. Morio. “Influence of input PDF parameters of a model on a failure probability estimation”. In: *Simulation Modelling Practice and Theory* 19.10 (2011), pp. 2244–2255.
- [153] J. Morio. “Non-parametric adaptive importance sampling for the probability estimation of a launcher impact position”. In: *Reliability Engineering & System Safety* 96.1 (2011). Special Issue on Safecomp 2008, pp. 178–183.
- [154] M. D. Morris. “Factorial Sampling Plans for Preliminary Computational Experiments”. In: *Technometrics* 33.2 (1991), pp. 161–174.
- [155] K. P. Murphy. *Machine learning : A probabilistic perspective*. MIT Press, 2013.
- [156] A. Narayan, J. Jakeman, and T. Zhou. “A Christoffel function weighted least squares algorithm for collocation approximations”. In: *Mathematics of Computation* 86.306 (2017), pp. 1913–1947.
- [157] A. Nataf. “Determination des distribution dont les marges sont données”. In: *Comptes Rendus de l’Académie des Sciences* 225 (1962), pp. 42–43.
- [158] R. B. Nelsen. *An Introduction to Copulas*. Springer Publishing Company, Incorporated, 2010.
- [159] J. von Neumann. “Various techniques used in connection with random digits”. In: *Monte Carlo Method*. Ed. by A. S. Householder, G. E. Forsythe, and H. H. Germond. Vol. 12. National Bureau of Standards Applied Mathematics Series. Washington, DC: US Government Printing Office, 1951. Chap. 13, pp. 36–38.
- [160] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. Society for Industrial and Applied Mathematics, 1992.
- [161] A. O’Hagan and J. E. Oakley. “Probability is perfect, but we can’t elicit it perfectly”. In: *Reliability Engineering & System Safety* 85.1 (2004), pp. 239–248.
- [162] A. O’Hagan. “Expert Knowledge Elicitation: Subjective but Scientific”. In: *The American Statistician* 73 (2019), pp. 69–81.
- [163] I. V. Oseledets and E. E. Tyrtyshnikov. “Breaking the Curse of Dimensionality, Or How to Use SVD in Many Dimensions.” In: *SIAM J. Sci. Comput.* 31.5 (2009), pp. 3744–3759.
- [164] A. Owen and Y. Zhou. “Safe and effective importance sampling”. In: *Journal of the American Statistical Association* 95.449 (2000), pp. 135–143.
- [165] A. B. Owen. “Controlling correlations in Latin hypercube samples”. In: *Journal of the American Statistical Association* 89.428 (1994), pp. 1517–1522.
- [166] A. B. Owen. *Monte Carlo theory, methods and examples*. 2013.
- [167] A. B. Owen. “Sobol’ Indices and Shapley Value”. In: *SIAM/ASA Journal on Uncertainty Quantification* 2.1 (2014), pp. 245–251.

- [168] A. B. Owen. “Variance Components and Generalized Sobol’ Indices”. In: *SIAM/ASA Journal on Uncertainty Quantification* 1.1 (2013), pp. 19–41.
- [169] Q. Pan and D. Dias. “Sliced inverse regression-based sparse polynomial chaos expansions for reliability analysis in high dimensions”. In: *Reliability Engineering & System Safety* 167 (2017). Special Section: Applications of Probabilistic Graphical Models in Dependability, Diagnosis and Prognosis, pp. 484–493.
- [170] V. Papadopoulos et al. “Accelerated subset simulation with neural networks for reliability analysis”. In: *Computer Methods in Applied Mechanics and Engineering* 223 (2012), pp. 70–80.
- [171] M. Papadrakakis, V. Papadopoulos, and N. D. Lagaros. “Structural reliability analysis of elastic-plastic structures using neural networks and Monte Carlo simulation”. In: *Computer Methods in Applied Mechanics and Engineering* 136.1-2 (1996), pp. 145–163.
- [172] I. Papaioannou, K. Breitung, and D. Straub. “Reliability sensitivity estimation with sequential importance sampling”. In: *Structural Safety* 75 (2018), pp. 24–34.
- [173] I. Papaioannou, M. Ehre, and D. Straub. “Efficient PCE representations for reliability analysis in high dimensions”. In: *Proceedings of the 19th working conference of the IFIP Working Group 7.5 on Reliability and Optimization of Structural Systems*. Ed. by J. Song. ETH Zürich, 2018.
- [174] I. Papaioannou, M. Ehre, and D. Straub. “PLS-based adaptation for efficient PCE representation in high dimensions”. In: *Journal of Computational Physics* 387 (2019), pp. 186–204.
- [175] I. Papaioannou, S. Geyer, and D. Straub. “Improved cross entropy-based importance sampling with a flexible mixture model”. In: *Reliability Engineering & System Safety* 191 (2019), p. 106564.
- [176] I. Papaioannou, C. Papadimitriou, and D. Straub. “Sequential importance sampling for structural reliability analysis”. In: *Structural Safety* 62 (2016), pp. 66–75.
- [177] I. Papaioannou and D. Straub. “Combination line sampling for structural reliability analysis”. In: *Structural Safety* 88 (2021), p. 102025.
- [178] I. Papaioannou and D. Straub. “Variance-based reliability sensitivity analysis and the FORM α -factors”. In: *Reliability Engineering & System Safety* 210 (2021), p. 107496.
- [179] I. Papaioannou et al. “MCMC algorithms for Subset Simulation”. In: *Probabilistic Engineering Mechanics* 41 (2015), pp. 89–103.
- [180] Y. C. Pati et al. “Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition”. In: *Proceedings of the 27th Annual Asilomar Conference on Signals, Systems, and Computers (1993)*. 1993, pp. 40–44.
- [181] B. Peherstorfer, B. Kramer, and K. Willcox. “Multifidelity Preconditioning of the Cross-Entropy Method for Rare Event Simulation and Failure Probability Estimation”. In: *SIAM/ASA Journal on Uncertainty Quantification* 6.2 (2018), pp. 737–761.
- [182] C. Prieur and S. Tarantola. “Variance-Based Sensitivity Analysis: Theory and Estimation Algorithms”. In: *Handbook of Uncertainty Quantification*. Ed. by R. Ghanem, D. Higdon, and H. Owhadi. Springer International Publishing, June 2017, pp. 1217–1239.
- [183] C. R.H. and W. Martin. “The orthogonal development of non-linear functionals in series of Fourier–Hermite functionals”. In: *Ann. Math.* 48 (1947), pp. 285–392.

- [184] R. Rackwitz and B. Fiessler. “Structural reliability under combined random load sequences”. In: *Computers & Structures* 9.5 (1978), pp. 489–494.
- [185] S. Rahman. “The f -Sensitivity Index”. In: *SIAM/ASA Journal on Uncertainty Quantification* 4.1 (2016), pp. 130–162.
- [186] H. Raiffa and R. Schlaifer. *Applied Statistical Decision Theory*. Harvard Business School Publications. Division of Research, Graduate School of Business Administration, Harvard University, 1961.
- [187] C. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. Cambridge, MA, USA: MIT Press, Jan. 2006, p. 248.
- [188] B. Reddy. *Introductory Functional Analysis: With Applications to Boundary Value Problems and Finite Elements*. Introductory Functional Analysis Series. Springer, 1998.
- [189] R. T. Rockafellar and S. Uryasev. “Optimization of Conditional Value-at-Risk”. In: *Journal of Risk* 2 (2000), pp. 21–41.
- [190] M. Rosenblatt. “Remarks on a multivariate transformation”. In: *The Annals of Mathematical Statistics* 23.3 (1952), pp. 470–472.
- [191] J. Rosenthal. *A First Look at Rigorous Probability Theory*. World Scientific, 2000.
- [192] R. Rosipal. “Nonlinear Partial Least Squares: An Overview”. In: *Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques* (Jan. 2010), pp. 169–189.
- [193] G. Rubino and B. Tuffin. *Rare Event Simulation using Monte Carlo Methods*. John Wiley & Sons, Ltd, 2009.
- [194] R. Y. Rubinstein. “Optimization of computer simulation models with rare events”. In: *European Journal of Operational Research* 99.1 (1997), pp. 89–112.
- [195] R. Y. Rubinstein. “The score function approach for sensitivity analysis of computer simulation models”. In: *Mathematics and Computers in Simulation* 28.5 (1986), pp. 351–379.
- [196] R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo Method*. 3rd. Wiley Publishing, 2017.
- [197] A. Saltelli, K. Chan, and E. Scott. *Sensitivity Analysis*. John Wiley & Sons, Inc., 2000.
- [198] A. Saltelli et al. *Sensitivity analysis in practice: a guide to assessing scientific models*. John Wiley & Sons, 2004.
- [199] R. Schöbi, B. Sudret, and S. Marelli. “Rare Event Estimation Using Polynomial-Chaos Kriging”. In: *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 3.2 (2017), p. D4016002.
- [200] R. Schöbi and B. Sudret. “Global sensitivity analysis in the context of imprecise probabilities (p-boxes) using sparse polynomial chaos expansions”. In: *Reliability Engineering & System Safety* 187 (2019). Sensitivity Analysis of Model Output, pp. 129–141.
- [201] G. Schuëller and R. Stix. “A critical appraisal of methods to determine failure probabilities”. In: *Structural Safety* 4.4 (1987), pp. 293–309.
- [202] L. Schueremans and D. V. Gemert. “Benefit of splines and neural networks in simulation based structural reliability analysis”. In: *Structural Safety* 27.3 (2005), pp. 246–261.
- [203] G. Seber and A. Lee. *Linear Regression Analysis*. Wiley Series in Probability and Statistics. Wiley, 2003.

- [204] G. A. Seber. *Nonlinear Regression*. John Wiley & Sons, 2005.
- [205] B. Settles. *Active Learning Literature Survey*. Computer Sciences Technical Report 1648. University of Wisconsin–Madison, 2009.
- [206] M. Shinozuka and G. Deodatis. “Simulation of stochastic processes by spectral representation.” In: *Applied Mechanics Reviews* 44.4 (1991), pp. 191–204.
- [207] J. Sickert et al. “Fuzzy probabilistic structural analysis considering fuzzy random functions”. In: *Proceedings of the 9th International Conference on Applications of Statistics and Probabilistics in Civil Engineering*. 2003, pp. 379–386.
- [208] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall, 1986.
- [209] M. Sklar. “Fonctions de répartition à n dimensions et leurs marges”. In: *Publ. inst. statist. univ. Paris* 8 (1959), pp. 229–231.
- [210] A. J. Smola and B. Schölkopf. “A tutorial on support vector regression”. In: *Statistics and Computing* 14.3 (2004), pp. 199–222.
- [211] S. A. Smolyak. “Quadrature and interpolation formulas for tensor products of certain class of functions”. In: *Dokl. Akad. Nauk SSSR* 148.5 (1963). Transl.: Soviet Math. Dokl. 4:240-243, 1963, pp. 1042–1053.
- [212] I. M. Sobol and G. A. “On an alternative global sensitivity estimators”. In: *Proc. of the Sensitivity Analysis on Model Output (SAMO) Conference 1995*. 1995.
- [213] I. Sobol’. “Sensitivity Estimates for Nonlinear Mathematical Models”. In: *Math. Modeling & Comp. Exp* 1 (1993), pp. 407–414.
- [214] C. Soize and R. Ghanem. “Physical Systems with Random Uncertainties: Chaos Representations with Arbitrary Probability Measure”. In: *SIAM J. Scientific Computing* 26 (Jan. 2004), pp. 395–410.
- [215] S. Song, Z. Lu, and H. Qiao. “Subset simulation for structural reliability sensitivity analysis”. In: *Reliability Engineering & System Safety* 94.2 (2009), pp. 658–665.
- [216] D. Straub, M. Ehre, and I. Papaioannou. *Decision-theoretic reliability sensitivity*. 2021.
- [217] B. Sudret. “Global sensitivity analysis using polynomial chaos expansions”. In: *Reliability Engineering & System Safety* 93.7 (2008), pp. 964–979.
- [218] B. Sudret. “Meta-models for Structural Reliability and Uncertainty Quantification”. In: *Proc. 5th Asian-Pacific Symp. Struct. Reliab. (APSSRA 2012)*, Singapore (Mar. 2012).
- [219] B. Sudret and A. Der Kiureghian. *Stochastic finite element methods and reliability: a state-of-the-art report*. Department of Civil and Environmental Engineering, University of California, 2000.
- [220] R. Tibshirani. “Regression Shrinkage and Selection via the Lasso”. In: *Journal of the Royal Statistical Society (Series B)* 58 (1996), pp. 267–288.
- [221] R. Tipireddy and R. Ghanem. “Basis adaptation in homogeneous chaos spaces”. In: *Journal of Computational Physics* 259 (2014), pp. 304–317.
- [222] M. E. Tipping. “Sparse Bayesian Learning and the Relevance Vector Machine”. In: *J. Mach. Learn. Res.* 1 (Sept. 2001), pp. 211–244.
- [223] E. Torre et al. “A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas”. In: *Probabilistic Engineering Mechanics* 55 (2019), pp. 1–16.

- [224] J. A. Tropp and A. C. Gilbert. “Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit”. In: *IEEE Trans. Inf. Theor.* 53.12 (2007), pp. 4655–4666.
- [225] P. Tsilifis et al. “Sparse Polynomial Chaos expansions using variational relevance vector machines”. In: *Journal of Computational Physics* 416 (2020), p. 109498.
- [226] L. R. Tucker. “Some mathematical notes on three-mode factor analysis”. In: *Psychometrika* 31.3 (1966), pp. 279–311.
- [227] E. Ullmann and I. Papaioannou. “Multilevel Estimation of Rare Events”. In: *SIAM/ASA Journal on Uncertainty Quantification* 3.1 (2015), pp. 922–953.
- [228] E. Vanmarcke and M. Grigoriu. “Stochastic Finite Element Analysis of Simple Beams”. In: *Journal of Engineering Mechanics* 109.5 (1983), pp. 1203–1214.
- [229] S. D. Veiga. “Global sensitivity analysis with dependence measures”. In: *Journal of Statistical Computation and Simulation* 85.7 (2015), pp. 1283–1305.
- [230] F. Wagner et al. “Multilevel Sequential Importance Sampling for Rare Event Estimation”. In: *SIAM Journal on Scientific Computing* 42.4 (2020), A2062–A2087.
- [231] P. Walley and W. Peter. *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1991.
- [232] P. Wang, Z. Lu, and Z. Tang. “A derivative based sensitivity measure of failure probability in the presence of epistemic and aleatory uncertainties”. In: *Computers & Mathematics with Applications* 65.1 (2013), pp. 89–101.
- [233] P. Wang, Z. Lu, and Z. Tang. “An application of the Kriging method in global sensitivity analysis with parameter uncertainty”. In: *Applied Mathematical Modelling* 37.9 (2013), pp. 6543–6555.
- [234] Z. Wang and J. Song. “Cross-entropy-based adaptive importance sampling using von Mises-Fisher mixture for high dimensional reliability analysis”. In: *Structural Safety* 59 (Mar. 2016), pp. 42–52.
- [235] P. Wiederkehr. *Global Sensitivity Analysis with Dependent Inputs*. 2018.
- [236] N. Wiener. “The homogeneous chaos”. In: *Amer. J. Math.* 60 (1938), pp. 897–936.
- [237] H. O. A. Wold. “Nonlinear estimation by iterative least squares procedures”. In: *Research Papers in statistics, Festschrift for J. Neyman*. Ed. by F. N. David. New York, New York: CRC Press, 1966, pp. 411–414.
- [238] S. Wold, N. Kettaneh-Wold, and B. Skagerberg. “Nonlinear PLS modeling”. In: *Chemometrics and Intelligent Laboratory Systems* 7.1 (1989). Proceedings of the First Scandinavian Symposium on Chemometrics, pp. 53–65.
- [239] S. Wold et al. “The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses”. In: *SIAM Journal on Scientific and Statistical Computing* 5.3 (1984), pp. 735–743.
- [240] Y.-T. Wu. “Computational methods for efficient structural reliability and reliability sensitivity analysis”. In: *AIAA journal* 32.8 (1994), pp. 1717–1723.
- [241] D. Xiu. “Efficient collocational approach for parametric uncertainty analysis”. In: *Communications in Computational Physics* 2 (Apr. 2007), pp. 293–309.
- [242] D. Xiu and G. E. Karniadakis. “The Wiener–Askey polynomial chaos for stochastic differential equations”. In: *SIAM Journal on Scientific Computing* 24.2 (2002), pp. 619–644.

- [243] J. Zhang and A. A. Taflanidis. “Adaptive Kriging Stochastic Sampling and Density Approximation and Its Application to Rare-Event Estimation”. In: *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 4.3 (2018), p. 04018021.
- [244] T. Zhang. “Adaptive Forward-Backward Greedy Algorithm for Sparse Learning with Linear Models”. In: *Advances in Neural Information Processing Systems*. Ed. by D. Koller et al. Vol. 21. Curran Associates, Inc., 2009.
- [245] X. Zhu and B. Sudret. *Emulation of stochastic simulators using generalized lambda models*. 2021.
- [246] K. M. Zuev et al. “Bayesian post-processor and other enhancements of Subset Simulation for estimating failure probabilities in high dimensions”. In: *Computers & Structures* 92-93 (2012), pp. 283–296.

PART II

PUBLISHED PAPERS

PLS-based adaptation for efficient PCE representation in high dimensions

Original Publication

I. Papaioannou, M. Ehre, and D. Straub. “PLS-based adaptation for efficient PCE representation in high dimensions”. In: *Journal of Computational Physics* 387 (2019), pp. 186–204.

Author’s contribution

Iason Papaioannou developed the idea of PLS-based PCE and implemented the algorithm with help from Max Ehre. Max Ehre carried out all numerical experiments. Iason Papaioannou drafted chapter 1-4 of the manuscript and Max Ehre drafted chapter 5. Iason Papaioannou and Max Ehre jointly drafted the conclusions in chapter 6 and Daniel Straub provided revisions on the manuscript.

Abstract

Uncertainty quantification of engineering systems modeled by computationally intensive numerical models remains a challenging task, despite the increase in computer power. Efficient uncertainty propagation of such models can be performed by use of surrogate models, such as polynomial chaos expansions (PCE). A major drawback of standard PCE is that its predictive ability decreases with increase of the problem dimension for a fixed computational budget. This is related to the fact that the number of terms in the expansion increases fast with the input variable dimension. To address this issue, Tipireddy and Ghanem (2014) introduced a sparse PCE representation based on a transformation of the coordinate system in Gaussian input variable spaces. In this contribution, we propose to identify the projection operator underlying this transformation and approximate the coefficients of the resulting PCE through partial least squares (PLS) analysis. The proposed

PCE-driven PLS algorithm identifies the directions with the largest predictive significance in the PCE representation based on a set of samples from the input random variables and corresponding response variable. This approach does not require gradient evaluations, which makes it efficient for high dimensional problems with black-box numerical models. We assess the proposed approach with three numerical examples in high-dimensional input spaces, comparing its performance with low-rank tensor approximations. These examples demonstrate that the PLS-based PCE method provides accurate representations even for strongly non-linear problems.

4.1 Introduction

In many domains of science and engineering, one employs models of physical systems that aim at representing accurately the behavior of the underlying system under future conditions. Input parameters of models and future conditions are subject to uncertainties. Uncertainties can be due to limited availability of data, limited understanding of the underlying physical process or the intrinsic randomness of a phenomenon, such as wind or earthquake. Proper quantification of uncertainties and their impact on the performance of the model is paramount for obtaining accurate predictions. Efficient uncertainty propagation of complex numerical models remains a challenge despite the increase in computer power. The challenge is two-fold: on the one hand, the analysis of complex systems often requires the use of computationally intensive deterministic solvers that are only available as *black boxes*, i.e. one does not have access to core routines of the computer code and hence cannot modify them. On the other hand, output quantities of interest are integrals over the space of uncertain inputs and numerical evaluation of these integrals suffers from the *curse of dimensionality*, i.e. the number of model evaluations increases geometrically with increase of the number of inputs for a fixed target accuracy.

Monte Carlo sampling can be easily coupled with black-box models and resolves the curse of dimensionality, but suffers from slow convergence rates. A possible remedy is to construct a surrogate model of the computationally intensive model using a simple mathematical form and then employ the surrogate model to perform uncertainty propagation. In particular, surrogate models based on polynomial chaos expansions (PCE) [18, 52] have enjoyed extensive application in uncertainty quantification due to their simplicity and guaranteed convergence property, among other reasons. The basic idea of PCE is to project the model output onto a space spanned by multivariate polynomials that are orthogonal with respect to the input probability measure. The projection can be performed by stochastic Galerkin schemes [18, 51, 28], which are intrusive in the sense that they require modification of existing deterministic solvers, or collocation-type methods [41, 7, 3, 49], which are non-intrusive and can be coupled with black-box deterministic solvers.

Non-intrusive PCE approaches estimate the coefficients of the expansion by numerical quadrature, interpolation or regression methods. Both numerical quadrature and interpolation techniques with tensor product grids suffer from the curse of dimensionality, i.e. their rate of convergence deteriorates drastically with increase of the dimension, e.g. [3, 49]. Approaches that employ sparse grids to delay the curse of dimensionality associated with integration/interpolation based on tensor product grids can be found in [50, 49, 31].

Accurate estimation of the PCE coefficients with regression requires an experimental design with size

equal to a multiple of the total number of PCE terms. The number of PCE terms in common truncation schemes increases fast with increase of the number of inputs. This implies that in problems with high-dimensional inputs, a large number of model evaluations is required for estimating accurately the PCE coefficients with regression. A possible solution is to construct a sparse polynomial basis through selecting the most significant terms in the PCE. This can be done through regularization techniques, such as the least-angle regression [8] or l_1 -minimization, also known as compressive sensing [16, 53]. Adaptive algorithms for determining sparse PCEs based on such techniques can be found in [8, 38, 22].

The quality of the PCE approximation obtained with regression can be potentially increased by choosing an appropriate experimental design set. A common choice is to use samples from the distribution of the input random variables obtained with standard Monte Carlo, stratified or quasi-random sampling schemes [9, 8]. Schemes based on the roots of the orthogonal polynomials are discussed in [7, 40] and randomized versions thereof in [55]. Alternative random sampling schemes that present optimal performance, especially in high-order PCEs, are discussed in [19, 30, 12]. Similar approaches have been proposed for use within compressive sensing-based sparse PCEs methods [53, 19, 23].

Another non-intrusive approach for surrogate modeling with polynomial bases is provided by canonical decompositions [15, 32, 17], a special case of low-rank tensor approximations (LRA). This approach is based on approximating the model response by a linear combination of rank-one approximations, obtained as products of univariate polynomial expansions. The coefficients can be determined adaptively, e.g. with a technique termed alternating least-squares regression [15, 11]. It is demonstrated in [25] that LRA performs better than sparse PCE in moderate dimensional problems and small experimental designs.

Recently, it has been proposed to reduce the number of terms in the classical PCE representation through performing a coordinate transformation in Gaussian space [42]. The transformed basis can be adapted to the output quantity of interest, e.g. through performing a small initial number of model evaluations. For example, the basis adaptation can be informed by evaluating the low-order coefficients with a sparse-grid numerical quadrature. This approach has been combined with compressive sensing for simultaneously determining the basis adaptation as well as the PCE coefficients in [43].

An alternative approach for identifying important directions in the space of uncertain inputs is the active subspace (AS) method [14, 13]. This approach is based on identifying the projection subspace of highest variability through decomposing the covariance matrix of the gradient of the model. Usually this matrix is estimated by a set of samples and corresponding gradient evaluations. Although AS can lead to vast dimensionality reduction, in high-dimensional problems with black box numerical models, the additional computational cost from the numerical evaluation of the gradients might be prohibitive. The AS method has been combined with PCEs in [44], wherein the covariance of the gradient vector is computed based on a low-order PCE.

In this contribution, we propose an approach that computes the basis transformation based on a given experimental design with a technique termed partial least squares (PLS) regression [48]. The proposed approach takes advantage of the structures in the covariance of the input parameters and model response to determine the directions with largest predictive significance in the PCE representation. Unlike the AS method, PLS does not require gradient evaluations at the samples,

and hence is ideally suited for application to high-dimensional problems with black box numerical models. The PLS method is widely used in the field of chemometrics, where regression with many variables but only a few observations is common [21, 47]. Here, we employ a nonlinear version of the PLS algorithm [46, 4] and modify it for use with PCE models. The proposed PCE-driven PLS algorithm identifies simultaneously a set of dominant directions and the corresponding PCE coefficients along each direction. The performance of the method is demonstrated with three high-dimensional examples, a linear elastic bar with stochastic axial rigidity, a hysteretic oscillator under random loading and a low-carbon steel plate with stochastic stiffness. The results are compared with the ones obtained by LRA and it is shown that the proposed PLS-based PCE approach performs consistently better than LRA for experimental design sizes in the order of the dimension of the random variable space.

4.2 PCE representations

Let \mathbf{X} be a random vector with outcome space \mathbb{R}^n and joint PDF $f_{\mathbf{X}}(\mathbf{x})$. Consider the Hilbert space \mathcal{H} of all functions from \mathbb{R}^n to \mathbb{R} with finite mean-square under the probability measure of \mathbf{X} . The inner product of two functions $g, h \in \mathcal{H}$ is defined as

$$\langle g, h \rangle_{\mathcal{H}} = \int_{\mathbb{R}^n} g(\mathbf{x})h(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} . \quad (4.1)$$

Consider the random variable $Y = \mathcal{M}(\mathbf{X})$ representing the response of an engineering model and assume that $\mathcal{M} \in \mathcal{H}$. Let $\{h_i(\mathbf{x}), i \in \mathbb{N}\}$ be a complete orthonormal basis of \mathcal{H} , thus satisfying

$$\langle h_i, h_j \rangle_{\mathcal{H}} = \delta_{ij} , \quad (4.2)$$

where δ_{ij} is the Kronecker symbol. Since $\{h_i(\mathbf{x}), i \in \mathbb{N}\}$ is a complete basis of \mathcal{H} , we can represent every element of \mathcal{H} as a linear combination of the functions $\{h_i(\mathbf{x}), i \in \mathbb{N}\}$. Therefore Y can be expressed as

$$Y = \mathcal{M}(\mathbf{X}) = \sum_{i=0}^{\infty} a_i h_i(\mathbf{X}) . \quad (4.3)$$

Truncating the representation of Eq. (4.3) after the first L terms, we get the following approximation of Y

$$\hat{Y} = \sum_{i=0}^L a_i h_i(\mathbf{X}) , \quad (4.4)$$

which converges to Y in the mean-square sense as $L \rightarrow \infty$. We now make the following assumption on the distribution of the vector \mathbf{X} .

Assumption 1. *The random vector \mathbf{X} follows the independent standard Gaussian distribution.*

In such case, we can construct an orthonormal polynomial basis of \mathcal{H} using products of one-dimensional normalized Hermite polynomials [18]

$$\Psi_{\mathbf{k}}(\mathbf{X}) = \prod_{i=1}^n \psi_{k_i}(X_i) , \quad (4.5)$$

where $\{\psi_i(X), i \in \mathbb{N}\}$ are the normalized (probabilist) Hermite polynomials and $\mathbf{k} = (k_1, \dots, k_n) \in \mathbb{N}^n$.

Remark. *In cases where the Gaussian restriction of assumption 1 does not apply, it is possible to express the random variable Y as a function of an underlying independent Gaussian input through performing an isoprobabilistic transformation [36].*

The p -th order total degree Hermite polynomial chaos expansion (PCE) of Y is the representation of Y on the space spanned by products of Hermite polynomials with total degree up to p

$$\hat{Y}_p = \sum_{|\mathbf{k}| \leq p} a_{\mathbf{k}} \Psi_{\mathbf{k}}(\mathbf{X}), \quad (4.6)$$

where $|\mathbf{k}| = \sum_{i=1}^n k_i$. The total number of terms in Eq. (4.6) is

$$P = \binom{n+p}{p}. \quad (4.7)$$

The coefficients $a_{\mathbf{k}}$ are found by projecting $\mathcal{M}(\mathbf{X})$ on the space spanned by $\{\Psi_{\mathbf{k}}, |\mathbf{k}| \leq p\}$. According to the projection theorem, this is equivalent to minimizing the norm of the truncation error of the PCE representation $\|Y - \hat{Y}_p\|_{\mathcal{H}} = \mathbb{E}[(Y - \hat{Y}_p)^2]^{1/2}$. Using a set of samples $\mathcal{X} = \{\mathbf{x}^{(i)}, i = 1, \dots, N\}$ from the distribution of \mathbf{X} and corresponding model evaluations $\mathcal{Y} = \{y^{(i)} = \mathcal{M}(\mathbf{x}^{(i)}), i = 1, \dots, N\}$, one can estimate the coefficients $a_{\mathbf{k}}$ through minimizing a sample estimate of $\mathbb{E}[(Y - \hat{Y}_p)^2]$, i.e. through solving

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a}: \mathbf{a} \in \mathbb{R}^P} \frac{1}{N} \sum_{i=1}^N \left[y^{(i)} - \sum_{|\mathbf{k}| \leq p} a_{\mathbf{k}} \Psi_{\mathbf{k}}(\mathbf{x}^{(i)}) \right]^2. \quad (4.8)$$

Eq. (4.8) corresponds to an ordinary least squares (OLS) regression problem [7].

Alternative choices of the experimental design set \mathcal{X} can lead to a weighted least squares problem, e.g. [12]. As seen in Eq. (4.7), the total number of terms P in the PCE representation increases factorially with increase of either the dimension n or the polynomial degree p . A typical requirement for obtaining reasonable estimates of the regression coefficients in OLS is that $N > 2P$, e.g. [2, 40]. It is noted that the minimum number of experimental points N depends on the sampling scheme used and on the probability measure of \mathbf{X} , and could potentially be significantly larger than $2P$. In fact, for standard Monte Carlo sampling it has been shown that stable solutions are obtained at best for $N \sim \mathcal{O}(P \log(P))$ and in some cases for $N \sim \mathcal{O}(P^2)$ [12, 30].

Therefore, for a fixed number of experimental points N , the number of terms in the expansion that can be computed with accuracy through solution of Eq. (4.8) is limited.

To avoid over-fitting and obtain reliable predictions even for relatively large P , several adaptive approaches have been proposed for selecting the most significant terms in the PCE based on solving regularized regression problems [8, 38, 22]. These approaches result in sparse PCE representations. Although regression-based sparse PCEs perform well in moderate dimensional problems, their applicability in high dimensions ($n > 100$) is limited. This is due to the fact that they require the evaluation of the full set of multi-indices \mathbf{k} of the orthogonal polynomials as well as the assembly and storage of the corresponding Vandermonde matrix. In particular, adaptive methods evaluate the multi-indices for a typically high maximum polynomial degree. For the total degree construction of Eq. (4.6), this requires a huge computing and storage capability in dimensions > 100 , when considering that common algorithms for ordering the multi-indices scale exponentially with the dimension,

e.g. [6]. Here, we focus on a different approach for reducing the number of coefficients in the PCE that performs a linear coordinate transformation of the Gaussian parameter space.

4.2.1 Transformed PCE basis

Consider the following coordinate transformation:

$$\mathbf{Z} = \mathbf{Q}^T \mathbf{X} , \quad (4.9)$$

where \mathbf{Q} is an $n \times n$ orthogonal matrix, i.e. satisfying $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$. The columns of matrix \mathbf{Q} are orthonormal vectors and define a complete basis in \mathbb{R}^n . Therefore, Eq. (4.9) defines a projection of \mathbf{X} on the coordinate system defined by the columns of \mathbf{Q} . Due to the rotational symmetry of the independent standard Gaussian distribution of \mathbf{X} , any orthogonal transformation of \mathbf{X} will also be independent standard Gaussian. Therefore, the polynomials defined in Eq. (4.5) with argument \mathbf{Z} form a complete basis in the transformed space defined by the matrix \mathbf{Q} . The p -th order PCE representation of Y on the transformed space reads

$$\hat{Y}_p^{\mathbf{Q}} = \sum_{|\mathbf{k}| \leq p} b_{\mathbf{k}} \Psi_{\mathbf{k}}(\mathbf{Z}) = \sum_{|\mathbf{k}| \leq p} b_{\mathbf{k}} \Psi_{\mathbf{k}}(\mathbf{Q}^T \mathbf{X}) . \quad (4.10)$$

The representations of Eq. (4.10) and Eq. (4.6) are equivalent, i.e. it is possible to express the coefficients $b_{\mathbf{k}}$ in terms of $a_{\mathbf{k}}$ [42]. The representation of Eq. (4.10), introduced by [42], provides an additional flexibility in constructing sparse PCE representations through identifying dominant effects in the form of linear combinations of input variables. For example, if a dominant direction in \mathbf{X} -space is known, then one possible construction of matrix \mathbf{Q} is to set its first column equal to the dominant direction and determine the remaining columns by the Gram-Schmidt process. In such case, it might be possible to obtain an accurate representation of Y by only including the terms in Eq. (4.10) for which $\{k_i = 0, i = 2, \dots, n\}$. In [42], it is suggested to determine the dominant direction by estimating the low-order PCE coefficients. As an example, the direction defined by the linear PCE coefficients is $\mathbf{q}_1 \propto [a_{(1, \dots, 0)}; \dots; a_{(0, \dots, 1)}]$.

Consider now the case where one has identified a set of m dominant directions. We formalize the process of retaining only the terms in Eq. (4.10) with non-zero indices in \mathbf{k} for $i \leq m$ by defining a reduced orthogonal matrix \mathbf{Q}_m of dimensions $n \times m$, whose columns correspond to the dominant directions. The corresponding PCE representation reads

$$\hat{Y}_p^{\mathbf{Q}_m} = \sum_{|\mathbf{k}_m| \leq p} b_{\mathbf{k}_m} \Psi_{\mathbf{k}_m}(\mathbf{Q}_m^T \mathbf{X}) , \quad (4.11)$$

where $\mathbf{k}_m \in \mathbb{N}^m$. In the following section we discuss an approach that determines directions in the input space with high predictive ability based on a set of experimental points. In the subsequent section, we employ this approach to construct the matrix \mathbf{Q}_m and compute the corresponding coefficients $\{b_{\mathbf{k}_m}, \mathbf{k}_m \in \mathbb{N}^m\}$.

4.3 Partial least squares

Partial Least Squares (PLS) is a modelling technique that attempts to find relations between observable variables using latent variables [48, 21, 47]. This approach was originally developed in the

field of chemometrics, where it is often the case that the number of independent variables in an experimental setting is significantly larger than the number of data points, whereas the underlying process is driven by a small number of latent (not directly observable) variables. The basic idea of PLS is to find uncorrelated linear transformations of the original predictor variables that have high covariance with the response variables.

Let \mathcal{X} be an $N \times n$ matrix of samples from the input random vector \mathbf{X} and let \mathcal{Y} be the corresponding $N \times 1$ vector of model responses. It is convenient to assume that both \mathcal{X} and \mathcal{Y} are centered around their means; centering implies the operation $\mathcal{X} \leftarrow \mathcal{X} - \bar{\mathcal{X}}$, where $\bar{\mathcal{X}}$ denotes the arithmetic mean of \mathcal{X} . Standard PLS projects the matrix \mathcal{X} to latent components \mathbf{t}_i of dimensions $N \times 1$ by sequentially maximizing the covariance between the response \mathcal{Y} and the latent components. After determining each \mathbf{t}_i , it assumes a linear relationship between \mathbf{t}_i and \mathcal{Y} and evaluates the coefficient of \mathbf{t}_i by OLS.

The procedure starts by evaluating the projection to the first latent component $\mathbf{t}_1 = \mathcal{X}\mathbf{w}_1$, where \mathbf{w}_1 has dimensions $n \times 1$, by maximizing the covariance between \mathbf{t}_1 and \mathcal{Y} under the constraint that $\|\mathbf{w}_1\| = 1$. The corresponding optimization problem is stated as

$$\mathbf{w}_1 = \arg \max_{\mathbf{w}: \mathbf{w} \in \mathbb{R}^n, \|\mathbf{w}\|=1} \text{cov}(\mathcal{Y}^T \mathcal{X} \mathbf{w}, \mathcal{Y}^T \mathcal{X} \mathbf{w}) . \quad (4.12)$$

The exact solution of Eq. (4.12) is given by

$$\mathbf{w}_1 = \frac{\mathcal{X}^T \mathcal{Y}}{\|\mathcal{X}^T \mathcal{Y}\|} . \quad (4.13)$$

The regression coefficient of \mathbf{t}_1 is then evaluated by OLS as

$$b_1 = \frac{\mathbf{t}_1^T \mathcal{Y}}{\mathbf{t}_1^T \mathbf{t}_1} . \quad (4.14)$$

To obtain the next latent component, the residual matrices \mathcal{E} and \mathcal{F} for the regressor and response matrices, \mathcal{X} and \mathcal{Y} , respectively, are evaluated by subtracting from \mathcal{X} and \mathcal{Y} their rank-one approximations based on \mathbf{t}_1

$$\mathcal{E} = \mathcal{X} - \mathbf{t}_1 \mathbf{p}_1^T , \quad (4.15)$$

$$\mathcal{F} = \mathcal{Y} - b_1 \mathbf{t}_1 , \quad (4.16)$$

where \mathbf{p}_1 is the load vector corresponding to \mathbf{t}_1 and defines the projection of the rows of \mathcal{X} on the first latent component. It is

$$\mathbf{p}_1 = \frac{\mathcal{X}^T \mathbf{t}_1}{\mathbf{t}_1^T \mathbf{t}_1} . \quad (4.17)$$

The procedure is continued by extracting the next component from the deflated matrices \mathcal{E} and \mathcal{F} , until a certain error criterion is satisfied. The latter is usually based on estimates of the mean-square error of the PLS prediction. One possible approximation of this error is obtained through the norm of the residual $\|\mathcal{F}\|$. Alternatively, a more robust estimate can be derived based on cross-validation, e.g. [48]. The PLS process leads to a total of $m \leq n$ latent components or scores \mathbf{t}_i and corresponding weight vectors \mathbf{w}_i and load vectors \mathbf{p}_i . The PLS algorithm is summarized in Alg. 1.

Remark. *The deflation of the response vector \mathcal{Y} in step 9 of Alg. 1 is not required for computing the PLS weights, i.e. it will not influence the resulting components and regression coefficients [21].*

Algorithm 1 PLS algorithm

-
- 1: **Input** Data matrix \mathcal{X} and response matrix \mathcal{Y}
 - 2: Center matrices: $\mathcal{X} \leftarrow \mathcal{X} - \bar{\mathcal{X}}, \mathcal{Y} \leftarrow \mathcal{Y} - \bar{\mathcal{Y}}$
 - 3: Set $\mathcal{E} = \mathcal{X}, \mathcal{F} = \mathcal{Y}, i = 1$
 - 4: **repeat**
 - 5: Compute weight: $\mathbf{w}_i = \mathcal{E}^T \mathcal{F} / \|\mathcal{E}^T \mathcal{F}\|$
 - 6: Compute score: $\mathbf{t}_i = \mathcal{E} \mathbf{w}_i$
 - 7: Compute load: $\mathbf{p}_i = \mathcal{E}^T \mathbf{t}_i / (\mathbf{t}_i^T \mathbf{t}_i)$
 - 8: Compute regressor: $b_i = \mathbf{t}_i^T \mathcal{F} / \mathbf{t}_i^T \mathbf{t}_i$
 - 9: Deflate: $\mathcal{E} \leftarrow \mathcal{E} - \mathbf{t}_i \mathbf{p}_i^T, \mathcal{F} \leftarrow \mathcal{F} - b_i \mathbf{t}_i$
 - 10: $i \leftarrow i + 1$
 - 11: **until** change in $\|\mathcal{F}\|$ is smaller than ϵ_y
-

Remark. Alg. 1 is often termed *PLS1* in the literature to distinguish it from the PLS algorithm for multivariate outputs, termed *PLS2*.

The scores, weights and loads computed by the PLS algorithm, can be gathered in matrices $\mathbf{T} = [\mathbf{t}_1, \dots, \mathbf{t}_m] \in \mathbb{R}^{N \times m}$, $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_m] \in \mathbb{R}^{n \times m}$ and $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_m] \in \mathbb{R}^{n \times m}$. The PLS algorithm determines each latent component as a linear combination of columns of the corresponding residual matrix \mathcal{E} . However, it is also possible to express \mathbf{t}_i as linear combinations of the data matrix \mathcal{X} [20, 35]. Define the matrix $\mathbf{R} \in \mathbb{R}^{n \times m}$ as follows

$$\mathbf{R} = \mathbf{W} (\mathbf{P}^T \mathbf{W})^{-1} . \quad (4.18)$$

The PLS scores can be expressed by projecting \mathcal{X} on the space defined by the columns of matrix \mathbf{R} , i.e.

$$\mathbf{T} = \mathcal{X} \mathbf{R} . \quad (4.19)$$

The columns of matrix $\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_m]$ can also be obtained through the following recursive relation [21]

$$\begin{aligned} \mathbf{r}_1 &= \mathbf{w}_1 , \\ \mathbf{r}_i &= \mathbf{w}_i - \mathbf{r}_{i-1} (\mathbf{p}_{i-1}^T \mathbf{w}_i) . \end{aligned} \quad (4.20)$$

The approximation $\hat{\mathcal{X}}$ of the data matrix \mathcal{X} using m PLS components can be expressed using the load matrix \mathbf{P} as follows

$$\hat{\mathcal{X}} = \mathbf{T} \mathbf{P}^T . \quad (4.21)$$

4.3.1 Properties of the PLS matrices

PLS identifies dominant directions in the input space that can be potentially used within the context of PCE in transformed basis. In order to proceed, it is useful to review the properties of the matrices derived by the PLS process.

The PLS algorithm identifies orthogonal directions in the data space, i.e. it holds [21]

$$\mathbf{t}_i^T \mathbf{t}_j = 0 \text{ for } i \neq j . \quad (4.22)$$

This implies that the PLS scores form an orthogonal basis in the space generated by the columns of matrix \mathcal{X} . Eq. (4.22) is a crucial ingredient of the PLS approach; it enables evaluation of the regression coefficients b_i one by one, as in step 8 of Alg. 1.

The columns of matrices \mathbf{W} , \mathbf{P} and \mathbf{R} define directions in the space generated by the rows of matrix \mathcal{X} ; each direction corresponds to a linear combination of the underlying random variables \mathbf{X} . The matrix \mathbf{R} contains the reduced basis in \mathbf{X} -space that defines the PLS components, cf. Eq. (4.19).

The matrix \mathbf{W} is orthogonal, i.e. it holds $\mathbf{W}^T \mathbf{W} = \mathbf{I}$ [21], and hence its columns form an orthonormal basis in the \mathbf{X} -space. However, the same cannot be said for matrices \mathbf{P} and \mathbf{R} . Therefore, if the PLS algorithm is applied not for prediction purposes but to determine orthogonal dominant directions then the orthogonal projection matrix \mathbf{W} is often the desired output, e.g. [54]. The matrices \mathbf{P} and \mathbf{R} are not necessarily orthogonal, but their columns are mutually orthogonal, i.e. it holds $\mathbf{P}^T \mathbf{R} = \mathbf{I}$ [35].

In the context of PCE representations in transformed basis, it is possible to define the reduced basis using the orthogonal directions of matrix \mathbf{W} , i.e. setting $\tilde{\mathbf{Q}}_m = \mathbf{W}$ in Eq. (4.11). Projecting the data on the columns of matrix \mathbf{W} would lead to scores $\tilde{\mathbf{T}} = \mathcal{X} \mathbf{W}$, that differ from the PLS scores given by Eq. (4.19). This is somewhat suboptimal as the PLS components are determined based on maximizing the covariance with the prediction error obtained from regressing the PLS scores \mathbf{T} (and not $\tilde{\mathbf{T}}$) with \mathcal{Y} . Moreover, unlike the PLS scores \mathbf{T} , the scores $\tilde{\mathbf{T}}$ are not necessarily orthogonal.

As will become clear in the next section, employing directly the directions of the PLS components, i.e. the columns of matrix \mathbf{R} , to define the PCE reduced basis is of particular benefit; it allows estimating simultaneously the reduced basis and PCE coefficients. However, as the matrix \mathbf{R} is in general not orthogonal, the projection of the random variables \mathbf{X} on the columns of \mathbf{R} will not be independent standard Gaussian. This poses a problem to the PCE representation of Eq. (4.11): the Hermite polynomial basis of Eq. (4.5) will not be orthogonal in the resulting transformed space. That is, if the matrix \mathbf{R} is not orthogonal then the orthogonal polynomial basis need to be determined for the problem at hand, e.g. by application of the Gram-Schmidt process.

Next we look more closely at the properties of matrix \mathbf{R} and determine a condition under which this matrix becomes orthogonal.

Lemma 4.3.1. *The matrix $\mathbf{R}^T \mathbf{W}$ is a lower triangular matrix with unit diagonal elements.*

Proof. The proof follows from the recursive relation of Eq. (4.20). Eq. (4.20) can be rewritten as follows

$$\mathbf{r}_i = \mathbf{w}_i + \sum_{j=1}^{i-1} c_j \mathbf{w}_j ,$$

where $c_j \in \mathbb{R}$ are coefficients, which are not used here. This gives

$$\mathbf{r}_i^T \mathbf{w}_k = 0 \text{ for } i < k$$

and

$$\mathbf{r}_i^T \mathbf{w}_i = 1 ,$$

where we have used that $\mathbf{w}_i^T \mathbf{w}_j = \delta_{ij}$. □

We now consider the special case where the columns of the centered data matrix \mathcal{X} have equal norm and are mutually orthogonal, i.e.

$$\mathcal{X}^T \mathcal{X} = \omega \mathbf{I}, \quad (4.23)$$

where $\omega \in \mathbb{R}_{>0}$ and \mathbf{I} is the $n \times n$ identity matrix.

Proposition 4.3.1. *If the data matrix \mathcal{X} satisfies Eq. (4.23), then the matrix \mathbf{R} is orthogonal, i.e. $\mathbf{R}^T \mathbf{R} = \mathbf{I}$.*

Proof. From Eq. (4.22) and expressing each PLS score as $\mathbf{t}_i = \mathcal{X} \mathbf{r}_i$, we get

$$\mathbf{r}_i^T (\mathcal{X}^T \mathcal{X}) \mathbf{r}_j = 0 \text{ for } i \neq k.$$

Using Eq. (4.23), we obtain

$$\mathbf{r}_i^T \mathbf{r}_j = 0 \text{ for } i \neq k.$$

From Eq. (4.20) and using lemma 4.3.1, we have

$$\mathbf{r}_i^T \mathbf{r}_i = \mathbf{r}_i^T \mathbf{w}_i - \mathbf{r}_i^T \mathbf{r}_{i-1} (\mathbf{p}_{i-1}^T \mathbf{w}_i) = \mathbf{r}_i^T \mathbf{w}_i = 1.$$

□

Proposition 4.3.1 implies that if the pair-wise sample correlation of the underlying random variables vanishes and the sample variances of all variables are equal, then the directions identified by PLS, the columns of \mathbf{R} , will be orthogonal and will have unit length. One interpretation of this result is that if the underlying variable space is uncorrelated and has equal variances, then the directions identified by PLS should contribute equally to the explained variance of \mathcal{X} .

Although Eq. (4.23) is not necessarily true, under assumption 1 it is asymptotically true as $N \rightarrow \infty$. Moreover, one can generate \mathcal{X} through sampling techniques that aim at approximately satisfying Eq. (4.23) for finite N . This is achieved by stratification techniques such as Latin hypercube sampling (LHS) [29], or a modified version that aims at minimizing the correlation between samples, e.g. [33].

4.4 PLS-based PCE

We now discuss how the PLS approach can be used in the context of the PCE representation in transformed basis. We discuss two versions of PLS-based PCE representations; the first is based on the linear PLS algorithm presented in Section 4.3 and the second is based on a nonlinear version of the PLS algorithm tailored for use with PCE models.

4.4.1 Linear PLS-based transformation

As mentioned in the previous section, it is possible to build a PCE representation of Y by selecting $Q_m = \mathbf{W}$ or \mathbf{R} in Eq. (4.11). The matrix of PLS weights \mathbf{W} is orthogonal, while the matrix \mathbf{R} defining the PLS components in \mathbf{X} -space is asymptotically orthogonal as $N \rightarrow \infty$. Having

determined the PLS directions with samples $\mathcal{X} = \{\mathbf{x}^{(i)}, i = 1, \dots, N\}$ from the distribution of \mathbf{X} and corresponding model evaluations $\mathcal{Y} = \{y^{(i)} = \mathcal{M}(\mathbf{x}^{(i)}), i = 1, \dots, N\}$, the same samples can be used to determine the PCE coefficients $\{b_{\mathbf{k}_m}, \mathbf{k}_m \in \mathbb{N}^m\}$ in Eq. (4.11) by OLS regression.

We note that for the first PLS component it holds $\mathbf{r}_1 = \mathbf{w}_1$, cf. Eq. (4.20). We make the following observation.

Proposition 4.4.1. *If the reduced basis is determined by a single direction ($m = 1$) as computed by Alg. 1, then this direction is asymptotically equivalent to the direction defined by the first-order PCE coefficients.*

Proof. The first linear PLS direction \mathbf{r}_1 is obtained as follows

$$\mathbf{r}_1 \propto \sum_{i=1}^N (\mathbf{x}^{(i)} - \bar{\mathcal{X}})(y^{(i)} - \bar{\mathcal{Y}}),$$

where $\bar{\mathcal{X}}$ and $\bar{\mathcal{Y}}$ are the arithmetic means of $\{\mathbf{x}^{(i)}, i = 1, \dots, N\}$ and $\{y^{(i)}, i = 1, \dots, N\}$, respectively. Under assumption 1, $\bar{\mathcal{X}} \rightarrow 0$ as $N \rightarrow \infty$. Therefore as $N \rightarrow \infty$, it is

$$\mathbf{r}_1 \propto \sum_{i=1}^N \mathbf{x}^{(i)} y^{(i)}.$$

The vector of first-order PCE coefficients \mathbf{a}_1 is obtained by projecting $Y = \mathcal{M}(\mathbf{X})$ on the linear Hermite polynomials $\{\Psi_{\mathbf{k}}, |\mathbf{k}| = 1\} = \{X_1, \dots, X_n\}$. It is

$$\mathbf{a}_1 = \mathbb{E}[\mathcal{M}(\mathbf{X})\mathbf{X}] = \mathbb{E}[Y\mathbf{X}].$$

As the samples $\mathbf{x}^{(i)}$ and $y^{(i)}$ follow respectively the distribution of \mathbf{X} and Y , the PLS direction \mathbf{r}_1 converges to the one defined by \mathbf{a}_1 as $N \rightarrow \infty$. \square

Proposition 4.4.1 implies that the linear PLS algorithm with $m = 1$ is asymptotically equivalent to the linear PCE-driven Gaussian adaptation proposed in Section 3.1 of [42].

The linear PLS algorithm does not always result in an optimal coordinate transformation because the matrices \mathbf{W} and \mathbf{R} are evaluated assuming a linear relationship between the response and each latent variable. Next we discuss an approach that adapts PLS for obtaining directions that are optimal for use in the representation of Eq. (4.11).

4.4.2 PCE-driven PLS-based transformation

Several variants of the classical PLS method have been proposed for addressing problems where the underlying process is nonlinear [37]. Here, we employ the approach proposed by [46] and later modified by [4]. This approach was originally proposed for quadratic models and later extended for use with neural networks [5]. Here, we adapt this approach for use in PCE representations.

The nonlinear PLS algorithm proceeds by obtaining a first approximation of each projection vector \mathbf{w}_i with standard PLS. Thereafter, it assumes a nonlinear relationship between the response and the

latent variable \mathbf{t}_i , which is fitted by OLS. Within the context of PCE representations, the nonlinear relationship in each latent variable is a one-dimensional Hermite polynomial expansion of order p

$$\hat{\mathcal{M}}_i^p(t) = \sum_{j=1}^p \hat{b}_{ij}^p \psi_j(t) . \quad (4.24)$$

For the first latent variable, the PCE regression problem is stated as

$$\mathcal{Y} = \sum_{j=1}^p b_{1j}^p \psi_j(\mathbf{t}_1) + \mathbf{e} , \quad (4.25)$$

wherein the operations are performed element-wise, $\mathbf{t}_1 = \mathcal{X}\mathbf{w}_1$ and \mathbf{e} is the vector of regression errors. Eq. (4.25) is solved for the coefficients $\{b_{1j}^p\}$. The vector \mathbf{w}_1 is then modified iteratively by means of a Newton-Raphson linearisation of Eq. (4.25), i.e. by performing a first-order Taylor series expansion of Eq. (4.25) with respect to \mathbf{w}_1 and then solving it for the increment $\Delta\mathbf{w}_1$. This gives

$$\Delta\mathbf{w}_1 = (\mathbf{A}^T \mathbf{A})^{-} \mathbf{A}^T \mathbf{e} , \quad (4.26)$$

where $(\cdot)^{-}$ denotes the generalized inverse of a matrix and \mathbf{A} is the gradient of the PCE model with respect to the weights $\mathbf{A} = \nabla_{\mathbf{w}} \hat{\mathcal{M}}_1^p(\mathcal{X}\mathbf{w})$. Thereafter, \mathbf{w}_1 is updated, $\mathbf{w}_1 = \mathbf{w}_1 + \Delta\mathbf{w}_1$, and normalized. The latent component \mathbf{t}_1 is then updated, Eq. (4.25) is fitted anew and the next increment $\Delta\mathbf{w}_1$ is evaluated. This iterative procedure is continued until $\Delta\mathbf{w}_1$ is sufficiently small.

To obtain the next latent component, the residual matrices \mathcal{E} and \mathcal{F} are evaluated by subtracting from \mathcal{X} its rank-one approximations based on \mathbf{t}_1 and from \mathcal{Y} its PCE approximation using the first direction $\hat{\mathcal{M}}_1^p(\mathbf{t}_1)$ and the same process is repeated using \mathcal{E} and \mathcal{F} as the new \mathcal{X} and \mathcal{Y} .

In order to obtain PLS directions that reflect the nonlinear nature of the underlying process while avoiding over-fitting, we choose the polynomial degree in Eq. (4.25) for each latent variable by evaluating each latent component for different polynomial degrees $q = \{1, \dots, p\}$ and retaining the one that results in the smallest modified leave-one-out error ϵ_{LOO}^q ; ϵ_{LOO}^q can be evaluated based on a single PCE built using the Vandermonde matrix of the OLS problem [10, 8]. The PCE-driven PLS algorithm is detailed in Alg. 2.

Remark. *The matrix \mathbf{A} required in step 13 of Alg. 2 can be computed analytically using the properties of the derivatives of the Hermite polynomials [1] and hence does not require additional model evaluations.*

The algorithm returns m quintuples $\{q_i, \hat{\mathcal{M}}_i^{q_i}, \mathbf{t}_i^{q_i}, \mathbf{w}_i^{q_i}, \mathbf{p}_i^{q_i}\}$, with q_i denoting the polynomial degree, $\hat{\mathcal{M}}_i^{q_i}$ the one-dimensional PCE representation, $\mathbf{t}_i^{q_i}$ the scores, $\mathbf{w}_i^{q_i}$ the weights and $\mathbf{p}_i^{q_i}$ the loads of the i -th PLS component. The PLS directions $\mathbf{r}_i^{q_i}$ can then be evaluated through Eq. (4.18) or Eq. (4.20). Using the one-dimensional fitted PCEs $\hat{\mathcal{M}}_i^{q_i}$ and the PLS directions $\mathbf{r}_i^{q_i}$, we obtain the following representation

$$\hat{Y}_m^{\text{PLS}} = \hat{\mathcal{M}}_m(\mathbf{X}) = b_0 + \sum_{i=1}^m \hat{\mathcal{M}}_i^{q_i} \left[(\mathbf{r}_i^{q_i})^T \tilde{\mathbf{X}} \right] , \quad (4.27)$$

where $b_0 = \bar{\mathcal{Y}}$ and $\tilde{\mathbf{X}} = \mathbf{X} - \bar{\mathcal{X}}$. Due to the asymptotic behavior of the matrix $\mathbf{R} = [\mathbf{r}_1^{q_1}, \dots, \mathbf{r}_m^{q_m}]$, described in Proposition 4.3.1, and because $\tilde{\mathbf{X}} \rightarrow \mathbf{X}$ as $N \rightarrow \infty$, the representation of Eq. (4.27) is

Algorithm 2 PCE-driven PLS algorithm

- 1: **Input** Data matrix \mathcal{X} and response matrix \mathcal{Y}
 - 2: Center matrices: $\mathcal{X} \leftarrow \mathcal{X} - \bar{\mathcal{X}}, \mathcal{Y} \leftarrow \mathcal{Y} - \bar{\mathcal{Y}}$
 - 3: Set $\mathcal{E} = \mathcal{X}, \mathcal{F} = \mathcal{Y}, i = 1$
 - 4: **repeat**
 - 5: Compute weight: $\mathbf{w}_i^0 = \mathcal{E}^T \mathcal{F} / \|\mathcal{E}^T \mathcal{F}\|$
 - 6: **for** $q \leftarrow 1, p$ **do**
 - 7: Set $\mathbf{w}_i^q = \mathbf{w}_i^0$
 - 8: **repeat**
 - 9: Compute score: $\mathbf{t}_i^q = \mathcal{E} \mathbf{w}_i^q$
 - 10: Fit a 1D PCE of order q : $\hat{\mathbf{b}}_i^q \leftarrow \text{fit} \left[\mathcal{F} = \sum_{j=1}^q b_{ij}^q \psi_j(\mathbf{t}_i^q) + \epsilon \right]$
 - 11: Set $\hat{\mathcal{M}}_i^q(t) = \sum_{j=1}^q \hat{b}_{ij}^q \psi_j(t)$
 - 12: Compute the error: $\hat{\mathcal{F}} = \hat{\mathcal{M}}_i^q(\mathbf{t}_i^q); \mathbf{e} = \mathcal{F} - \hat{\mathcal{F}}$
 - 13: Compute: $\Delta \mathbf{w}_i^q = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{e}$ with $\mathbf{A} = \nabla_{\mathbf{w}} \hat{\mathcal{M}}_i^q(\mathcal{E} \mathbf{w})$
 - 14: Set: $\mathbf{w}_i^q \leftarrow \mathbf{w}_i^q + \Delta \mathbf{w}_i^q$
 - 15: Normalize: $\mathbf{w}_i^q \leftarrow \mathbf{w}_i^q / \|\mathbf{w}_i^q\|$
 - 16: **until** $\|\Delta \mathbf{w}_i^q\|$ is smaller than ϵ_w
 - 17: Evaluate the relative leave-one-out error ϵ_{LOO}^q as in [8]
 - 18: Set $\{q_i, \hat{\mathcal{M}}_i^{q_i}, \mathbf{w}_i^{q_i}\}$ as the triple $\{q, \hat{\mathcal{M}}_i^q, \mathbf{w}_i^q\}$ with the smallest ϵ_{LOO}^q
 - 19: Compute score: $\mathbf{t}_i^{q_i} = \mathcal{E} \mathbf{w}_i^{q_i}$
 - 20: Compute load: $\mathbf{p}_i^{q_i} = \mathcal{E}^T \mathbf{t}_i^{q_i} / ((\mathbf{t}_i^{q_i})^T \mathbf{t}_i^{q_i})$
 - 21: Deflate: $\mathcal{E} \leftarrow \mathcal{E} - \mathbf{t}_i^{q_i} (\mathbf{p}_i^{q_i})^T, \mathcal{F} \leftarrow \mathcal{F} - \hat{\mathcal{M}}_i^{q_i}(\mathbf{t}_i^{q_i})$
 - 22: $i \leftarrow i + 1$
 - 23: **until** change in $\|\mathcal{F}\|$ is smaller than ϵ_y
 - 24: **return** $\{q_i, \hat{\mathcal{M}}_i^{q_i}, \mathbf{t}_i^{q_i}, \mathbf{w}_i^{q_i}, \mathbf{p}_i^{q_i}\}, i = 1 \dots, m$.
-

asymptotically equivalent to the PCE representation of Eq. (4.11) with $\mathbf{Q}_m = \mathbf{R}$, for the case where only the main-effects in the transformed coordinate system are considered.

The PCE-driven PLS algorithm identifies simultaneously the PLS directions and the coefficients of the one-dimensional PCEs in each PLS component. A disadvantage of this approach is that the matrix \mathbf{R} defining the PLS components is only asymptotically orthogonal. This implies that for finite N the multivariate Hermite polynomials are not orthogonal with respect to the distribution of the latent variables and hence the polynomial basis loses its optimality. However, as mentioned in Subsection 4.3.1, tailored sampling techniques can be used to obtain approximately orthogonal PLS directions for finite N .

An alternative approach would be to employ the orthogonal matrix $\mathbf{W} = [\mathbf{w}_1^{q_1}, \dots, \mathbf{w}_m^{q_m}]$ obtained from Alg. 2 to define the orthogonal projection. In such case, the PCE coefficients need to be evaluated anew through setting $\mathbf{Q}_m = \mathbf{W}$ and regressing Eq. (4.11) with the responses \mathcal{Y} .

4.5 Examples

In this section, we evaluate the proposed method with three numerical examples in high dimensions. We investigate the performance of both the linear PLS-based approach of Subsection 4.4.1 and the PCE-driven PLS algorithm of Subsection 4.4.2. We compare the linear PLS approach with a single latent component to the linear PCE-driven Gaussian adaptation of Tipireddy and Ghanem (TG) [42], computed with a quadrature-based pseudo-spectral projection, to numerically verify Proposition 4.4.1. We employ a sufficient number of quadrature points in the TG approach to ensure accuracy of the result. We note that the proposed PLS methods are based on model evaluations at a set of samples from the distribution of the input variables, whereas the TG approach applies numerical quadrature to evaluate the PCE coefficients. Hence, a direct comparison of the computational cost of PLS methods with the TG approach would be difficult to set up¹. Instead, we compare the performance of the PCE-driven PLS method against polynomial-basis low-rank approximations (LRA), which can be constructed based on the same set of model evaluations. We use the LRA implementation of UQlab [27], which employs alternating least-squares to fit the LRA with an adaptive scheme for the rank selection while considering every polynomial order up to p within the selected ranks. We choose to compare our method to the particular implementation of LRA, as it has been shown that it performs better than sparse PCEs in moderate dimensional problems and small experimental designs [25]. In addition, the computational cost of building an LRA in high dimensions is feasible as its construction is based on products of univariate polynomial expansions. In contrast, sparse PCEs, e.g. based on compressive sensing, require evaluation and storage of the full Vandermonde matrix for high polynomial orders, which is prohibitive in very high dimensions (see the relevant discussion in Section 4.2). For all methods, we choose a maximum polynomial degree of $p = 10$ and for the LRA a maximum rank of $R = 10$.

The reduced basis identified by the columns of the \mathbf{R} -matrix, $\{\mathbf{r}_i^{q_i}, i = 1, \dots, m\}$, need not be orthogonal for finite sample size N ; this compromises the optimality of the Hermite PCE basis as the basis will not be orthogonal in the transformed space. In order to quantify this error, we consider

¹The first direction obtained by the linear PLS-based PCE can be viewed as a Monte Carlo approximation of the linear PCE-based transformation of TG. Therefore, the linear PLS-based approach is itself a way of approximating TG.

the Gramian of \mathbf{R}

$$\mathbf{G}(\mathbf{R}) = \mathbf{R}^T \mathbf{R} . \quad (4.28)$$

If \mathbf{R} is orthogonal, then it is $\mathbf{G}(\mathbf{R}) = \mathbf{I}$. We therefore use $\|\mathbf{I} - \mathbf{G}(\mathbf{R})\|_2$ as the error measure, where $\|\cdot\|_2$ denotes the 2-norm of a matrix.

We compare the proposed method against LRA in terms of the errors in the mean and variance of the output quantity of interest as well as the generalization error err_G . The latter is defined as the mean-square of the residual

$$\text{err}_G = \text{E}[(Y - \hat{Y})^2] , \quad (4.29)$$

where \hat{Y} denotes the output response of the surrogate model. An estimate $\widehat{\text{err}}_G$ of err_G can be obtained using a large set of samples \mathcal{X}_{val} , termed validation set.

The experimental designs for both PLS-PCE and LRA are generated via LHS with sample decorrelation, using the built-in Matlab function `lhsdesign` with correlation criteria; this function iteratively generates samples with LHS to find the ones with the smaller sample correlation. For each example, the analysis is performed 100 times to obtain confidence intervals (CI) on the predictive quantities. Reference solutions are obtained with Monte Carlo simulation with 2×10^5 samples.

4.5.1 Linear elastic bar

The first example consists of a linear elastic bar of length $L = 1\text{m}$, as shown in Fig. 4.1. The displacement of the bar $u(x)$ satisfies the following differential equation

$$-\frac{d}{dx} \left(D(x) \frac{d}{dx} u \right) = q(x) \text{ in } [0, L] . \quad (4.30)$$

The axial resistance of the bar $D(x) = EA(x)$ is described by a homogeneous random field with

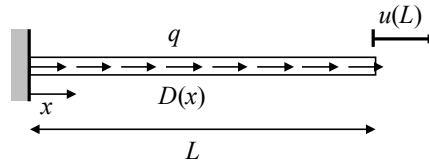


Figure 4.1: Linear elastic bar with random axial rigidity.

lognormal marginal distribution with mean $\mu_D = 100\text{kN}$ and standard deviation $\sigma_D = 10\text{kN}$. The autocorrelation function of the underlying Gaussian random field $\ln D$ is $\rho_{\ln D}(\Delta x) = \exp(-|\Delta x|/l)$ with correlation length $l = 0.04\text{m}$. The random field $\ln D$ is represented by a Karhunen-Loève (KL) expansion [18] with 100 terms, which captures 95% of the variability of $\ln D$. This leads to an input random vector consisting of 100 independent standard Gaussian random variables. The bar is subjected to a deterministic load $q = 1\text{kN/m}$. Eq. (4.30) is solved by the finite element method with 100 piecewise linear finite elements. The output quantity is the displacement at the tip of the bar $Y = u(L)$. Fig. 4.2 compares the log-densities obtained with the linear PLS-based PCE with $m = 1$ for increasing experimental design sizes to the one obtained with the linear PCE-based adaptation of TG. The latter compares well with the reference solution computed with direct Monte Carlo, which

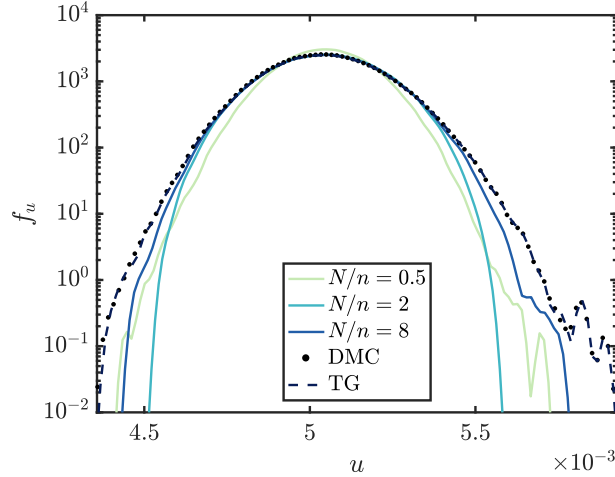


Figure 4.2: Linear elastic bar: log-densities of response. Comparison of linear PLS-PCE (with $m = 1$) with various design sizes with the linear PCE-based adaptation of Tipireddy and Ghanem (TG) [42] and direct Monte Carlo (DMC) using 2×10^5 samples.

implies that a single latent component suffices to describe the behavior of the model response. It is also shown that the linear PLS-based PCE approaches the solution of TG as the number of samples N increases, which verifies Proposition 4.4.1. Fig. 4.3 compares the generalization error obtained

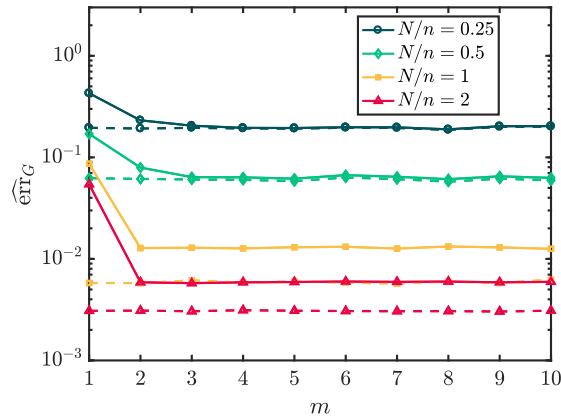


Figure 4.3: Linear elastic bar: comparison of the linear PLS-based (solid line) and nonlinear (PCE-driven) PLS-based (dashed line) adaptation at various design sizes and number of latent components.

with the linear and PCE-driven PLS methods for increasing number of latent components m . It is seen that the PCE-driven PLS method using a single latent component yields consistently lower errors than the linear PLS, whereas using additional components does not improve the results. This is to be expected as the nonlinear PLS algorithm employed within the PCE-driven PLS approach identifies the directions that minimize the residual in the PCE approximation. As the linear PLS assumes a linear relationship between input and output and, hence, is not optimized for use with higher order PCEs, it requires more latent components to capture the behavior of the model.

We now compare the performance of the nonlinear PLS-PCE method with the LRA surrogate based on the same experimental design scheme. Both surrogates capture the response PDF increasingly well as the number of points in the experimental design rises (Fig. 4.4). The PLS-based PCE

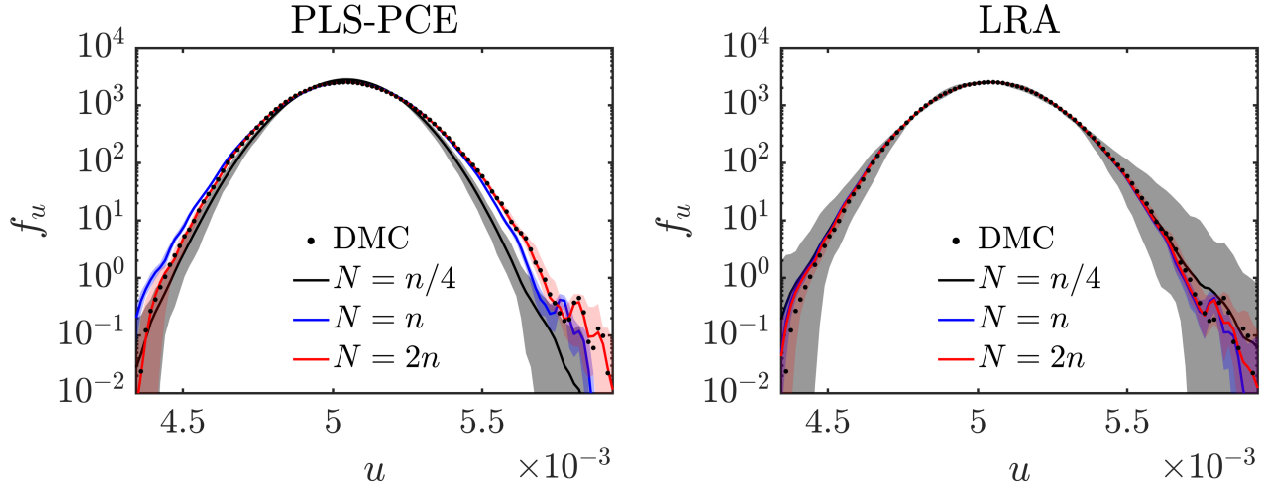


Figure 4.4: Linear elastic bar: log-densities of response at various design sizes with 95 % CI. Results of (nonlinear) PLS-PCE and LRA are compared with direct Monte Carlo (DMC) results using 2×10^5 samples.

model exhibits a bias at low experimental designs, which vanishes as the size of experimental design increases. The variability of the PDF estimates obtained by the LRA model is in general larger than the one of the PLS-based PCE model. Fig. 4.5 shows the one-dimensional PCE model along the first latent direction. The nonlinearity (polynomial degree) of the PLS-based PCE model in general increases with the increase of the experimental design size. At $N/n = 2$, the PLS-based PCE model exhibits good convergence while the LRA model still produces occasional outliers (Fig. 4.6). In terms of generalization error, the LRA error is slightly smaller for small experimental designs ($N/n \leq 0.5$). The PLS-based PCE error continuously decreases as we add points to the experimental design while the LRA error seems to stagnate from $N/n = 1$ on (Fig. 4.7, top left). Both methods yield virtually identical mean errors, while the variance error exhibits similar behaviour as the generalization error (Fig. 4.7, bottom right & left). Across all N , the PLS-based PCE reduced space is constructed with $m = 2$ directions, which exhibit relatively low orthogonality error (Fig. 4.7, top right).

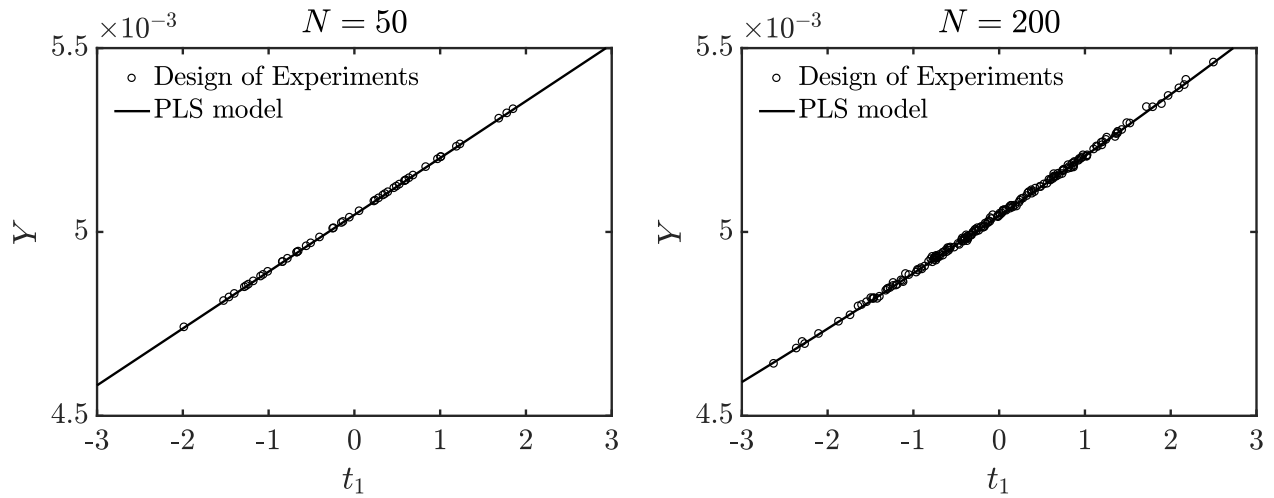


Figure 4.5: Linear elastic bar: 1-D Surrogate model along first PLS direction t_1 vs. the design of experiments at $N/n = 0.5$ (left) and $N/n = 2$ (right).

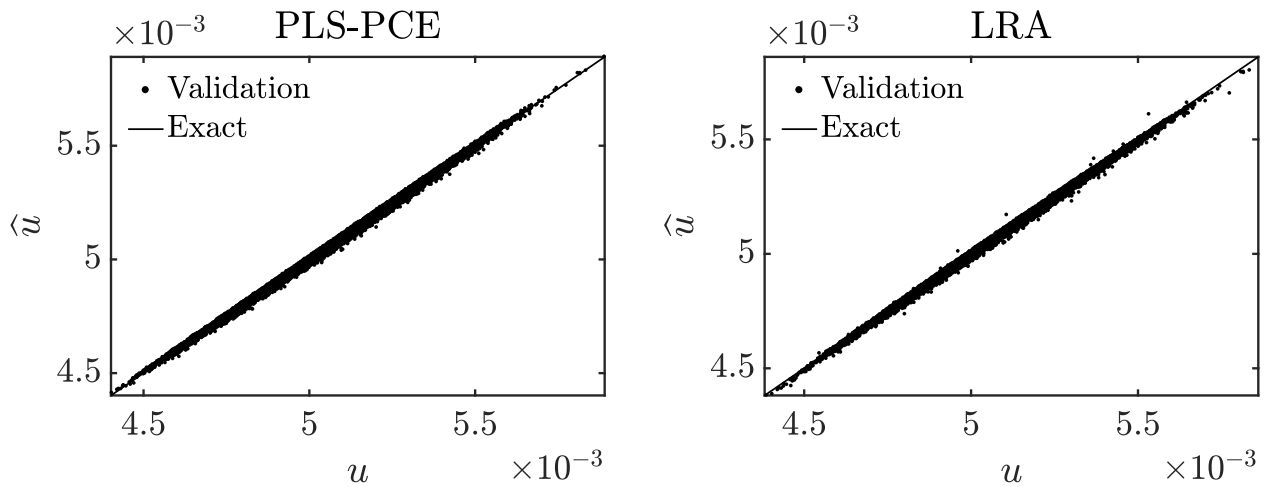


Figure 4.6: Linear elastic bar: scatter plots of (nonlinear) PLS-based PCE (left) and LRA (right) response vs original response at $N/n = 2$ using a validation set of 2×10^5 samples.

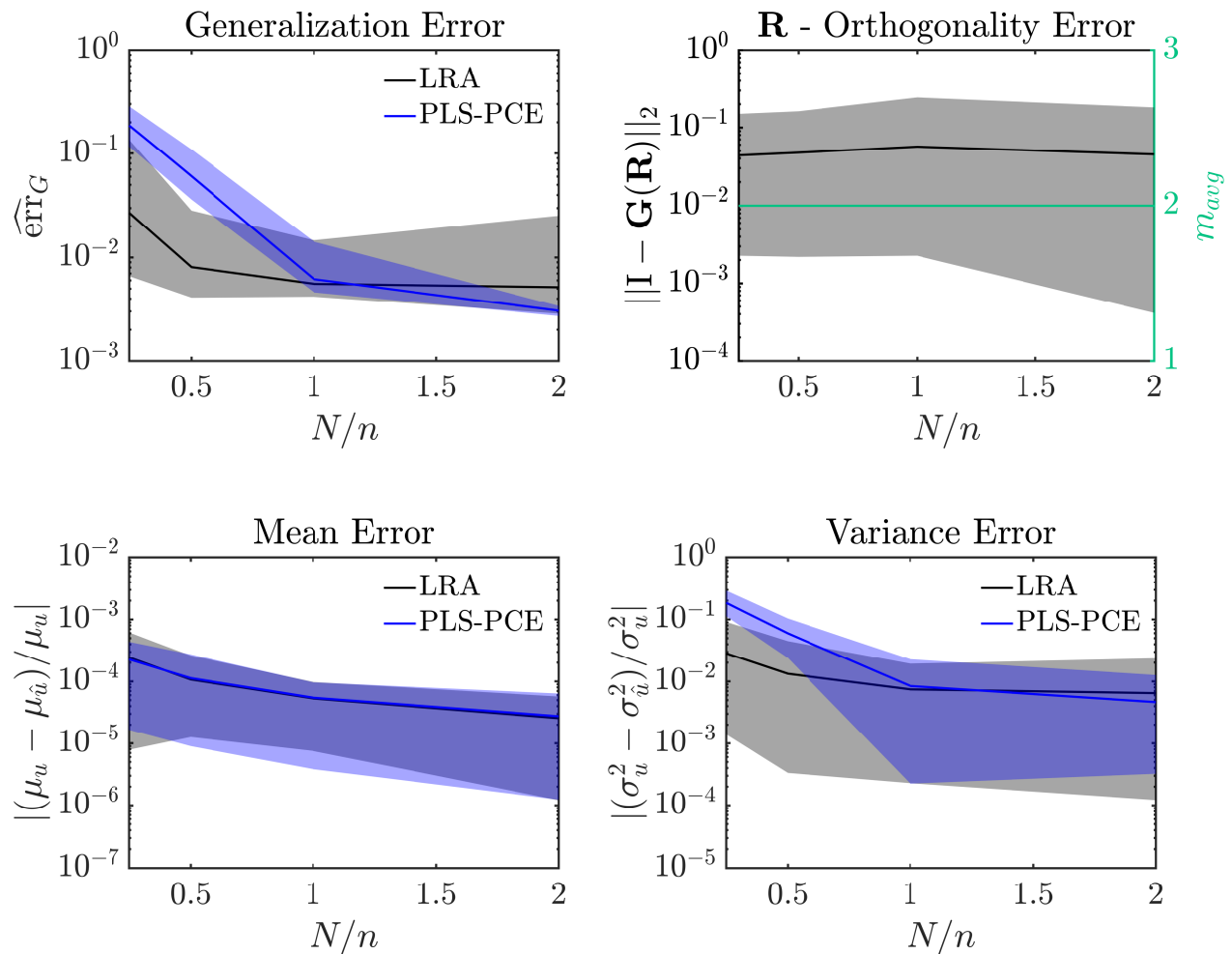


Figure 4.7: Linear elastic bar: error measures with 95 % confidence intervals obtained from 100 repeated analyses with different experimental designs; reference solution obtained with 2×10^5 samples. The top right panel includes the average number of latent components (PLS directions) m_{avg} .

4.5.2 Nonlinear oscillator

The second example, adapted from [45], is a hysteretic oscillator under random loading, defined by the following differential equation:

$$m\ddot{u}(t) + c\dot{u}(t) + k[\alpha u(t) + (1 - \alpha)u_y z(t)] = f(t) , \quad (4.31)$$

where $u(t)$, $\dot{u}(t)$ and $\ddot{u}(t)$ denote the displacement, velocity and acceleration of the oscillator. The mass, stiffness and damping of the oscillator are $m = 6 \times 10^4 \text{kg}$, $k = 5 \times 10^6 \text{N/m}$, $c = 2m\zeta\sqrt{k/m}$ with $\zeta = 5\%$, and the yielding displacement is $u_y = 0.04 \text{m}$. The parameter α , which controls the degree of hysteresis is set to $\alpha = 0.1$. The parameter $z(t)$ follows the Bouc-Wen hysteresis law

$$\dot{z}(t) = \frac{1}{u_y} [A\dot{u}(t) - \beta|\dot{u}(t)||z(t)|^{\bar{n}-1}z(t) - \gamma\dot{u}(t)|z(t)|^{\bar{n}}] , \quad (4.32)$$

with $\beta = \gamma = 0.5$, $A = 1$ and $\bar{n} = 3$. The loading $f(t)$ is a seismic load process modelled by a white noise ground acceleration and discretized in the frequency domain as follows [39]

$$f(t) = -m\sigma \sum_{i=1}^{n/2} [X_i \cos(\omega_i t) + X_{(n/2+i)} \sin(\omega_i t)] , \quad (4.33)$$

where $X_i, i = 1, \dots, n$, are independent standard Gaussian random variables, $\omega_i = i\Delta\omega$, $\Delta\omega = 30\pi/n$ (the cut-off frequency is $\omega_{\text{cut}} = 15\pi$) and $\sigma = \sqrt{2S\Delta\omega}$, where $S = 0.03 \text{m}^2/\text{s}^3$ is the intensity of the white noise. We use $n = 300$ terms in Eq. (4.33), which leads to an input random vector \mathbf{X} of dimension 300. We are interested in approximating the displacement of the oscillator at $t = 8\text{s}$, $u(8\text{s})$. As in example 1, convergence of the linear PLS-based PCE with $m = 1$ to the

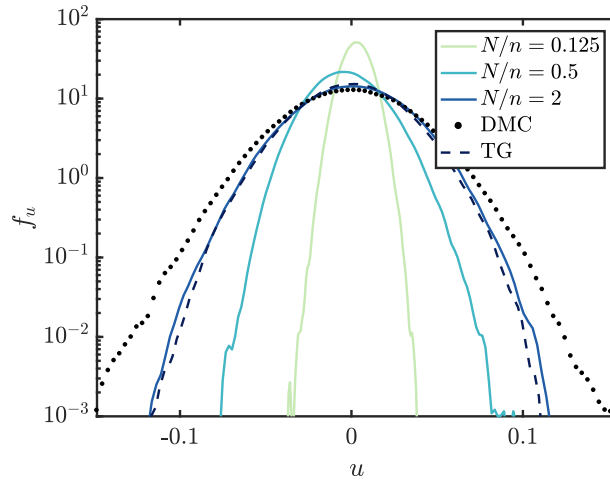


Figure 4.8: Nonlinear oscillator: log-densities of response. Comparison of linear PLS-PCE (with $m = 1$) with various design sizes with the linear PCE-based adaptation of Tipireddy and Ghanem (TG) [42] and direct Monte Carlo (DMC) using 2×10^5 samples.

TG solution is again observed. However, here the TG approach with a single linear PCE-based component cannot capture the behavior of the nonlinear oscillator (Fig. 4.8). Evidently, the strong nonlinearity present in the hysteretic oscillator model is more challenging to both PLS-based PCE and LRA compared to the bar example. In particular, we were unable to obtain non-diverging

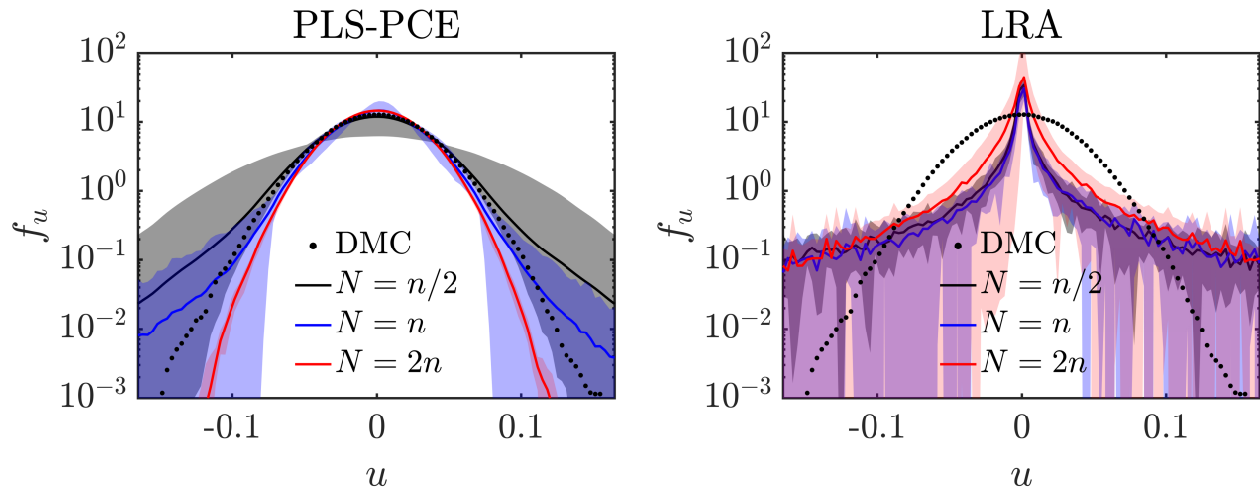


Figure 4.9: Nonlinear oscillator: log-densities of response at various design sizes with 95 % CI. Results of (nonlinear) PLS-PCE and LRA are compared with direct Monte Carlo (DMC) results using 2×10^5 samples.

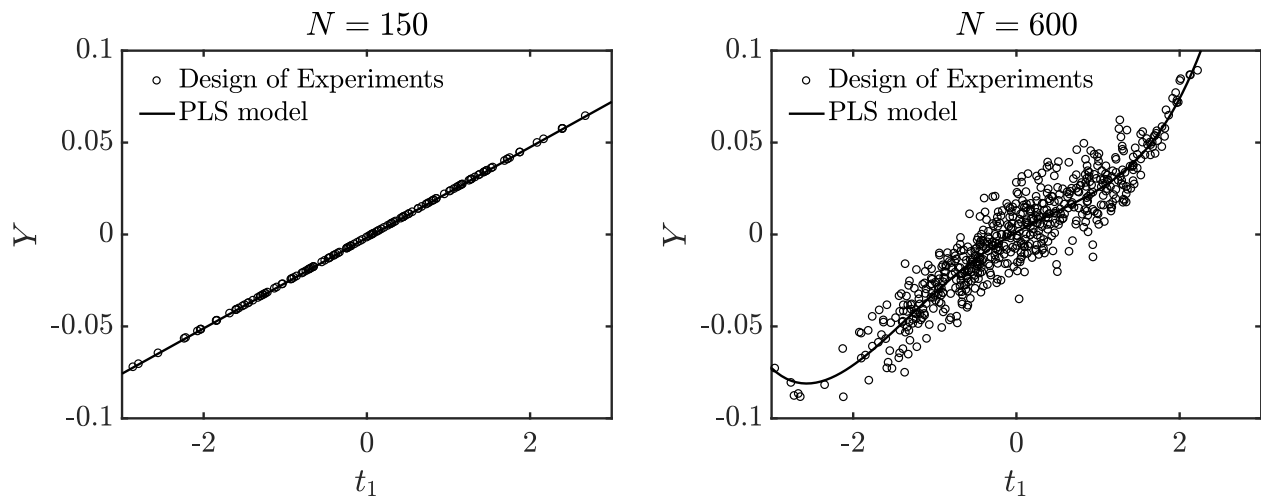


Figure 4.10: Nonlinear oscillator: 1-D Surrogate model along first PLS direction t_1 vs. the design of experiments at $N/n = 0.5$ (left) and $N/n = 2$ (right).

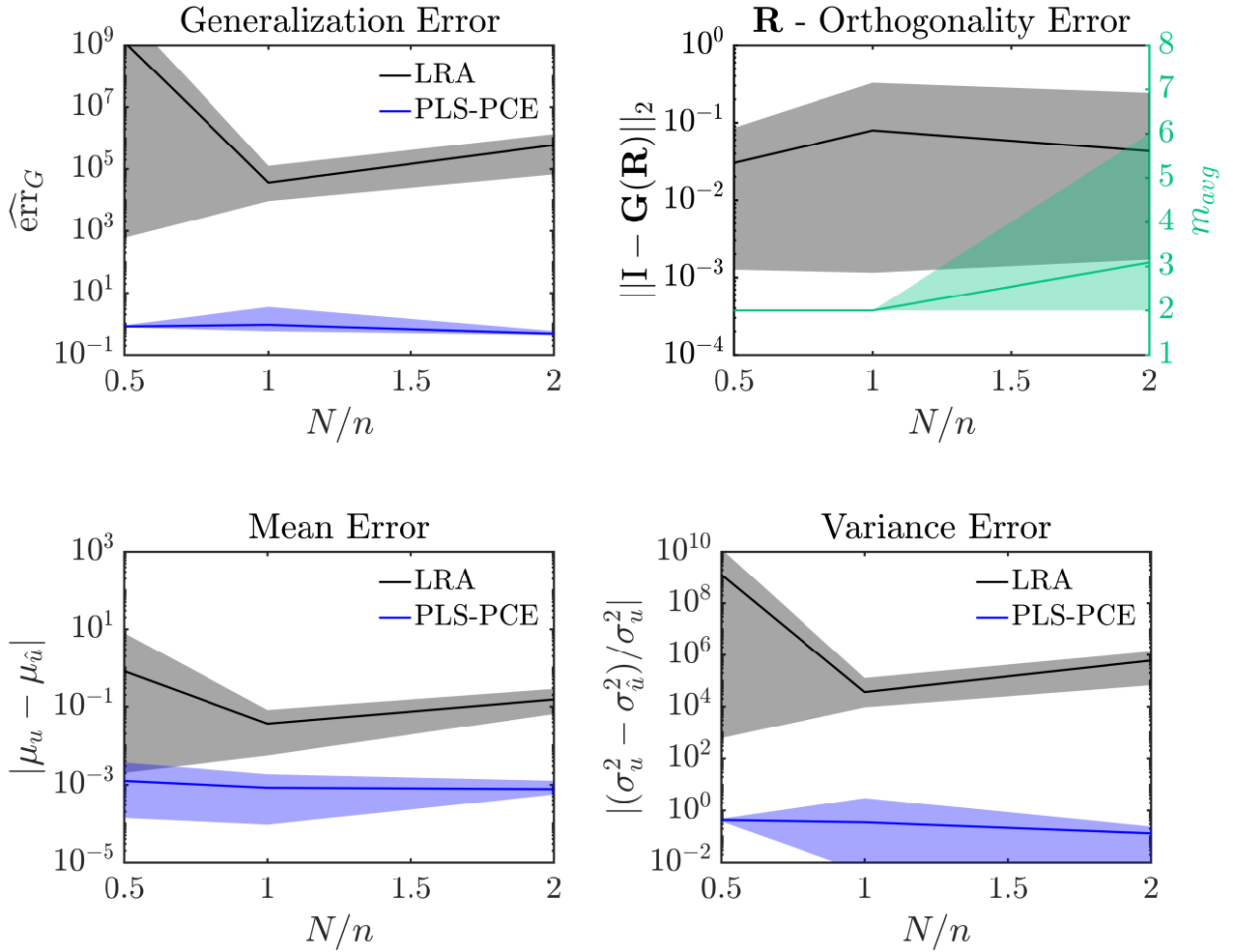


Figure 4.11: Nonlinear oscillator: error measures with 95 % confidence intervals obtained from 100 repeated analyses with different experimental designs; reference solution obtained with 2×10^5 samples. The top right panel includes the average number of latent components (PLS directions) m_{avg} .

surrogate models with the LRA as can be seen from Fig. 4.9. In contrast, PLS-based PCE yields a series of converging surrogate models as N increases. That is, the response PDF is approximated increasingly well and the generalization error decreases monotonously (Fig. 4.11, top left). The polynomial degree of the one-dimensional PCE identified for the first latent directions increases on average with the size of the experimental design, while the percentage of explained variance by the first latent component decreases (Fig. 4.10). Therefore, the number of PLS components increases with increase of the experimental design size (Fig. 4.11, top right). The response mean is captured well by the PLS-based PCE model even with the smallest investigated N (Fig. 4.11, bottom left). Note, that for this example, Fig. 4.11 depicts the unscaled absolute mean error since $\mu_u = 0$. In this example, depending on the experimental design, between 2–4 reduced space directions are included by the PLS algorithm as $N \geq n$. The orthogonality error increases slightly when $N \geq n$, yet it remains reasonably low for all N . Fig. 4.12 depicts the three Gramian matrices corresponding to $m = \{2, 3, 4\}$. The first two directions which explain most of the output variance enclose angles

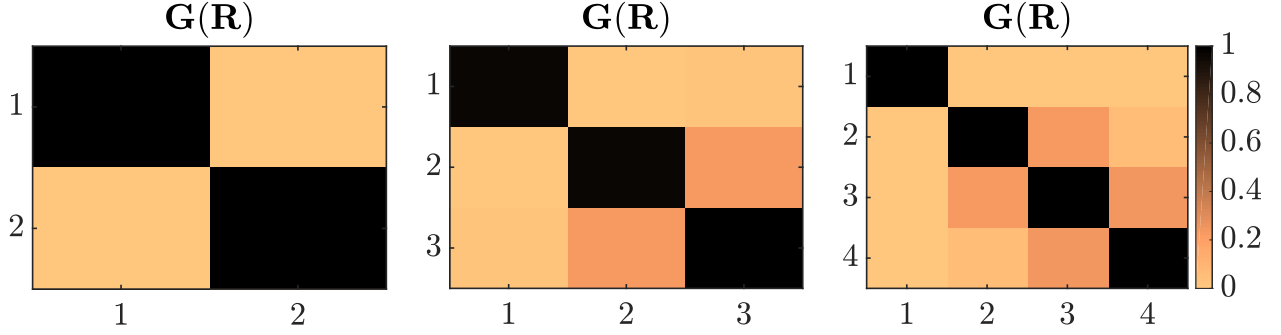


Figure 4.12: Depiction of Gramian matrices of the nonlinear PLS algorithm resulting from $m = 2$ (left), 3 (center) and 4 (right).

close to $\pi/2$ in all three cases, whereas combinations of less important directions are responsible for most of the orthogonality error. Thus, using \mathbf{R} to define the transformation does not compromise significantly the optimality of the Hermite polynomial basis.

4.5.3 Steel plate

For a third example, we consider a modified version of the example given in [26], which consists of a low-carbon steel plate of length 0.32 m, width 0.32 m, thickness $t = 0.01$ m, and a hole of radius 0.02 m located at the center. The Poisson ratio is set to $\nu = 0.29$ and the density of the plate is $\rho = 7850$ kg/m³. The horizontal and vertical displacements are constrained at the left edge. The plate is subjected to a fixed surface load of $q = 96$ MPa, which acts on the right narrow plate side. The Young's modulus $E(x, y)$ is considered uncertain and spatially variable. It is described by a homogeneous random field with lognormal marginal distribution, mean value $\mu_E = 2 \times 10^5$ MPa and standard deviation $\sigma_E = 3 \times 10^4$ MPa. The autocorrelation function of the underlying Gaussian field $\ln E$ is modeled by the isotropic exponential model, $\rho_{\ln E}(\Delta x, \Delta y) = \exp(-\sqrt{\Delta x^2 + \Delta y^2}/l)$ with correlation length $l = 0.04$ m. The random field $\ln D$ is discretized by a KL expansion with $M = 1000$ terms, which yields a global relative variance error of 7%. The stress ($\boldsymbol{\sigma}(x, y) = [\sigma_x(x, y), \sigma_y(x, y), \tau_{xy}(x, y)]^T$), strain ($\boldsymbol{\epsilon}(x, y) = [\epsilon_x(x, y), \epsilon_y(x, y), \gamma_{xy}(x, y)]^T$) and displacement ($\mathbf{u}(x, y) = [u_x(x, y), u_y(x, y)]^T$) fields of the plate are given through elasticity theory, namely the Cauchy-Navier equations [24]. Given the configuration of the plate, the model can be simplified under the plane stress hypothesis, which yields

$$G(x, y)\nabla^2 \mathbf{u}(x, y) + \frac{E(x, y)}{2(1 - \nu)}\nabla(\nabla \cdot \mathbf{u}(x, y)) + \mathbf{b} = 0. \quad (4.34)$$

Therein, $G(x, y) := E(x, y)/(2(1 + \nu))$ is the shear modulus, and $\mathbf{b} = [b_x, b_y]^T$ is the vector of body forces acting on the plate. Eq. (4.34) is discretized with a finite-element method. That is, the spatial domain of the plate is discretized into 282 eight-noded quadrilateral elements, as shown in Fig. 4.13. The scalar model output is the first principal plane stress

$$\sigma_1 = 0.5(\sigma_x + \sigma_y) + \sqrt{[0.5(\sigma_x + \sigma_y)]^2 + \tau_{xy}^2}$$

at node 11 (see green marker Fig. 4.13), which is where maximum plane stresses occur typically in this example. Fig. 4.14 shows that the TG method with linear PCE-based adaptation performs

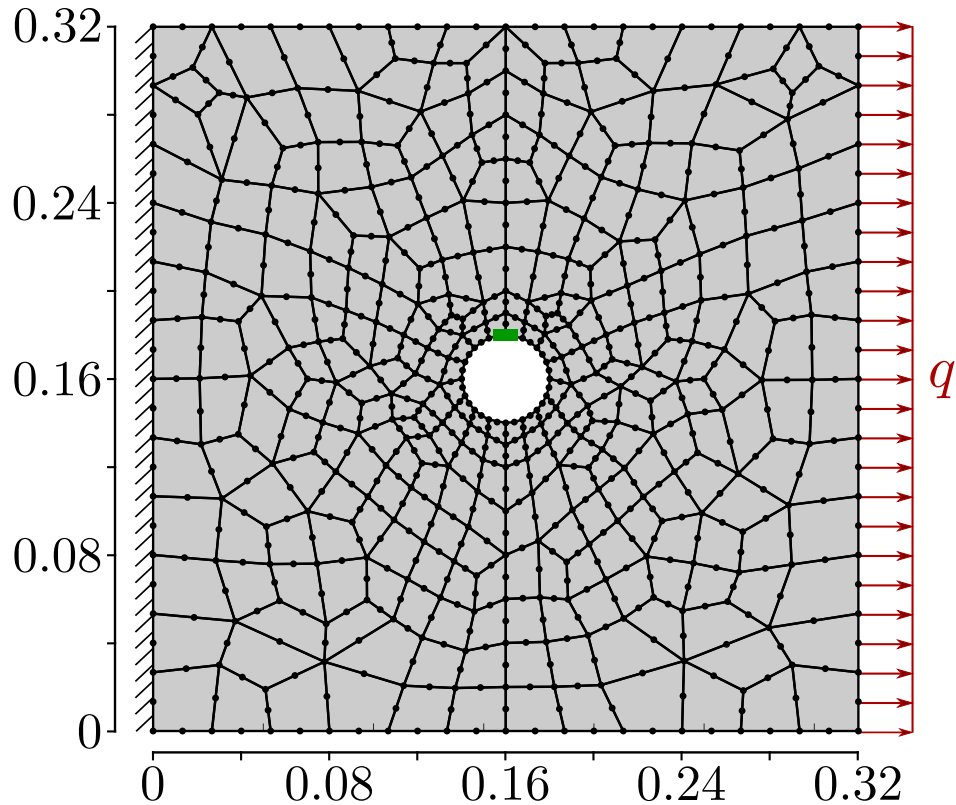


Figure 4.13: FE-mesh of 2D-plate model. Green marker: location of maximum first principal stress σ_1 .

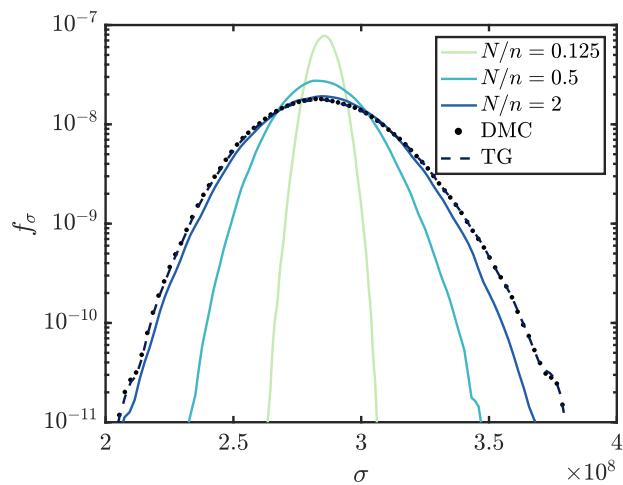


Figure 4.14: Steel plate: log-densities of response. Comparison of linear PLS-PCE (with $m = 1$) with various design sizes with the linear PCE-based adaptation of Tipireddy and Ghanem (TG) [42] and direct Monte Carlo (DMC) using 2×10^5 samples.

well for this example. Again, the linear PLS-based PCE with $m = 1$ approaches the TG solution with increase of N . Comparing the linear and nonlinear PLS-PCE methods, we see again that the nonlinear PCE-driven PLS algorithm gives consistently lower generalization errors with fewer number of latent components than the linear PLS (Fig. 4.15). Comparing the performance of the nonlinear

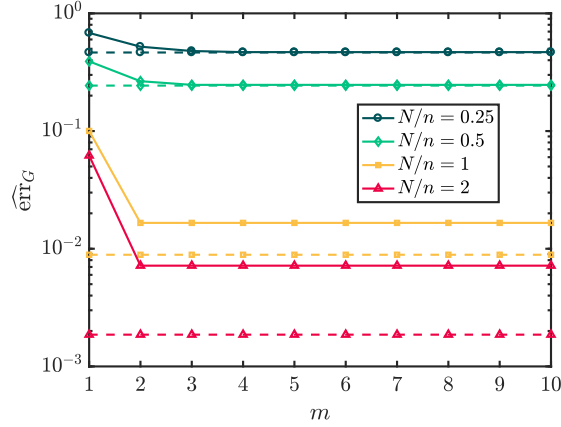


Figure 4.15: Steel plate: comparison of the linear PLS-based (solid line) and nonlinear (PCE-driven) PLS-based (dashed line) adaptation at various design sizes and number of latent components.

PLS-PCE and LRA surrogates, we see that both methods are capable of producing a converging approximation of the numerical model as the number of points in the experimental design increases (Figs. 4.16 and 4.17). Fig. 4.17 indicates that the PLS-based PCE model represents the model response more accurately in the tails. Moreover, the variability associated with the random choice of the experimental design is very small across all N . Similar to the bar example, all PLS-based PCE surrogates are constructed with $m = 2$ with two quasi-orthogonal directions in the reduced space (Fig. 4.18, top right). The PLS-based PCE mean and generalization errors become smaller than the corresponding LRA errors for $N/n = 2$ (Fig. 4.18, bottom & top left).

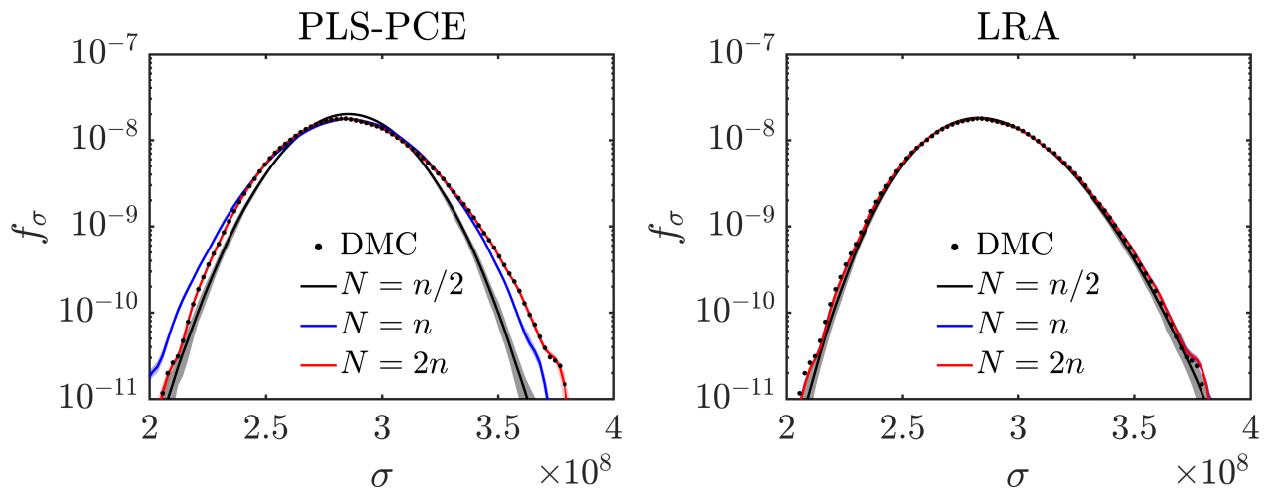


Figure 4.16: Steel plate: log-densities of response at various design sizes with 95 % CI. Results of (nonlinear) PLS-PCE and LRA are compared with direct Monte Carlo (DMC) results using 2×10^5 samples.

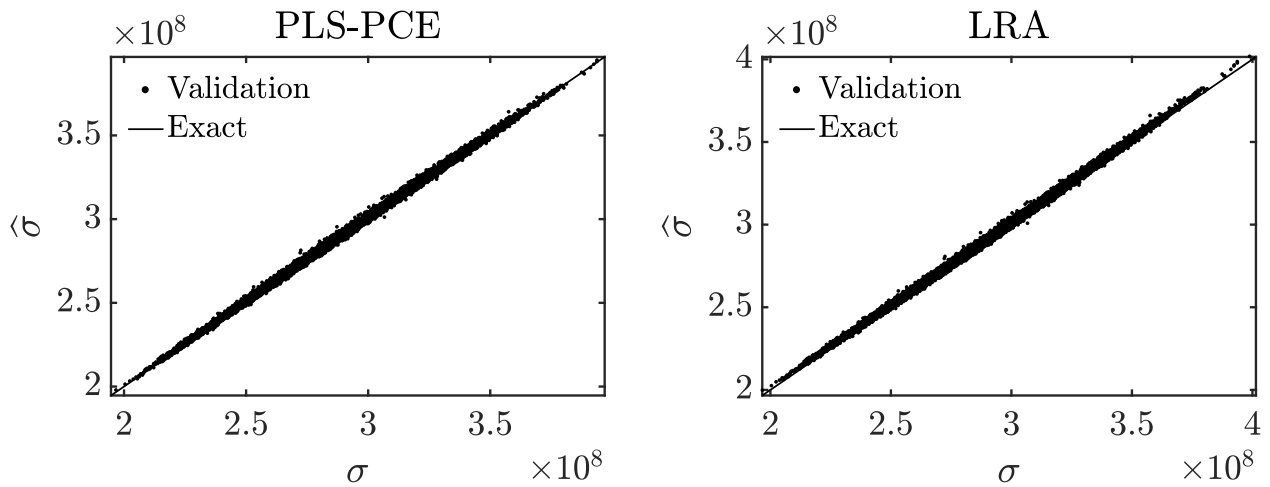


Figure 4.17: Steel plate: scatter plots of PLS-based PCE (left) and LRA (right) response vs original response at $N/n = 2$ using a validation set of 2×10^5 samples.

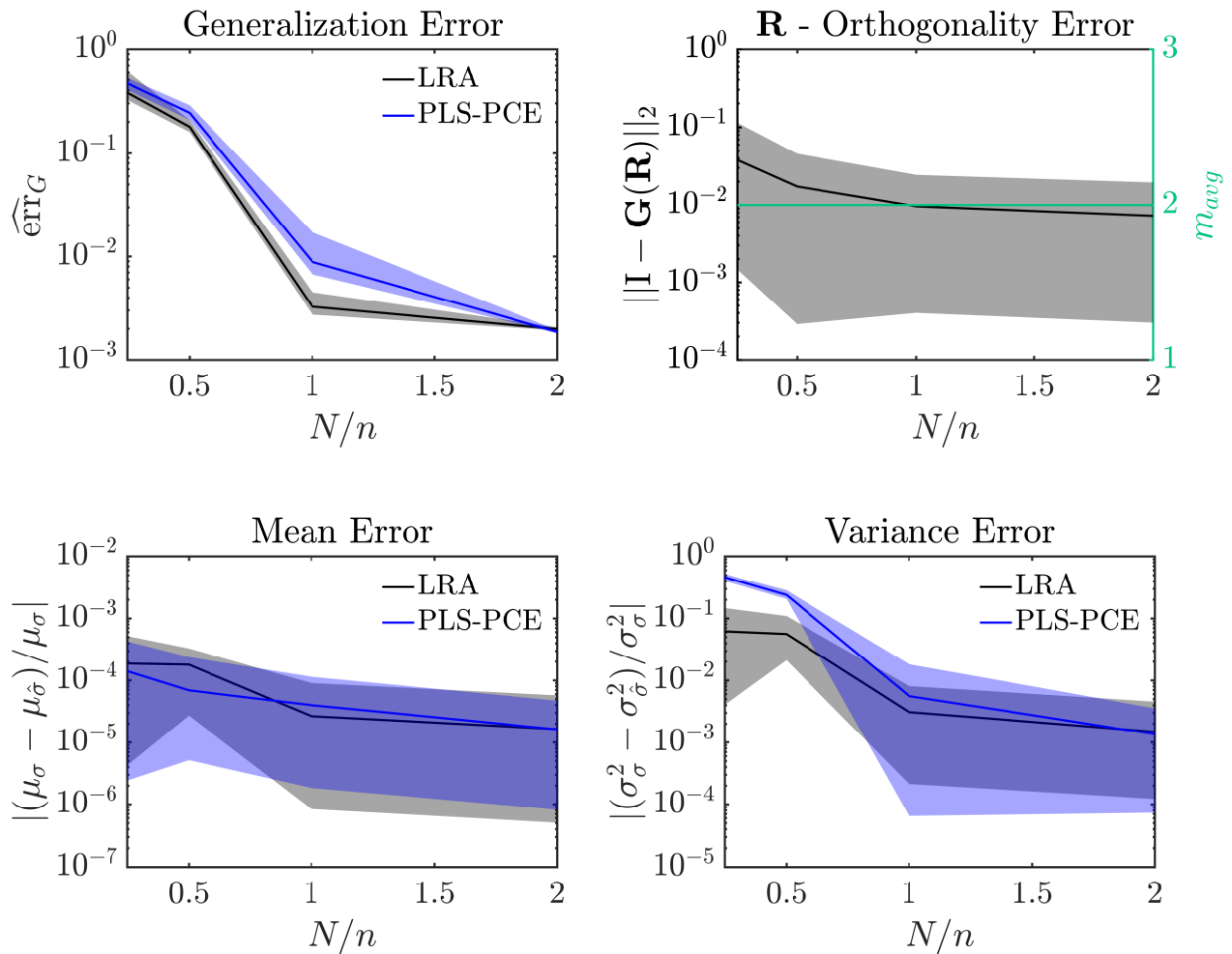


Figure 4.18: Steel plate: error measures with 95 % confidence intervals obtained from 100 repeated analyses with different experimental designs; reference solution obtained with 2×10^5 samples. The top right panel includes the average number of latent components (PLS directions) m_{avg} .

4.6 Concluding remarks

This paper presented a novel sparse polynomial chaos expansion (PCE) representation based on a transformation of the coordinate system in Gaussian space using a number of dominant directions. These directions are identified based on partial least squares (PLS) analysis using a set of experimental points. Two PLS algorithms were investigated; the standard linear PLS algorithm and a novel PCE-driven nonlinear PLS algorithm. It was shown that the linear PLS with a single latent component is asymptotically (as $N \rightarrow \infty$) equivalent to the linear PCE-based adaptation of [42]. The proposed PCE-driven PLS algorithm is able to simultaneously determine the dominant directions in input space and the PCE coefficients in the transformed space. Three numerical examples demonstrated the ability of the method to provide accurate estimates of the moments and PDF of quantities of interest in problems with high-dimensional input spaces, provided that the behavior of the model is governed by a small number of latent variables. The method presented comparable performance to the low-rank tensor approximation (LRA) surrogate in moderately nonlinear problems for experimental design sizes in the order of the input dimension. In addition, the proposed method was able to provide an accurate representation of the response of a highly nonlinear oscillator, whereas the LRA model was unable to obtain converging results for the same example.

4.6.1 Discussion on computational complexity

The linear PLS algorithm (Alg. 1) has complexity approximately $\mathcal{O}(m n N)$, with m being the number of latent components. Solution of the OLS problem in the reduced space has complexity $\mathcal{O}(P_m^2 N)$, with P_m denoting the number of terms in the total degree expansion in reduced space. The complexity of the nonlinear PCE-driven PLS algorithm is dominated by the Newton-Raphson iteration (step 13 in Alg. 2), which has approximate cost $\mathcal{O}(n^2 N)$ (for $N > n$). This step is repeated at most l times, with l denoting the maximum number of Newton-Raphson iterations. Since the nonlinear PLS process is repeated for each candidate polynomial degree and each latent component, the computational complexity of the algorithm is approximately $\mathcal{O}(m p l n^2 N)$, with p being the maximum polynomial degree considered. Although the cost of Alg. 2 is often much higher than the one of Alg. 1, it remains polynomial in the dimension n . It is noted that the computational complexity of ordinary or regularized regression-based PCE in high dimensions is governed by the complexity of the algorithm used to order the multi-indices of the orthogonal polynomials, which is asymptotically $2^{\mathcal{O}(n)}$. This does not pose a problem to PLS-PCE methods as typically $m \ll n$. Hence, the proposed algorithms are orders of magnitude more efficient than sparse PCEs based on compressive sensing as $n \rightarrow \infty$. It is also worth mentioning that the two PLS-PCE methods have low memory requirements, as opposed to regularized regression algorithms that require storage of the full Vandermonde matrix of dimensions $N \times P$, with P increasing factorially with n for a total degree construction.

4.6.2 Outlook

The size of the experimental design for obtaining accurate predictions depends on the nonlinearity of the problem and on its effective dimension, i.e. the number of latent variables that have significant

effect on the behavior of the output quantity. One possible future improvement of the method is to estimate the size of the experimental design adaptively, for example based on a cross validation error criterion. In the present implementation, the experimental design is generated with Latin hypercube sampling with sample decorrelation, to minimize the error in the orthogonality of the polynomials in transformed space. The performance of alternative sampling schemes will be investigated in future studies. Additionally, the method can potentially be extended to the case of multivariate output quantities. This could be enabled by application of the multivariate version of the PLS method, the PLS2 algorithm.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the priority program SPP 1886 by grant STR 1140/6-1.

References

- [1] M. Abramowitz and I. A. Stegun. *Handbook of mathematical function: with formulas, graphs and mathematical tables*. New York: Dover Publications, 1965.
- [2] P. C. Austin and E. W. Steyerberg. “The number of subjects per variable required in linear regression analyses”. In: *Journal of Clinical Epidemiology* 68.6 (2015), pp. 627–636.
- [3] I. Babuska, F. Nobile, and R. Tempone. “A Stochastic Collocation Method for Elliptic Partial Differential Equations with Random Input Data”. In: *SIAM J. Numerical Analysis* 45 (2007), pp. 1005–1034.
- [4] G. Baffi, E. Martin, and A. Morris. “Non-linear projection to latent structures revisited (the neural network PLS algorithm)”. In: *Computers & Chemical Engineering* 23.9 (1999), pp. 1293–1307.
- [5] G. Baffi, E. Martin, and A. Morris. “Non-linear projection to latent structures revisited (the neural network PLS algorithm)”. In: *Computers & Chemical Engineering* 23.9 (1999), pp. 1293–1307.
- [6] M. Bardet, J.-C. Faugère, and B. Salvy. “On the complexity of the F5 Gröbner basis algorithm”. In: *Journal of Symbolic Computation* 70 (2015), pp. 49–70.
- [7] M. Berveiller, B. Sudret, and M. Lemaire. “Stochastic finite element: a non intrusive approach by regression”. In: *European Journal of Computational Mechanics* 15.1-3 (2006), pp. 81–92.
- [8] G. Blatman and B. Sudret. “Adaptive sparse polynomial chaos expansion based on least-angle regression”. In: *Journal of Computational Physics* 230.6 (2011), pp. 2345–2367.
- [9] G. Blatman and B. Sudret. “An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis”. In: *Probabilistic Engineering Mechanics* 25.2 (2010), pp. 183–197.
- [10] O. Chapelle, V. Vapnik, and Y. Bengio. “Model selection for small sample regression”. In: *Machine Learning* 48.1-3 (2002), pp. 9–23.

- [11] M. Chevreuil et al. “A Least-Squares Method for Sparse Low Rank Approximation of Multivariate Functions”. In: *SIAM/ASA Journal on Uncertainty Quantification* 3.1 (2015), pp. 897–921.
- [12] A. Cohen and G. Migliorati. “Optimal weighted least-squares methods”. In: *SMAI Journal of Computational Mathematics* 3 (2017), pp. 181–203.
- [13] P. Constantine et al. “Exploiting active subspaces to quantify uncertainty in the numerical simulation of the HyShot II scramjet”. In: *Journal of Computational Physics* 302 (2015), pp. 1–20.
- [14] P. G. Constantine, E. Dow, and Q. Wang. “Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces”. In: *SIAM Journal on Scientific Computing* 36.4 (2014), A1500–A1524.
- [15] A. Doostan and G. Iaccarino. “A least-squares approximation of partial differential equations with high-dimensional random inputs”. In: *Journal of Computational Physics* 228.12 (2009), pp. 4332–4345.
- [16] A. Doostan and H. Owhadi. “A non-adapted sparse approximation of PDEs with stochastic inputs”. In: *Journal of Computational Physics* 230.8 (2011), pp. 3015–3034.
- [17] A. Doostan, A. Validi, and G. Iaccarino. “Non-intrusive low-rank separated approximation of high-dimensional stochastic models”. In: *Computer Methods in Applied Mechanics and Engineering* 263 (2013), pp. 42–55.
- [18] R. Ghanem and P. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer-Verlag, Berlin, 1991.
- [19] J. Hampton and A. Doostan. “Coherence motivated sampling and convergence analysis of least squares polynomial chaos regression”. In: *Computer Methods in Applied Mechanics and Engineering* 290 (2015), pp. 73–97.
- [20] I. S. Helland. “On the structure of partial least squares regression”. In: *Communications in Statistics-Simulation and Computation* 17.2 (1988), pp. 581–607.
- [21] A. Höskuldsson. “PLS regression methods”. In: *Journal of Chemometrics* 2.3 (1988), pp. 211–228.
- [22] J. D. Jakeman, M. S. Eldred, and K. Sargsyan. “Enhancing l_1 -minimization estimates of polynomial chaos expansions using basis selection”. In: *Journal of Computational Physics* 289 (2015), pp. 18–34.
- [23] J. D. Jakeman, A. Narayan, and T. Zhou. “A generalized sampling and preconditioning scheme for sparse approximation of polynomial chaos expansions”. In: *SIAM Journal on Scientific Computing* 39.3 (2017), A1114–A1144.
- [24] C. Johnson. *Numerical solution of partial differential equations by the finite element method*. Dover Publications, 2009. 288 pp.
- [25] K. Konakli and B. Sudret. “Reliability analysis of high-dimensional models using low-rank tensor approximations”. In: *Probabilistic Engineering Mechanics* 46 (2016), pp. 18–36.
- [26] P.-L. Liu and K.-G. Liu. “Selection of random field mesh in finite element reliability analysis”. In: *Journal of Engineering Mechanics* 119.4 (1993), pp. 667–680.
- [27] S. Marelli and B. Sudret. “UQLab: A framework for uncertainty quantification in Matlab”. In: *Proc. 2nd Int. Conf. on Vulnerability, Risk Analysis and Management (ICVRAM2014)*. 2014, pp. 2554–2563.

- [28] H. G. Matthies and A. Keese. “Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations”. In: *Computer Methods in Applied Mechanics and Engineering* 194.12-16 (2005), pp. 1295–1331.
- [29] M. D. McKay, R. J. Beckman, and W. J. Conover. “Comparison of three methods for selecting values of input variables in the analysis of output from a computer code”. In: *Technometrics* 21.2 (1979), pp. 239–245.
- [30] A. Narayan, J. Jakeman, and T. Zhou. “A Christoffel function weighted least squares algorithm for collocation approximations”. In: *Mathematics of Computation* 86.306 (2017), pp. 1913–1947.
- [31] F. Nobile, R. Tempone, and C. G. Webster. “A sparse grid stochastic collocation method for partial differential equations with random input data”. In: *SIAM Journal on Numerical Analysis* 46.5 (2008), pp. 2309–2345.
- [32] A. Nouy. “Proper generalized decompositions and separated representations for the numerical solution of high dimensional stochastic problems”. In: *Archives of Computational Methods in Engineering* 17.4 (2010), pp. 403–434.
- [33] A. B. Owen. “Controlling correlations in Latin hypercube samples”. In: *Journal of the American Statistical Association* 89.428 (1994), pp. 1517–1522.
- [34] I. Papaioannou, M. Ehre, and D. Straub. “PLS-based adaptation for efficient PCE representation in high dimensions”. In: *Journal of Computational Physics* 387 (2019), pp. 186–204.
- [35] A. Phatak and S. De Jong. “The geometry of partial least squares”. In: *Journal of Chemometrics: A Journal of the Chemometrics Society* 11.4 (1997), pp. 311–338.
- [36] M. Rosenblatt. “Remarks on a multivariate transformation”. In: *The Annals of Mathematical Statistics* 23.3 (1952), pp. 470–472.
- [37] R. Rosipal. “Nonlinear Partial Least Squares: An Overview”. In: *Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques* (Jan. 2010), pp. 169–189.
- [38] K. Sargsyan et al. “Dimensionality reduction for complex models via Bayesian compressive sensing”. In: *International Journal for Uncertainty Quantification* 4.1 (2014), pp. 63–93.
- [39] M. Shinozuka and G. Deodatis. “Simulation of stochastic processes by spectral representation.” In: *Applied Mechanics Reviews* 44.4 (1991), pp. 191–204.
- [40] B. Sudret. “Global sensitivity analysis using polynomial chaos expansions”. In: *Reliability Engineering & System Safety* 93.7 (2008), pp. 964–979.
- [41] M. A. Tatang et al. “An efficient method for parametric uncertainty analysis of numerical geophysical models”. In: *Journal of Geophysical Research: Atmospheres* 102.D18 (1997), pp. 21925–21932.
- [42] R. Tipireddy and R. Ghanem. “Basis adaptation in homogeneous chaos spaces”. In: *Journal of Computational Physics* 259 (2014), pp. 304–317.
- [43] P. Tsilifis et al. “Compressive sensing adaptation for polynomial chaos expansions”. In: *Journal of Computational Physics* 380 (2019), pp. 29–47.
- [44] P. A. Tsilifis. “Gradient-informed basis adaptation for Legendre chaos expansions”. In: *Journal of Verification, Validation and Uncertainty Quantification* 3.1 (2018), p. 011005.

- [45] Z. Wang and J. Song. “Cross-entropy-based adaptive importance sampling using von Mises-Fisher mixture for high dimensional reliability analysis”. In: *Structural Safety* 59 (Mar. 2016), pp. 42–52.
- [46] S. Wold, N. Kettaneh-Wold, and B. Skagerberg. “Nonlinear PLS modeling”. In: *Chemometrics and Intelligent Laboratory Systems* 7.1 (1989). Proceedings of the First Scandinavian Symposium on Chemometrics, pp. 53–65.
- [47] S. Wold, M. Sjöström, and L. Eriksson. “PLS-regression: a basic tool of chemometrics”. In: *Chemometrics and Intelligent Laboratory Systems* 58.2 (2001), pp. 109–130.
- [48] S. Wold et al. “The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses”. In: *SIAM Journal on Scientific and Statistical Computing* 5.3 (1984), pp. 735–743.
- [49] D. Xiu. “Efficient collocational approach for parametric uncertainty analysis”. In: *Communications in Computational Physics* 2 (Apr. 2007), pp. 293–309.
- [50] D. Xiu and J. S. Hesthaven. “High-order collocation methods for differential equations with random inputs”. In: *SIAM Journal on Scientific Computing* 27.3 (2005), pp. 1118–1139.
- [51] D. Xiu and G. E. Karniadakis. “Modeling uncertainty in flow simulations via generalized polynomial chaos”. In: *Journal of Computational Physics* 187.1 (2003), pp. 137–167.
- [52] D. Xiu and G. E. Karniadakis. “The Wiener–Askey polynomial chaos for stochastic differential equations”. In: *SIAM Journal on Scientific Computing* 24.2 (2002), pp. 619–644.
- [53] L. Yan, L. Guo, and D. Xiu. “Stochastic collocation algorithms using l1-minimization”. In: *International Journal for Uncertainty Quantification* 2.3 (2012), pp. 279–293.
- [54] X.-Q. Zeng and G.-Z. Li. “Incremental partial least squares analysis of big streaming data”. In: *Pattern Recognition* 47.11 (2014), pp. 3726–3735.
- [55] C. Zhou, Z. Lu, and W. Li. “Sparse grid integration based solutions for moment-independent importance measures”. In: *Probabilistic Engineering Mechanics* 39 (2015), pp. 46–55.

Global sensitivity analysis in high dimensions with PLS-PCE

Original Publication

M. Ehre, I. Papaioannou, and D. Straub. “Global sensitivity analysis in high dimensions with PLS-PCE”. in: *Reliability Engineering & System Safety* 198 (2020), p. 106861.

Author’s contribution

Max Ehre developed the idea underlying the method and implemented the algorithm. Max Ehre carried out all numerical experiments and drafted the manuscript under the guidance of Iason Papaioannou and Daniel Straub.

Abstract

Global sensitivity analysis is a central part of uncertainty quantification with engineering models. Variance-based sensitivity measures such as Sobol’ and total-effect indices are amongst the most popular and commonly used tools for global sensitivity analysis. Multiple sampling-based estimators of these measures are available, but they often come at considerable computational cost due to the large number of required model evaluations. If the computational model is expensive to evaluate, these approaches are quickly rendered infeasible. An alternative is the use of surrogate models, which reduce the computational cost per sample significantly. This contribution focuses on a recently introduced latent-variable-based polynomial chaos expansion (PCE) based on partial least squares (PLS) analysis, which is particularly suitable for high-dimensional problems. We develop an efficient way of computing variance-based sensitivities with the PLS-PCE surrogate. By back-transforming the surrogate model from its latent variable space-basis to the original input variable space-basis, we

derive analytical expressions for the sought sensitivities. These expressions depend on the surrogate model coefficients exclusively. Thus, once the surrogate model is built, the variance-based sensitivities can be computed at negligible computational cost and no additional sampling is required. The accuracy of the method is demonstrated with two numerical experiments of an elastic truss and a thin steel plate.

5.1 Introduction

Surrogate models have received much attention due to their potential of alleviating computational cost significantly in applications requiring elaborate and expensive numerical models, see e.g. [18, 16, 47]. A common example is the propagation of uncertainties through computationally intensive numerical models. The general concept of surrogate modeling techniques is to establish an abstract, parametrized input-output-relation that has similar properties as the original model. The parameters of the surrogate model are determined based on a finite set of original model evaluations, to maximize similarity between the surrogate and the original model according to a suitable criterion. Subsequently, the surrogate model can be used to cheaply approximate the original model and, in the context of uncertainty quantification, compute statistics of the output or a quantity of interest derived thereof.

In many scenarios, a statistical characterization of the pure model output is less important than an analysis of its sensitivity with respect to changes and variability in the model inputs. Surrogate models have also proven useful in efficiently performing model sensitivity analysis - an otherwise computationally intensive task. Sensitivity analysis is a collection of measures and tools designed to determine how the random inputs and/or deterministic model parameters of a model influence its output or a quantity of interest derived from the output. Amongst these, one can discern local (derivative-based) [25, 1, 15, 35, 36] and global [41, 19, 24, 51, 8] sensitivity measures. Local sensitivities are suitable to determine the impact an input has on the output in the vicinity of a nominal value, by virtue of the model structure. However, they neglect the global significance of the input. Global sensitivity measures on the other hand take into account the entire input variable support as well as the variability of the inputs over their support. Regression-based sensitivity measures aim at linearly regressing the output on its inputs to identify global sensitivity indices; this approach works well if the output depends approximately linearly on the inputs [24, 51]. A second category of global sensitivity measures is referred to as ANOVA (ANalysis Of VAriance) [13]. Generally speaking, these measures aim at quantifying an input variable's influence (or that of a combination of inputs) through identifying the fraction of output variance it causes. A recent, third category of global sensitivity measures can be summarized under the terms 'moment-independent' or 'distribution-based' [7, 34, 32, 6, 57, 17]. The underlying idea is to quantify the sensitivity of the output to a given input through the distance between the output density conditional on the given input from its unconditional counterpart. Our work focusses on variance-based sensitivity measures. The most commonly used variance-based measures are the Sobol' index [45] and the total-effect index [21], which can be computed using Monte Carlo methods [41, 26, 44, 29, 42], Fourier analysis [11, 2] or surrogate models as in [46, 28]. The works of [46, 28] have derived variance-based sensitivity measures directly from the model coefficients of conventional polynomial chaos expansions (PCE) [55] and polynomial-based low-rank approximations (LRA) [10], respectively.

Along these lines, we derive global, variance-based sensitivity measures for the model output from the coefficients of basis-adapted PCEs [49]. The basic idea of basis adaptation is to identify a low dimensional latent variable space and construct a PCE in this space. We focus on a recently introduced approach for identifying the latent variables and computing the corresponding PCE coefficients termed partial least squares-driven polynomial chaos expansion (PLS-PCE) [37]. PLS-PCE allows application of PCEs in very high dimensions. By back-transforming the PCE from the latent variable space to the original input variable space, we enable estimation of the sensitivity indices as with the standard PCE model [46].

The paper is structured as follows: in Section 5.2, we review the PLS-PCE surrogate model, its construction and some important properties. In Section 5.3, we give a brief introduction to variance-based sensitivity analysis and its application in the context of polynomial basis surrogate models. In Section 5.4, we develop the methodology to compute sensitivities based on the model coefficients. In Section 5.5, we demonstrate the new method based on two numerical examples and in Section 5.6 we provide some concluding remarks.

5.2 Partial least squares and polynomial chaos expansions

Let \mathbf{X} be a random vector on the outcome space \mathbb{R}^d with joint cumulative distribution function (CDF) $F_{\mathbf{X}}$ and $\mathcal{Y}(\mathbf{X}) = Y \in \mathbb{R}$. If \mathcal{Y} is square-integrable, i.e. $\mathbb{E}_{\mathbf{X}}[\mathcal{Y}(\mathbf{X})^2] < \infty$, it belongs to a Hilbert space \mathcal{H} with inner product of any two functions $g, h \in \mathcal{H}$

$$\langle g, h \rangle_{\mathcal{H}} = \mathbb{E}_{\mathbf{X}}[g(\mathbf{X})h(\mathbf{X})] \quad (5.1)$$

$$= \int_{\mathbb{R}^d} g(\mathbf{x})h(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}, \quad (5.2)$$

where $f_{\mathbf{X}}(\mathbf{x})$ is the joint probability density function (PDF) of \mathbf{X} . g and h are orthogonal if

$$\langle g, h \rangle_{\mathcal{H}} = \mathbb{E}_{\mathbf{X}}[g(\mathbf{X})h(\mathbf{X})] = 0. \quad (5.3)$$

Note, that if g and h can be written as products of univariate functions g_i and h_i , $i = 1, \dots, d$, of the components of \mathbf{X} and these components are statistically independent, the following holds:

$$\langle g, h \rangle_{\mathcal{H}} = \prod_{i=1}^d \mathbb{E}_{X_i}[g_i(X_i)h_i(X_i)]. \quad (5.4)$$

5.2.1 Polynomial Chaos Expansion

Given a complete and orthonormal basis of \mathcal{H} , $\{h_i(\mathbf{X}), i \in \mathbb{N}\}$, Y may be expressed as a linear combination of the basis functions:

$$Y = \mathcal{Y}(\mathbf{X}) = \sum_{i=0}^{\infty} b_i h_i(\mathbf{X}). \quad (5.5)$$

Then, since $\mathcal{Y} \in \mathcal{H}$, the approximation

$$\hat{Y}_n = \hat{\mathcal{Y}}(\mathbf{X}) = \sum_{i=0}^n b_i h_i(\mathbf{X}) \quad (5.6)$$

asymptotically ($n \rightarrow \infty$) converges to Y in the mean-square sense. Henceforth, without loss of generality, we will consider the case $F_{\mathbf{X}} = \Phi_d$, where Φ_d denotes the d -variate independent standard-normal CDF. If the joint PDF of \mathbf{X} is known, one can express \mathbf{X} as a function of standard normal random variables through an iso-probabilistic transformation [39]. Then, one can construct an orthonormal polynomial basis of \mathcal{H} using products of one-dimensional normalized Hermite polynomials

$$\Psi_{\mathbf{k}}(\mathbf{X}) = \prod_{i=1}^d \psi_{k_i}(X_i) \quad (5.7)$$

where $\{\psi_i(X), i \in \mathbb{N}\}$ are the normalized (probabilist) Hermite polynomials and $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{N}^d$. The PCE of maximum total order p reads

$$\hat{Y}_p = \sum_{|\mathbf{k}| \leq p} b_{\mathbf{k}} \Psi_{\mathbf{k}}(\mathbf{X}). \quad (5.8)$$

The total number of basis functions in the PCE P is given combinatorially in terms of the dimensions d and the maximum total polynomial order p :

$$P = \binom{d+p}{p}. \quad (5.9)$$

The coefficients \mathbf{b} are computed through a projection of \mathcal{Y} onto the space spanned by $\{\Psi_{\mathbf{k}}, |\mathbf{k}| \leq p\}$, where the projection can be transformed into an equivalent ordinary least squares (OLS) problem [4]. Eq. (5.9) indicates a fast growth of the associated regression problem with increasing dimension d , rendering PCEs intractable for high-dimensional problems. Sparse PCE methods have been proposed to relax this constraint by solving a modified, \mathcal{L}_1 -regularized least-squares problem, which penalizes the number of terms in the expansion and thus reduces P [5]. This is also known under the term 'compressive sampling/sensing' [12, 56]. Nevertheless, the computation of a sparse PCE still requires computing the entirety of all possible basis elements, which can become a second (combinatorial) bottleneck in addition to the solution of the regression problem.

5.2.2 Basis adaptation

In order to address this problem, one may rotate the PCE representation onto a new basis defined by the new variables $\mathbf{Z} = \mathbf{Q}^T \mathbf{X}$, where $\mathbf{Q} \in \mathbb{R}^{d \times d}$ and $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$, with \mathbf{I} denoting the identity matrix. Then, an equivalent PCE representation is given by [49]

$$\hat{Y}_p^{\mathbf{Q}} = \sum_{|\mathbf{k}| \leq p} a_{\mathbf{k}} \Psi_{\mathbf{k}}(\mathbf{Z}) = \sum_{|\mathbf{k}| \leq p} a_{\mathbf{k}} \Psi_{\mathbf{k}}(\mathbf{Q}^T \mathbf{X}). \quad (5.10)$$

The coordinate transformation allows for the construction of PCEs along important directions of the problem input space. These directions are defined by linear combinations of the original variable

vector \mathbf{X} , the coefficients of which are stored in the rows of \mathbf{Q} . Then, by retaining only the $m < d$ most important directions in \mathbf{Q} , one obtains a matrix \mathbf{Q}_m and the corresponding PCE reads

$$\hat{Y}_p^{\mathbf{Q}_m} = \sum_{|\tilde{\mathbf{k}}| \leq p} a_{\tilde{\mathbf{k}}} \Psi_{\tilde{\mathbf{k}}}(\mathbf{Q}_m^T \mathbf{X}), \quad (5.11)$$

where $\tilde{\mathbf{k}} \in \mathbb{N}^m$. [49] compute the basis adaptation \mathbf{Q}_m by evaluating first- or second-order PCE coefficients only with a sparse-grid numerical quadrature. [50] couple this approach with compressive sensing to simultaneously identify \mathbf{Q}_m and the PCE coefficients in the latent space. In [37], we show that important directions can be identified efficiently based on a set of original function evaluations via partial least squares (PLS). The next section summarizes the approach.

5.2.3 Partial least squares-based PCE

PLS finds a relationship between variables \mathbf{X} and Y based on N observations of both quantities [54, 22, 53]. $\mathcal{X} \in \mathbb{R}^{N \times d}$ stores observations from \mathbf{X} and $\mathcal{Y} \in \mathbb{R}^{N \times 1}$ stores the corresponding responses. PLS sequentially identifies latent components $\mathbf{t}_i \in \mathbb{R}^{N \times 1}$ such that they have maximum covariance with \mathcal{Y} . After determining each \mathbf{t}_i , PLS assumes a linear relationship between \mathbf{t}_i and \mathcal{Y} and evaluates the corresponding coefficient a_i of \mathbf{t}_i by OLS. After each sequence, the matrices \mathcal{X} and \mathcal{Y} are deflated by the contribution of the i -th PLS-component. Components are extracted until a certain error criterion is met, which can be formulated e.g. through the norm of the residual response vector or via cross-validation.

The nonlinear version of PLS in turn relaxes the assumption of a linear relationship between latent component and the response. A number of nonlinear PLS algorithms have been proposed [40]. Here we employ the approach of [52, 3], which introduces an additional loop into the algorithm for running a Newton-Raphson procedure iterating between the current latent component and the response. In the context of PCE, the nonlinear relationship between the $\{\mathbf{t}_i\}_{i=1, \dots, m}$ and the response is a one-dimensional Hermite polynomial expansion [37]. The coefficients of the PLS-driven PCE can be computed simultaneously with the latent variable structure as a byproduct of the PLS algorithm. Ultimately, the nonlinear PCE-driven PLS algorithm, which is developed in [37] and summarized in Alg. 3, identifies m latent components. For each component, it returns the direction \mathbf{r}_i and the 1-dimensional PCE along this direction, which is defined by its polynomial order q_i and the coefficient vector \mathbf{a}_i . The polynomial order is identified with leave-one-out cross validation. For each (i -th) latent component, the nonlinear PLS iteration is repeated for different polynomial orders and q_i is chosen as the order minimizing the leave-one-out error. The PLS-PCE reads

$$\hat{Y}_m^{\text{PLS}} = a_0 + \sum_{i=1}^m (\mathbf{a}_i^{q_i})^T \boldsymbol{\psi}_{q_i} \left[(\mathbf{r}_i)^T \tilde{\mathbf{X}} \right], \quad (5.12)$$

where $a_0 = \widehat{\mathbb{E}}[\mathcal{Y}]$, $\boldsymbol{\psi}_{q_i}(\mathbf{X})$ is a vector function assembling the evaluations of the one-dimensional Hermite polynomials up to order q_i and $\tilde{\mathbf{X}} = \mathbf{X} - \boldsymbol{\mu}_{\mathcal{X}}$, where $\boldsymbol{\mu}_{\mathcal{X}}$ is the columnwise sample mean of the training data \mathcal{X} . The PLS directions \mathbf{r}_i can be evaluated in terms of the PLS weights \mathbf{w}_i and loads \mathbf{p}_i computed by Alg. 3 through the following recursive relation [23]:

$$\begin{aligned} \mathbf{r}_1 &= \mathbf{w}_1 \\ \mathbf{r}_i &= \mathbf{w}_i - \mathbf{r}_{i-1} (\mathbf{p}_{i-1}^T \mathbf{w}_i). \end{aligned} \quad (5.13)$$

Algorithm 3 PCE-driven PLS algorithm [37]

-
- 1: **Input** Data matrix \mathcal{X} and response matrix \mathcal{Y}
 - 2: Center matrices: $\mathcal{X} \leftarrow \mathcal{X} - \bar{\mathcal{X}}, \mathcal{Y} \leftarrow \mathcal{Y} - \bar{\mathcal{Y}}$
 - 3: Set $\mathcal{E} = \mathcal{X}, \mathcal{F} = \mathcal{Y}, i = 1$
 - 4: **repeat**
 - 5: Compute weight: $\mathbf{w}_i^0 = \mathcal{E}^T \mathcal{F} / \|\mathcal{E}^T \mathcal{F}\|$
 - 6: **for** $q \leftarrow 1, p$ **do**
 - 7: Set $\mathbf{w}_i^q = \mathbf{w}_i^0$
 - 8: **repeat**
 - 9: Compute score: $\mathbf{t}_i^q = \mathcal{E} \mathbf{w}_i^q$
 - 10: Fit a 1D PCE of order q : $\hat{\mathbf{a}}_i^q \leftarrow \text{fit} [\mathcal{F} = (\mathbf{a}_i^q)^T \boldsymbol{\psi}_q(\mathbf{t}_i^q) + \boldsymbol{\epsilon}]$
 - 11: Set $\hat{\mathcal{M}}_i^q(t) = (\hat{\mathbf{a}}_i^q)^T \boldsymbol{\psi}_q(\mathbf{t}_i^q)(t)$
 - 12: Compute the error: $\hat{\mathcal{F}} = (\hat{\mathbf{a}}_i^q)^T \boldsymbol{\psi}_q(\mathbf{t}_i^q); \mathbf{e} = \mathcal{F} - \hat{\mathcal{F}}$
 - 13: Compute: $\Delta \mathbf{w}_i^q = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{e}$ with $\mathbf{A} = \nabla_{\mathbf{w}} (\hat{\mathbf{a}}_i^q)^T \boldsymbol{\psi}_q(\mathcal{E} \mathbf{w})$
 - 14: Set: $\mathbf{w}_i^q \leftarrow \mathbf{w}_i^q + \Delta \mathbf{w}_i^q$
 - 15: Normalize: $\mathbf{w}_i^q \leftarrow \mathbf{w}_i^q / \|\mathbf{w}_i^q\|$
 - 16: **until** $\|\Delta \mathbf{w}_i^q\|$ is smaller than ϵ_w
 - 17: Evaluate the relative leave-one-out error ϵ_{LOO}^q as in [5]
 - 18: Set $\{q_i, \hat{\mathbf{a}}_i^{q_i}, \mathbf{w}_i^{q_i}\}$ as the triple $\{q, \hat{\mathbf{a}}_i^q, \mathbf{w}_i^q\}$ with the smallest ϵ_{LOO}^q
 - 19: Compute score: $\mathbf{t}_i^{q_i} = \mathcal{E} \mathbf{w}_i^{q_i}$
 - 20: Compute load: $\mathbf{p}_i^{q_i} = \mathcal{E}^T \mathbf{t}_i^{q_i} / ((\mathbf{t}_i^{q_i})^T \mathbf{t}_i^{q_i})$
 - 21: Deflate: $\mathcal{E} \leftarrow \mathcal{E} - \mathbf{t}_i^{q_i} (\mathbf{p}_i^{q_i})^T, \mathcal{F} \leftarrow \mathcal{F} - (\hat{\mathbf{a}}_i^{q_i})^T \boldsymbol{\psi}_{q_i}(\mathbf{t}_i^{q_i})$
 - 22: $i \leftarrow i + 1$
 - 23: **until** change in $\|\mathcal{F}\|$ is smaller than ϵ_y
 - 24: **return** $\{q_i, \hat{\mathbf{a}}_i^{q_i}, \mathbf{t}_i^{q_i}, \mathbf{w}_i^{q_i}, \mathbf{p}_i^{q_i}\}, i = 1 \dots, m$.
-

Let $\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_m] \in \mathbb{R}^{d \times m}$ be the matrix that collects all PLS directions. The matrix \mathbf{R} is not necessarily orthogonal, i.e. in general $\mathbf{R}^T \mathbf{R} \neq \mathbf{I}$. However, in [37] it is shown that $\mathbf{R}^T \mathbf{R} \rightarrow \mathbf{I}$ as $N \rightarrow \infty$ and hence Eq. (5.12) is asymptotically equivalent to a PCE of the form of Eq. (5.11), where only the main effects in the latent components are considered.

5.3 Global sensitivity analysis

5.3.1 Variance-based sensitivity analysis

The idea behind variance-based sensitivity analysis for model outputs Y is to decompose $\mathbb{V}[Y]$ into partial variances that are attributable to variable combinations in the input \mathbf{X} . If \mathbf{X} is jointly uniform on $[0, 1]$ and its components are independent, this is accomplished by projecting Y onto a unique, orthogonal basis with respect to the uniform joint density. The representation of Y is then the Sobol'-Hoeffding decomposition [45], which reads:

$$f(\mathbf{X}) = f_0 + \sum_{i=1}^d f_i(X_i) + \sum_{i=1}^d \sum_{j=i+1}^d f_{ij}(X_i, X_j) + \dots + f_{12\dots d}(\mathbf{X}). \quad (5.14)$$

Each summand in Eq. (5.14) represents the influence of a distinct variable subset of \mathbf{X} , $\mathbf{X}_{\mathcal{A}}$, and due to the orthogonality property, the partial variance associated with \mathcal{A} is given immediately by $\mathbb{V}[f_{\mathcal{A}}]$. The Sobol' index is then the ratio of the partial variance due to $f_{\mathcal{A}}$ and the total variance [45]:

$$S_{Y,\mathcal{A}} = \mathbb{V}[f_{\mathcal{A}}]/\mathbb{V}[Y]. \quad (5.15)$$

Alternatively, one can utilize the closed Sobol' index [38], which is based on the partial variance contributed by $\mathbf{X}_{\mathcal{A}}$ and any subset of $\mathbf{X}_{\mathcal{A}}$, i.e.,

$$S_{Y,\mathcal{A}}^{\text{clo}} = \sum_{B \subseteq \mathcal{A}} \mathbb{V}[f_B]/\mathbb{V}[Y]. \quad (5.16)$$

While the Sobol' and closed Sobol' indices are identical for single variables, i.e., $\text{card}(\mathcal{A}) = 1$, the former represents the net interaction in between all elements of $\mathbf{X}_{\mathcal{A}}$ and the latter represents the total contribution of all elements of $\mathbf{X}_{\mathcal{A}}$. Finally, the total-effect index S^T [21, 38] is based on the partial variance contributed by all variable combinations containing any element from $\mathbf{X}_{\mathcal{A}}$, such that

$$S_{Y,\mathcal{A}}^T = \sum_{\mathcal{A} \cap \mathcal{B} \neq \emptyset} \mathbb{V}[f_{\mathcal{B}}]/\mathbb{V}[Y]. \quad (5.17)$$

The decomposition of Eq. (5.14) is generalizable to arbitrary joint distributions with independent components through an iso-probabilistic transformation.

5.3.2 PCE-based sensitivity analysis

A major benefit of representing a response $Y \in \mathcal{H}$ as \hat{Y} with an orthogonal basis of \mathcal{H} lies in the simplicity of finding statistical properties of \hat{Y} and thus - if the model accurately represents Y -

approximately of Y . Given the model representation in Eq. (5.8) with P terms, e.g. the first two moments can be computed as

$$\mathbb{E}[\hat{Y}] = \mathbf{b}_0, \quad \mathbb{V}[\hat{Y}] = \sum_{0 < |\mathbf{k}| \leq p} b_{\mathbf{k}}^2. \quad (5.18)$$

Moreover, [46] showed that the indices $S_{\hat{Y}, \mathcal{A}}$ and $S_{\hat{Y}, \mathcal{A}}^T$ of the representation in Eq. (5.8) can also be found merely by post-processing its coefficients \mathbf{b} . For a given subset of the input variables denoted by the index set \mathcal{A} , we define a boolean index vector $\mathcal{I}^{\mathcal{A}} \in \{0, 1\}^{d \times 1}$ s.t. $\mathcal{I}_i^{\mathcal{A}} = 0$ if $i \notin \mathcal{A}$ and $\mathcal{I}_i^{\mathcal{A}} = 1$ if $i \in \mathcal{A}$. In the same way, we define such a boolean vector for the multi-index \mathbf{k} s.t. $\mathcal{I}_i^{\mathbf{k}} = 0$ if $k_i = 0$ and $\mathcal{I}_i^{\mathbf{k}} = 1$ if $k_i > 0$. Then, the PCE-based sensitivity indices read

$$\hat{S}_{\hat{Y}, \mathcal{A}} = \frac{1}{\mathbb{V}[\hat{Y}]} \sum_{\substack{\mathcal{I}^{\mathcal{A}} = \mathcal{I}^{\mathbf{k}}, \\ 0 < |\mathbf{k}| \leq p}} b_{\mathbf{k}}^2, \quad \hat{S}_{\hat{Y}, \mathcal{A}}^{\text{clo}} = \frac{1}{\mathbb{V}[\hat{Y}]} \sum_{\substack{\mathcal{I}^{\mathcal{A}} - \mathcal{I}^{\mathbf{k}} \geq \mathbf{0}, \\ 0 < |\mathbf{k}| \leq p}} b_{\mathbf{k}}^2, \quad \hat{S}_{\hat{Y}, \mathcal{A}}^T = \frac{1}{\mathbb{V}[\hat{Y}]} \sum_{\substack{(\mathcal{I}^{\mathcal{A}})^T \mathcal{I}^{\mathbf{k}} \neq \mathbf{0}, \\ 0 < |\mathbf{k}| \leq p}} b_{\mathbf{k}}^2. \quad (5.19)$$

5.4 Global sensitivity analysis with PLS-PCE

Here we derive expressions for $S_{\hat{Y}}$ and $S_{\hat{Y}}^T$ for \hat{Y} of the form in Eq. (5.12). Note, that if the columns of \mathbf{R} form an orthonormal basis, i.e., if $\mathbf{R}^T \mathbf{R} = \mathbf{I}$, the sensitivity indices of any latent variable component $Z_i = \mathbf{r}_i^T \tilde{\mathbf{X}}$ can be obtained immediately as

$$S_{\hat{Y}_m^{\text{PLS}}, Z_i} = S_{\hat{Y}_m^{\text{PLS}}, Z_i}^{\text{clo}} = S_{\hat{Y}_m^{\text{PLS}}, Z_i}^T = \frac{\sum_{j=1}^{q_i} (a_{ij}^{q_i})^2}{\sum_{i=1}^m \sum_{j=1}^{q_i} (a_{ij}^{q_i})^2}. \quad (5.20)$$

However, as discussed in Section 5.2, the condition on \mathbf{R} only holds asymptotically as $N \rightarrow \infty$. In practice, the sensitivity indices associated with the latent variables are of less interest than those of the original inputs. That is, one is interested in computing sensitivities of \hat{Y}_m^{PLS} to the original input vector \mathbf{X} rather than \mathbf{Z} . For convenience, we restate the format of \hat{Y}_m^{PLS} :

$$\hat{Y}_m^{\text{PLS}} = a_0 + \sum_{i=1}^m (\mathbf{a}_i^{q_i})^T \boldsymbol{\psi}_{q_i} [(\mathbf{r}_i)^T (\mathbf{X} - \boldsymbol{\mu}_{\mathcal{X}})]. \quad (5.21)$$

In the following two subsections we derive expressions and state corresponding algorithms for computing the Sobol' and total-effect indices of the PLS-PCE model response with respect to \mathbf{X} for both large and small sample sizes.

5.4.1 Computation in the asymptotic limit $N \rightarrow \infty$

Asymptotically, i.e. for $N \rightarrow \infty$, we have

$$\lim_{N \rightarrow \infty} \boldsymbol{\mu}_{\mathcal{X}} = \mathbf{0}$$

and [37] proves that

$$\lim_{N \rightarrow \infty} \|\mathbf{r}_i\| = 1 \quad i = 2, \dots, m,$$

while the first PLS-direction $\|\mathbf{r}_1\|$ always has length 1. [9] provides a multinomial theorem for non-normalized probabilist's Hermite polynomial of order k , which is restated in the following for the normalized polynomials.

Theorem 5.4.1. *Let $j \in \mathbb{N}_0$ be the polynomial order of the normalized probabilist's Hermite polynomial ψ_j , $d \in \mathbb{N}$ and $\mathbf{X} \in \mathbb{R}^{d \times 1}$. Let further $\mathbf{k} \in \mathbb{N}_0^d$ be an index set and $\mathbf{s} \in \mathbb{R}^{d \times 1}$ such that $\sum_{i=1}^d s_i^2 = 1$. Then,*

$$\begin{aligned} \psi_j(\mathbf{s}^\top \mathbf{X}) &= \sqrt{j!} \sum_{|\mathbf{k}|=j} \prod_{l=1}^d \frac{s_l^{k_l}}{\sqrt{k_l!}} \psi_{k_l}(X_l) \\ &= \sqrt{j!} \sum_{|\mathbf{k}|=j} \frac{s_1^{k_1} \cdot s_2^{k_2} \cdot \dots \cdot s_d^{k_d}}{\sqrt{k_1! \cdot k_2! \cdot \dots \cdot k_d!}} \Psi_{\mathbf{k}}(\mathbf{X}). \end{aligned} \quad (5.22)$$

Therefore, in the asymptotic limit we can use Eq. (5.22) to write

$$\hat{Y}_m^{\text{PLS}} = a_0 + \sum_{i=1}^m \sum_{|\mathbf{k}| \leq q_i} a_{i|\mathbf{k}|}^{q_i} \sqrt{|\mathbf{k}|!} \frac{r_{i1}^{k_1} \cdot r_{i2}^{k_2} \cdot \dots \cdot r_{id}^{k_d}}{\sqrt{k_1! \cdot k_2! \cdot \dots \cdot k_d!}} \Psi_{\mathbf{k}}(\mathbf{X}). \quad (5.23)$$

In practice, the sample mean decays towards 0 relatively fast, such that the approximation error introduced by neglecting the variable centering in Eq. (5.23) is typically orders of magnitude smaller than the leading error introduced by the surrogate model itself. The error due to $\|\mathbf{r}_i\| \neq 1$ grows with the number of included components m (with $m = 1$, the representation is exact since $\|\mathbf{r}_1\| = 1$ always). Alternatively, it is possible to derive exact expressions with respect to both non-zero sample mean and non-unit-length component directions and we will do so in Subsection 5.4.2.

Eq. (5.23) is merely a linear combination of m standard PCEs, each representing a latent component in standard PCE format, so that we can write

$$\hat{Y}_m^{\text{PLS}} = a_0 + \sum_{|\mathbf{k}| \leq q_{\max}} c_{\mathbf{k}} \Psi_{\mathbf{k}}(\mathbf{X}), \quad (5.24)$$

where

$$q_{\max} = \max_{i \in \{1, \dots, m\}} (q_i). \quad (5.25)$$

The equivalent PCE coefficients \mathbf{c} read

$$c_{\mathbf{k}} = \sum_{i=1}^m a_{i|\mathbf{k}|}^{q_i} \sqrt{|\mathbf{k}|!} \prod_{l=1}^d \frac{r_{il}^{k_l}}{\sqrt{k_l!}}, \quad (5.26)$$

where $\{a_{i|\mathbf{k}|}^{q_i} : q_i < |\mathbf{k}|\} = 0$. Thus, one can apply the standard post-processing defined by Eq. (5.19) to the format in Eq. (5.24) in order to obtain variance-based sensitivity indices based on [46]. The corresponding subroutine is referred to as `PCE_sensitivites`.

Alg. 4 efficiently determines the Sobol' and total-effect indices of a PLS-PCE model. The algorithm requires to compute the set of multi-indices $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{N}^d$ that satisfy $|\mathbf{k}| \leq q_{\max}$. This multi-index set is only computed once at the beginning using a routine termed `multi_index`. Various methodologies such as the ball-box-algorithm [48] have been proposed for computing the set

Algorithm 4 PLS-based sensitivities - asymptotic case

```

1: Input PLS components  $q_i, \mathbf{a}^{q_i}, \mathbf{r}_i \forall i = 1, \dots, m$ 
2: Initialize  $\mathbf{c} = \mathbf{0}$ 
3:  $q_{max} = \max_{i \in \{1, \dots, m\}} (q_i)$ 
4: Compute PCE multi-index set, e.g.:  $\boldsymbol{\alpha} \leftarrow \text{multi\_index}(d, q_{max}) \in \mathbb{N}^{P \times d}$ 
5: for  $i \leftarrow 1, m$  do
6:    $j = 1$ 
7:   while  $|\boldsymbol{\alpha}_j| \leq q_i$  do
8:     Set current multi-index  $\mathbf{k} \leftarrow \boldsymbol{\alpha}_j \in \mathbb{N}^{1 \times d}$ 
9:     Compute coefficients  $c_{\mathbf{k}} \leftarrow a_{i|\mathbf{k}|}^{q_i} \sqrt{|\mathbf{k}|!} \prod_{l=1}^d \frac{r_{il}^{k_l}}{\sqrt{k_l!}}$ 
10:    Augment:  $c_j \leftarrow c_j + c_{\mathbf{k}}$ 
11:     $j \leftarrow j + 1$ 
12: Compute sensitivities:  $S, S^{\text{clo}}, S^T \leftarrow \text{PCE\_sensitivities}(\boldsymbol{\alpha}, \mathbf{c})$ 
13: Return  $S, S^{\text{clo}}, S^T$ 

```

of multi-indices of a PCE. We observe that the index set \mathbf{k} required for the PLS-based sensitivity indices is equivalent to that of a full PCE formulation of maximum polynomial order q_{max} . Fortunately, the additional degrees of freedom emerging from the latent variable formulation (i.e. the \mathbf{r}_i) lead to significantly smaller required polynomial degrees in PLS-PCE compared to sparse and classical PCE models. That is, the computational bottleneck of computing \mathbf{k} is not critical in most applications.

5.4.2 Corrections for small samples sizes

The presented methodology to extract sensitivities from a PLS-PCE model is asymptotically exact. However, PLS-PCE is a surrogate modeling technique, which is particularly suitable when the number of samples is small compared to the problem dimension. In this case, the exact sensitivities accounting for non-zero sample mean $\boldsymbol{\mu}_{\mathcal{X}}$ and non-unit-length PLS-directions \mathbf{r}_i can still be derived. Rewriting Eq. (5.21), we have

$$\hat{Y}_m^{\text{PLS}} = a_0 + \sum_{i=1}^m (\mathbf{a}_i^{q_i})^T \psi_{q_i} (\|\mathbf{r}_i\| (\tilde{\mathbf{r}}_i)^T (\mathbf{X} - \boldsymbol{\mu}_{\mathcal{X}})), \quad (5.27)$$

where $\tilde{\mathbf{r}}_i = \mathbf{r}_i / \|\mathbf{r}_i\|$. We can view the argument of each polynomial ψ_j in Eq. (5.27) as an affine transformation of the PLS-component $z = \tilde{\mathbf{r}}_i^T \mathbf{X}$. Then, expressing each $\psi_j(\beta z + \gamma)$ in Eq. (5.27) in terms of $\psi_j(z)$, where $\beta = \|\mathbf{r}_i\|$ and $\gamma = -(\mathbf{r}_i)^T \boldsymbol{\mu}_{\mathcal{X}}$, provides a representation of the PLS-PCE model that can be exactly transformed in the form of Eq. (5.23) even for small sample sizes (but in the asymptotic limit as well). We start with the following well-known product theorem [20, 43]:

$$\psi_j(\beta z) = \sqrt{j!} \sum_{l=0}^{\lfloor j/2 \rfloor} \frac{\beta^{j-2l} (\beta^2 - 1)^l}{\sqrt{(j-2l)! 2^l l!}} \psi_{j-2l}(z). \quad (5.28)$$

Moreover, expanding $\psi_j(z + \gamma)$ in a Taylor series around z yields

$$\psi_j(z + \gamma) = \sum_{t=0}^j \binom{j}{t} \sqrt{\frac{t!}{j!}} \gamma^{j-t} \psi_t(z). \quad (5.29)$$

Consequently,

$$\begin{aligned}
 \psi_j(\|\mathbf{r}_i\|(\tilde{\mathbf{r}}_i)^\top(\mathbf{X} - \boldsymbol{\mu}_X)) &\stackrel{(5.28)}{=} \sqrt{j!} \sum_{l=0}^{\lfloor j/2 \rfloor} \zeta_1(j, l) \psi_{j-2l}((\tilde{\mathbf{r}}_i)^\top(\mathbf{X} - \boldsymbol{\mu}_X)) \\
 &\stackrel{(5.29)}{=} \sqrt{j!} \sum_{l=0}^{\lfloor j/2 \rfloor} \zeta_1(j, l) \sum_{t=0}^{j-2l} \zeta_2(j, l, t) \psi_t((\tilde{\mathbf{r}}_i)^\top \mathbf{X}) \\
 &\stackrel{(5.22)}{=} j! \sum_{l=0}^{\lfloor j/2 \rfloor} \zeta_1(j, l) \sum_{t=0}^{j-2l} \zeta_2(j, l, t) \sum_{|\mathbf{k}|=t} \zeta_3(\mathbf{k}) \Psi_{\mathbf{k}}(\mathbf{X}),
 \end{aligned} \tag{5.30}$$

where

$$\zeta_1(j, l) = \frac{\|\mathbf{r}_i\|^{j-2l} (\|\mathbf{r}_i\|^2 - 1)^l}{\sqrt{(j-2l)! 2^l l!}} \tag{5.31}$$

$$\zeta_2(j, l, t) = \binom{j-2l}{t} \sqrt{\frac{t!}{(j-2l)!}} (-\mathbf{r}_i)^\top \boldsymbol{\mu}_X^{j-2l-t} \tag{5.32}$$

$$\zeta_3(\mathbf{k}) = \frac{\tilde{r}_{i1}^{k_1} \cdot \tilde{r}_{i2}^{k_2} \cdot \dots \cdot \tilde{r}_{id}^{k_d}}{\sqrt{k_1! \cdot k_2! \cdot \dots \cdot k_d!}}. \tag{5.33}$$

In the same way as for the asymptotic formulation, this yields cumulative coefficients for a given multivariate basis function associated with \mathbf{k} . An efficient way to compute the equivalent PCE-coefficients, which ensures the index set $\boldsymbol{\alpha}$ only has to be computed once, is presented by Alg. 5.

Algorithm 5 PLS-based sensitivities - non-asymptotic case

- 1: **Input** PLS components $\mathbf{a}^i, q_i, \mathbf{r}_i \forall i = 1, \dots, m$
 - 2: **Initialize** $\mathbf{c} = \mathbf{0}$
 - 3: $q_{max} = \max_{i \in \{1, \dots, m\}} (q_i)$
 - 4: Compute PCE multi-index set, e.g.: $\boldsymbol{\alpha} \leftarrow \text{multi_index}(d, q_{max}) \in \mathbb{N}^{P \times d}$
 - 5: **for** $i \leftarrow 1, m$ **do**
 - 6: **for** $j \leftarrow 0, q_i$ **do**
 - 7: **for** $l \leftarrow 0, \lfloor j/2 \rfloor$ **do**
 - 8: Compute $\zeta_1(j, l)$ from Eq. (5.31)
 - 9: **for** $t \leftarrow 0, j - 2l$ **do**
 - 10: Compute $\zeta_2(j, l, t)$ from Eq. (5.32)
 - 11: Get the multi-index subset of length t : $\mathbf{A} = \{\boldsymbol{\alpha}_p : |\boldsymbol{\alpha}_p| = t\}$
 - 12: **for** $p \leftarrow 1, \text{len}(\mathbf{A})$ **do**
 - 13: Get current multi-index $\mathbf{k} \leftarrow \boldsymbol{\alpha}_p \in \mathbb{N}^{1 \times d}$
 - 14: Compute $\tilde{\zeta}_3(\mathbf{k})$ from Eq. (5.33)
 - 15: Augment: $c_p \leftarrow c_p + j! a_{it}^{q_i} \zeta_1 \zeta_2 \zeta_3$
 - 16: $p \leftarrow p + 1$
 - 17: Compute sensitivities: $S, S^{\text{clo}}, S^T \leftarrow \text{PCE_sensitivities}(\boldsymbol{\alpha}, \mathbf{c})$
 - 18: **Return** S, S^{clo}, S^T
-

5.4.3 A comment on variance-based sensitivity analysis with the standard basis adaptation format

In the previous two subsections, we have derived expressions for the Sobol' and total-effect index of a basis-adapted PCE of the format in Eq. (5.12), which differs from the standard basis adaptation format in Eq. (5.11). While the former is a summation of m univariate PCEs of m latent variables, the latter represents an m -variate PCE of the latent variables. In the following we state an expression for the standard basis-adapted PCE format in terms of Hermite polynomials of the original inputs only. We write $\tilde{\mathbf{k}} \in \mathbb{N}^m$ and $\mathbf{k} \in \mathbb{N}^d$ to express multi-indices in the latent and original input spaces, respectively. Then,

$$\begin{aligned}
\hat{Y}_p^{\mathbf{Q}_m} &\stackrel{(5.11)}{=} \sum_{|\tilde{\mathbf{k}}| \leq p} a_{\tilde{\mathbf{k}}} \Psi_{\tilde{\mathbf{k}}}(\mathbf{Q}_m^T \mathbf{X}) \\
&\stackrel{(5.22)}{=} \sum_{|\tilde{\mathbf{k}}| \leq p} a_{\tilde{\mathbf{k}}} \prod_{j=1}^m \sum_{|\mathbf{k}|=\tilde{k}_j} \sqrt{|\mathbf{k}|!} \prod_{\ell=1}^d \frac{Q_{m,\ell j}^{k_\ell}}{\sqrt{k_\ell!}} \psi_\ell(X_\ell) \\
&= \sum_{|\tilde{\mathbf{k}}| \leq p} a_{\tilde{\mathbf{k}}} \prod_{j=1}^m \sum_{|\mathbf{k}|=\tilde{k}_j} \zeta_{\mathbf{k}} \Psi_{\mathbf{k}}(\mathbf{X}), \quad \zeta_{\mathbf{k}} = \sqrt{|\mathbf{k}|!} \prod_{\ell=1}^d \frac{Q_{m,\ell j}^{k_\ell}}{\sqrt{k_\ell!}}.
\end{aligned} \tag{5.34}$$

Back-transforming Eq. (5.34) to a standard PCE is non-trivial due to the inner summation, which induces multiple occurrences of the \mathbf{k} -th multivariate basis polynomial $\Psi_{\mathbf{k}}$ in the full expansion. Instead of back-transforming Eq. (5.34) to a standard PCE, one may plug it directly in the definitions of the variance-based sensitivity indices in Eqs. (5.15) and (5.17) and collect the partial variance contributions associated with any $\Psi_{\mathbf{k}}$. Such an approach is chosen in [28] for deriving variance-based sensitivities of canonical LRAs. We leave this task for a future work. We remark that for the case where the basis-adapted format of Eq. (5.11) does not consider mixed effects, i.e. it is $|\tilde{\mathbf{k}}| = \max(\tilde{k}_i)$, the coefficients of the equivalent standard PCE representation will be identical to the ones defined in Eq. (5.26) with r_{ij} set to $Q_{m,ij}$. In such case, Alg. 4 is directly applicable for computing the sensitivity indices.

5.5 Numerical experiments

In this section, we evaluate the proposed methods with one low-dimensional and one high-dimensional numerical experiment. We examine the performance of both the asymptotic approximation proposed in Alg. 4 and the exact computation performed with Alg. 5. The results are compared to those obtained with other surrogate modelling techniques, namely sparse PCE based on least-angle regression [5] and LRAs in the canonical polyadics format [28]. For all polynomial bases (sparse PCE, LRA, PLS-PCE) we prescribe a maximum degree of $p = 10$, for the LRA a maximum rank of $R = 10$ and for the PLS-PCE the same maximum number of components, i.e. $m = 10$. When computing LRAs, we use an adaptive scheme for the rank selection while considering every polynomial order up to p within the selected ranks [10]. All experimental designs are generated via latin hypercube sampling. In both examples, we draw 100 random experimental designs to quantify the relative error mean and standard deviation, where the error for a quantity Q with respect to its reference solution

Q_{ref} is defined as their difference:

$$\epsilon_Q = Q - Q_{ref}.$$

In our studies, Q refers to the Sobol' and the Total-effect indices and the respective reference solutions Q_{ref} are obtained with double-loop Monte Carlo and the estimators stated in [26, 42] depending on the example. The latter estimators are based on drawing a single set of n independent input samples of dimensions d and splitting them in two equally sized subsets. Permuting columns between the subsets yields a dependent sample of size $n(d+2)/2$ based on which all Sobol' and total-effect indices can be estimated. We refer to these as permutation matrix estimators.

5.5.1 Elastic truss

We consider an elastic truss that consists of 23 rods as depicted in Fig. 5.1 [30]. Horizontal and diagonal rods have cross-sections A_1 , A_2 and Young's moduli E_1 , E_2 , respectively. The truss sustains 6 vertical point loads $P_1 - P_6$. The input variable definitions are provided in Tab. 5.1. We compute

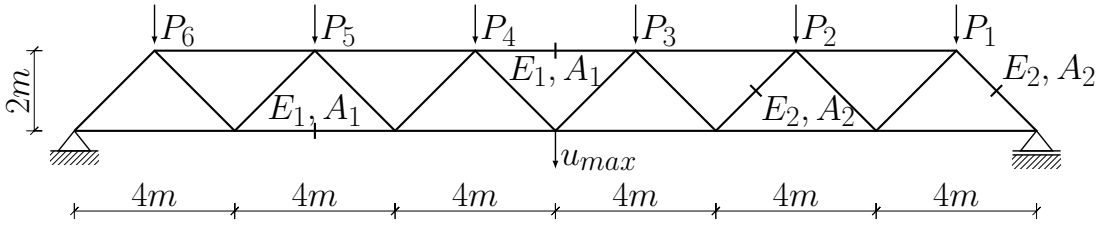


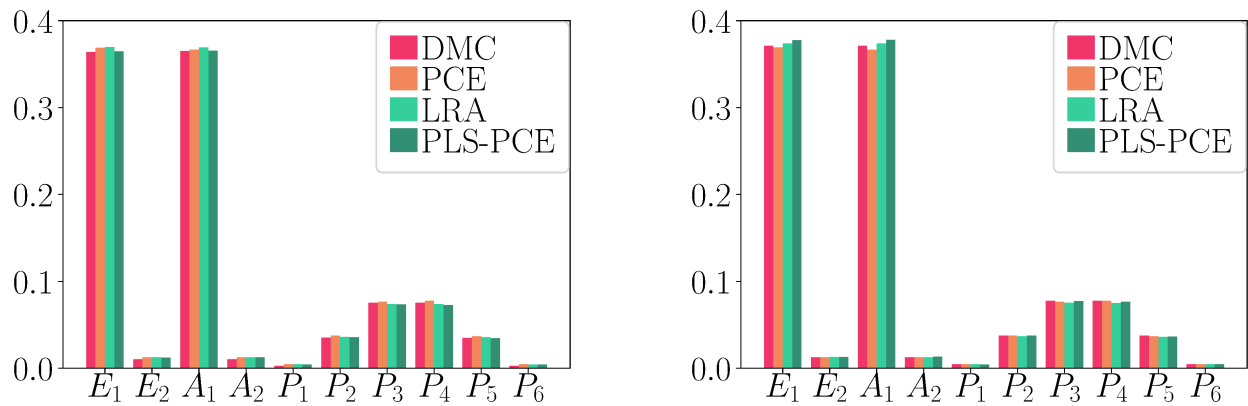
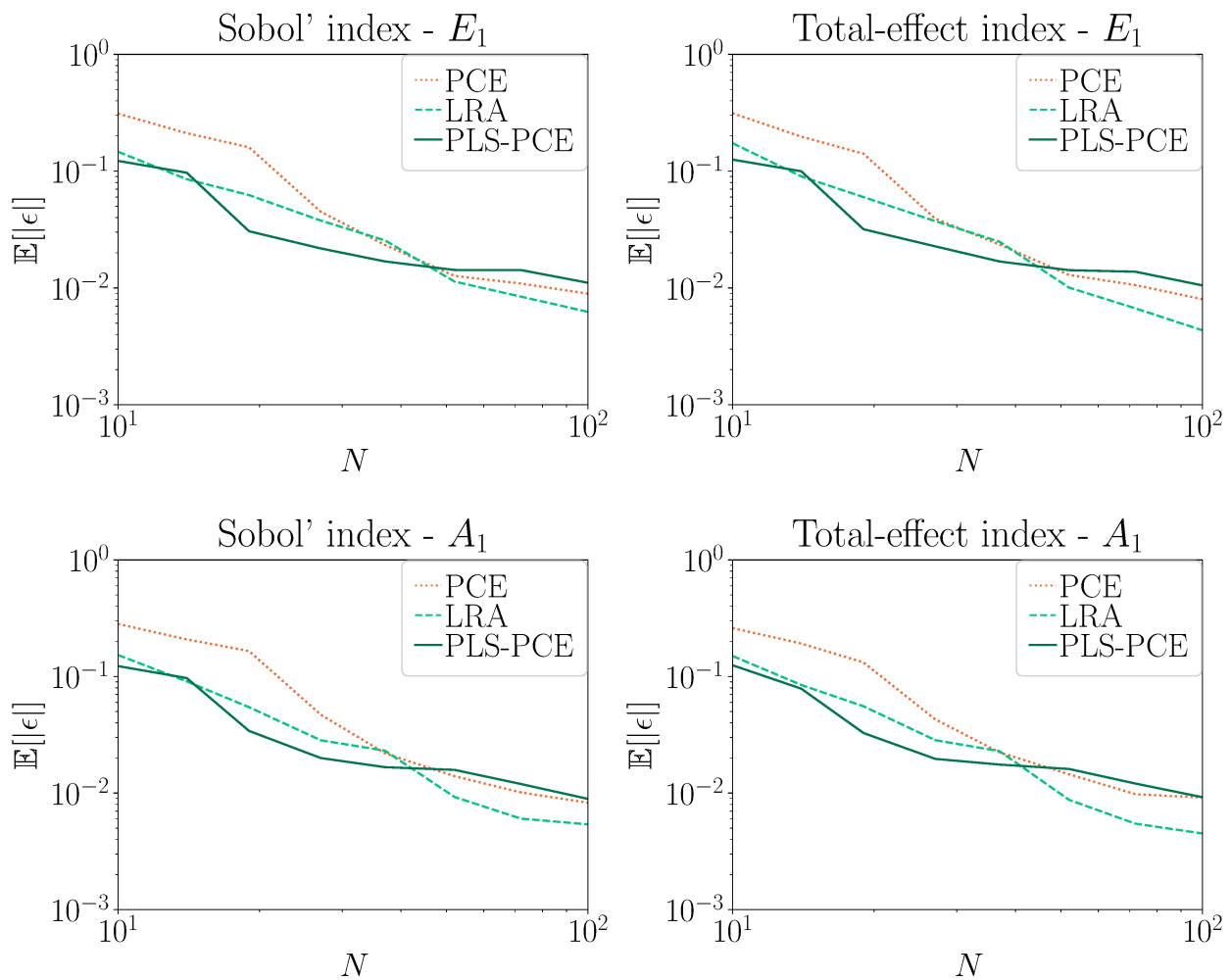
Figure 5.1: 2-D truss example.

Table 5.1: Input variable definitions of the truss example.

Random Variable	Distribution	Mean	Standard deviation
A_1 [m^2]	Log-Normal	$2 \cdot 10^{-3}$	$2 \cdot 10^{-4}$
A_2 [m^2]	Log-Normal	$1 \cdot 10^{-3}$	$1 \cdot 10^{-4}$
E_1, E_2 [Pa]	Log-Normal	$2.1 \cdot 10^{11}$	$2.1 \cdot 10^{10}$
$P_1 - P_6$ [N]	Gumbel	$5.0 \cdot 10^4$	$7.5 \cdot 10^3$

Sobol' and total-effect indices for the maximum truss deflection u_{max} using Alg. 4. Reference solutions (direct Monte Carlo - DMC) are obtained based on $n = 10^6$ independent samples (Fig. 5.2) with the permutation matrix estimators.

Fig. 5.2 indicates good agreement of the PLS-PCE-based sensitivities with the reference solution. Fig. 5.3 shows that all three surrogate-based sensitivity indices are estimated with similar mean relative error and convergence rate as N increases. Fig. 5.4 indicates the same for the relative error variance. Fig. 5.5 shows a performance comparison of Algs. 4 and 5 for the truss. Here, the

Figure 5.2: Sobol' (left) and total-effect (right) indices of u_{max} obtained with $N = 100$.Figure 5.3: Mean relative errors for the two most influential inputs E_1 and A_1 , computed with PCE, LRA and PLS-PCE.

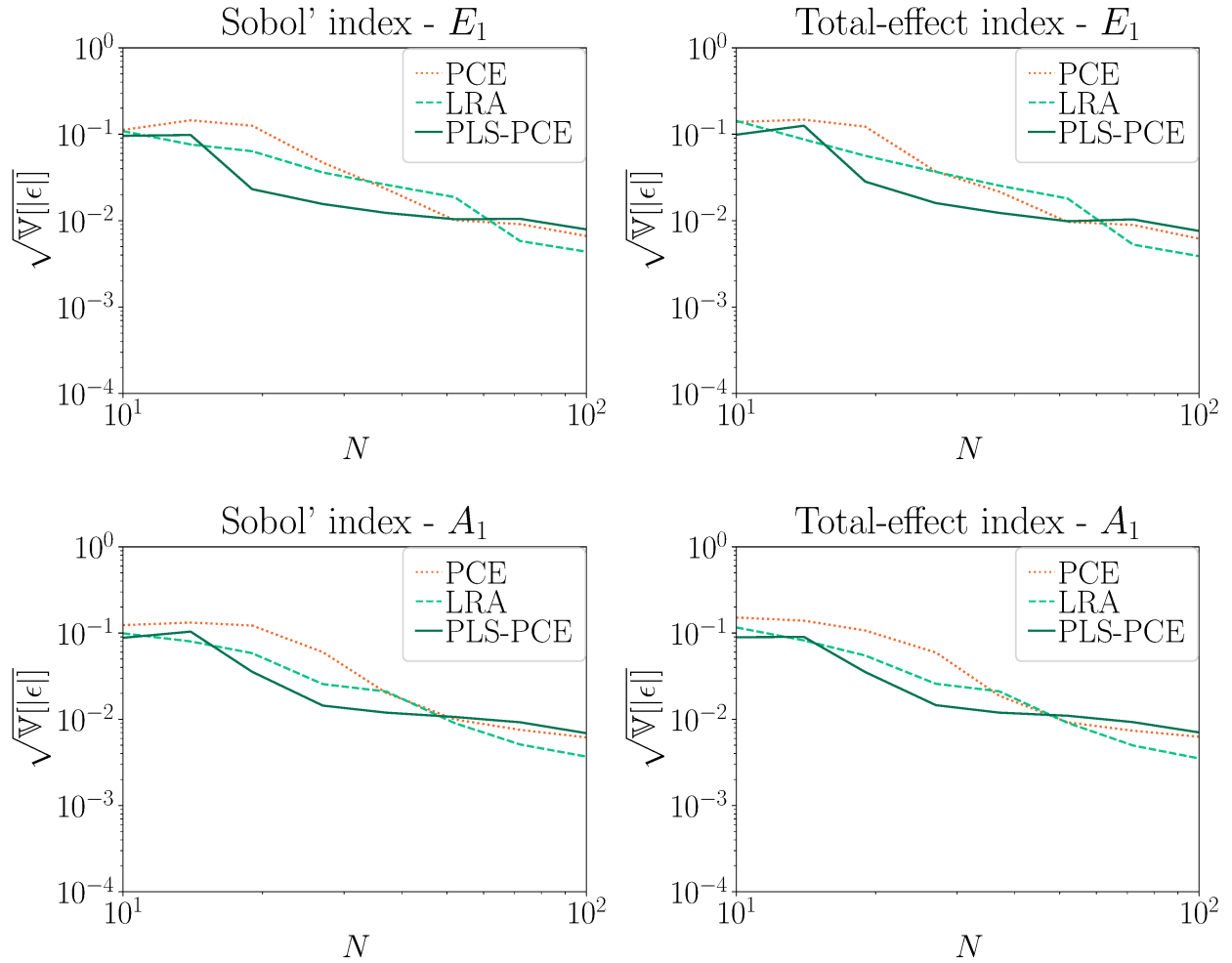


Figure 5.4: Relative error standard deviation for the two most influential inputs E_1 and A_1 , computed with sparse PCE, LRA and PLS-PCE.

asymptotic approximation introduces only negligible error into the sensitivity estimate. For variables of little significance, which yield small absolute values for the corresponding sensitivity indices (e.g. E_2 , A_2), the surrogate modelling error is much larger than the error due to the asymptotic approximation of the sensitivity indices so that there is no visible difference left in the error plots.

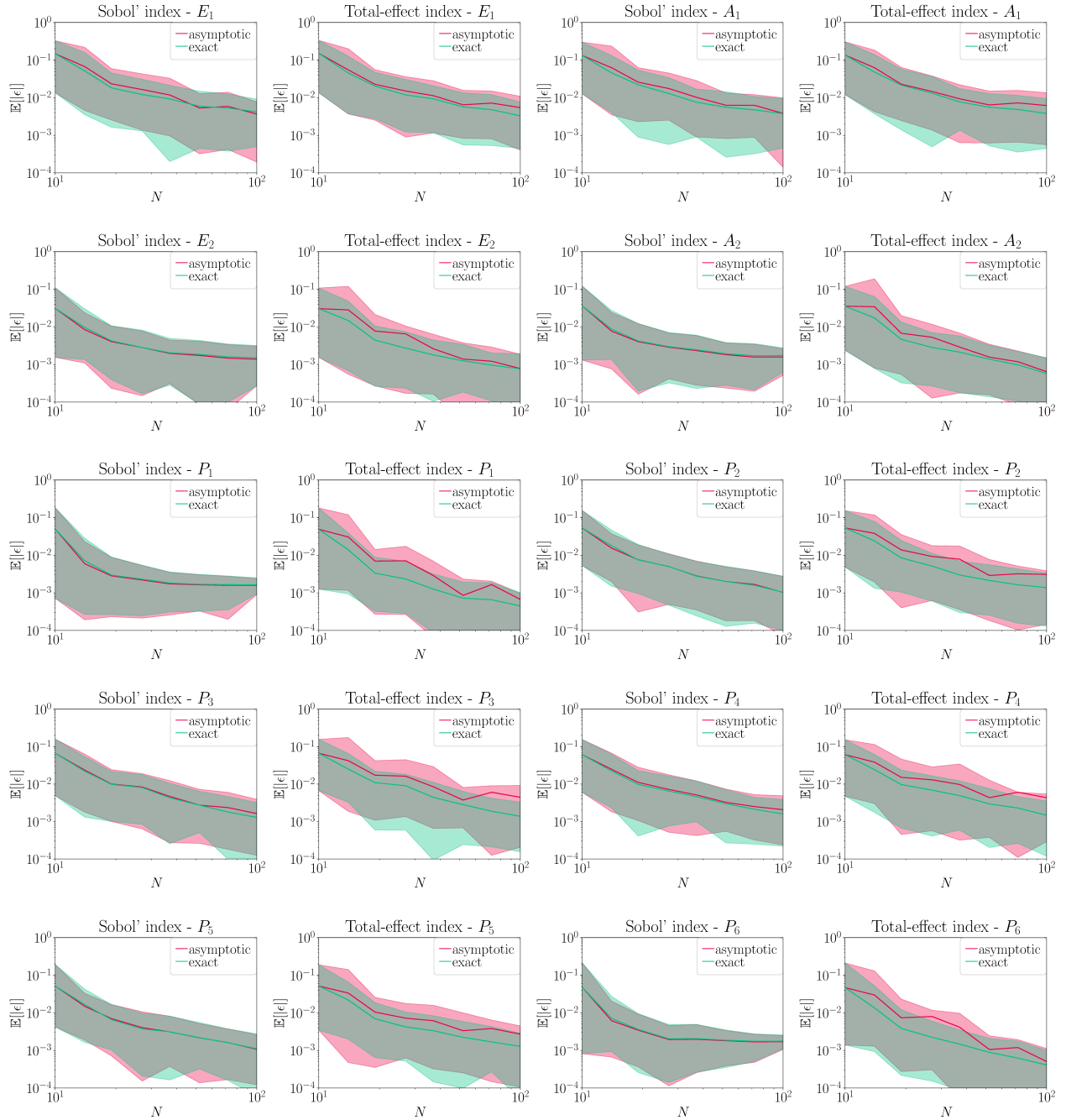


Figure 5.5: Mean relative errors with 90% CI, computed with asymptotic approximation (Alg. 4) and exactly (Alg. 5).

5.5.2 Steel plate

For the second example, we consider a modified version of the example given in [33], which consists of a low-carbon steel plate of length 0.32 m, width 0.32 m, thickness $t = 0.01$ m, and a hole of radius 0.02 m located at the center. The Poisson ratio is set to $\nu = 0.29$ and the density of the plate is $\rho = 7850$ kg/m³. The horizontal and vertical displacements are constrained at the left edge. Both the surface load q , which acts on the right plate side, and the plate's Young's modulus $E(x, y)$ are considered uncertain and spatially variable. Both are described by homogeneous random fields in two and one spatial dimension, respectively. E has log-normal marginal distribution, mean value $\mu_E = 2 \times 10^5$ MPa and standard deviation $\sigma_E = 3 \times 10^4$ MPa. The autocorrelation function of the underlying Gaussian field $\ln E$ is modeled by the isotropic exponential model

$$\rho_{\ln E}(\Delta x, \Delta y) = \exp(-\sqrt{\Delta x^2 + \Delta y^2}/l_E) \quad (5.35)$$

with correlation length $l_E = 0.08$ m. The random field $\ln E$ is discretized by a Karhunen-Loève-expansion (KLE), i.e.

$$E(x, y) = \exp \left\{ \mu_{\ln E} + \sigma_{\ln E} \sum_{i=1}^{d_E} \sqrt{\lambda_i^E} \varphi_i^E(x, y) \xi_i^E \right\}, \quad (5.36)$$

where $\mu_{\ln E}$ and $\sigma_{\ln E}$ are the parameters of the log-normal marginal distribution of E , $\{\lambda_i^E, \varphi_i^E\}$ are the eigenpairs of the correlation kernel Eq. (5.35) and $\boldsymbol{\xi}^E \in \mathbb{R}^{d_E \times 1}$ is a standard-normal random vector. The number of terms in the expansion d_E is chosen such that 90% of the random field variance is represented by the discretization in Eq. (5.36), which yields $d_E = 169$. Selected eigenfunctions of $\rho_{\ln E}$ are shown in Fig. 5.7. q also has log-normal marginal distribution with mean value $\mu_q = 60$ MPa and standard deviation $\sigma_q = 12$ MPa. The autocorrelation function of the underlying Gaussian field $\ln q$ is also modeled by an isotropic exponential model,

$$\rho_{\ln q}(\Delta y) = \exp(-|\Delta y|/l_q) \quad (5.37)$$

with correlation length $l_q = 0.02$ m. The random field $\ln q$ is also discretized by a KLE, s.t.

$$q(y) = \exp \left\{ \mu_{\ln q} + \sigma_{\ln q} \sum_{i=1}^{d_q} \sqrt{\lambda_i^q} \varphi_i^q(y) \xi_i^q \right\}, \quad (5.38)$$

where $\mu_{\ln q}$ and $\sigma_{\ln q}$ are the parameters of the log-normal marginal distribution of q , $\{\lambda_i^q, \varphi_i^q\}$ are the eigenpairs of the load correlation kernel Eq. (5.37) and $\boldsymbol{\xi}^q \in \mathbb{R}^{d_q \times 1}$ is a standard-normal random vector constituting the model input of the plate together with $\boldsymbol{\xi}^E$, i.e. $\mathbf{X} = [\boldsymbol{\xi}^q, \boldsymbol{\xi}^E]^T$. With the same accuracy criterion on the represented random field variance as for E ($> 90\%$), one obtains $d_q = 32$. The first 4 eigenfunctions of $\rho_{\ln q}$ are shown in Fig. 5.8.

The stress field

$$\boldsymbol{\sigma}(x, y) = [\sigma_x(x, y), \sigma_y(x, y), \tau_{xy}(x, y)]^T,$$

strain field

$$\boldsymbol{\epsilon}(x, y) = [\epsilon_x(x, y), \epsilon_y(x, y), \gamma_{xy}(x, y)]^T$$

and displacement field

$$\mathbf{u}(x, y) = [u_x(x, y), u_y(x, y)]^T$$

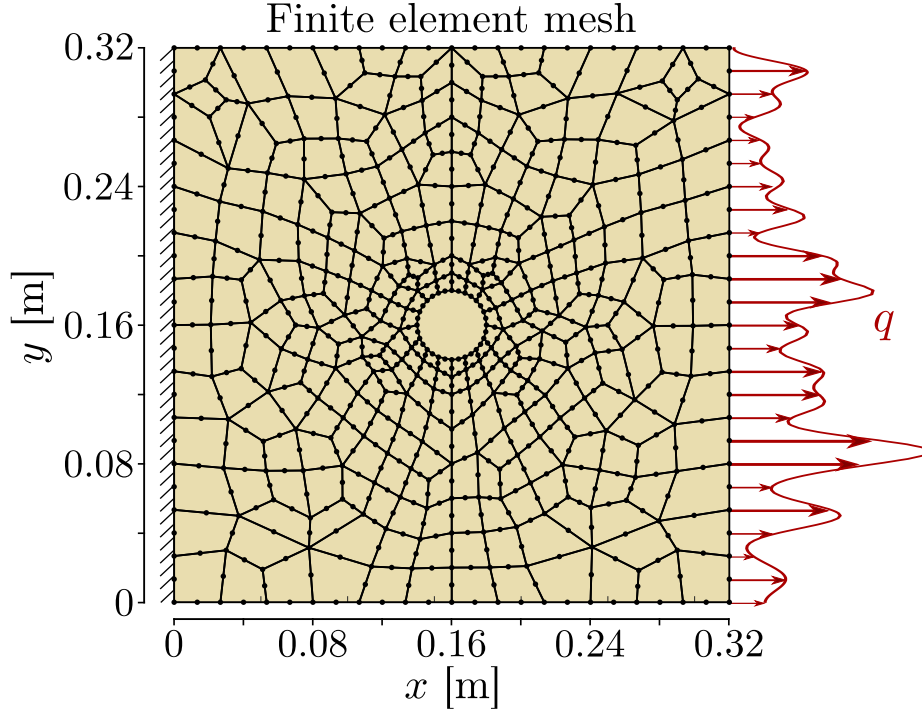


Figure 5.6: FE-mesh of 2D-plate model with uncertain Young's modulus E subject to random surface load q .

of the plate are given through elasticity theory, namely the Cauchy-Navier equations [27]. Given the configuration of the plate, the model can be simplified under the plane stress hypothesis, which yields

$$G(x, y) \nabla^2 \mathbf{u}(x, y) + \frac{E(x, y)}{2(1 - \nu)} \nabla(\nabla \cdot \mathbf{u}(x, y)) + \mathbf{b} = 0. \quad (5.39)$$

Therein, $G(x, y) = E(x, y)/(2(1 + \nu))$ is the shear modulus, and $\mathbf{b} = [b_x, b_y]^T$ is the vector of body forces acting on the plate. Eq. (5.39) is discretized with a finite-element method. That is, the spatial domain of the plate is discretized into 282 eight-noded quadrilateral elements, as shown in Fig. 5.6. The scalar model output is the maximally occurring first principal plane stress

$$\sigma_1 = 0.5(\sigma_x + \sigma_y) + \sqrt{[0.5(\sigma_x + \sigma_y)]^2 + \tau_{xy}^2}.$$

The FE-model of the plate with random inputs is illustrated in Fig. 5.6. We compute sensitivity indices for both the random variables characterizing the uncertainty associated with the single modes of the KL-decompositions as well as for the two random fields as such. The random field sensitivity analysis can be understood as interpreting each random field as a single input. Thus the variance decomposition of the model output with respect to its inputs is computed once with respect to each single random variable input and once with respect to two subgroups of these random variables each characterizing one of the random fields in the problem description. Random field-oriented Sobol' indices are always of closed form (see Eq. (5.16)) as the classical Sobol' index would conceal most of the contributed variance. The PLS-PCE-based sensitivity indices are computed with Alg. 4 and compared against LRA-based sensitivities and Monte Carlo reference solutions. The latter (DMC) are obtained with $4 \cdot 10^6$ samples using the double-loop ($2 \cdot 10^3$ samples per stage) to compute the

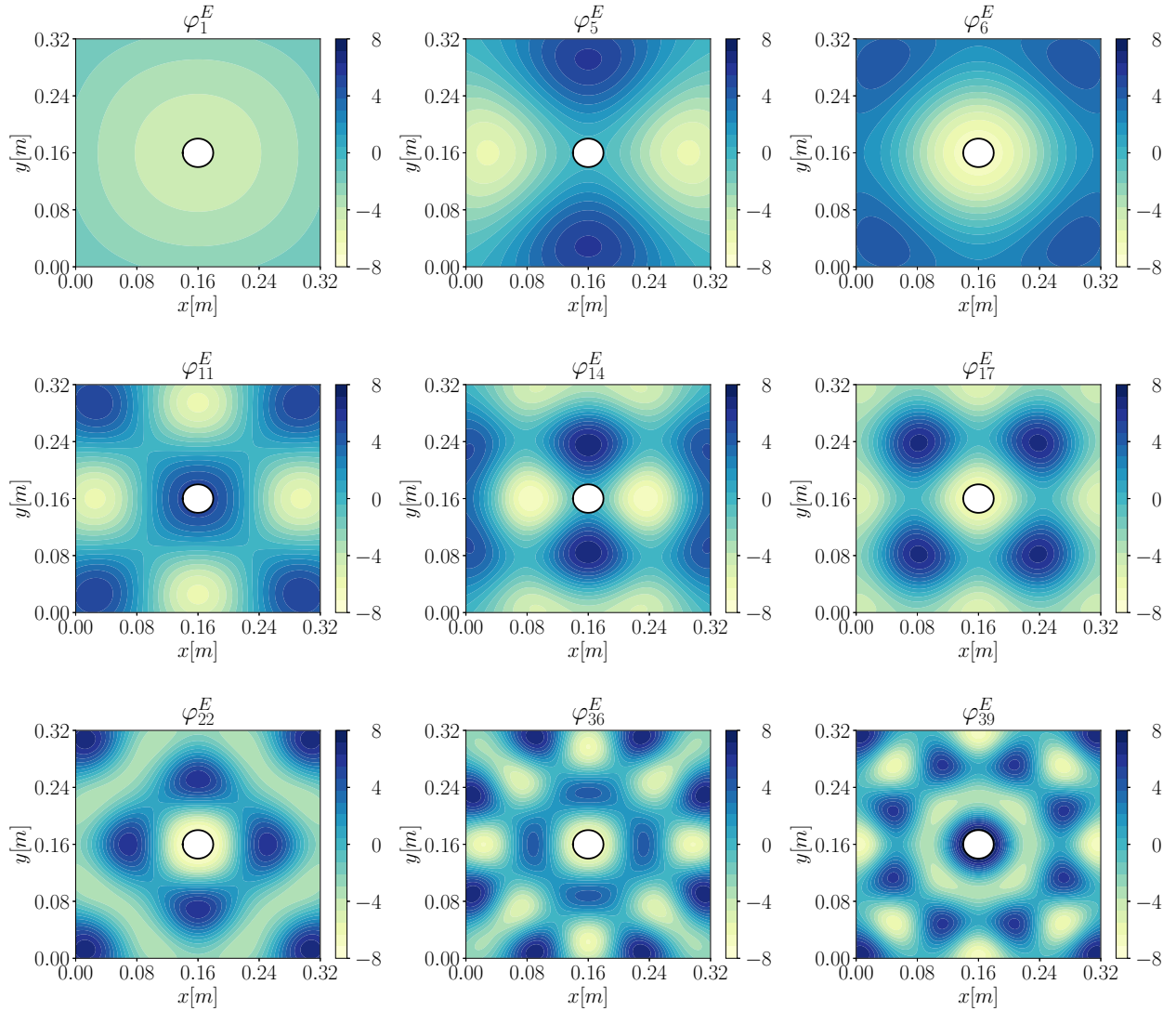


Figure 5.7: The nine most important eigenfunctions of the exponential, isotropic Young's modulus correlation kernel $\rho_{\ln E}$.

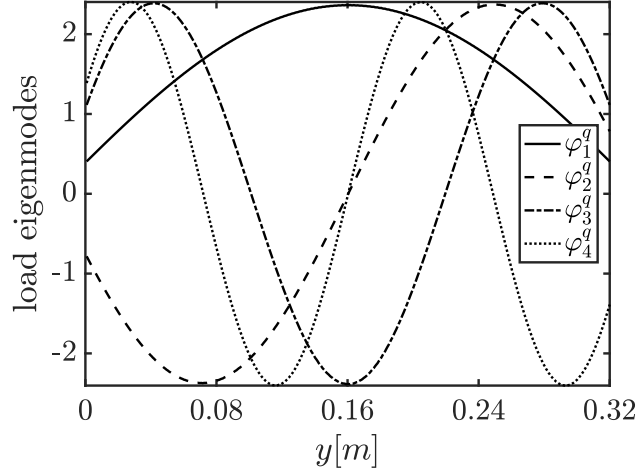


Figure 5.8: First four eigenfunctions of exponential load correlation kernel $\rho_{\ln q}, \varphi_1^q(y) - \varphi_4^q(y)$.

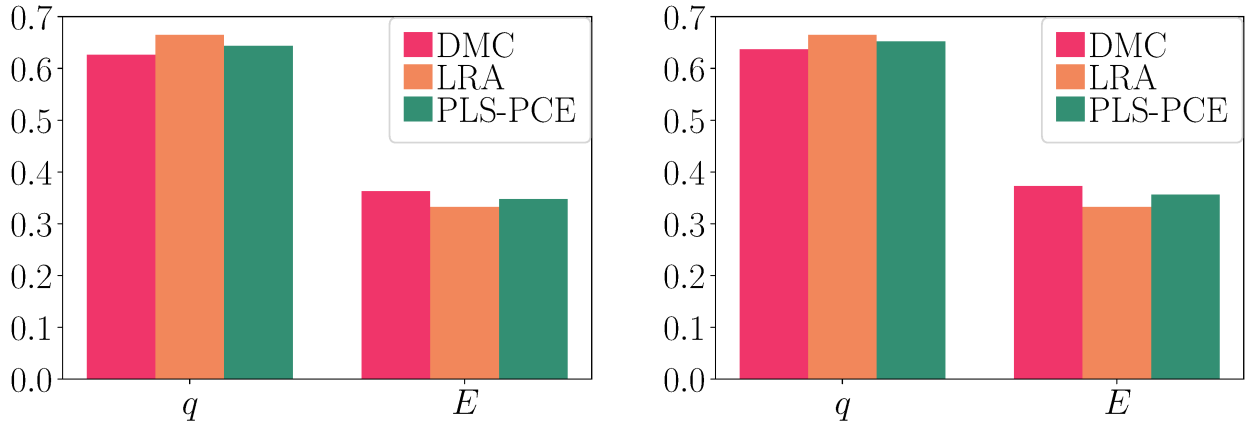


Figure 5.9: Closed Sobol' (left) and total-effect (right) indices for the two random fields q and E obtained with $N = 10^3$.

random field sensitivities and with $n = 2 \times 10^4$ independent samples using the permutation matrix estimators to compute the random variable sensitivities. For this application, sparse PCE surrogates are difficult to obtain beyond relatively low total polynomial degrees ($p \leq 3$) due to the large input dimension d . PCE-based sensitivity indices are thus not considered in this study.

The random field-oriented sensitivity index means are plotted in Fig. 5.9 and attribute a larger influence to the random load field $q(y)$ compared to the material parameter field $E(x, y)$. The PLS-PCE-based indices are consistently closer to the reference than the LRA-based indices while both surrogate-based indices approximate the reference well. According to Fig. 5.10, both surrogate indices converge in standard deviation with increasing N while the LRA-based mean relative error does not decrease further beyond a certain sample size. The PLS-PCEs are superior in this respect as the corresponding mean relative error decreases further as $N \rightarrow 10^3$. The means of the sensitivity indices of the random variables corresponding to the KL eigenmodes (the 10 most important ones) are plotted in Fig. 5.11. The random variable corresponding to the first mode of the load random field is by far the most important input. The next 9 random variable inputs in the ranking all

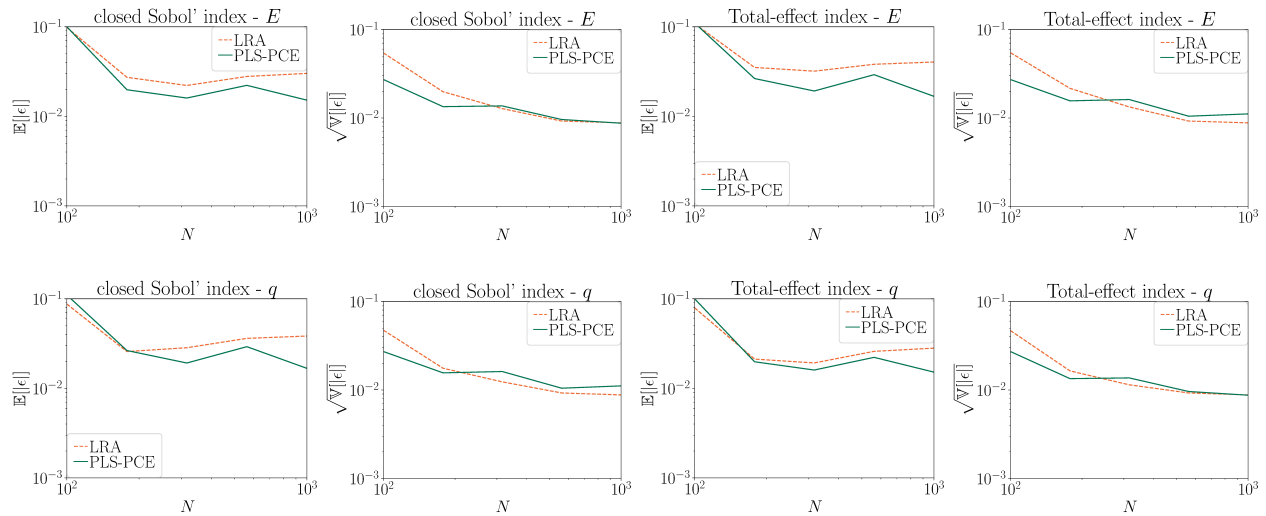


Figure 5.10: Relative error mean and standard deviation for sensitivity estimates of the maximum first main stress to q and E .

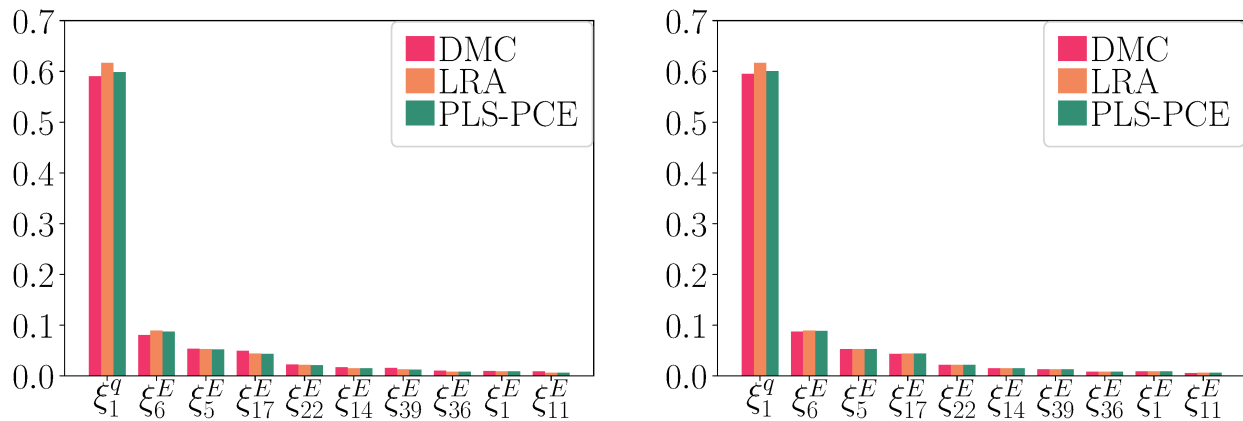


Figure 5.11: Sobol' (left) and total-effect (right) indices for the ten most important random field modes obtained with $N = 10^3$.

correspond to modes of the material parameter field $E(x, y)$ and their associated eigenfunctions are plotted in Fig. 5.7. For the largest mode-oriented sensitivity indices, the PLS-PCE yields slightly better approximations of the reference than the LRA while both approximations are fairly accurate.

Figs. 5.12 and 5.13 show the convergence study of the mode-oriented Sobol' and total-effect indices as N increases. The dominant random variable ξ_1^q reproduces the convergence behaviour of the random field-oriented sensitivity study. Namely, the mean relative error of the LRA-based sensitivity reaches a plateau and increases again after reaching a certain experimental design size ($N \approx 300$). The remainder of the 10 most important inputs (all E -modes) show consistent convergence in the relative error mean and standard deviation as N increases.

5.6 Conclusion

This paper derives analytical expressions for variance-based sensitivity indices of PLS-PCE surrogate model outputs and formulates two algorithms for their efficient computation. The expressions for the sensitivities involve only the surrogate model coefficients. Thus, these sensitivities can be computed with negligible additional computational effort once the surrogate model is identified by a mere post-processing of the model coefficients. For the first algorithm, a multinomial theorem for Hermite polynomials is applied to derive expressions for the sensitivity measures based on the model coefficients, which is asymptotically exact. That is, with the number of samples $N \rightarrow \infty$, these estimators exactly match the theoretical Sobol' and total-effect indices of the surrogate model. For the second algorithm, corrected expressions are derived, which are exact also when N is small. These can be used when the experimental design is small. The sensitivity estimates obtained with both algorithms are compared to Monte-Carlo-based reference solutions as well as estimates based on sparse PCEs and LRAs for two different numerical examples: A low-dimensional ($d = 10$) truss example as well as a high-dimensional plate example ($d = 201$). For both examples, the PLS-PCE-based sensitivity estimates approximate the reference solution well and perform at least as good as the two alternative surrogate-based estimators. Finally, we comment on the applicability of our approach to general basis adaptation formats: we recast the standard basis adapted format in terms of the original input but find that a back-transformation to a standard PCE format based on the stated expression is non-trivial. Instead, sampling-free variance-based sensitivity indices may be computed for standard basis-adapted PCEs by analysing the stated expression term by term as done for canonical LRAs in [28]. We leave this to future research. The presented procedure can be extended to multivariate output in combination with the PLS2 algorithm [31].

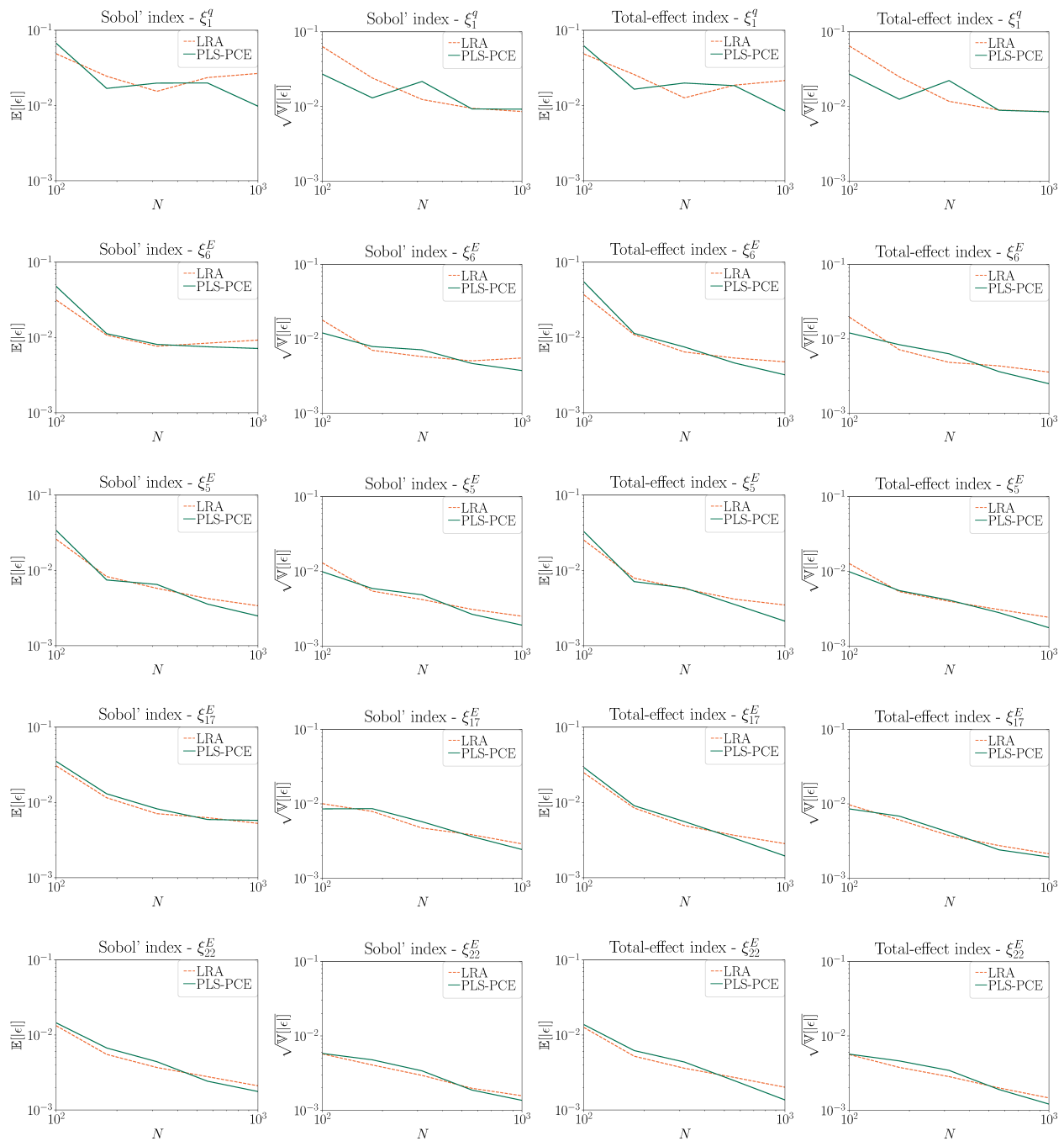


Figure 5.12: Relative error mean and standard deviation for the Sobol' and total-effect indices of the ten (1-5.) most influential model inputs computed with LRA and PLS-PCE over varying experimental design size N .

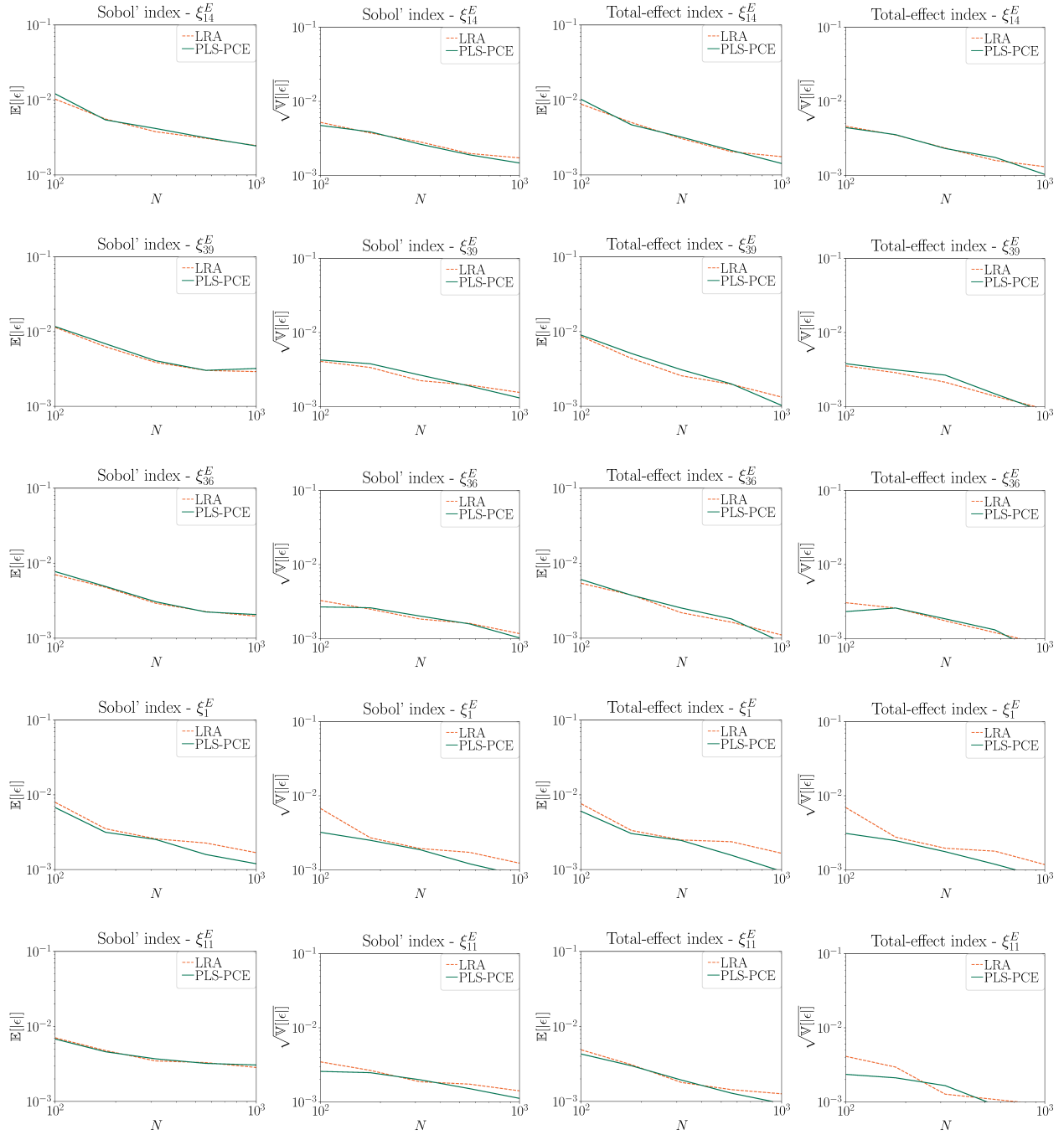


Figure 5.13: Relative error mean and standard deviation for the Sobol' and total-effect indices of the ten (6-10.) most influential model inputs computed with LRA and PLS-PCE over varying experimental design size N .

5.7 Acknowledgment

This project was supported by the German Research Foundation (DFG) through Grant STR 1140/6-1 under SPP 1886.

References

- [1] F. M. Alam, K. R. McNaught, and T. J. Ringrose. “Using Morris’ randomized OAT design as a factor screening method for developing simulation metamodels”. In: *Simulation Conference, 2004. Proceedings of the 2004 Winter*. Vol. 1. IEEE. 2004.
- [2] G. Archer, A. Saltelli, and I. Sobol. “Sensitivity measures, ANOVA-like techniques and the use of bootstrap”. In: *Journal of Statistical Computation and Simulation* 58.2 (1997), pp. 99–120.
- [3] G. Baffi, E. Martin, and A. Morris. “Non-linear projection to latent structures revisited (the neural network PLS algorithm)”. In: *Computers & Chemical Engineering* 23.9 (1999), pp. 1293–1307.
- [4] M. Berveiller, B. Sudret, and M. Lemaire. “Stochastic finite element: a non intrusive approach by regression”. In: *European Journal of Computational Mechanics* 15.1-3 (2006), pp. 81–92.
- [5] G. Blatman and B. Sudret. “Adaptive sparse polynomial chaos expansion based on least-angle regression”. In: *Journal of Computational Physics* 230.6 (2011), pp. 2345–2367.
- [6] E. Borgonovo et al. “Transformations and invariance in the sensitivity analysis of computer experiments”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76.5 (2014), pp. 925–947.
- [7] E. Borgonovo. “A new uncertainty importance measure”. In: *Reliability Engineering & System Safety* 92.6 (2007), pp. 771–784.
- [8] E. Borgonovo and E. Plischke. “Sensitivity analysis: a review of recent advances”. In: *European Journal of Operational Research* 248.3 (2016), pp. 869–887.
- [9] F. Buet-Golfouse. “A Multinomial Theorem for Hermite Polynomials and Financial Applications”. In: *Applied Mathematics* 06 (2015), pp. 1017–1030.
- [10] M. Chevreuil et al. “A Least-Squares Method for Sparse Low Rank Approximation of Multivariate Functions”. In: *SIAM/ASA Journal on Uncertainty Quantification* 3.1 (2015), pp. 897–921.
- [11] R. Cukier et al. “Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory”. In: *The Journal of chemical physics* 59.8 (1973), pp. 3873–3878.
- [12] A. Doostan and H. Owhadi. “A non-adapted sparse approximation of PDEs with stochastic inputs”. In: *Journal of Computational Physics* 230.8 (2011), pp. 3015–3034.
- [13] B. Efron and C. Stein. “The jackknife estimate of variance”. In: *The Annals of Statistics* (1981), pp. 586–596.
- [14] M. Ehre, I. Papaioannou, and D. Straub. “Global sensitivity analysis in high dimensions with PLS-PCE”. In: *Reliability Engineering & System Safety* 198 (2020), p. 106861.

- [15] P.-A. Ekström and R. Broed. “Sensitivity analysis methods and a biosphere test case implemented in EIKOS”. In: *Posiva Working Report 31* (2006), p. 84.
- [16] A. I. J. Forrester, A. Sobester, and A. J. Keane. *Engineering Design via Surrogate Modelling - A Practical Guide*. Wiley, 2008.
- [17] G. Greengar and C. Manohar. “Global response sensitivity analysis of uncertain structures”. In: *Structural Safety* 58 (2016), pp. 94–104.
- [18] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., 2001.
- [19] J. C. Helton et al. “Survey of sampling-based methods for uncertainty and sensitivity analysis”. In: *Reliability Engineering & System Safety* 91.10-11 (2006), pp. 1175–1209.
- [20] T. Hida et al. *White noise : an infinite dimensional calculus*. Mathematics and Its Applications 253. Dordrecht: Springer, 1993.
- [21] T. Homma and A. Saltelli. “Importance measures in global sensitivity analysis of nonlinear models”. In: *Reliability Engineering & System Safety* 52.1 (1996), pp. 1–17.
- [22] A. Höskuldsson. “PLS regression methods”. In: *Journal of Chemometrics* 2.3 (1988), pp. 211–228.
- [23] A. Höskuldsson. “PLS regression methods”. In: *Journal of Chemometrics* 2.3 (1988), pp. 211–228.
- [24] B. Iooss and P. Lemaître. “A review on global sensitivity analysis methods”. In: *Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications*. Ed. by C. Meloni and G. Dellino. Springer, 2015.
- [25] J. Iott, R. T. Haftka, and H. M. Adelman. “Selecting step sizes in sensitivity analysis by finite differences”. In: *NASA Technical Memorandum 86382* (Aug. 1985).
- [26] M. J. Jansen. “Analysis of variance designs for model output”. In: *Computer Physics Communications* 117.1 (1999), pp. 35–43.
- [27] C. Johnson. *Numerical solution of partial differential equations by the finite element method*. Dover Publications, 2009. 288 pp.
- [28] K. Konakli and B. Sudret. “Global sensitivity analysis using low-rank tensor approximations”. In: *Reliability Engineering & System Safety* 156.Supplement C (2016), pp. 64–83.
- [29] S. Kucherenko et al. “Global sensitivity indices for nonlinear mathematical models, Review”. In: *Wilmott Mag* 1 (2005), pp. 56–61.
- [30] S. H. Lee and B. M. Kwak. “Response surface augmented moment method for efficient reliability analysis”. In: *Structural Safety* 28.3 (2006), pp. 261–272.
- [31] L. Li, I. Papaioannou, and D. Straub. *Partial least squares-based polynomial chaos expansion for global sensitivity analysis of dynamic models in high dimensions*. Manuscript. 2019.
- [32] L. Li et al. “Moment-independent importance measure of basic variable and its state dependent parameter solution”. In: *Structural Safety* 38 (2012), pp. 40–47.
- [33] P.-L. Liu and K.-G. Liu. “Selection of random field mesh in finite element reliability analysis”. In: *Journal of Engineering Mechanics* 119.4 (1993), pp. 667–680.
- [34] Q. Liu and T. Homma. “A new importance measure for sensitivity analysis”. In: *Journal of Nuclear Science and Technology* 47.1 (2010), pp. 53–61.

- [35] I. Papaioannou, K. Breitung, and D. Straub. “Reliability sensitivity analysis with Monte Carlo methods”. In: *Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures*. CRC Press, 2014, pp. 5335–5342.
- [36] I. Papaioannou, K. Breitung, and D. Straub. “Reliability sensitivity estimation with sequential importance sampling”. In: *Structural Safety* 75 (2018), pp. 24–34.
- [37] I. Papaioannou, M. Ehre, and D. Straub. “PLS-based adaptation for efficient PCE representation in high dimensions”. In: *Journal of Computational Physics* 387 (2019), pp. 186–204.
- [38] C. Prieur and S. Tarantola. “Variance-Based Sensitivity Analysis: Theory and Estimation Algorithms”. In: *Handbook of Uncertainty Quantification*. Ed. by R. Ghanem, D. Higdon, and H. Owhadi. Springer International Publishing, June 2017, pp. 1217–1239.
- [39] M. Rosenblatt. “Remarks on a multivariate transformation”. In: *The Annals of Mathematical Statistics* 23.3 (1952), pp. 470–472.
- [40] R. Rosipal. “Nonlinear Partial Least Squares: An Overview”. In: *Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques* (Jan. 2010), pp. 169–189.
- [41] A. Saltelli, K. Chan, and E. Scott. *Sensitivity Analysis*. John Wiley & Sons, Inc., 2000.
- [42] A. Saltelli et al. “Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index”. In: *Computer Physics Communications* 181 (2010), pp. 259–270.
- [43] E. Schlögl. “Option pricing where the underlying assets follow a Gram/Charlier density of arbitrary order”. In: *Journal of Economic Dynamics and Control* 37.3 (2013), pp. 611–632.
- [44] I. M. Sobol. “Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates”. In: *Mathematics and computers in simulation* 55.1-3 (2001), pp. 271–280.
- [45] I. Sobol’. “Sensitivity Estimates for Nonlinear Mathematical Models”. In: *Math. Modeling & Comp. Exp* 1 (1993), pp. 407–414.
- [46] B. Sudret. “Global sensitivity analysis using polynomial chaos expansions”. In: *Reliability Engineering & System Safety* 93.7 (2008), pp. 964–979.
- [47] B. Sudret. “Meta-models for Structural Reliability and Uncertainty Quantification”. In: *Proc. 5th Asian-Pacific Symp. Struct. Reliab. (APSSRA 2012)*, Singapore (Mar. 2012).
- [48] B. Sudret and A. Der Kiureghian. *Stochastic finite element methods and reliability: a state-of-the-art report*. Department of Civil and Environmental Engineering, University of California, 2000.
- [49] R. Tipireddy and R. Ghanem. “Basis adaptation in homogeneous chaos spaces”. In: *Journal of Computational Physics* 259 (2014), pp. 304–317.
- [50] P. Tsilifis et al. “Compressive sensing adaptation for polynomial chaos expansions”. In: *Journal of Computational Physics* 380 (2019), pp. 29–47.
- [51] P. Wei, Z. Lu, and J. Song. “Variable importance analysis: a comprehensive review”. In: *Reliability Engineering & System Safety* 142 (2015), pp. 399–432.
- [52] S. Wold, N. Kettaneh-Wold, and B. Skagerberg. “Nonlinear PLS modeling”. In: *Chemometrics and Intelligent Laboratory Systems* 7.1 (1989). Proceedings of the First Scandinavian Symposium on Chemometrics, pp. 53–65.

- [53] S. Wold, M. Sjöström, and L. Eriksson. “PLS-regression: a basic tool of chemometrics”. In: *Chemometrics and Intelligent Laboratory Systems* 58.2 (2001), pp. 109–130.
- [54] S. Wold et al. “The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses”. In: *SIAM Journal on Scientific and Statistical Computing* 5.3 (1984), pp. 735–743.
- [55] D. Xiu and G. E. Karniadakis. “The Wiener–Askey polynomial chaos for stochastic differential equations”. In: *SIAM Journal on Scientific Computing* 24.2 (2002), pp. 619–644.
- [56] L. Yan, L. Guo, and D. Xiu. “Stochastic collocation algorithms using ℓ_1 -minimization”. In: *International Journal for Uncertainty Quantification* 2 (Jan. 2012), pp. 279–293.
- [57] C. Zhou, Z. Lu, and W. Li. “Sparse grid integration based solutions for moment-independent importance measures”. In: *Probabilistic Engineering Mechanics* 39 (2015), pp. 46–55.

Sequential, active learning of low-dimensional model representations for reliability analysis

Original Publication

M. Ehre, I. Papaioannou, B. Sudret, and D. Straub. *Active sequential learning of low-dimensional model representations for reliability analysis*. Submitted to SIAM J. Comp. Sci. 2021.

Author's contribution

Max Ehre and Iason Papaioannou developed the concepts underlying the method. Max Ehre developed and implemented the algorithm and carried out all numerical experiments. Max Ehre drafted the manuscript under the guidance of Iason Papaioannou, Daniel Straub and Bruno Sudret.

Abstract

To date, the analysis of high-dimensional, computationally expensive engineering models remains a difficult challenge in risk and reliability engineering. We use a combination of dimensionality reduction and surrogate modelling termed partial least squares-driven polynomial chaos expansion (PLS-PCE) to render such problems feasible. Standalone surrogate models typically perform poorly for reliability analysis. Therefore, in a previous work, we have used PLS-PCEs to reconstruct the intermediate densities of a sequential importance sampling approach to reliability analysis. Here, we extend this approach with an active learning procedure that allows for improved error control at each importance sampling level. To this end, we formulate an estimate of the combined estimation error for both the subspace identified in the dimension reduction step and surrogate model constructed therein. With this, it is possible to adapt the training set so as to optimally learn the subspace representation and the surrogate model constructed therein. The approach is gradient-free and thus

can be directly applied to black box-type models. We demonstrate the performance of this approach with a series of low- (2 dimensions) to high- (869 dimensions) dimensional example problems featuring a number of well-known caveats for reliability methods besides high dimensions and expensive computational models: strongly nonlinear limit-state functions, multiple relevant failure regions and small probabilities of failure.

6.1 Introduction and previous work

An important challenge in the design, analysis and maintenance of engineering systems is the management of the associated uncertainties. It is common practice to analyse engineering systems by employing computational models that aim at representing the physical processes relevant to the system in consideration. These computational models take the form of an input-output mapping. Therein, uncertainty is represented by equipping the model input with an appropriate probabilistic model. Undesirable system responses are defined through a limit-state function (LSF). Reliability analysis is concerned with quantifying the probability of failure, which can be expressed as a d -fold integral of the input probability mass over the failure domain defined by non-positive values of the LSF, where d is the number of uncertain model inputs (see Section 6.2). In engineering, target failure probabilities are typically small; hence, reliability analysis requires the estimation of rare event probabilities. Reliability analysis approaches can be categorized into approximation (e.g. the first- and second-order reliability methods FORM and SORM [67, 28, 18]) and simulation methods. If the LSF is only weakly nonlinear and the input dimension of the model is moderate, FORM and SORM perform well even for small failure probabilities. The simplest simulation method is the Monte Carlo (MC) method [55]. The MC method performs well independent of the problem input dimension, however its performance deteriorates as the failure probability decreases if the computational budget is fixed. Various techniques such as importance sampling (IS) [13, 25, 2] and line-sampling [30, 40] have been proposed to mitigate this dependence on the magnitude of the failure probability. More recently, sequential MC methods such as subset simulation [3] and IS-based sequential methods [42, 43, 84, 62, 69, 61] have been used successfully to efficiently solve high-dimensional reliability problems with small failure probabilities. If the computational model is expensive and a hierarchy of increasingly coarse and cheap models is accessible, multilevel and multi-fidelity [64] MC methods can help alleviate computational cost by performing most model evaluations on the cheaper models (e.g., a discretized differential equation with coarser resolution). In [80], multilevel MC is combined with subset simulation and recently [83] have introduced multilevel sequential IS based on the sequential IS approach in [62]. All of the above-mentioned approaches are designed to work with the probabilistic computational model directly. However, often this model encompasses a numerical solver for (sets of) partial differential equations such that a model evaluation is computationally expensive.

This has increasingly lead researchers to turn towards surrogate model-based reliability methods. Such methods attempt to approximate the expensive computational model with a cheap surrogate model, whose coefficients are identified based on a set of original model evaluations: the training set. [26] uses a polynomial response surface method for performing reliability analysis as early as 1989. [29] proposes an improved version of the response surface method. Since then, a variety of surrogate modelling techniques has been applied in the context of reliability analysis such as artificial neural networks [58, 35, 72], support vector machines [34, 12, 11], Gaussian process regression-based models [22, 21] and projection to polynomial bases including polynomial chaos expansions (PCE) [48, 46,

45, 74] and low-rank tensor approximations [39].

Static, global surrogate models suffer from a decrease in accuracy in the tails of the model response distribution such that they are of limited use for reliability analysis. In this context, *static* refers to surrogate models that are constructed based on a fixed training set and *global* refers to surrogate models that are trained and evaluated on the entire input space (as opposed to locally con- and re-fined models). Thus, one can distinguish two strategies to overcome this limitation:

- *Locality*: Surrogate models are coupled with sequential sampling techniques which serve to focus the training set and accuracy in the relevant regions around the failure hypersurface [57, 12, 11, 6, 59].
- *Adaptivity* (in the training set): The training set is augmented with points that are most informative with respect to the failure probability estimate according to an 'in-fill criterion'. The refined surrogate model is then used to estimate the probability of failure with a sampling method and a large number of cheap samples. Such procedures are summarized under the terms active learning (AL) or optimal experimental design. AL in combination with crude MC have been applied in reliability-based optimization and reliability analysis in [22, 54, 8, 66]. [72] investigates the performance of splines and neural networks in combination with directional sampling and IS and [21, 14] combine Gaussian process models with IS. [71] proposes a crude MC procedure relying on a Gaussian process surrogate model with PCE-based mean trend (PCE-Kriging) along with a novel termination criterion for the AL.

Often, both AL and sequential sampling techniques are combined using various combinations of in-fill criteria and sequential sampling techniques such as adaptive IS [5] and subset simulation [12, 33, 6, 11]. [53] turns away from surrogate models that naturally provide a measure of prediction uncertainty such as Gaussian processes or support vector machines and demonstrate how an AL algorithm can be realized with PCE using a bootstrap estimator of the PCE prediction uncertainty.

In spite of a plethora of existing approaches to surrogate-assisted reliability analysis, the literature on high-dimensional problems ($d \geq 100$) in this context is scarce. [37, 47] propose to perform reliability analysis with a static, global Kriging model constructed in a low-dimensional linear subspace of the original model input space, which is identified by the active subspaces method [16] and autoencoders, respectively. Both [37, 47] apply their methods to moderate-dimensional problems with up to $d = 20$ and $d = 40$ input variables, respectively. [56] uses sliced inverse regression to identify a linear low-dimensional subspace and construct a static, global PCE in this space based on which they perform reliability analysis directly. [90] develops these ideas further by combining the active subspace-Kriging model with an AL approach and applies this combination to a high-dimensional analytical problem of $d = 300$ that possesses a perfectly linear low-dimensional structure.

In this work, we propose an importance sampler based on a dimensionality-reducing surrogate model termed partial least squares-driven PCE (PLS-PCE) [60] to efficiently solve high-dimensional reliability problems with underlying computationally expensive, nonlinear models and small target probabilities ($\mathcal{O}(10^{-9})$). Similar to sliced inverse regression and active subspaces, PLS-PCE achieves dimensionality reduction by identifying a low-dimensional linear subspace of the original input space. Our method is based on [59] but introduces AL to refine the PLS-PCE approximation in each sequence of the IS procedure. In [59], PLS-PCE models are reconstructed in each level of a

sequential importance sampling (SIS) scheme that is used to gradually shift the importance density towards the optimal importance density. In this work, we augment this approach with two novel contributions to rare event simulation of computationally expensive, potentially (but not necessarily) high-dimensional and nonlinear models:

1. We demonstrate how to perform active learning with PCE models by deriving an in-fill criterion from large-sample properties of the PCE coefficient estimates.
2. We use projection to linear subspaces to construct efficient surrogate models for high-dimensional problems and include the subspace estimation error in the in-fill criterion. This means, we are not only learning the surrogate model but also the subspace itself.

Using AL in the context of PLS-PCE-based SIS provides effective error control and benefits from the local confinement of the learning procedure of each subspace/surrogate model combination to the support of the current importance density. Constructing local variance estimates for polynomial models in the way we propose here creates new possibilities to design goal-oriented surrogate modelling approaches that are driven by adaptive sampling based on such models (where so far, Gaussian processes were the dominant tool).

In Section 6.2, we set up the reliability problem and discuss the crude MC sampler of the probability of failure. Section 6.3 reviews IS and a variant of SIS [62] that is at the base of our approach. Section 6.4 introduces PLS-PCE models and their construction. Subsection 6.5.2 details the theoretical foundations of active learning of PLS-PCE models within SIS and summarizes our approach. In Section 6.6, we present comprehensive investigations of the method's performance in two engineering examples and provide a detailed discussion of the results. Conclusions are given in Section 6.7.

6.2 Reliability analysis

Consider a system represented by the computational model $\mathcal{Y} : \mathbb{D}_{\mathbf{X}} \rightarrow \mathbb{R}$ with d -dimensional continuous random input vector $\mathbf{X} : \Omega \rightarrow \mathbb{D}_{\mathbf{X}} \subseteq \mathbb{R}^d$, where Ω is the sample space of \mathbf{X} and by $F_{\mathbf{X}}(\mathbf{x})$, we denote its joint cumulative distribution function (CDF). \mathcal{Y} maps to the system response $Y = \mathcal{Y}(\mathbf{x})$ with the model input $\mathbf{x} \in \mathbb{D}_{\mathbf{X}}$. Based on the response Y , unacceptable system states are defined by means of the limit-state function (LSF) $\tilde{g}(Y)$. Defining $g(\mathbf{x}) = \tilde{g} \circ \mathcal{Y}(\mathbf{x})$ and introducing the convention

$$g(\mathbf{x}) = \begin{cases} \leq 0, & \text{Failure} \\ > 0, & \text{Safety,} \end{cases}$$

the failure event of the system is defined as $F = \{\mathbf{x} \in \mathbb{D}_{\mathbf{X}} : g(\mathbf{x}) \leq 0\}$. The probability of failure is given by [19]

$$p = \mathbb{P}(F) = \int_{\mathbb{D}_{\mathbf{X}}} \mathbb{I}[g(\mathbf{x}) \leq 0] f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \mathbb{E}_{f_{\mathbf{X}}} [\mathbb{I}(g(\mathbf{X}) \leq 0)], \quad (6.1)$$

where $f_{\mathbf{X}}(\mathbf{x}) = \partial^d F / (\partial x_1 \dots \partial x_d) |_{\mathbf{x}}$ is the joint probability density function (PDF) of \mathbf{X} and the indicator function $\mathbb{I}[\cdot]$ equals 1 if the condition in the argument is true and 0 otherwise. Without loss of generality, one may formulate an equivalent reliability problem with respect to the standard-normal probability space using the random vector $\mathbf{U} : \Omega \rightarrow \mathbb{R}^d$. Given an isoprobabilistic transformation

$T : \mathbb{D}_{\mathbf{X}} \rightarrow \mathbb{R}^d$, such that $\mathbf{U} = T(\mathbf{X})$, see, e.g., [31, 49], and defining $G(\mathbf{U}) = g(T^{-1}(\mathbf{U}))$, one can write Eq. (6.1) as

$$p = \int_{\mathbb{R}^d} \mathbb{I}[G(\mathbf{u}) \leq 0] \varphi_d(\mathbf{u}) \, d\mathbf{u} = \mathbb{E}_{\varphi_d} [\mathbb{I}(G(\mathbf{U}) \leq 0)], \quad (6.2)$$

where φ_d denotes the d -dimensional independent standard-normal PDF. The crude MC estimate of Eq. (6.2) is

$$\hat{p}_{\text{MC}} = \frac{1}{n} \sum_{k=1}^n \mathbb{I}[G(\mathbf{u}^k) \leq 0], \quad \mathbf{u}^k \stackrel{i.i.d.}{\sim} \varphi_d, \quad (6.3)$$

where $\mathbf{u}^k \stackrel{i.i.d.}{\sim} \varphi_d$ means that $\{\mathbf{u}^k\}_{k=1}^n$ are n samples that are independent and identically distributed according to φ_d . This estimate is unbiased and has coefficient of variation (CoV)

$$\delta_{\text{MC}} = \sqrt{\frac{1-p}{np}}. \quad (6.4)$$

The number of samples required to compute \hat{p}_{MC} at a prescribed CoV δ_0 reads

$$n_0 = \frac{1-p}{\delta_0^2 p} \stackrel{p \ll 1}{\approx} \frac{1}{\delta_0^2 p}. \quad (6.5)$$

Therefore, crude MC is inefficient for estimating rare event probabilities as, by definition, $p \ll 1$ and thus n_0 becomes large.

6.3 Sequential importance sampling for rare event estimation

Variance reduction techniques can be used to reduce the CoV of the probability estimate at a fixed budget of samples compared to crude MC. One of the most commonly used variance reduction methods is the IS method. Let h be a density, such that $h(\mathbf{u}) > 0$ whenever $G(\mathbf{u}) \leq 0$. Then, one can rewrite Eq. (6.2)

$$p = \int_{\mathbb{R}^d} \mathbb{I}(G(\mathbf{u}) \leq 0) \overbrace{\frac{\varphi_d(\mathbf{u})}{h(\mathbf{u})}}^{\omega(\mathbf{u})} h(\mathbf{u}) \, d\mathbf{u} = \mathbb{E}_h [\mathbb{I}(G(\mathbf{U}) \leq 0) \omega(\mathbf{U})], \quad (6.6)$$

which leads to the (unbiased) importance sampling estimator

$$\hat{p}_{\text{IS}} = \frac{1}{n} \sum_{k=1}^n \mathbb{I}[G(\mathbf{u}^k) \leq 0] \omega(\mathbf{u}^k), \quad \mathbf{u}^k \stackrel{i.i.d.}{\sim} h. \quad (6.7)$$

The efficiency of IS depends intimately on the choice of the IS density h and numerous techniques to construct it have been put forward. There exists an optimal importance density h^* in the sense that it leads to $\mathbb{V}[\hat{p}_{\text{IS}}] = 0$:

$$h^*(\mathbf{u}) = \frac{1}{p} \mathbb{I}[G(\mathbf{u}) \leq 0] \varphi_d(\mathbf{u}). \quad (6.8)$$

While this result is not immediately useful in estimating p as it requires knowledge of p , it can be used to guide the selection of a suitable IS function h .

The SIS method proposed in [62] selects the IS density sequentially starting from a known distribution h_0 that is easy to sample from. It relies on a sequence of distributions $\{h_i(\mathbf{u})\}_{i=0}^M$,

$$h_i(\mathbf{u}) = \frac{\eta_i(\mathbf{u})}{p_i}, \quad i = 1, \dots, M, \quad (6.9)$$

where $\{\eta_i(\mathbf{u})\}_{i=0}^M$ are non-normalized versions of $\{h_i(\mathbf{u})\}_{i=0}^M$ and $\{p_i\}_{i=0}^M$ are the respective normalizing constants. The goal is to arrive at h_M , which is sufficiently close to h^* based on some criterion, and perform importance sampling with h_M . To this end, it is necessary to estimate p_M and obtain samples from h_M . Based on the likelihood ratio of two succeeding non-normalized distributions $\omega_i(\mathbf{u}) = \eta_i(\mathbf{u})/\eta_{i-1}(\mathbf{u})$, we have

$$s_i = \frac{p_i}{p_{i-1}} = \int_{\mathbb{R}^d} \frac{\eta_i(\mathbf{u})}{\eta_{i-1}(\mathbf{u})} h_{i-1}(\mathbf{u}) d\mathbf{u} = \mathbb{E}_{h_{i-1}}[\omega_i(\mathbf{u})]. \quad (6.10)$$

Therefore, an estimate of p_M is given by

$$\hat{p}_M = \prod_{i=1}^M \hat{s}_i \quad \text{with} \quad \hat{s}_i = \frac{1}{n} \sum_{k=1}^n \omega_i(\mathbf{u}^k), \quad \mathbf{u}^k \stackrel{i.i.d.}{\sim} h_{i-1}. \quad (6.11)$$

Samples from h_i can be obtained using Markov Chain Monte Carlo (MCMC) methods given samples from h_{i-1} . More precisely, [62] proposes a resample-move scheme in which Markov chain seeds are obtained as samples from h_{i-1} that are then reweighted (resampled with weights) according to $\omega_i(\mathbf{u})$. In this way, the seed samples are already approximately distributed according to the stationary distribution of the Markov chain h_i and long burn-in periods can be avoided. We adopt an adaptive conditional MCMC sampler (aCS) to perform the move step due to its robust performance in high-dimensional settings. Details can be found in [62].

The h_i are chosen as smooth approximations of h^* using the standard-normal CDF $\Phi(\cdot)$ (compare Fig. 6.1):

$$h_i(\mathbf{u}) = \frac{1}{p_i} \Phi\left(-\frac{G(\mathbf{u})}{\sigma_i}\right) \varphi_d(\mathbf{u}) = \frac{1}{p_i} \eta_i(\mathbf{u}), \quad (6.12)$$

where $p_i = \mathbb{E}_{\varphi_d}[\Phi(-G(\mathbf{U})/\sigma_i)]$ is a normalizing constant and σ_i is the smoothing parameter. Prescribing $\sigma_0 > \sigma_1 > \dots > \sigma_M$ ensures that the sequence $\{h_i(\mathbf{u})\}_{i=0}^M$ approaches h^* . In each level, to avoid degeneration of the weights ω_i (meaning ω_i assuming values close to 0 at all current samples), $h_{i-1}(\mathbf{u})$ and $h_i(\mathbf{u})$ cannot be too different in the sense that they share no support regions on which both have considerable probability mass. This is avoided by prescribing an upper bound for the estimated coefficient of variation of the weights $\hat{\delta}_{w,i} = \widehat{\text{COV}}[\omega_i(\mathbf{U})]$, which provides a criterion for determining σ_i :

$$\sigma_i = \arg \min_{\sigma \in [0, \sigma_{i-1}]} \left(\hat{\delta}_{w,i}(\sigma) - \delta_{\text{target}} \right)^2. \quad (6.13)$$

[62] recommends $\delta_{\text{target}} = 1.5$. The algorithm terminates when h_i is close enough to h^* in the sense that

$$\widehat{\text{COV}} \left[\frac{h^*(\mathbf{U})}{h_i(\mathbf{U})} \right] = \widehat{\text{COV}} \left[\frac{\varphi_d(\mathbf{U}) \mathbf{I}(G(\mathbf{U}) \leq 0)}{\varphi_d(\mathbf{U}) \Phi(-G(\mathbf{U})/\sigma_i)} \right] = \widehat{\text{COV}} \left[\frac{\mathbf{I}(G(\mathbf{U}) \leq 0)}{\Phi(-G(\mathbf{u})/\sigma_i)} \right] \leq \delta_{\text{target}}. \quad (6.14)$$

The final estimate of $\mathbb{P}(\mathbf{F})$ reads

$$\hat{p}_{\text{SIS}} = \hat{p}_M \hat{\mathbb{E}}_{\varphi_d} \left[\frac{\mathbf{I}(G(\mathbf{U}) \leq 0)}{\eta_M(\mathbf{U})} \right] = \left(\prod_{i=1}^M \hat{s}_i \right) \frac{1}{n} \sum_{k=1}^n \frac{\mathbf{I}(G(\mathbf{u}^k) \leq 0)}{\Phi(-G(\mathbf{u}^k)/\sigma_M)}, \quad \mathbf{u}^k \stackrel{i.i.d.}{\sim} h_M. \quad (6.15)$$

Alg. 6 summarizes the complete SIS-aCS procedure.

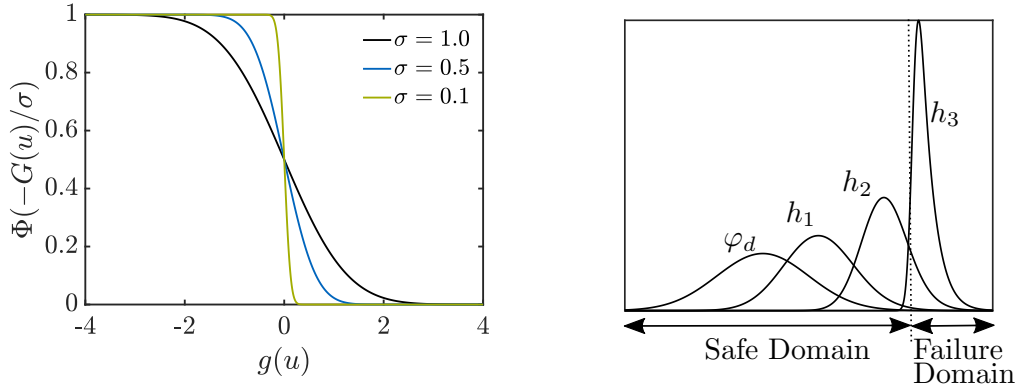


Figure 6.1: Smooth approximations to the indicator function $I(g(\mathbf{u}) \leq 0)$ (left) and importance densities $h_i(\mathbf{u}) \propto \Phi(-G(\mathbf{u})/\sigma_i) \varphi_d(\mathbf{u})$ based on this approximation (right).

6.4 Partial least squares-based polynomial chaos expansions

6.4.1 Polynomial Chaos Expansions

Polynomial chaos expansions (PCEs) are a tool for forward modelling the relationship between an input \mathbf{X} and an output $Y = \mathcal{Y}(\mathbf{X})$. With \mathcal{H} , we denote the Hilbert space of functions that are square-integrable with respect to $f_{\mathbf{X}}$, i.e., $\{v : \mathbb{E}_{f_{\mathbf{X}}}[v(\mathbf{X})^2] < \infty\}$. \mathcal{H} admits an inner product of two functions $v, w \in \mathcal{H}$:

$$\langle v, w \rangle_{\mathcal{H}} = \mathbb{E}_{f_{\mathbf{X}}(\mathbf{x})}[v(\mathbf{X})w(\mathbf{X})] = \int_{\mathbb{R}^d} v(\mathbf{x})w(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}. \quad (6.16)$$

Let $\{v_j(\mathbf{X}), j \in \mathbb{N}\}$ be a complete and orthonormal basis of \mathcal{H} so that $\langle v_j, v_\ell \rangle_{\mathcal{H}} = \delta_{j\ell}$ and let $\mathcal{Y} \in \mathcal{H}$. Then,

$$\mathcal{Y}(\mathbf{X}) = \sum_{j=0}^{\infty} b_j v_j(\mathbf{X}), \quad (6.17)$$

where the coefficients b_j are defined by projecting \mathcal{Y} on the basis:

$$b_j = \langle \mathcal{Y}, v_j \rangle_{\mathcal{H}}, \quad j \in \mathbb{N}. \quad (6.18)$$

Since $\mathcal{Y} \in \mathcal{H}$, the truncation

$$\widehat{\mathcal{Y}}_n(\mathbf{X}) = \sum_{j=0}^n b_j v_j(\mathbf{X}) \quad (6.19)$$

asymptotically converges to \mathcal{Y} as $n \rightarrow \infty$ in the mean square sense. [88] demonstrates how to construct complete orthonormal bases of \mathcal{H} as polynomial families for various standard input distribution types. In particular, if $F_{\mathbf{X}}(\mathbf{x}) = \Phi_d(\mathbf{x})$, where Φ_d denotes the d -variate independent standard-normal CDF, the tensorized, normalized probabilist's Hermite polynomials

$$\Psi_{\mathbf{k}}(\mathbf{U}) = \prod_{i=1}^d \psi_{k_i}(U_i) \quad (6.20)$$

Algorithm 6 SIS-aCS [62]

- 1: **Input** LSF $G(\mathbf{u})$, target CoV δ_{target} , samples per level n , input dimension d , burn-in period b , max. iterations i_{max}
 - 2:
 - 3: Set $i = 0$, $\sigma_0 = \infty$, $h_0(\mathbf{u}) = \varphi_d(\mathbf{u})$
 - 4: Sample $\mathbf{U}_0 = \{\mathbf{u}^k, k = 1, \dots, n\} \in \mathbb{R}^{n \times d}$ $\triangleright \mathbf{u}^k \stackrel{i.i.d.}{\sim} h_0(\mathbf{u})$
 - 5: Compute $\mathbf{G}_0 = G(\mathbf{U}_0) \in \mathbb{R}^{n \times 1}$
 - 6: **for** $i \leftarrow 1, i_{\text{max}}$ **do**
 - 7: $i \leftarrow i + 1$
 - 8: Compute σ_i according to Eq. (6.13)
 - 9: Compute weights $\boldsymbol{\omega}_i = \{\Phi[-\mathbf{G}_{i-1}/\sigma_i]/\Phi[-\mathbf{G}_{i-1}/\sigma_{i-1}], k = 1, \dots, n\} \in \mathbb{R}^{n \times 1}$
 - 10: Compute \hat{s}_i according to Eq. (6.11).
 - 11: $\mathbf{U}_{i-1} \leftarrow$ draw weighted resample from \mathbf{U}_{i-1} with weights $\boldsymbol{\omega}_i$ \triangleright sample with replacement
 - 12: $(\mathbf{U}_i, \mathbf{G}_i) =$ MCMC-aCS($\mathbf{U}_{i-1}, \mathbf{G}_{i-1}, b$) \triangleright Details on MCMC-aCS in [62]
 - 13: **if** Eq. (6.14) **then**
 - 14: **break**
 - 15: Set $M \leftarrow i$
 - 16: Estimate failure probability $\hat{p}_{\text{SIS}} = \left(\prod_{i=1}^M \hat{s}_i\right) \frac{1}{n} \sum_{k=1}^n \frac{I(\mathbf{G}_M^k \leq 0)}{\Phi(-\mathbf{G}_M^k/\sigma_M)}$ \triangleright Eq. (6.15)
 - 17: **return** $\mathbf{U}_M, \mathbf{G}_M, \hat{p}_{\text{SIS}}$.
-

form a complete orthonormal basis of \mathcal{H} . $\{\psi_j(U), j \in \mathbb{N}\}$ are the univariate, normalized (probabilist's) Hermite polynomials and $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{N}^d$. By means of the isoprobabilistic transformation $T : \mathbf{X} \rightarrow \mathbf{U}$ introduced in the previous section, we define PCEs in standard-normal space for the remainder of the paper. The PCE of maximum total order p reads

$$\hat{\mathcal{Y}}_p(\mathbf{U}) = \sum_{|\mathbf{k}| \leq p} b_{\mathbf{k}} \Psi_{\mathbf{k}}(\mathbf{U}). \quad (6.21)$$

The total number of basis functions in the PCE, P , depends on the input dimension d and the maximum total polynomial order p :

$$P = \binom{d+p}{p}. \quad (6.22)$$

The projection in Eq. (6.18) can be transformed into an equivalent ordinary least squares (OLS) problem [7]. PCEs become computationally intractable if d is large, i.e., they cannot be used for problems with high-dimensional input due to the sheer number of basis functions and corresponding coefficients. In particular, the computation is rendered infeasible by the necessary number of operations to compute the set of P multi-indices and the necessary number of model evaluations to obtain meaningful estimates of the coefficients. Solution strategies to overcome these limitations (at least partially) include a hyperbolic truncation of the index set (this means to replace the condition on the ℓ_1 -norm in Eq. (6.21), $|\mathbf{k}| \leq p$, with one on a general ℓ_q -norm of $|\mathbf{k}|_\alpha = (\sum_{i=1}^d p_i^q)^{1/q} \leq p$ with $q < 1$) or enforcing a maximum interaction order (i.e., a maximum number of non-zero entries in \mathbf{k}) [10]. These approaches result in more parsimonious models and allow for PCEs to be applied in higher-dimensional problems, however do so at the cost of decreased model expressivity. Sparsity-inducing solvers have been proposed to relax the dimensionality constraint imposed by the size of the regression problem. Approaches may be based on a variety of solvers for the ℓ_1 -regularized least squares problem such as least-angle regression (LARS) that is used for PCEs in [9], compressive

sensing [89] and orthogonal matching pursuit [63, 77, 20] as well as sparse Bayesian learning methods [76, 36, 70, 79]. For a comprehensive overview, the reader is referred to the recent literature review and benchmark study [52, 51].

6.4.2 Basis adaptation via partial least squares

In order to obtain a parsimonious yet expressive model, we turn to low-dimensional model representations rather than sparse solutions to the full-dimensional model. To achieve this, the PCE representation is rotated onto a new basis defined by the variables $\mathbf{Z} = \mathbf{Q}^T \mathbf{U}$, where $\mathbf{Q} \in \mathbb{R}^{d \times d}$ and $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$, with \mathbf{I} denoting the identity matrix. This has first been proposed in [75]. The PCE with respect to the novel basis reads

$$\hat{y}_p^{\mathbf{Q}}(\mathbf{U}) = \sum_{|\mathbf{k}| \leq p} a_{\mathbf{k}} \Psi_{\mathbf{k}}(\mathbf{Z}) = \sum_{|\mathbf{k}| \leq p} a_{\mathbf{k}} \Psi_{\mathbf{k}}(\mathbf{Q}^T \mathbf{U}). \quad (6.23)$$

With \mathbf{U} a standard-normal random vector and \mathbf{Q} an orthogonal matrix, \mathbf{Z} is a standard-normal random vector. Therefore, both original and transformed input space possess the same PCE basis, namely the probabilist's Hermite polynomials. Merely, a new set of coefficients $a_{\mathbf{k}}$ enters the formulation in the adapted basis. The columns of \mathbf{Q} define linear combinations of the original input. We seek to choose \mathbf{Q} such that most of the relevant information to construct an accurate surrogate \mathcal{Y} is captured in the first m directions, where $m < d$ leads to dimensionality reduction. We retain only these first m columns of \mathbf{Q} in the matrix \mathbf{Q}_m and define a corresponding PCE of reduced dimension as

$$\hat{y}_p^{\mathbf{Q}_m}(\mathbf{U}) = \sum_{|\mathbf{k}| \leq p} a_{\mathbf{k}} \Psi_{\mathbf{k}}(\mathbf{Q}_m^T \mathbf{U}), \quad (6.24)$$

where $\mathbf{k} \in \mathbb{N}^m$. [75] computes the basis adaptation \mathbf{Q}_m by evaluating first- or second-order PCE coefficients only with a sparse-grid numerical quadrature. [78] couples this approach with compressive sensing to simultaneously identify \mathbf{Q}_m and the PCE coefficients in the subspace. In [60], we show that important directions can be identified efficiently based on a set of original function evaluations via partial least squares (PLS).

PLS establishes a linear relationship between variables \mathbf{U} and Y based on $n_{\mathcal{E}}$ observations of both quantities [86]. By $\mathbf{U}_{\mathcal{E}} \in \mathbb{R}^{n_{\mathcal{E}} \times d}$, we denote the matrix of $n_{\mathcal{E}}$ observations of \mathbf{U} and by $\mathbf{Y}_{\mathcal{E}} \in \mathbb{R}^{n_{\mathcal{E}} \times 1}$ we denote the corresponding vector of scalar responses. PLS sequentially identifies m latent components $\{\mathbf{t}_j\}_{j=1}^m$, where $\mathbf{t}_j \in \mathbb{R}^{n_{\mathcal{E}} \times 1}$ such that they have maximum covariance with $\mathbf{Y}_{\mathcal{E}}$. After determining each \mathbf{t}_j , PLS assumes a linear relationship between \mathbf{t}_j and $\mathbf{Y}_{\mathcal{E}}$ and evaluates the corresponding coefficient a_j of \mathbf{t}_j by OLS. After each iteration, the matrices $\mathbf{U}_{\mathcal{E}}$ and $\mathbf{Y}_{\mathcal{E}}$ are deflated by the contribution of the j -th PLS-component. Components are extracted until a certain error criterion is met, which can be formulated, e.g., through the norm of the residual response vector or via cross-validation. Dimensionality-reducing regression methods such als PLS-based regression are known to shrink the regression coefficients towards zero to produce biased estimates in exchange for reducing the estimator variances (bias-variance-tradeoff). In this way, these dimensionality-reducing methods are able to produce smaller overall mean squared estimation errors. (see, e.g., [17] for PLS).

The nonlinear version of PLS in turn relaxes the assumption of a linear relationship between latent component and the response. A number of nonlinear PLS algorithms have been proposed [68]. Here we employ the approach of Refs. [85, 4] that introduces an additional loop into the algorithm

for running a Newton-Raphson procedure iterating between the current latent component and the response. Ultimately, we are interested in computing the orthogonal transformation matrix \mathbf{Q}_m in Eq. (6.24). PLS produces two different matrices \mathbf{R} and \mathbf{W} that are suitable to this end, which motivates two different flavors of PLS-PCE. In PLS-PCE-R as proposed in [60] (see Subsection 6.4.3), each nonlinear relationship between the $\{\mathbf{t}_j\}_{j=1}^m$ and the response is modelled as a univariate PCE. The coefficients of these univariate PCEs are computed simultaneously with the latent structure and the resulting model is a sum of univariate PCEs. Alternatively, the univariate PCEs are discarded after the PLS-PCE algorithm terminates and a multivariate (sparse) PCE is constructed in the subspace formed by the so-called weights $\{\mathbf{w}_j\}_{j=1}^m$ leading to PLS-PCE-W (see Subsection 6.4.4).

6.4.3 PLS-PCE-R

PLS-PCE-R identifies m latent components and for each component, it returns the direction \mathbf{r}_j and the univariate PCE along this direction. The univariate PCEs are defined by their polynomial orders $\{q_j\}_{j=1}^m$ and the associated coefficient vectors $\{\mathbf{a}_j\}_{j=1}^m$. The polynomial order is identified with leave-one-out cross validation [15]. For each (j -th) latent component, the nonlinear PLS iteration is repeated for different polynomial orders and q_j is chosen as the order minimizing the leave-one-out error. The PLS-PCE-R model reads

$$\hat{\mathcal{Y}}(\mathbf{u}) = \hat{a}_0 + \sum_{j=1}^m \left(\hat{\mathbf{a}}_j^{q_j} \right)^T \boldsymbol{\psi}_{q_j} \left[\mathbf{r}_j^T (\mathbf{u} - \boldsymbol{\mu}_{\mathbf{U}}) \right], \quad (6.25)$$

where $\hat{a}_0 = \hat{\mathbb{E}}[\mathbf{Y}]$, $\boldsymbol{\psi}_{q_j}(\mathbf{U})$ is a vector function assembling the evaluations of the one-dimensional Hermite polynomials up to order q_j and $\boldsymbol{\mu}_{\mathbf{U}}$ is the columnwise sample mean of $\mathbf{U}_{\mathcal{E}}$. The model structure is illustrated in Fig. 6.2. The PLS directions \mathbf{r}_j can be evaluated in terms of the PLS weights \mathbf{w}_j and loads \mathbf{p}_j through the following recursive relation [32]

$$\begin{aligned} \mathbf{r}_1 &= \mathbf{w}_1 \\ \mathbf{r}_j &= \mathbf{w}_j - \mathbf{r}_{j-1} (\mathbf{p}_{j-1}^T \mathbf{w}_j). \end{aligned} \quad (6.26)$$

$\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_m] \in \mathbb{R}^{d \times m}$ is a matrix collecting all PLS directions. \mathbf{R} is not necessarily orthogonal, i.e., in general $\mathbf{R}^T \mathbf{R} \neq \mathbf{I}$. However, in [60] it is shown that $\mathbf{R}^T \mathbf{R} \approx \mathbf{I}$ when $n_{\mathcal{E}}$ is large and $\mathbf{U}_{\mathcal{E}}^T \mathbf{U}_{\mathcal{E}}$ is diagonal, which is the case if $\mathbf{U}_{\mathcal{E}}$ is drawn from φ_d . In this case, Eq. (6.25) is equivalent to a PCE of the form Eq. (6.24), where only main effects in the latent components are considered.

6.4.4 PLS-PCE-W

PLS-PCE-W defines \mathbf{W} as basis of the subspace rather than \mathbf{R} , where $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_m] \in \mathbb{R}^{d \times m}$. Within linear PLS, the columns of \mathbf{W} form an orthogonal basis. Within nonlinear PLS, the Newton-Raphson step may introduce deviations from orthogonality, which are however negligible in all tested examples. The univariate PCEs obtained through the Newton-Raphson step will be optimal with respect to \mathbf{R} , not \mathbf{W} . Thus, in PLS-PCE-W these univariate polynomials are discarded once \mathbf{W} is identified and a multivariate (sparse) PCE is constructed in the subspace defined by \mathbf{W} using least-angle regression and a hyperbolic truncation scheme for the multivariate PCE basis as proposed by [9]. In this way PLS-PCE-W achieves more flexibility compared to PLS-PCE-R by including

interactions of the latent components in exchange for a departure from optimality in the match between latent component and surrogate model. In analogy to Eq. (6.24), the PLS-PCE-W model reads

$$\hat{\mathcal{Y}}(\mathbf{u}) = \hat{a}_0 + \sum_{\mathbf{k} \in \boldsymbol{\alpha}} \hat{a}_{\mathbf{k}} \boldsymbol{\Psi}_{\mathbf{k}} [\mathbf{W}^T (\mathbf{u} - \boldsymbol{\mu}_{\mathcal{U}})], \quad (6.27)$$

where $\boldsymbol{\alpha} \in \mathbb{N}^{P \times d}$ is the multi-index set, which indicates the polynomial orders of the d univariate polynomials in each of the P multivariate polynomials as obtained with LARS. Both PLS-PCE-R

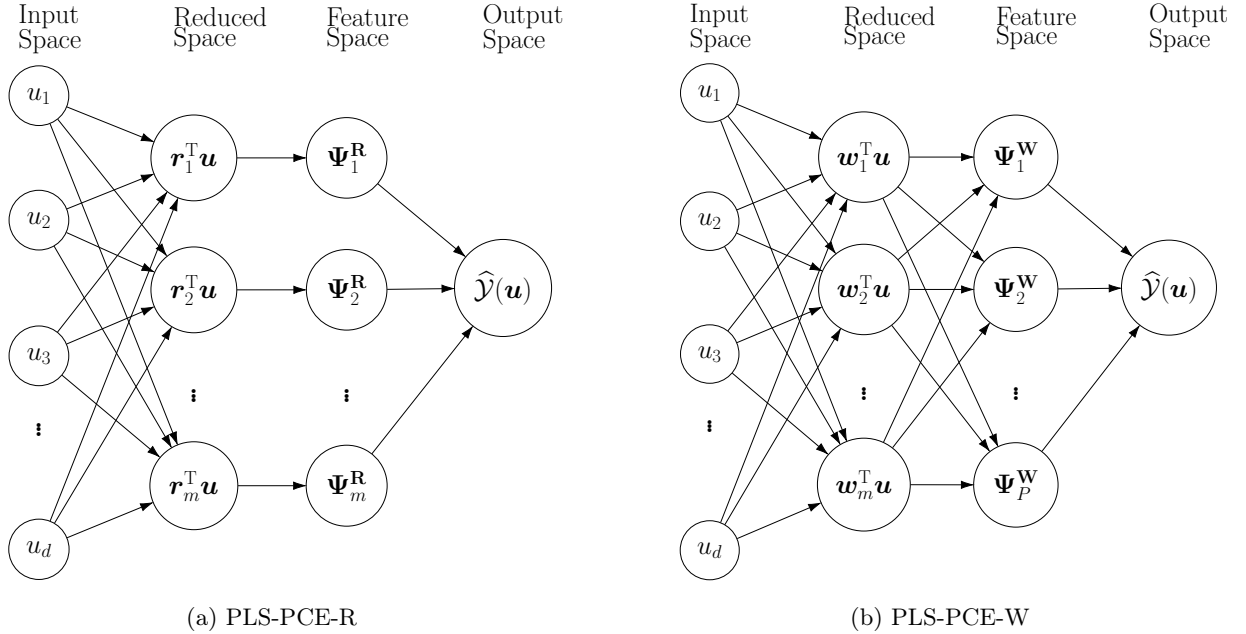


Figure 6.2: Structure of two different PLS-PCE models, where $\boldsymbol{\Psi}_j^{\mathbf{W}} = \boldsymbol{\Psi}_{\boldsymbol{\alpha}_j}$ as defined in Eq. (6.27) and $\boldsymbol{\Psi}_j^{\mathbf{R}} = (\hat{\mathbf{a}}_j^{q_j})^T \boldsymbol{\psi}_{q_j}$ as seen from Eq. (6.25). Essential differences exist in the choice of the reduced space basis (layer 2) and the modelling of cross-terms when mapping from reduced to feature space (layers 2 & 3) with PLS-PCE-W (b).

and PLS-PCE-W are summarized in Alg. 7. In the following, we will use the PLS-PCE-W model as we observed a superior performance for this model compared to PLS-PCE-R models in the context of the proposed approach.

6.5 Learning PLS-PCE models in each SIS level

6.5.1 The sequential subspace importance sampler

We recently proposed to reconstruct low-dimensional PLS-PCE-W models in each level of SIS to improve the tractability of high-dimensional reliability analysis with computationally expensive models [59]. We term this approach sequential subspace importance sampling or SSIS. The efficiency of SIS benefits from surrogate modelling through a considerable reduction of required model evaluations.

Algorithm 7 PCE-driven PLS algorithm [60]

- 1: **Input** Input matrix $\mathbf{U}_{\mathcal{E}}$ and output vector $\mathbf{Y}_{\mathcal{E}}$, maximum polynomial order p
 - 2:
 - 3: Set $\mathbf{E} = \mathbf{U}_{\mathcal{E}} - \boldsymbol{\mu}_{\mathbf{U}}$, $\mathbf{F} = \mathbf{Y}_{\mathcal{E}} - \boldsymbol{\mu}_{\mathbf{Y}}$, $\epsilon_w = 10^{-3}$, $\epsilon_y = 10^{-3}$, $j = 1$
 - 4: **repeat**
 - 5: Compute weight $\mathbf{w}_j^0 = \mathbf{E}^T \mathbf{F} / \|\mathbf{E}^T \mathbf{F}\|$
 - 6: **for** $q \leftarrow 1, p$ **do**
 - 7: Set $\mathbf{w}_j^q = \mathbf{w}_j^0$
 - 8: **repeat**
 - 9: Compute score $\mathbf{t}_j^q = \mathbf{E} \mathbf{w}_j^q$
 - 10: Fit a 1D PCE of order q $\hat{\mathbf{a}}_j^q \leftarrow \text{fit} \left[\mathbf{F} = (\mathbf{a}_j^q)^T \boldsymbol{\psi}_q(\mathbf{t}_j^q) + \boldsymbol{\epsilon} \right]$
 - 11: Set $\widehat{\mathcal{M}}_j^q(t) = (\hat{\mathbf{a}}_j^q)^T \boldsymbol{\psi}_q(\mathbf{t}_j^q)(t)$
 - 12: Compute the error $\mathbf{e} = \mathbf{F} - (\hat{\mathbf{a}}_j^q)^T \boldsymbol{\psi}_q(\mathbf{t}_j^q)$
 - 13: Compute $\Delta \mathbf{w}_j^q = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{e}$ with $\mathbf{A} = \nabla_{\mathbf{w}} (\hat{\mathbf{a}}_j^q)^T \boldsymbol{\psi}_q(\mathbf{E} \mathbf{w})$
 - 14: Set $\mathbf{w}_j^q \leftarrow \mathbf{w}_j^q + \Delta \mathbf{w}_j^q$
 - 15: Normalize $\mathbf{w}_j^q \leftarrow \mathbf{w}_j^q / \|\mathbf{w}_j^q\|$
 - 16: **until** $\|\Delta \mathbf{w}_j^q\|$ is smaller than ϵ_w
 - 17: Evaluate the relative leave-one-out error ϵ_{LOO}^q as in [9]
 - 18: Set $\{q_j, \hat{\mathbf{a}}_j^{q_j}, \mathbf{w}_j^{q_j}\}$ as the triple $\{q, \hat{\mathbf{a}}_j^q, \mathbf{w}_j^q\}$ with the smallest ϵ_{LOO}^q
 - 19: Compute score: $\mathbf{t}_j^{q_j} = \mathbf{E} \mathbf{w}_j^{q_j}$
 - 20: Compute load: $\mathbf{p}_j^{q_j} = \mathbf{E}^T \mathbf{t}_j^{q_j} / ((\mathbf{t}_j^{q_j})^T \mathbf{t}_j^{q_j})$
 - 21: Deflate: $\mathbf{E} \leftarrow \mathbf{E} - \mathbf{t}_j^{q_j} (\mathbf{p}_j^{q_j})^T$, $\mathbf{F} \leftarrow \mathbf{F} - (\hat{\mathbf{a}}_j^{q_j})^T \boldsymbol{\psi}_{q_j}(\mathbf{t}_j^{q_j})$
 - 22: $j \leftarrow j + 1$
 - 23: **until** change in $\|\mathbf{F}\|$ is smaller than ϵ_y
 - 24: Compute $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_m]$ according to Eq. (6.26) \triangleright For the R -based version of PLS-PCE
 - 25: Build $\hat{\mathcal{Y}}(\mathbf{u})$ according to Eq. (6.25)
 - 26: Gather $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m]$ \triangleright For the W -based version of PLS-PCE
 - 27: Build $\hat{\mathcal{Y}}(\mathbf{u})$ according to Eq. (6.27) and [9]
 - 28: **return** \mathbf{R}/\mathbf{W} , $\hat{\mathcal{Y}}(\mathbf{u})$
-

The PLS-PCE model alone, being a global surrogate model, is a relatively limited tool for reliability analysis. Combining it with SIS provides the means to sequentially move the training set towards relevant regions in the input space and thereby renders difficult reliability problems accessible to surrogate modelling. At the i -th SSIS level, a new *local* training set is sampled from the current importance density h_i through a resampling step on the N available samples from h_i . The new *local* training set is appended to the *global* training set comprising earlier designs from levels 1 through $i - 1$. Based on the updated *global* training set, a new PLS-PCE model is constructed and SIS is rerun for $i + 1$ levels from h_0 to obtain samples for the next *local* training set. Due to this restart, it is sensible to let previously used local training sets remain in the global training set such that the i -th surrogate model accurately predicts the LSF output along the entire path of samples moving from the nominal distribution h_0 to h_i . The restart itself incurs no additional LSF evaluations and serves to stabilize the method: Without restart, the computation of σ_{i+1} according to Eq. (6.13) is based on two different surrogate models: the most recent model constructed in level i appears in the numerator of the sample CoV of the weights and the model constructed in level $i - 1$ appears in the denominator. These models may however be too different from one another to admit a solution in Eq. (6.13), i.e., to achieve the prescribed CoV δ_{target} between two subsequent IS densities.

In an additional step, before propagating the intermediate importance density to the next level of the SSIS algorithm, we introduce AL. This ensures a prescribed surrogate model accuracy in regions of high probability mass of the current sampling density. In turn, this refined surrogate model is used to propagate samples to the next level. When the underlying SIS algorithm reaches convergence, a final AL procedure, performed over samples of the final importance density, ensures that the probability of failure is estimated with a surrogate model that captures the failure hypersurface well. This approach is termed adaptive sequential subspace importance sampling or ASSIS.

Active learning has emerged in the late 1980s as a subfield of machine learning [73] and was known in the statistical theory of regression as optimal experimental design since the early 1970s [27]. At its heart is the idea that supervised learning algorithms can perform better if allowed to choose their training data. We consider a 'pool-based sampling' variant of active learning, in which a large pool of unlabeled data points are made available to the algorithm. Within SIS, one has n samples from h_i available in the i -th level. The algorithm then selects n_{add} points that are labeled (i.e. for which the LSF is evaluated) and added to the training set based on a measure of information gain. This measure typically takes the form of a learning function \mathcal{L} that is maximized over the sample pool to perform selection. The learning function employed in the context of SSIS is discussed in Subsection 6.5.2.

The probability of failure estimator for SSIS/ASSIS is analogous to Eq. (6.15) with the difference that SIS is performed with an LSF approximation \widehat{G} that is based on the final surrogate model:

$$\widehat{p} = \left(\prod_{i=1}^M \widehat{s}_i \right) \frac{1}{n} \sum_{k=1}^n \frac{\mathbb{I}(\widehat{G}(\mathbf{u}^k) \leq 0) \varphi_d(\mathbf{u}^k)}{\eta_M(\mathbf{u}^k)}, \quad \mathbf{u}^k \stackrel{i.i.d.}{\sim} h_M. \quad (6.28)$$

The ratio of normalizing constants $\{\widehat{s}_i\}_{i=1}^M$ are estimated as

$$\widehat{s}_i = \frac{1}{n} \sum_{k=1}^n \widehat{\omega}_i(\mathbf{u}^k) = \frac{1}{n} \sum_{k=1}^n \frac{\Phi(-\widehat{G}(\mathbf{u}^k)/\sigma_i)}{\Phi(-\widehat{G}(\mathbf{u}^k)/\sigma_{i-1})}, \quad \mathbf{u}^k \stackrel{i.i.d.}{\sim} h_i. \quad (6.29)$$

The SSIS/ASSIS algorithms are stopped based on a similar criterion as for SIS given in Eq. (6.14):

$$\widehat{\text{COV}} \left[\frac{\mathbf{I}(\widehat{G}(\mathbf{U}) \leq 0)}{\Phi(-\widehat{G}(\mathbf{U})/\sigma_i)} \right] \leq \delta_{\text{target}}. \quad (6.30)$$

Fig. 6.3 depicts flow diagrams of the SSIS and ASSIS algorithms.

6.5.2 Active learning of low-dimensional model representations

In the context of SSIS, the learning function \mathcal{L} should express the prediction uncertainty at each sample of the current IS density for a given PLS-PCE-W surrogate. This prediction uncertainty is due to the estimation of both the subspace and the surrogate model with a finite-sized training set. We describe this uncertainty with the variance of the LSF based on the surrogate model conditional on \mathbf{u} , $\mathbb{V}[\widehat{G}|\mathbf{U} = \mathbf{u}]$. Note that, whenever the distribution with respect to which $\mathbb{E}[\cdot]$ or $\mathbb{V}[\cdot]$ are evaluated is not made explicit as a subscript, it is implicitly assumed as the distribution of the argument. For example, $\mathbb{V}[\widehat{G}|\mathbf{U} = \mathbf{u}] = \mathbb{V}_{f_{\widehat{G}|\mathbf{u}}}[\widehat{G}|\mathbf{U} = \mathbf{u}]$.

Let $\boldsymbol{\xi}_0 = \mathbf{a} \in \mathbb{R}^{P \times 1}$ and $\boldsymbol{\xi}_j = \mathbf{w}_j \in \mathbb{R}^{d \times 1}$, $j = 1, \dots, m$, such that $\boldsymbol{\xi} = [\boldsymbol{\xi}_0^T, \boldsymbol{\xi}_1^T, \dots, \boldsymbol{\xi}_m^T]^T \in \mathbb{R}^{(md+P) \times 1}$ is the collection of all $md + P$ model parameters. Further, let $\boldsymbol{\xi}^*$ denote their corresponding point estimates returned by Alg. 7. The first-order expansion of $\widehat{\mathbb{V}}[\widehat{G}|\mathbf{u}]$ around $\boldsymbol{\xi}^*$ reads

$$\widehat{\sigma}_G^2(\mathbf{u}) = \widehat{\mathbb{V}}[\widehat{G}|\mathbf{u}] \approx \left[\frac{\partial \widehat{G}}{\partial \boldsymbol{\xi}} \right]_{\boldsymbol{\xi}=\boldsymbol{\xi}^*}^T \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}\boldsymbol{\xi}} \left[\frac{\partial \widehat{G}}{\partial \boldsymbol{\xi}} \right]_{\boldsymbol{\xi}=\boldsymbol{\xi}^*}, \quad (6.31)$$

where $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}\boldsymbol{\xi}}$ is an estimate of the parameter covariance matrix. Next, we neglect the pairwise cross-covariance of PCE coefficients \mathbf{a} and the subspace components \mathbf{w}_j and consider

$$\widehat{\sigma}_G^2(\mathbf{u}) = \widehat{\mathbb{V}}[\widehat{G}|\mathbf{u}] \approx \sum_{j=0}^m \left[\frac{\partial \widehat{G}(\mathbf{u}, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}_j} \right]_{\boldsymbol{\xi}_j=\boldsymbol{\xi}_j^*}^T \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}_j \boldsymbol{\xi}_j} \left[\frac{\partial \widehat{G}(\mathbf{u}, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}_j} \right]_{\boldsymbol{\xi}_j=\boldsymbol{\xi}_j^*} \quad (6.32)$$

This significantly reduces the number of $\boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\xi}}$ -entries that have to be estimated, namely from $P^2 + 2Pmd + m^2d^2$ to $P^2 + md^2$. More importantly, the coefficients of the PCE, $\boldsymbol{\xi}_0$, are obtained with linear regression while the subspace, $\{\boldsymbol{\xi}_j\}_{j=1}^m$, is obtained in the inner loop of Alg. 7 with nonlinear regression. Due to this sequential estimation of the $\{\boldsymbol{\xi}_j\}_{j=0}^m$, there is no straightforward way of obtaining an estimate of the full covariance matrix. In particular, we are not aware of such an estimate for the parameters of nonlinear PLS. Hence, this simplification is not only convenient but also necessary in practice. We do observe, however, that the off-diagonal elements of the estimated component-wise cross-covariance matrices $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}_j \boldsymbol{\xi}_j}$ are several orders of magnitude smaller compared to the the main diagonal elements. This indicates that the model uncertainty estimate is dominated by parameter variances. In fact, in a more radical approach that remains unexplored in this work, one may consider parameter variances only (i.e., only $P + md$ entries of the full covariance matrix are retained). Such an approach is, e.g., used in [65]. Under some regularity conditions, the estimator $\boldsymbol{\xi}_j^*$ is consistent [87] and converges in distribution to a multivariate Gaussian distribution with mean $\boldsymbol{\xi}_j$ and covariance $\boldsymbol{\Sigma}_{\boldsymbol{\xi}_j \boldsymbol{\xi}_j}$. In analogy with linear regression, an estimate of $\boldsymbol{\Sigma}_{\boldsymbol{\xi}_j \boldsymbol{\xi}_j}$ is given through

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}_j \boldsymbol{\xi}_j} = \widehat{\sigma}_\epsilon^2 (\mathbf{A}_j^T \mathbf{A}_j)^{-1} \quad (6.33)$$

with

$$\mathbf{A}_j = \left[\frac{\partial \widehat{\mathcal{Y}}(\mathbf{u}, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}_j} \right]_{\substack{\boldsymbol{\xi}=\boldsymbol{\xi}^* \\ \mathbf{u}=\mathbf{U}_\mathcal{E}}} \in \mathbb{R}^{n_\mathcal{E} \times d} \quad \text{and} \quad \widehat{\sigma}_\epsilon^2 = \frac{1}{n_\mathcal{E} - md - P} \sum_{k=1}^{n_\mathcal{E}} [\mathbf{Y}_\mathcal{E}^k - \widehat{\mathcal{Y}}(\mathbf{U}_\mathcal{E}^k)]^2. \quad (6.34)$$

$\widehat{\sigma}_\epsilon^2$ is the standard estimator for the error variance of the surrogate model. \mathbf{A}_j is the gradient of the surrogate model \mathcal{Y} with respect to the model parameters evaluated at each of the $n_\mathcal{E}$ points in the training set $\mathbf{U}_\mathcal{E}$. \mathbf{A}_0 is merely the design matrix and does not require the computation of any derivatives. Note that computing the gradients $\{\mathbf{A}_j\}_{j=0}^m$ does not require any model evaluations. For $j = 0$, it is

$$\frac{\partial \widehat{\mathcal{Y}}(\mathbf{u}, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}_0} = [\boldsymbol{\Psi}_i(\mathbf{W}^\top(\mathbf{u} - \boldsymbol{\mu}_\mathbf{U}))]_{i=1}^{P-1} \quad \text{with} \quad \mathbf{W} = [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_m]. \quad (6.35)$$

For $j > 0$ and recalling $\mathbf{z} = \mathbf{W}^\top(\mathbf{u} - \boldsymbol{\mu}_\mathbf{U})$, we have

$$\begin{aligned} \frac{\partial \Psi_{\mathbf{k}}(\mathbf{z})}{\partial \boldsymbol{\xi}_j} &= \frac{\partial}{\partial \mathbf{w}_j} \Psi_{\mathbf{k}}(\mathbf{W}^\top(\mathbf{u} - \boldsymbol{\mu}_\mathbf{U})) \\ &= (\mathbf{u} - \boldsymbol{\mu}_\mathbf{U}) \frac{\partial \Psi_{\mathbf{k}}(z_j)}{\partial z_j} \\ &= (\mathbf{u} - \boldsymbol{\mu}_\mathbf{U}) \left(\prod_{\substack{i=1 \\ i \neq j}}^m \psi_{k_i}(\mathbf{w}_i^\top \mathbf{u}) \right) \frac{\partial \psi_{k_j}(\mathbf{w}_j^\top \mathbf{u})}{\partial z_j} \\ &= (\mathbf{u} - \boldsymbol{\mu}_\mathbf{U}) \left(\prod_{\substack{i=1 \\ i \neq j}}^m \psi_{k_i}(\mathbf{w}_i^\top \mathbf{u}) \right) \sqrt{k_j} \psi_{k_j-1}(\mathbf{w}_j^\top \mathbf{u}). \end{aligned} \quad (6.36)$$

In the last equality, we have used the following expression for derivatives of univariate normalized Hermite polynomials:

$$\frac{d\psi_n(x)}{dx} = \sqrt{n} \psi_{n-1}(x). \quad (6.37)$$

$\partial \widehat{\mathcal{Y}}(\mathbf{u}, \boldsymbol{\xi}) / \partial \boldsymbol{\xi}_j$ for $j > 0$ follows as

$$\frac{\partial \widehat{\mathcal{Y}}(\mathbf{u}, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}_j} = \frac{\partial \widehat{\mathcal{Y}}(\mathbf{z})}{\partial \mathbf{w}_j} = \sum_{\mathbf{k} \in \boldsymbol{\alpha}} \widehat{a}_{\mathbf{k}} \frac{\partial \Psi_{\mathbf{k}}(\mathbf{z})}{\partial \boldsymbol{\xi}_j}, \quad j > 0. \quad (6.38)$$

The partial derivative $\partial \widehat{G} / \partial \boldsymbol{\xi}_j$ in Eq. (6.32) can be evaluated using the chain rule of differentiation, which yields

$$\frac{\partial \widehat{G}}{\partial \boldsymbol{\xi}_j} = \frac{\partial \widehat{G}}{\partial \widehat{\mathcal{Y}}} \frac{\partial \widehat{\mathcal{Y}}}{\partial \boldsymbol{\xi}_j}. \quad (6.39)$$

The first term on the right-hand side is typically easy to compute and often equals ± 1 (the sign is irrelevant as the gradient enters the quadratic form in Eq. (6.32)) if the LSF returns the difference between the model output and a prescribed threshold. In this case, the first factor on the right-hand side of Eq. (6.39) drops out. If, however, the LSF is not continuously differentiable with respect to the model, we may construct a surrogate model of G directly by using a training set containing LSF evaluations rather than model evaluations in Alg. 7. The second term on the right-hand side can be

obtained reusing the gradients from the \mathbf{A}_j in Eq. (6.34) that — in this case — are not evaluated at the training set and thus are functions of \mathbf{u} .

When setting up the learning function, there is a distinction to be made between an intermediate SIS level and the final SIS level: In the intermediate level, the goal is to accurately estimate the ratios of normalizing constants and to propagate the samples to the next level. In the final level, the goal is to build the probability of failure estimator and thus to accurately approximate the true limit-state hypersurface. With this in mind, the learning functions for adapting the surrogate models in levels $i = 1, \dots, M$, and after the final level are readily stated as

$$\mathcal{L}_G(\mathbf{u}) = \begin{cases} \sigma_{\widehat{G}}(\mathbf{u}), & \text{intermediate SIS level} \\ \sigma_{\widehat{G}}(\mathbf{u})/|\widehat{G}(\mathbf{u})|, & \text{after final SIS level.} \end{cases} \quad (6.40)$$

After the final level, SIS has converged and we are using samples from the final biasing density h_M to refit a surrogate model that captures the failure hypersurface well. The learning function in this case is defined in the spirit of the learning function put forward in [22]. The denominator penalizes samples whose image under \widehat{G} is far away from 0 assuming that therefore they are themselves far away from the failure hypersurface. Such samples are unlikely to be misclassified as safe if located in the failure domain or vice versa. In all previous levels of SIS, there is no failure hypersurface to be approximated but only importance weights and the resulting ratio of normalizing constants. Here, the denominator in the learning function is dropped as there is no benefit to penalizing samples with large absolute image values under \widehat{G} .

In each AL iteration, the pool is searched for one or several points maximizing $\mathcal{L}(\mathbf{u})$. If $n_{\text{add}} > 1$ new points are added per AL iteration, the current sample pool is transformed to the low-dimensional subspace defined by \mathbf{W} in order to identify n_{add} clusters (e.g., with k-means). Clustering in the subspace circumvents the performance deterioration most clustering methods experience in high dimensions [41]. The point maximising Eq. (6.40) in each cluster is added to the training set. In this way, the algorithm avoids a local concentration of the training set in a single region and is also able to handle problems with multiple disconnected failure domains as long as these are contained in the subspace.

The active learning is terminated based on the maximum local standard deviation relative to the target average in the intermediate levels or based on the relative change of the probability of failure estimate after the final level:

$$\left\{ \begin{array}{l} \max_{k=1, \dots, n} \left(\frac{\sigma_{\widehat{G}}(\mathbf{u}_k)}{\mathbb{E}[\widehat{G}(\mathbf{U})]} \right) \leq \epsilon_{\text{AL}}, \quad \text{intermediate SIS level} \\ \frac{\widehat{p} - \widehat{p}_{\text{last}}}{\widehat{p}} \leq \epsilon_{\text{AL}}, \quad \text{after final SIS level} \end{array} \right\}, \quad (6.41)$$

where appropriate choices for ϵ_{AL} lie in $[10^{-2}, 10^{-1}]$. \widehat{p} and $\widehat{p}_{\text{last}}$ denote the probability of failure estimate based on the current and the last training set within the AL loop. The probability of failure is estimated with a surrogate model-based run of SIS-aCS in each AL iteration. This causes no additional cost in terms of original model evaluations and ensures a reliable evaluation of the criterion even for extremely small failure probabilities. The active learning procedure is detailed in Alg. 8 and the complete method is detailed in Alg. 9.

6.6 Numerical experiments

Algorithm 8 Active Learning

```

1: Input LSF  $G(\mathbf{u})$ , AL error level  $\epsilon_{\text{AL}}$ , # of AL clusters  $n_{\text{add}}$ , Polynomial order  $p$ , training set
    $\{\mathbf{U}_{\mathcal{E}}, \mathbf{G}_{\mathcal{E}}\}$ , Sample pool  $\mathbf{U}_{\text{pool}}$ 
2:
3: while true do ▷ Active learning loop
4:   Run  $[\mathbf{W}, \widehat{G}] = \text{PLS-PCE}(\mathbf{U}_{\mathcal{E}}, \mathbf{G}_{\mathcal{E}}, p, 'W')$  ▷ Alg. 7
5:   if Eq. (6.41) then
6:     break
7:   Identify  $n_{\text{add}}$  clusters among  $\mathbf{U}_{\text{pool}}\mathbf{W}$  ▷ Clustering performed in the subspace defined by  $\mathbf{W}$ 
8:   for each cluster do
9:      $\mathbf{U}_{\text{cluster}} = \{\mathbf{u} \in \mathbf{U}_{\text{pool}} : \mathbf{u} \in \text{cluster}\}$ 
10:    Evaluate  $\mathbf{u}^* = \text{argmax}[\mathcal{L}(\mathbf{U}_{\text{cluster}})]$  according to Eqs. (6.32) to (6.34), (6.39) and (6.40).
11:    Append  $\mathbf{U}_{\mathcal{E}} \leftarrow [\mathbf{U}_{\mathcal{E}}, \mathbf{u}^*]$ 
12:    Append  $\mathbf{G}_{\mathcal{E}} \leftarrow [\mathbf{G}_{\mathcal{E}}, G(\mathbf{u}^*)]$ 
13:    Remove  $\mathbf{u}^*$  from  $\mathbf{U}_{\text{pool}}$ 
14: return  $\mathbf{U}_{\mathcal{E}}, \mathbf{G}_{\mathcal{E}}, \widehat{G}$ .

```

Algorithm 9 ASSIS (with PLS-PCE-W)

```

1: Input LSF  $G(\mathbf{u})$ , Target CoV  $\delta_{\text{target}}$ , Samples per level  $n$ , Input dimension  $d$ , training set size
    $n_{\mathcal{E}}$ , AL error level  $\epsilon_{\text{AL}}$ , # of AL clusters  $n_{\text{add}}$ , Polynomial order  $p$ ,
2:
3: Set  $i = 0$ ,  $\sigma_i = \infty$ ,  $h_i(\mathbf{u}) = \varphi_d(\mathbf{u})$ 
4: Initialize  $\mathbf{U}_{\mathcal{E}} = [\cdot]$ ,  $\mathbf{G}_{\mathcal{E}} = [\cdot]$ 
5: Sample  $\mathbf{U}_0 = \{\mathbf{u}^k\}_{k=1}^n \in \mathbb{R}^{n \times d}$  ▷  $\mathbf{u}^k \stackrel{i.i.d.}{\sim} h_i(\mathbf{u})$ 
6: while true do ▷ Sequential importance sampling loop
7:    $i \leftarrow i + 1$ 
8:   Sample  $\mathbf{U}_{\text{tmp}} = \{\mathbf{u}^k\}_{k=1}^{n_{\mathcal{E}}} \in \mathbb{R}^{n_{\mathcal{E}} \times d}$  ▷  $\mathbf{u}^k \stackrel{i.i.d.}{\sim} h_i(\mathbf{u})$ 
9:   Compute  $\mathbf{G}_{\text{tmp}} = G(\mathbf{U}_{\text{tmp}}) \in \mathbb{R}^{n_{\mathcal{E}} \times 1}$ 
10:  Append  $\mathbf{U}_{\mathcal{E}} \leftarrow [\mathbf{U}_{\mathcal{E}}, \mathbf{U}_{\text{tmp}}]$ 
11:  Append  $\mathbf{G}_{\mathcal{E}} \leftarrow [\mathbf{G}_{\mathcal{E}}, \mathbf{G}_{\text{tmp}}]$ 
12:  if  $i > 1$  then
13:    Run  $\widehat{G} = \text{PLS-PCE}(\mathbf{U}_{\mathcal{E}}, \mathbf{G}_{\mathcal{E}}, p, 'W')$  ▷ Alg. 7
14:    Run  $\mathbf{U}_{i-1}, \mathbf{G}_{i-1} = \text{SIS-aCS}(\widehat{G}, \delta_{\text{target}}, n, d, i - 1)$  ▷ Alg. 6
15:    Run  $\mathbf{U}_{\mathcal{E}}, \mathbf{G}_{\mathcal{E}}, \widehat{G} = \text{Active Learning}(G(\mathbf{u}), \epsilon_{\text{AL}}, n_{\text{add}}, p, \mathbf{U}_{\mathcal{E}}, \mathbf{G}_{\mathcal{E}}, \mathbf{U}_{i-1})$  ▷ Alg. 8
16:    Compute  $\mathbf{G}_{i-1} = \widehat{G}(\mathbf{U}_{i-1}) \in \mathbb{R}^{n \times 1}$ 
17:    Compute  $\sigma_i$  according to Eq. (6.13)
18:    Compute  $\widehat{\omega}_i$  and  $\widehat{s}_i$  according to Eq. (6.29)
19:     $\mathbf{U}_{i-1}, \mathbf{G}_{i-1} \leftarrow \text{resample from } \mathbf{U}_{i-1}, \mathbf{G}_{i-1} \text{ with weights } \widehat{\omega}_i(\mathbf{U}_{i-1})$  ▷ sample with replacement
20:    Run  $\mathbf{U}_i, \mathbf{G}_i = \text{SIS-aCS}(\mathbf{U}_{i-1}, \mathbf{G}_{i-1})$  ▷ Perform a single MCMC step
21:    if Eq. (6.30) then
22:      Set  $M \leftarrow i$ 
23:      Run  $\mathbf{U}_{\mathcal{E}}, \mathbf{G}_{\mathcal{E}}, \widehat{G} = \text{Active Learning}(G(\mathbf{u}), \epsilon_{\text{AL}}, n_{\text{add}}, p, \mathbf{U}_{\mathcal{E}}, \mathbf{G}_{\mathcal{E}}, \mathbf{U}_{i-1})$  ▷ Alg. 8
24:      break
25: Run  $(\mathbf{U}_M, \mathbf{G}_M, \widehat{p}_{\text{ASSIS}}) = \text{SIS-aCS}(\widehat{G}_M, \delta_{\text{target}}, n, d, M)$  ▷ Alg. 6
26: return  $M, \mathbf{U}_M, \mathbf{G}_M, \widehat{p}_{\text{ASSIS}}$ .

```

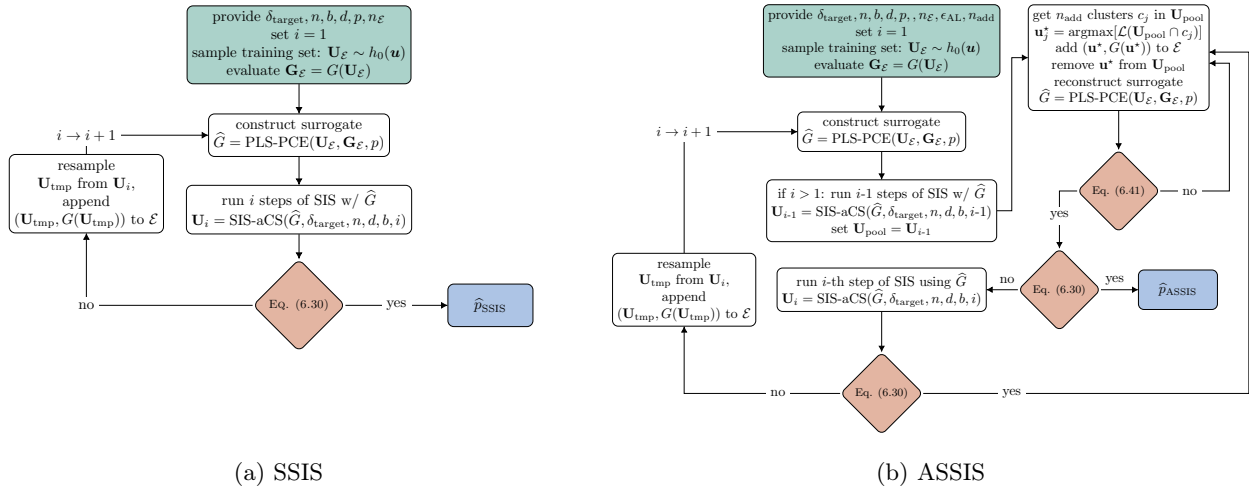


Figure 6.3: Comparison of SIS-PLS-PCE with (right) and without (left) active learning.

6.6.1 Error measures

In the following, we examine a series of examples of low to high input dimensionality characterized by varying degrees of nonlinearity of the LSF and varying number of disconnected failure regions. The computational cost of each approach is measured with the total number of required calls to the underlying computational model. The accuracy of the estimator is measured in terms of relative bias and CoV

$$\text{relative Bias} = \frac{p - \mathbb{E}[\hat{p}]}{p} \quad (6.42)$$

$$\text{CoV} = \frac{\sqrt{\mathbb{V}[\hat{p}]}}{\mathbb{E}[\hat{p}]}, \quad (6.43)$$

where p is the known exact probability of failure or a reference solution computed with a large number of samples as reported in the corresponding references in Tab. 6.1. Further, we compute the relative root mean squared error (RMSE) of the probability of any failure estimate \hat{p} , which combines bias and variability of the estimator as

$$\text{relative RMSE} = \sqrt{\frac{\mathbb{E}[(p - \hat{p})^2]}{p^2}} = \sqrt{\text{relative Bias}^2 + \left(\frac{\mathbb{E}[\hat{p}]}{p}\right)^2 \text{CoV}^2} \quad (6.44)$$

The expectation and variance operators in the above equations are approximated by repeating each analysis 100 times. Additionally, the relative estimation error is defined as

$$\text{relative error} = \frac{\hat{p}}{p}. \quad (6.45)$$

6.6.2 Low- and medium-dimensional examples

The subspace importance sampler is designed to tackle high-dimensional problems, yet its performance should not deteriorate as the problem dimension decreases. We first investigate its perfor-

Table 6.1: Low- to medium-dimensional investigated benchmark problems.

Problem	Failure probability	Inputs	Input Variables	Properties	References
Hat	$1.037 \cdot 10^{-4}$	2	standard-normal	Strongly nonlinear	[71]
Cantilever	$3.94 \cdot 10^{-6}$	2	Gaussian	Strongly nonlinear	[6]
4-Branch (acc. to [6])	$5.60 \cdot 10^{-9}$	2	standard-normal	Multiple failure regions; extremely rare event	[6, 82]
Borehole ($276.7 \frac{m^3}{year}$)	$1 \cdot 10^{-5}$	8	Log-normal, Uniform	Strongly nonlinear, No underlying low-dimensional structure	[1]
Truss (0.12m)	$1.6 \cdot 10^{-3}$	10	Log-normal, Gumbel	mildly nonlinear	[44]
Rare Truss (0.18m)	$1.02 \cdot 10^{-8}$	10	Log-normal, Gumbel	Extremely rare event; nonlinear	[44] (modified)
Quadratic ($\kappa = 5$)	$6.62 \cdot 10^{-6}$	10	standard-normal	Strongly nonlinear; Underlying low-dimensional structure	[25, 81]
Quadratic ($\kappa = 5$)	$6.62 \cdot 10^{-6}$	100	standard-normal	Strongly nonlinear; Underlying low-dimensional structure	[25, 81]

mance in eight exemplary problems with dimension $2 \leq d \leq 100$. We demonstrate how both SSIS and ASSIS cope with multiple failure domains, strong nonlinearities and extremely small target failure probabilities. In the interest of brevity, the examples are listed in Tab. 6.1 along with the problem dimension, target probability of failure and key characteristics of the problem. The references provided in Tab. 6.1 may be consulted for detailed descriptions of the problem setups.

We solve the example problems with SIS-aCS with $n = 2 \cdot 10^3$ samples per level and a burn-in period of $b = 5$ samples within each MCMC chain. As suggested in [62], we choose $\delta_{\text{target}} = 1.5$ for the exit criterion Eq. (6.14) for SIS-aCS as well as our surrogate-based samplers. We compare this reference to SSIS and ASSIS for which we use an initial sample size of $n_{\mathcal{E}} = 5d$. All underlying PLS-PCE-W models are computed with a maximum number of subspace directions of $m = 10$ and a maximum total polynomial degree of $|q|_{\ell_q} \leq 7$, where $q = 0.75$. To achieve a fair comparison between ASSIS and SSIS, we first run ASSIS and then SSIS with $n_{\mathcal{E}}$ for the latter chosen such that both methods use an approximately equal number of LSF evaluations. For both SSIS and ASSIS, we choose $n = 10^4$ with a burn-in period of $b = 30$. For ASSIS, we set $\epsilon_{\text{AL}} = 0.1$. Within SSIS/ASSIS many samples per level and long burn-in periods are affordable as sampling is performed with the surrogate model. For ASSIS we select $n_{\text{add}} = 1$ unless prior knowledge of the problem structure suggests otherwise (the only exception in the set of examples considered here is the 4-branch function for which we select $n_{\text{add}} = 4$ as it features four relevant failure regions in the input space). Fig. 6.4 displays the performance of SIS, SSIS and ASSIS for the examples in Tab. 6.1 in terms of the error measures defined in Eqs. (6.42) to (6.44) and the total number of LSF evaluations (with the original model).

For all showcased examples, ASSIS yields equally or more accurate estimates compared to SSIS

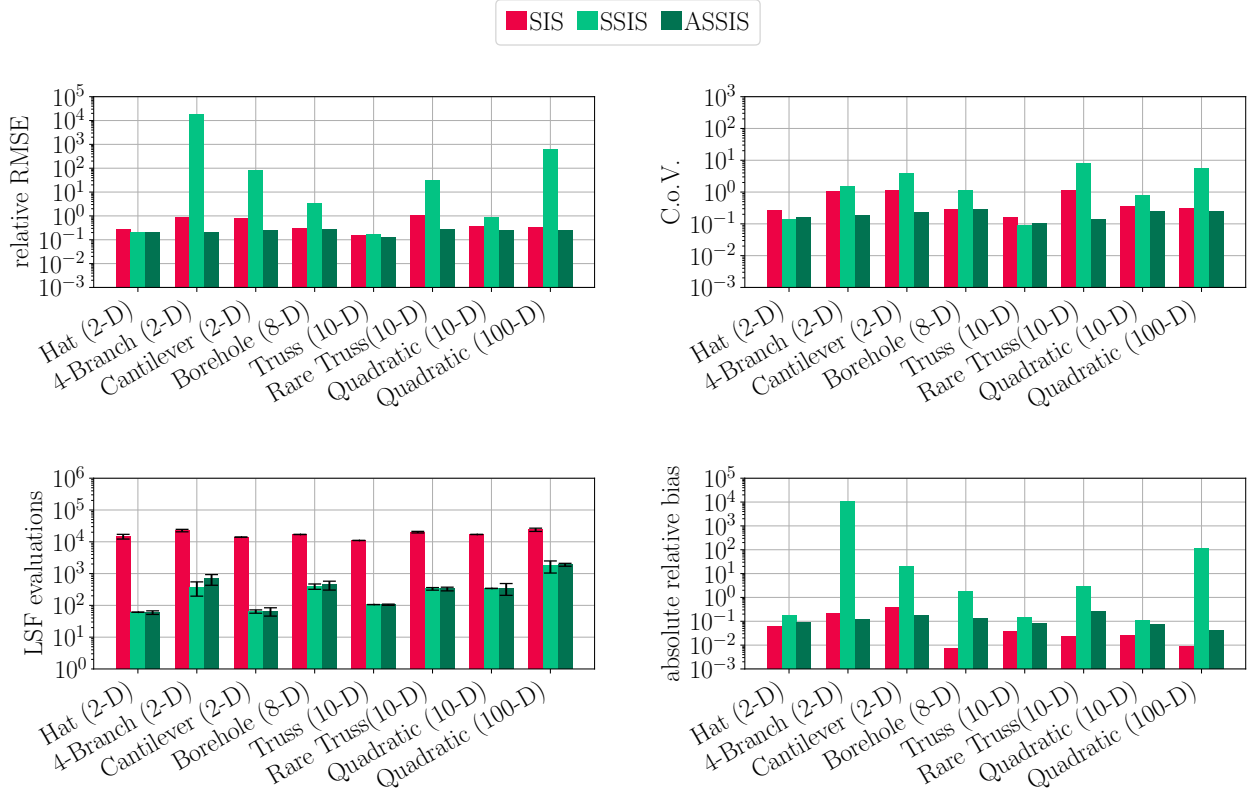


Figure 6.4: Low- and medium-dimensional examples: accuracy and cost comparison. Cost error bars include ± 2 standard deviations.

at equal cost. It also either matches or outperforms SIS at significantly reduced costs. Except for the easiest problems, i.e., those featuring well-behaved (truss) or low-dimensional (2D hat) LSFs associated with comparatively large failure probabilities, the in-level adaptivity of ASSIS leads to significant bias correction (Fig. 6.4, bottom right) and variance reduction (Fig. 6.4, top right).

[62] discusses the choice of the MCMC sampler for SIS and find that aCS as employed here is outperformed by a Gaussian mixture proposal in low-dimensional problems, while the latter is the preferred choice as the problem dimension grows. Our method is designed for the solution of high-dimensional reliability problems and we thus consistently use aCS.

Comparing the truss and the rare truss models, the additional number of SIS levels required in the solution of the latter evidently leads to a deterioration of the SSIS estimate (Fig. 6.4, top left). This is due to single runs (less than 10 %) among the 100 repetitions in which the sampled training sets lead to extreme outliers in the failure probability estimates (Fig. 6.5). While this effect vanishes when increasing the number of samples in the training set, ASSIS offers a more cost-effective alternative to avoid such outliers by actively learning an informative augmentation of adverse training sets. In this way, subspace identification and surrogate modelling errors cannot propagate and accumulate across the levels of SIS as they are controlled by the AL procedure. In fact, the phenomenon of rather rare but all the more severe outliers deteriorating the error mean and variability is a problem SSIS is facing not only in the rare truss example but also in the cantilever and both quadratic examples. Conversely, it is seen that in the 4-branch example, SSIS consistently and considerably

overestimates the probability of failure while ASSIS captures the probability of failure rather well.

The two quadratic LSF models with 10 and 100 input dimensions demonstrate how the required

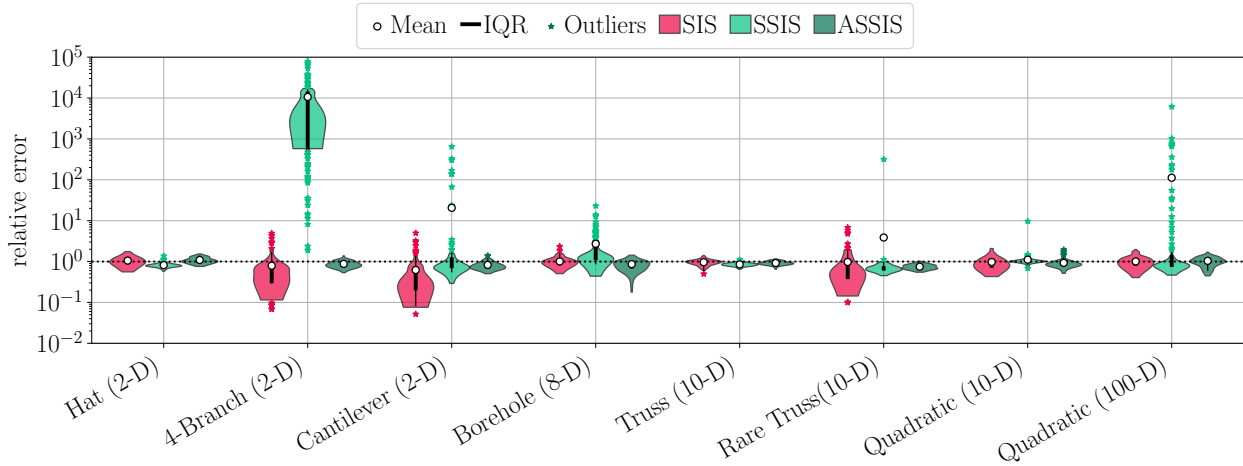


Figure 6.5: Low- and medium-dimensional examples: violin plots of the relative error along with means, inter-quartile ranges (IQR) and outliers. For the sake of clarity, kernel density estimates are computed after excluding outliers based on the relative distance to the data median.

number of LSF evaluations depends on the problem dimension in both surrogate-based approaches. This is due to the fact that the PLS-PCE model requires at least d (often more) samples to identify a suitable subspace. Thus, as described above, we choose $n_{\mathcal{E}}$ as a multiple of d . Since the surrogate-free version of SIS-aCS does not possess such a dependence on a problem dimension at all, the ratio of computational cost associated with SIS and ASSIS decreases as d increases. This observation also indicates that if d grows large enough, SIS-aCS will outperform any surrogate-based approach. This is expected for cases with $d = \mathcal{O}(10^5)$ and above; therefore, this observation is of little practical relevance for most engineering models, where ASSIS will likely be the most cost-effective choice.

6.6.3 High-dimensional example: Steel plate

We consider a modified version of the example given in [81, 50], which consists of a low-carbon steel plate of length 0.32 m, width 0.32 m, thickness $t = 0.01$ m, and a hole of radius 0.02 m located at the center. The Poisson ratio is set to $\nu = 0.29$ and the density of the plate is $\rho = 7850$ kg/m³. The horizontal and vertical displacements are constrained at the left edge. The plate is subjected to a random surface load that acts on the right narrow plate side. The load is modelled as a log-normal random variable with mean $\mu_q = 60$ MPa and $\sigma_q = 12$ MPa. The Young's modulus $E(x, y)$ is considered uncertain and spatially variable. It is described by a homogeneous random field with lognormal marginal distribution, mean value $\mu_E = 2 \times 10^5$ MPa and standard deviation $\sigma_E = 3 \times 10^4$ MPa. The autocorrelation function of the underlying Gaussian field $\ln E$ is modeled by the isotropic exponential model

$$\rho_{\ln E}(\Delta x, \Delta y) = \exp \left\{ -\frac{\sqrt{\Delta x^2 + \Delta y^2}}{l_E} \right\} \quad (6.46)$$

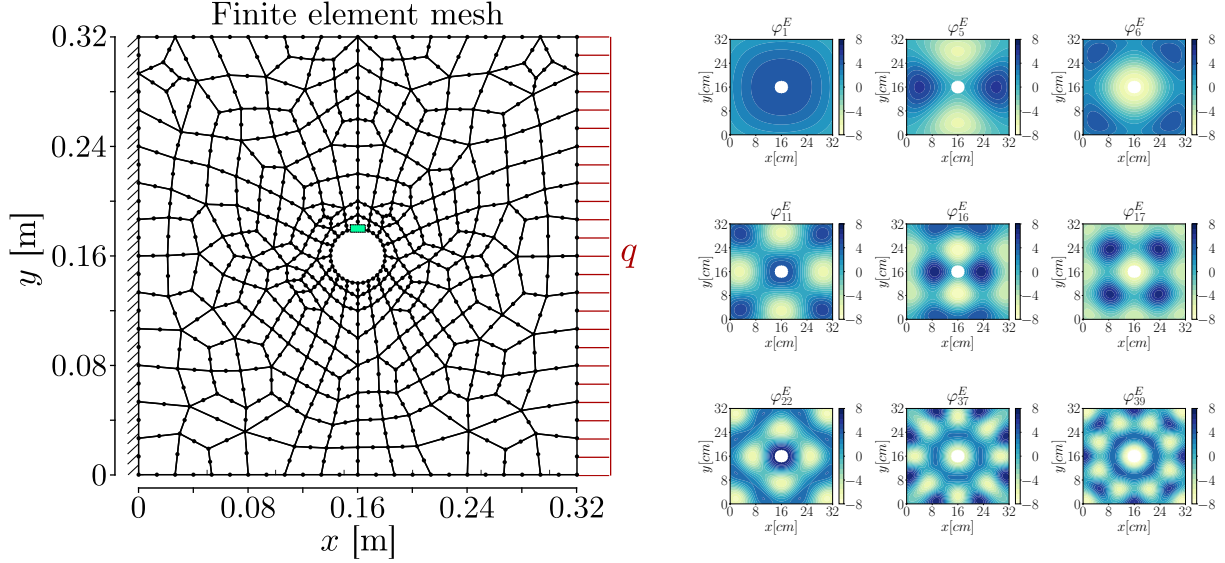


Figure 6.6: Left: FE-mesh of 2D-plate model with control node of the first principal stress σ_1 .

with correlation length $l_{\ln E} = 0.04\text{m}$. The Gaussian random field $\ln E$ is discretized by a Karhunen-Loève-expansion (KLE) with $d_E = 868$, which yields a mean error variance of 7.5% and reads

$$E(x, y) = \exp \left\{ \mu_{\ln E} + \sigma_{\ln E} \sum_{i=1}^{d_E} \sqrt{\lambda_i^E} \varphi_i^E(x, y) \xi_i \right\}. \quad (6.47)$$

$\mu_{\ln E}$ and $\sigma_{\ln E}$ are the parameters of the log-normal marginal distribution of E , $\{\lambda_i^q, \varphi_i^E\}$ are the eigenpairs of the correlation kernel in Eq . (6.46) and $\boldsymbol{\xi} \in \mathbb{R}^{d \times 1}$ is a standard-normal random vector. The most influential eigenfunctions (based on a global output-oriented sensitivity analysis of the plate model performed in [23]) are shown in Fig. 6.6 on the right.

The stress ($\boldsymbol{\sigma}(x, y) = [\sigma_x(x, y), \sigma_y(x, y), \tau_{xy}(x, y)]^T$), strain ($\boldsymbol{\epsilon}(x, y) = [\epsilon_x(x, y), \epsilon_y(x, y), \gamma_{xy}(x, y)]^T$) and displacement ($\mathbf{u}(x, y) = [u_x(x, y), u_y(x, y)]^T$) fields of the plate are given through elasticity theory, namely the Cauchy-Navier equations [38]. Given the configuration of the plate, the model can be simplified under the plane stress hypothesis, which yields

$$G(x, y) \nabla^2 \mathbf{u}(x, y) + \frac{E(x, y)}{2(1 - \nu)} \nabla(\nabla \cdot \mathbf{u}(x, y)) + \mathbf{b} = 0. \quad (6.48)$$

Therein, $G(x, y) := E(x, y)/(2(1 + \nu))$ is the shear modulus, and $\mathbf{b} = [b_x, b_y]^T$ is the vector of body forces acting on the plate. Eq. (6.48) is discretized with a finite-element method. That is, the spatial domain of the plate is discretized into 282 eight-noded quadrilateral elements, as shown in Fig. 6.6. In a grid independence study, the plate's probability of failure was found to slightly increase with decreasing mesh element size, which is likely due to the reduction of averaging effects when integrating higher-order KL-terms. However, for the purpose of testing ASSIS, the model is sufficiently accurate and features two important properties: 1. It possesses a low-dimensional structure that can be exploited with dimensionality-reducing surrogates. 2. It is truly high-dimensional in the sense that the solution does not only depend on a small subset of the input variables (i.e., the low-dimensional structure is not a trivial subspace of the original input space). The LSF is defined

Table 6.2: Accuracy and cost of SIS, SSIS & ASSIS for the plate example based on 100 repetitions of the analysis. The reference $p_{\text{ref}} = 4.23 \cdot 10^{-6}$ is computed with 100 repeated runs of subset simulation with 10^4 samples per level with $\text{CoV} = 0.0119$ for the mean estimate.

Method	$\mathbb{E}[p]$	relative RMSE	CoV	relative bias	avg. # LSF evaluations
SIS-aCS	$3.88 \cdot 10^{-6}$	0.576	0.625	0.083	17000
SSIS	$3.99 \cdot 10^{-6}$	0.061	0.021	0.058	1300
ASSIS	$4.10 \cdot 10^{-6}$	0.036	0.021	0.030	1318

by means of a threshold for the the first principal plane stress

$$\sigma_1 = 0.5(\sigma_x + \sigma_y) + \sqrt{[0.5(\sigma_x + \sigma_y)]^2 + \tau_{xy}^2}$$

evaluated at node 11 (see green marker Fig. 6.6, left). Node 11 indicates a location where maximum plane stresses occur frequently in this example. The LSF reads

$$g(\mathbf{U}) = \sigma_{\text{threshold}} - \sigma_1(\mathbf{U}), \quad (6.49)$$

where $\sigma_{\text{threshold}} = 450$ MPa. The target probability of failure is determined to $p = 4.23 \cdot 10^{-6}$ with $\text{CoV} = 0.0119$ as the average of 100 repeated runs of subset simulation [3] with 10^4 samples per level.

SIS-aCS is run with $n = 2 \cdot 10^3$ samples per level and a burn-in period of $b = 5$ samples within each MCMC chain. SSIS and ASSIS are run with $n = 10^5$ samples per SIS level, a burn-in period $b = 30$ and an AL threshold of $\epsilon_{\text{AL}} = 0.1$. In the first level $n_{\mathcal{E}} = 900$ and in each additional level only $n_{\mathcal{E}} = 100$ samples are added in the initial sampling phase. Tab. 6.2 lists the average estimated probabilities of failure along with error measures and average number of required LSF evaluations. It is seen that both SSIS and ASSIS alleviate computational cost by more than an order of magnitude while at the same time reducing the relative RMSE by at least an order of magnitude. The decomposition of the RMSE in CoV and relative bias reveals that this is mostly due to variance reduction as SIS-aCS already yields a small bias.

A parameter study of important 'tweakable' parameters of ASSIS is depicted in Fig. 6.7. Parameters that are not subject to a parametric study are chosen as above, with the exception of $n = 10^4$ instead of $n = 10^5$. The estimation error and computational cost of ASSIS are analyzed for varying AL thresholds ϵ_{AL} , number of samples in the training set $n_{\mathcal{E}}$, the number of samples per SIS level n and the target CoV δ_{target} used for the SIS procedure. The scaling of 10% between the initial training set and all subsequent training samples is kept constant.

The parameters ϵ_{AL} and $n_{\mathcal{E}}$ describe the behaviour of the surrogate modelling and active learning procedures while n and δ_{target} describe SIS itself. Fig. 6.7 shows that increasing the target coefficient of variation leads to a reduced number of levels in the SIS procedure, which is directly associated with a reduction in computational cost. The reduction is relatively small here as most of the samples are added in the first level. By design, the number of required samples remains unaffected by varying the number of samples per SIS level, while the estimation error depends reciprocally on it. Conversely, and also by design, the computational cost depends monotonically on the choice of $n_{\mathcal{E}}$. If a majority of the used original LSF evaluations are added during an AL procedure, this relationship may be nonlinear. For the plate example, however, the initially drawn

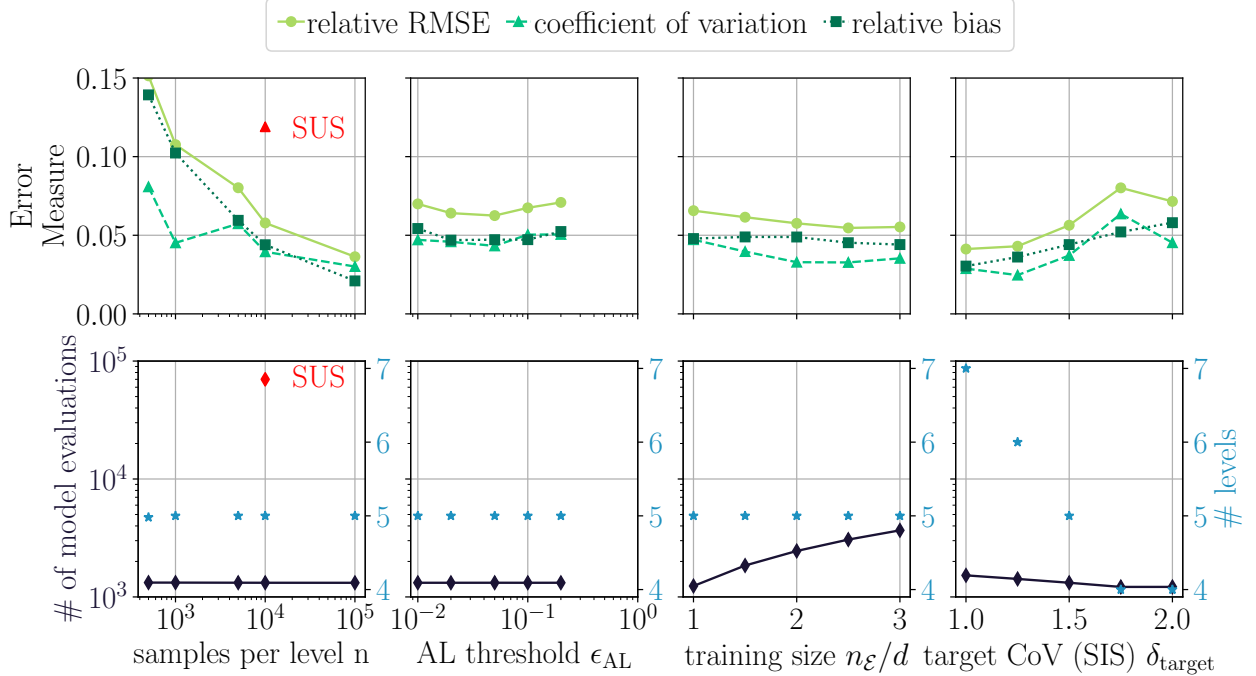


Figure 6.7: Steel plate reliability using ASSIS: parameter influence studies. Top: Error measures as defined in Eqs. (6.42) to (6.44) for ASSIS (green lines w/ markers). Bottom: Computational cost in terms of total number of limit-state function evaluations with the true computational model (left y-axis; black solid lines with diamond markers) and number of SIS levels to convergence (right y-axis; blue star markers). Top left: CoV of a subset simulation reference run with $n = 10^4$ samples per level (red triangle marker). Bottom left: total number of required limit-state function evaluations of a subset simulation reference run with $n = 10^4$ samples per level (red triangle marker).

training samples at each level makes up for the majority of used original LSF evaluations, hence the linear dependency. The estimation errors decrease slightly with increasing training set size, although the effect is limited as high accuracy is already achieved with the first training set of the lowest investigated size. The fact that the subspace does not change significantly with increasing SIS level leaves little to be learned by adding more LSF evaluations to the training set. This is also the reason for the competitive performance of SSIS in this example. The estimation errors (as well as the computational cost in this case) remain unaffected by varying AL thresholds ϵ_{AL} , which is in line with the observation that a large fraction of the computational budget is spent on sampling the initial training set rather than the AL-based training set augmentation.

6.7 Concluding remarks

This paper proposes a method for the cost-efficient solution of high-dimensional reliability problems. We build on a recently introduced dimensionality reducing surrogate modelling technique termed partial least squares-driven polynomial chaos expansion (PLS-PCE) [60] and previous work, in which we use PLS-PCE surrogates to reconstruct biasing densities within a sequential importance sampling scheme [59] (sequential subspace importance sampling: SSIS). We refine this approach by devising

an active learning procedure in each SIS level, which serves to effectively control the estimation error introduced by the surrogate-based importance density reconstructions. The learning procedure, i.e., the selection of new points for the training set, is driven by an estimate of both the subspace and surrogate model estimation error. This criterion can be generally used in polynomial chaos expansion-based active learning procedures.

We showcase the performance of SSIS and ASSIS in nine example applications with input dimensionality ranging from $d = 2$ to 869. The examples feature different typical caveats for reliability methods such as multiple failure domains, strongly nonlinear limit-state functions and extremely small target probabilities of failure. Depending on the example, we achieve a cost reduction of one to over two orders of magnitude with ASSIS compared to the reference method (sequential importance sampling with the original model) at equal or lower estimation errors. It is shown that SSIS is susceptible to the randomness of the initial training set occasionally producing outliers if the training set is adverse. The active learning procedure (ASSIS) remedies this drawback and stabilizes the estimator by augmenting potentially adverse training sets with informative additional samples.

The million dollar question, as with any surrogate model, is on the method's ability to generalize. Certainly, there exist examples that do not possess a suitable linear subspace as required by PLS-PCE modelling. Further, cases of model misspecification may arise if the computational model cannot be represented with PCEs (e.g., if it is a rational function). Then, the probability of failure estimate produced by ASSIS will be neither consistent nor unbiased. However, by means of coupling PLS-PCE with sequential importance sampling, both requirements are relaxed somewhat as only a locally accurate surrogate model is required to propagate samples from one intermediate biasing density to the next. Hence, ASSIS can still be expected to perform well if the computational model may be represented in terms of a sequence of local linear subspaces on which the model can be approximated well with polynomials. Relaxing the orthogonality or even the linearity assumption on the latent space transformation likely bears potential to improve the performance of dimensionality-reduced PCEs. Doing so will require methods to track the appropriate PCE basis upon determining the law of the transformed input random vector (as these will not be standard-normal if the latent space transformation is no longer subject to the orthogonality constraint).

6.8 Acknowledgment

This project was supported by the German Research Foundation (DFG) through Grant STR 1140/6-1 under SPP 1886.

References

- [1] J. An and A. Owen. "Quasi-regression". In: *Journal of Complexity* 17.4 (2001), pp. 588–607.
- [2] S. Au and J. Beck. "A new adaptive importance sampling scheme for reliability calculations". In: *Structural Safety* 21.2 (1999), pp. 135–158.

- [3] S.-K. Au and J. L. Beck. “Estimation of small failure probabilities in high dimensions by subset simulation”. In: *Probabilistic Engineering Mechanics* 16.4 (2001), pp. 263–277.
- [4] G. Baffi, E. Martin, and A. Morris. “Non-linear projection to latent structures revisited (the neural network PLS algorithm)”. In: *Computers & Chemical Engineering* 23.9 (1999), pp. 1293–1307.
- [5] M. Balesdent, J. Morio, and J. Marzat. “Kriging-based adaptive Importance Sampling algorithms for rare event estimation”. In: *Structural Safety* 44.Supplement C (2013), pp. 1–10.
- [6] J. Bect, L. Li, and E. Vazquez. “Bayesian Subset Simulation”. In: *SIAM/ASA Journal on Uncertainty Quantification* 5.1 (2017), pp. 762–786.
- [7] M. Berveiller, B. Sudret, and M. Lemaire. “Stochastic finite element: a non intrusive approach by regression”. In: *European Journal of Computational Mechanics* 15.1-3 (2006), pp. 81–92.
- [8] B. J. Bichon et al. “Efficient Global Reliability Analysis for Nonlinear Implicit Performance Functions”. In: *AIAA Journal* 46.10 (2008), pp. 2459–2468.
- [9] G. Blatman and B. Sudret. “Adaptive sparse polynomial chaos expansion based on least-angle regression”. In: *Journal of Computational Physics* 230.6 (2011), pp. 2345–2367.
- [10] G. Blatman and B. Sudret. “Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach”. In: *Comptes Rendus Mécanique* 336.6 (2008), pp. 518–523.
- [11] J.-M. Bourinet. “Rare-event probability estimation with adaptive support vector regression surrogates”. In: *Reliability Engineering & System Safety* 150 (2016), pp. 210–221.
- [12] J.-M. Bourinet, F. Deheeger, and M. Lemaire. “Assessing small failure probabilities by combined subset simulation and Support Vector Machines”. In: *Structural Safety* 33.6 (2011), pp. 343–353.
- [13] C. G. Bucher. “Adaptive sampling — an iterative fast Monte Carlo procedure”. In: *Structural Safety* 5.2 (1988), pp. 119–126.
- [14] F. Cadini, F. Santos, and E. Zio. “An improved adaptive kriging-based importance technique for sampling multiple failure regions of low probability”. In: *Reliability Engineering & System Safety* 131 (2014), pp. 109–117.
- [15] O. Chapelle, V. Vapnik, and Y. Bengio. “Model selection for small sample regression”. In: *Machine Learning* 48.1-3 (2002), pp. 9–23.
- [16] P. G. Constantine, E. Dow, and Q. Wang. “Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces”. In: *SIAM Journal on Scientific Computing* 36.4 (2014), A1500–A1524.
- [17] S. De Jong. “PLS shrinks”. In: *Journal of Chemometrics* 9.4 (1995), pp. 323–326.
- [18] A. Der Kiureghian. “First-and second-order reliability methods”. In: *Engineering Design Reliability Handbook*. Ed. by E. Nikolaidis, D. M. Ghiocel, and S. Singhal. Boca Raton, FL: CRC Press, 2005. Chap. 14.
- [19] O. Ditlevsen and H. O. Madsen. *Structural reliability methods*. John Wiley & Sons Ltd, 1996.
- [20] A. Doostan and H. Owhadi. “A non-adapted sparse approximation of PDEs with stochastic inputs”. In: *Journal of Computational Physics* 230.8 (2011), pp. 3015–3034.
- [21] V. Dubourg, B. Sudret, and F. Deheeger. “Metamodel-based importance sampling for structural reliability analysis”. In: *Probabilistic Engineering Mechanics* 33 (2013), pp. 47–57.

- [22] B. Echard, N. Gayton, and M. Lemaire. “AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation”. In: *Structural Safety* 33.2 (2011), pp. 145–154.
- [23] M. Ehre, I. Papaioannou, and D. Straub. “Global sensitivity analysis in high dimensions with PLS-PCE”. In: *Reliability Engineering & System Safety* 198 (2020), p. 106861.
- [24] M. Ehre et al. *Active sequential learning of low-dimensional model representations for reliability analysis*. Submitted to SIAM J. Comp. Sci. 2021.
- [25] S. Engelund and R. Rackwitz. “A benchmark study on importance sampling techniques in structural reliability”. In: *Structural Safety* 12.4 (1993), pp. 255–276.
- [26] L. Faravelli. “Response Surface Approach for Reliability Analysis”. In: *Journal of Engineering Mechanics* 115.12 (1989), pp. 2763–2781.
- [27] V. Fedorov. *Theory of Optimal Experiments*. Probability and Mathematical Statistics. Academic Press, Jan. 1972. 306 pp.
- [28] B. Fiessler, R. Rackwitz, and H.-J. Neumann. “Quadratic limit states in structural reliability”. In: *Journal of the Engineering Mechanics Division* 105.4 (1979), pp. 661–676.
- [29] X. Guan and R. Melchers. “Effect of response surface parameter variation on structural reliability estimates”. In: *Structural Safety* 23.4 (2001), pp. 429–444.
- [30] M. Hohenbichler and R. Rackwitz. “Improvement Of Second-Order Reliability Estimates by Importance Sampling”. In: *Journal of Engineering Mechanics* 114.12 (1988), pp. 2195–2199.
- [31] M. Hohenbichler and R. Rackwitz. “Non-normal dependent vectors in structural safety”. In: *Journal of the Engineering Mechanics Division* 107.6 (1981), pp. 1227–1238.
- [32] A. Höskuldsson. “PLS regression methods”. In: *Journal of Chemometrics* 2.3 (1988), pp. 211–228.
- [33] X. Huang, J. Chen, and H. Zhu. “Assessing small failure probabilities by AK-SS: An active learning method combining Kriging and Subset Simulation”. In: *Structural Safety* 59 (2016), pp. 86–95.
- [34] J. E. Hurtado. “Filtered importance sampling with support vector margin: a powerful method for structural reliability analysis”. In: *Structural Safety* 29.1 (2007), pp. 2–15.
- [35] J. E. Hurtado and D. A. Alvarez. “Neural-network-based reliability analysis: a comparative study”. In: *Computer Methods in Applied Mechanics and Engineering* 191.1 (2001), pp. 113–132.
- [36] S. Ji, Y. Xue, and L. Carin. “Bayesian Compressive Sensing”. In: *IEEE Transactions on Signal Processing* 56.6 (2008), pp. 2346–2356.
- [37] Z. Jiang and J. Li. “High dimensional structural reliability with dimension reduction”. In: *Structural Safety* 69 (2017), pp. 35–46.
- [38] C. Johnson. *Numerical solution of partial differential equations by the finite element method*. Dover Publications, 2009. 288 pp.
- [39] K. Konakli and B. Sudret. “Reliability analysis of high-dimensional models using low-rank tensor approximations”. In: *Probabilistic Engineering Mechanics* 46 (2016), pp. 18–36.
- [40] P. Koutsourelakis, H. Pradlwarter, and G. Schuëller. “Reliability of structures in high dimensions, part I: algorithms and applications”. In: *Probabilistic Engineering Mechanics* 19.4 (2004), pp. 409–417.

- [41] H. Kriegel, P. Kröger, and A. Zimek. “Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering”. In: *ACM Trans. Knowl. Discov. Data* 3 (2009), 1:1–1:58.
- [42] D. P. Kroese, R. Y. Rubinstein, and P. W. Glynn. “Chapter 2 - The Cross-Entropy Method for Estimation”. In: *Handbook of Statistics: Machine Learning: Theory and Applications*. Vol. 31. Handbook of Statistics. Elsevier, 2013, pp. 19–34.
- [43] N. Kurtz and J. Song. “Cross-entropy-based adaptive importance sampling using Gaussian mixture”. In: *Structural Safety* 42 (2013), pp. 35–44.
- [44] S. H. Lee and B. M. Kwak. “Response surface augmented moment method for efficient reliability analysis”. In: *Structural Safety* 28.3 (2006), pp. 261–272.
- [45] J. Li, J. Li, and D. Xiu. “An efficient surrogate-based method for computing rare failure probability”. In: *Journal of Computational Physics* 230.24 (2011), pp. 8683–8697.
- [46] J. Li and D. Xiu. “Evaluation of failure probability via surrogate models”. In: *Journal of Computational Physics* 229.23 (2010), pp. 8966–8980.
- [47] M. Li and Z. Wang. “Deep learning for high-dimensional reliability analysis”. In: *Mechanical Systems and Signal Processing* 139 (2020), p. 106399.
- [48] R. Li and R. Ghanem. “Adaptive polynomial chaos expansions applied to statistics of extremes in nonlinear random vibration”. In: *Probabilistic Engineering Mechanics* 13.2 (1998), pp. 125–136.
- [49] P.-L. Liu and A. Der Kiureghian. “Multivariate distribution models with prescribed marginals and covariances”. In: *Probabilistic Engineering Mechanics* 1.2 (1986), pp. 105–112.
- [50] P.-L. Liu and K.-G. Liu. “Selection of random field mesh in finite element reliability analysis”. In: *Journal of Engineering Mechanics* 119.4 (1993), pp. 667–680.
- [51] N. Lüthen, S. Marelli, and B. Sudret. *A benchmark of basis-adaptive sparse polynomial chaos expansions for engineering regression problems*. 2021.
- [52] N. Lüthen, S. Marelli, and B. Sudret. *Sparse Polynomial Chaos Expansions: Literature Survey and Benchmark*. 2021.
- [53] S. Marelli and B. Sudret. “An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis”. In: *Structural Safety* 75 (2018), pp. 67–74.
- [54] J. Oakley. “Estimating percentiles of uncertain computer code outputs”. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 53.1 (2004), pp. 83–93.
- [55] A. B. Owen. *Monte Carlo theory, methods and examples*. 2013.
- [56] Q. Pan and D. Dias. “Sliced inverse regression-based sparse polynomial chaos expansions for reliability analysis in high dimensions”. In: *Reliability Engineering & System Safety* 167 (2017). Special Section: Applications of Probabilistic Graphical Models in Dependability, Diagnosis and Prognosis, pp. 484–493.
- [57] V. Papadopoulos et al. “Accelerated subset simulation with neural networks for reliability analysis”. In: *Computer Methods in Applied Mechanics and Engineering* 223 (2012), pp. 70–80.
- [58] M. Papadrakakis, V. Papadopoulos, and N. D. Lagaros. “Structural reliability analysis of elastic-plastic structures using neural networks and Monte Carlo simulation”. In: *Computer Methods in Applied Mechanics and Engineering* 136.1-2 (1996), pp. 145–163.

- [59] I. Papaioannou, M. Ehre, and D. Straub. “Efficient PCE representations for reliability analysis in high dimensions”. In: *Proceedings of the 19th working conference of the IFIP Working Group 7.5 on Reliability and Optimization of Structural Systems*. Ed. by J. Song. ETH Zürich, 2018.
- [60] I. Papaioannou, M. Ehre, and D. Straub. “PLS-based adaptation for efficient PCE representation in high dimensions”. In: *Journal of Computational Physics* 387 (2019), pp. 186–204.
- [61] I. Papaioannou, S. Geyer, and D. Straub. “Improved cross entropy-based importance sampling with a flexible mixture model”. In: *Reliability Engineering & System Safety* 191 (2019), p. 106564.
- [62] I. Papaioannou, C. Papadimitriou, and D. Straub. “Sequential importance sampling for structural reliability analysis”. In: *Structural Safety* 62 (2016), pp. 66–75.
- [63] Y. C. Pati et al. “Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition”. In: *Proceedings of the 27th Annual Asilomar Conference on Signals, Systems, and Computers (1993)*. 1993, pp. 40–44.
- [64] B. Peherstorfer, B. Kramer, and K. Willcox. “Multifidelity Preconditioning of the Cross-Entropy Method for Rare Event Simulation and Failure Probability Estimation”. In: *SIAM/ASA Journal on Uncertainty Quantification* 6.2 (2018), pp. 737–761.
- [65] G. Perrin. “Adaptive calibration of a computer code with time-series output”. In: *Reliability Engineering & System Safety* 196 (2020), p. 106728.
- [66] V. Picheny et al. “Adaptive Designs of Experiments for Accurate Approximation of a Target Region”. In: *Journal of Mechanical Design* 132.7 (June 2010).
- [67] R. Rackwitz and B. Fiessler. “Structural reliability under combined random load sequences”. In: *Computers & Structures* 9.5 (1978), pp. 489–494.
- [68] R. Rosipal. “Nonlinear Partial Least Squares: An Overview”. In: *Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques* (Jan. 2010), pp. 169–189.
- [69] R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo Method*. 3rd. Wiley Publishing, 2017.
- [70] K. Sargsyan et al. “Dimensionality reduction for complex models via Bayesian compressive sensing”. In: *International Journal for Uncertainty Quantification* 4.1 (2014), pp. 63–93.
- [71] R. Schöbi, B. Sudret, and S. Marelli. “Rare Event Estimation Using Polynomial-Chaos Kriging”. In: *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 3.2 (2017), p. D4016002.
- [72] L. Schueremans and D. V. Gemert. “Benefit of splines and neural networks in simulation based structural reliability analysis”. In: *Structural Safety* 27.3 (2005), pp. 246–261.
- [73] B. Settles. *Active Learning Literature Survey*. Computer Sciences Technical Report 1648. University of Wisconsin–Madison, 2009.
- [74] B. Sudret, G. Blatman, and M. Berveiller. “Response surfaces based on polynomial chaos expansions”. In: *Construction reliability: Safety, Variability and Sustainability* (2013), pp. 147–167.
- [75] R. Tipireddy and R. Ghanem. “Basis adaptation in homogeneous chaos spaces”. In: *Journal of Computational Physics* 259 (2014), pp. 304–317.

- [76] M. E. Tipping. “Sparse Bayesian Learning and the Relevance Vector Machine”. In: *J. Mach. Learn. Res.* 1 (Sept. 2001), pp. 211–244.
- [77] J. A. Tropp and A. C. Gilbert. “Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit”. In: *IEEE Trans. Inf. Theor.* 53.12 (2007), pp. 4655–4666.
- [78] P. Tsilifis et al. “Compressive sensing adaptation for polynomial chaos expansions”. In: *Journal of Computational Physics* 380 (2019), pp. 29–47.
- [79] P. Tsilifis et al. “Sparse Polynomial Chaos expansions using variational relevance vector machines”. In: *Journal of Computational Physics* 416 (2020), p. 109498.
- [80] E. Ullmann and I. Papaioannou. “Multilevel Estimation of Rare Events”. In: *SIAM/ASA Journal on Uncertainty Quantification* 3.1 (2015), pp. 922–953.
- [81] F. Uribe et al. “Cross-entropy-based importance sampling with failure-informed dimension reduction for rare event simulation”. In: *arXiv:2006.05496 preprint* (June 2020), pp. 1–24.
- [82] P. Waarts. “Structural Reliability Using Finite Element Analysis: An Appraisal of DARS: Directional Adaptive Response Surface Sampling”. PhD thesis. 2000.
- [83] F. Wagner et al. “Multilevel Sequential Importance Sampling for Rare Event Estimation”. In: *SIAM Journal on Scientific Computing* 42.4 (2020), A2062–A2087.
- [84] Z. Wang and J. Song. “Cross-entropy-based adaptive importance sampling using von Mises-Fisher mixture for high dimensional reliability analysis”. In: *Structural Safety* 59 (Mar. 2016), pp. 42–52.
- [85] S. Wold, N. Kettaneh-Wold, and B. Skagerberg. “Nonlinear PLS modeling”. In: *Chemometrics and Intelligent Laboratory Systems* 7.1 (1989). Proceedings of the First Scandinavian Symposium on Chemometrics, pp. 53–65.
- [86] S. Wold et al. “The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses”. In: *SIAM Journal on Scientific and Statistical Computing* 5.3 (1984), pp. 735–743.
- [87] C.-F. Wu. “Asymptotic Theory of Nonlinear Least Squares Estimation”. In: *The Annals of Statistics* 9.3 (1981), pp. 501–513.
- [88] D. Xiu and G. E. Karniadakis. “The Wiener–Askey polynomial chaos for stochastic differential equations”. In: *SIAM Journal on Scientific Computing* 24.2 (2002), pp. 619–644.
- [89] L. Yan, L. Guo, and D. Xiu. “Stochastic collocation algorithms using l1-minimization”. In: *International Journal for Uncertainty Quantification* 2.3 (2012), pp. 279–293.
- [90] T. Zhou and Y. Peng. “Structural reliability analysis via dimension reduction, adaptive sampling, and Monte Carlo simulation”. In: *Structural and Multidisciplinary Optimization* (2020), pp. 2629–2651.

Conditional reliability analysis in high dimensions based on controlled mixture importance sampling and information reuse

Original Publication

M. Ehre, I. Papaioannou, K. E. Willcox, and D. Straub. “Conditional reliability analysis in high dimensions based on controlled mixture importance sampling and information reuse”. In: *Computer Methods in Applied Mechanics and Engineering* 381 (2021), p. 113826.

Author’s contribution

Max Ehre and Iason Papaioannou developed the concepts underlying the method. Max Ehre developed and implemented the algorithm and carried out all numerical experiments. Max Ehre drafted the manuscript under the guidance of Iason Papaioannou, Daniel Straub and Karen Willcox.

Abstract

In many contexts, it is of interest to assess the impact of selected parameters on the failure probability of a physical system. To this end, one can perform conditional reliability analysis, in which the probability of failure becomes a function of these parameters. Computing conditional reliability requires recomputing failure probabilities for a sample sequence of the parameters, which strongly increases the already high computational cost of conventional reliability analysis. We alleviate these costs by reusing information from previous reliability computations in each subsequent reliability analysis of the sequence. The method is designed using two variants of importance sampling and performs information transfer by reusing importance densities from previous reliability analyses in the current one. We put forward a criterion for selecting the most informative importance densities,

which is robust with respect to the input space dimension, and use a recently proposed density mixture model for constructing effective importance densities in high dimensions. The method controls the estimator coefficient of variation to achieve a prescribed accuracy. We demonstrate its performance by means of two engineering examples featuring a number of pitfall features such as strong non-linearity, high dimensionality and small failure probabilities (10^{-5} to 10^{-9}).

7.1 Introduction

In order to accurately predict model behaviour with confidence, it is vital to account for uncertainties influencing the model and its output. Reliability analysis is concerned with quantifying the extremal behaviour of a model under uncertainty by computing its probability of failure, i.e., the probability of an unacceptable model response. Often it is of interest to repeatedly perform the analysis on a series of parametrised reliability problems. Such situations arise in reliability-based design optimization (RBDO), where the parametrisation is given by the design parameters, or whenever it is desirable to separate the model inputs into two categories. Hereafter, we refer to these categories as type A and type B. One example of such a separation is in reducible (epistemic) and irreducible (aleatory) uncertainty [22, 36, 23, 25, 13]. The general goal of separating inputs in this way is to establish a distinct relationship between type B variables and the probability of failure conditional on type B variables. By conditioning the probability of failure on type B variables, one obtains a measure for the influence of these variables on the probability of failure. In general, the concept applies to any target that can be cast in terms of an expected value. Such a formulation is useful in many contexts — it may, e.g., be used to provide estimates of credibility bounds, dispersion measures or the distribution of the probability of failure conditional on type B variables thus quantifying lack of knowledge/confidence caused by these variables. It gives rise to global sensitivity measures of the conditional failure probability with respect to the type B variables (e.g., Sobol’ indices [16]) and facilitates the computation of the partial value of (im-)perfect information for eliciting optimal decisions based on conditional failure probabilities [37]. We have demonstrated how to construct surrogate models mapping the type B-variables to the conditional probability of failure to obtain global sensitivity measures at significantly reduced computational cost [16].

Conventional reliability analysis is a challenging task as failure probabilities are typically associated with rare events [14, 33] and thus assume small values. Simultaneously, it has received considerable attention due to its relevance to engineering and financial applications. Structural reliability methods (SRM) can be categorized into approximation-based methods, such as the first- (FORM) and second-order reliability method (SORM) [12] and sampling-based methods (importance sampling [8, 20, 1], sequential importance sampling [5, 43], subset simulation [3], cross-entropy importance sampling [50, 31, 32], line-sampling [26, 30], multi-level Monte Carlo (MC) [51] and multi-fidelity MC [44]). Conditional reliability analysis is considerably more expensive compared to its conventional counterpart as it requires the solution of a sequence of reliability problems rather than a single one. A number of sampling approaches have been developed for computing the probability of multiple correlated failure events efficiently. Ref. [27] introduces parallel subset simulation to estimate failure probabilities of several failure events simultaneously by defining a principle variable that is correlated with each failure event. In Refs. [34, 4] subset simulation is applied to a parallel system with each system component representing one reliability problem in the sequence. The corresponding failure probabilities can be estimated based on the failure probability of the system and the conditional

samples from each subset. Refs. [27, 34, 4] are efficient if all considered failure events are strongly correlated, but will encounter difficulties if there are failure events occurring in the sequence that are not correlated with any of the other events. This implies that these methods are not suited for conditional reliability analysis if the number of type B-samples is large and/or if the type B-samples contribute a large fraction of variability to the conditional probability of failure, since in either case, the probability of disjunct failure events in the sequence is considerable. Ref. [35] proposes a method for robust optimization problems, i.e., design optimization under probabilistic constraints that include the mean and variance of the model response. Control variates are used to recycle information stemming from the previous optimizer iteration to accelerate the MC constraint computation in each optimization step (but the first). The design parameters, which parametrize the model and which change in each design iteration, can be viewed as a deterministic counterpart to type B-variables (see Subsection 7.3.2). The same is true for RBDO, where the design optimization is carried out under constraints on system reliability rather than response moments. Ref. [5] proposes bridge importance sampling to solve RBDO: importance densities of reliability computations at previous steps in the design optimization are used to initialize a bridging step towards the current optimal importance density. Suitable density candidates are identified based on a heuristic that has inspired an earlier approach to conditional reliability analysis [17] as well as this contribution (Subsection 7.3.3). Ref. [10] proposes to solve RBDO with importance sampling where information from previous design iterations is incorporated in the choice of the importance density of the current reliability problem.

Here, we propose a method for solving the conditional reliability problem in high dimensions efficiently through information reuse. In Section 7.2, we briefly recap conventional reliability analysis and popular solution approaches before formally introducing the conditional reliability problem in Section 7.3. We then discuss our approach to information reuse that consists of a selection strategy for informative importance densities and two importance samplers (mixed and controlled) that serve to exploit the selected densities in the current reliability estimate. In Section 7.4, we present comprehensive investigations of the method’s performance in two engineering examples and provide a detailed discussion of the results. Conclusions are given in Section 7.5.

7.2 Conventional reliability analysis

In this section, we set up the reliability problem formulation and discuss well-established approaches to its solution. In the second part, we discuss the cross-entropy method (CE) [50, 49] and a recently introduced improved version thereof (iCE) [42].

7.2.1 Problem Statement

Consider a system that is modelled by $\mathcal{Y} : \mathbb{D}_{\Theta} \rightarrow \mathbb{R}$ with d -dimensional random input vector $\Theta : \Omega \rightarrow \mathbb{D}_{\Theta} \subseteq \mathbb{R}^d$, where Ω is the sample space of Θ . F_{Θ} is the joint cumulative distribution function (CDF) of Θ . \mathcal{Y} maps to the system response $Y = \mathcal{Y}(\theta)$ with the model input $\theta \in \mathbb{D}_{\Theta}$. Based on the response Y , unacceptable system states are defined by means of the limit-state function

LSF $\tilde{g}(Y)$. Defining $g(\boldsymbol{\theta}) = \tilde{g} \circ \mathcal{Y}(\boldsymbol{\theta})$ and introducing the convention

$$g(\boldsymbol{\theta}) = \begin{cases} \leq 0, \text{ Failure} \\ > 0, \text{ Safety,} \end{cases}$$

the failure event of the system is defined as $F = \{\boldsymbol{\theta} \in \mathbb{D}_{\boldsymbol{\Theta}} : g(\boldsymbol{\theta}) \leq 0\}$. The probability of failure is given by [14]

$$P = \mathbb{P}(F) = \int_{\mathbb{D}_{\boldsymbol{\Theta}}} \mathbb{I}[g(\boldsymbol{\theta}) \leq 0] f(\boldsymbol{\theta}) d\boldsymbol{\theta} = \mathbb{E}[\mathbb{I}(g(\boldsymbol{\Theta}) \leq 0)], \quad (7.1)$$

where f is the joint probability density function (PDF) of $\boldsymbol{\Theta}$ and the indicator function $\mathbb{I}[\cdot]$ equals 1 if true and 0 otherwise. Without loss of generality, one may formulate an equivalent reliability problem with respect to the standard-normal probability space using the random vector $\mathbf{U} : \Omega \rightarrow \mathbb{R}^d$. Given an isoprobabilistic transformation $T : \mathbb{D}_{\boldsymbol{\Theta}} \rightarrow \mathbb{R}^d$, such that $\mathbf{U} = T(\boldsymbol{\Theta})$ [48] and defining $G = g(T^{-1}(\mathbf{U}))$, one can write Eq. (7.1) as

$$P = \int_{\mathbb{R}^d} \mathbb{I}[G(\mathbf{u}) \leq 0] \varphi_d(\mathbf{u}) d\mathbf{u} = \mathbb{E}[\mathbb{I}(G(\mathbf{U}) \leq 0)], \quad (7.2)$$

where φ_d denotes the d -dimensional independent standard-normal PDF.

7.2.2 Standard MC

The standard MC estimate of the integral in Eq. (7.2) reads

$$\hat{p}_{\text{MC}} = \frac{1}{n} \sum_{k=1}^n \mathbb{I}[G(\mathbf{u}^k) \leq 0], \quad \mathbf{u}^k \stackrel{i.i.d.}{\sim} \varphi_d.$$

This estimate is unbiased and has coefficient of variation (CoV)

$$\delta_{\text{MC}} = \sqrt{\frac{1-P}{nP}}.$$

Its costs in terms of g -evaluations ($= n$) are independent of the model dimension d . If $P \ll 1$, δ_{MC} scales approximately inversely with the square root of the failure probability and n becomes large for small values of P . Namely, for a target δ_0 , at least n_0 evaluations of G are required, where

$$n_0 = \frac{1-P}{\delta_0^2 P}.$$

Thus, while independent of the model input dimension, the standard MC estimate is not suited for estimating rare events if evaluating \mathcal{Y} is not cheap.

7.2.3 Importance sampling

One of the most commonly used techniques to alleviate the above restriction on the sample size while achieving a prescribed CoV is the importance sampling (IS) method. Let h be a density, such

that $h(\mathbf{u}) > 0$ whenever $G(\mathbf{u}) \leq 0$. Then, we can rewrite Eq. (7.1)

$$P = \int_{\mathbb{R}^d} \mathbf{I}(G(\mathbf{u}) \leq 0) \overbrace{\frac{\varphi_d(\mathbf{u})}{h(\mathbf{u})}}^{w(\mathbf{u})} h(\mathbf{u}) d\mathbf{u} = \mathbb{E}_h [\mathbf{I}(G(\mathbf{U}) \leq 0)w(\mathbf{U})], \quad (7.3)$$

where h is termed the *importance, auxiliary, instrumental* or *biasing* density and w is the *likelihood ratio* or *IS weight*. In the context of importance sampling, φ_d is often referred to as the *nominal* density. The corresponding estimate of the probability of failure is given by

$$\hat{p}_{\text{IS}} = \frac{1}{n} \sum_{k=1}^n \mathbf{I}[G(\mathbf{u}^k) \leq 0]w(\mathbf{u}^k), \quad \mathbf{u}^k \stackrel{i.i.d.}{\sim} h. \quad (7.4)$$

\hat{p}_{IS} is an unbiased estimate of P and its variance is

$$\mathbb{V}[\hat{p}_{\text{IS}}] = \frac{1}{n} \mathbb{V}_h [\mathbf{I}[G(\mathbf{U}) \leq 0]w(\mathbf{U})].$$

Estimating the above based on a set of samples drawn from h , we obtain an estimate for the CoV of \hat{p}_{IS} as

$$\hat{\delta}_{\text{IS}} = \frac{1}{\hat{p}_{\text{IS}}} \sqrt{\frac{1}{n(n-1)} \sum_{k=1}^n (\mathbf{I}[G(\mathbf{u}^k) \leq 0]w(\mathbf{u}^k) - \hat{p}_{\text{IS}})^2}, \quad \mathbf{u}^k \stackrel{i.i.d.}{\sim} h. \quad (7.5)$$

There exists an optimal importance density h^* such that $\mathbb{V}[\hat{p}_{\text{IS}}] = 0$:

$$h^*(\mathbf{u}) = \frac{1}{P} \mathbf{I}[G(\mathbf{u}) \leq 0] \varphi_d(\mathbf{u}). \quad (7.6)$$

Note, that h^* requires knowledge of the target quantity P . Thus it cannot be used immediately to compute the integral in Eq. (7.3). However, it gives rise to a variety of approaches that aim at approximating h^* by propagating a sequence of distributions from φ_d towards h^* , e.g., via conditional sampling using Markov Chain Monte Carlo [20, 43] or through fitting parametric density models [8, 50, 49].

7.2.4 The iCE method (iCE)

Here, we discuss a recently proposed version of sequential importance sampling that is based on the classical CE method proposed in Ref. [49] and has been demonstrated to work well in high dimensions [42]. Consider a parametric version of the importance density $h(\mathbf{u}, \mathbf{v})$, which is defined by the parameter vector $\mathbf{v} \in \mathcal{V}$. The parameter space \mathcal{V} contains \mathbf{v}_0 , where $h(\mathbf{u}, \mathbf{v}_0) = \varphi_d(\mathbf{u})$. The standard CE method aims at minimizing the Kullback-Leibler (KL) divergence $D_{KL}(h^*(\mathbf{u})||h(\mathbf{u}, \mathbf{v}))$ between $h^*(\mathbf{u})$ and $h(\mathbf{u}, \mathbf{v})$ over the parameter space \mathcal{V} , which is defined as [50]

$$D_{KL}(h^*(\mathbf{u})||h(\mathbf{u}, \mathbf{v})) = \mathbb{E}_{h^*} \left[\ln \left(\frac{h^*(\mathbf{u})}{h(\mathbf{u}, \mathbf{v})} \right) \right] \quad (7.7)$$

$$\stackrel{\text{Eq. (7.6)}}{=} \frac{1}{P} \mathbb{E}_{\varphi_d} [\mathbf{I}[G(\mathbf{u}) \leq 0] \ln(h^*(\mathbf{u}))] - \frac{1}{P} \mathbb{E}_{\varphi_d} [\mathbf{I}[G(\mathbf{u}) \leq 0]h(\mathbf{u}, \mathbf{v})].$$

The first summand on the right-hand side of Eq. (7.7) is not a function of \mathbf{v} , so that minimizing $D_{KL}(h^*(\mathbf{u})||h(\mathbf{u}, \mathbf{v}))$ can be expressed as

$$\mathbf{v}^* = \arg \max_{\mathbf{v} \in \mathcal{V}} \mathbb{E}_{\varphi_d} [\mathbb{I}[G(\mathbf{U}) \leq 0] \ln(h(\mathbf{U}, \mathbf{v}))] \quad (7.8)$$

and its sample-based approximation reads

$$\hat{\mathbf{v}}^* = \arg \max_{\mathbf{v} \in \mathcal{V}} \frac{1}{n} \sum_{k=1}^n [\mathbb{I}[G(\mathbf{u}^k) \leq 0] \ln(h(\mathbf{u}^k, \mathbf{v}))], \quad \mathbf{u}^k \stackrel{i.i.d.}{\sim} \varphi_d. \quad (7.9)$$

For fixed \mathbf{v} , the objective function in Eq. (7.9) is equivalent to a weighted version of \hat{p}_{MC} . That is, to approximate \mathbf{v}^* well with $\hat{\mathbf{v}}^*$, n has to be large if F is a rare event. The CE method circumvents this problem by approaching h^* stepwise with a sequence of parametric distributions defined by $\{\mathbf{v}_i, i = 1, \dots, m\}$. The failure event F is represented by a series of more probable intermediate events $\{F_i, i = 1, \dots, m\}$ that are defined by manipulating their associated threshold ξ_i s.t. $F_i = \{\mathbf{u} \in \mathbb{R}^d : G(\mathbf{u}) \leq \xi_i\}$, where $\xi_1 > \xi_2 > \dots > \xi_{m-1} > \xi_m$. Starting from $h(\mathbf{u}, \mathbf{v}_0) = \varphi_d(\mathbf{u})$, the threshold ξ_i is determined as the lower ρ -quantile of the LSF based on samples from the parametric density associated with F_{i-1} , $h(\mathbf{u}, \mathbf{v}_{i-1})$, with typical choices for the quantile value $\rho = [10^{-2}, 10^{-1}]$. The i -th parametric density is then found through minimizing the KL divergence between $h_i(\mathbf{u})$ and $h(\mathbf{u}, \mathbf{v}_i)$, where $h_i(\mathbf{u})$ is the optimal importance sampling density associated with the threshold ξ_i . Once all $n\rho$ new samples lie within the failure domain, i.e., G evaluated at these samples is always negative, the algorithm is stopped. Solving Eq. (7.9) based on samples from $h(\mathbf{u}, \hat{\mathbf{v}}_{i-1})$ rather than φ_d introduces the weight $W(\mathbf{u}, \hat{\mathbf{v}}_{i-1}) = \frac{\varphi_d(\mathbf{u})}{h(\mathbf{u}, \hat{\mathbf{v}}_{i-1})}$:

$$\hat{\mathbf{v}}_i = \arg \max_{\mathbf{v} \in \mathcal{V}} \frac{1}{n} \sum_{k=1}^n \mathbb{I}[G(\mathbf{u}^k) \leq \xi_i] \ln(h(\mathbf{u}^k, \mathbf{v})) W(\mathbf{u}^k, \hat{\mathbf{v}}_{i-1}), \quad \mathbf{u}^k \stackrel{i.i.d.}{\sim} h(\mathbf{u}, \hat{\mathbf{v}}_{i-1}). \quad (7.10)$$

Computing the new parameter set $\hat{\mathbf{v}}_i$ based on $n\rho$ samples in each iteration effectively leaves a fraction of $1 - \rho$ (90 – 99%) of the samples unused and motivates the first of two major points of departure of iCE from CE: Within iCE, h_i is re-parametrized using a smooth approximation of $\mathbb{I}[g \leq 0]$ based on the standard-normal CDF $\Phi(\cdot)$:

$$h_i(\mathbf{u}) = \frac{1}{P_i} \Phi\left(-\frac{G(\mathbf{u})}{\sigma_i}\right) \varphi_d(\mathbf{u}) = \frac{1}{P_i} \eta_i(\mathbf{u}), \quad (7.11)$$

where $P_i = \mathbb{E}_{\varphi_d}[\Phi(-G(\mathbf{U})/\sigma_i)]$ is a normalizing constant and σ_i is a smoothing parameter. This distribution sequence has been used to construct adaptive importance sampling-based approaches to reliability analysis [43, 15], reliability sensitivity analysis [39] and RBDO [5]. The CE-program now reads

$$\hat{\mathbf{v}}_i = \arg \max_{\mathbf{v} \in \mathcal{V}} \frac{1}{n} \sum_{k=1}^n \ln(h(\mathbf{u}^k, \mathbf{v})) W(\mathbf{u}^k, \hat{\mathbf{v}}_{i-1}), \quad \mathbf{u}^k \stackrel{i.i.d.}{\sim} h(\mathbf{u}, \hat{\mathbf{v}}_{i-1}), \quad (7.12)$$

where $W(\mathbf{u}, \hat{\mathbf{v}}_{i-1}) = \frac{\eta_i(\mathbf{u})}{h(\mathbf{u}, \hat{\mathbf{v}}_{i-1})}$. In program Eq. (7.12), all samples available from $h(\mathbf{u}, \hat{\mathbf{v}}_{i-1})$ will be used with their respective modified weight. Then, in each step, the current σ_i is identified such that the sample CoV of the weights $\{W(\mathbf{u}^k, \hat{\mathbf{v}}_{i-1}), k = 1, \dots, n\}$, $\hat{\delta}_W(\sigma)$, adheres to a target value δ_{target} :

$$\sigma_i = \arg \min_{\sigma \in [0, \sigma_{i-1}]} \left(\hat{\delta}_W(\sigma) - \delta_{\text{target}} \right)^2. \quad (7.13)$$

$\widehat{\delta}_W(\sigma)$ is a measure of dissimilarity between two subsequent importance densities h_i and $h(\mathbf{u}, \widehat{\mathbf{v}}_{i-1})$ and the choice of δ_{target} needs to balance two conflicting targets: on the one hand, large δ_{target} leads to inaccurate solutions of program Eq. (7.12) as the employed samples from $h(\mathbf{u}, \widehat{\mathbf{v}}_{i-1})$ cannot represent h_i well. On the other hand, with small δ_{target} , two subsequent intermediate importance sampling densities will be similar such that the method progresses slowly and many iterations are required to achieve convergence. [42] suggests $\delta_{\text{target}} = 1.5$, which is employed here as well. The algorithm terminates, when the approximated and optimal importance density, h_i and h^* , are sufficiently close in the sense that $\text{CoV}[I[G(\mathbf{u})]/\Phi(-g(\mathbf{u})/\sigma_i)] \leq \delta_{\text{target}}$ with $\delta_{\text{target}} = 1.5$ a typical choice [42]. Note, that δ_{target} is computed with $h_i(\mathbf{u})$ rather than the parametric $h(\mathbf{u}, \widehat{\mathbf{v}}_i)$, since it is more robust with respect to the flexibility of the parametric model. After termination, additional samples can be drawn from the final parametric importance density to achieve a prescribed estimator CoV according to Eq. (7.5). Alg. 10 describes the iCE procedure in detail.

Algorithm 10 The iCE method

Input LSF $G(\mathbf{u})$, input space dimension d , target CoV δ_{target} , samples per level N

Output estimate \widehat{p}_{iCE} , compute estimate CoV $\widehat{\delta}_{\text{iCE}}$, no. of levels m ,
importance densities $\{h(\mathbf{u}, \widehat{\mathbf{v}}_i), i = 1, \dots, m\}$

```

1: procedure ICE( $g, \delta_{\text{target}}, N, d$ )
2:   Set converged = false
3:   Set  $i = 1$ 
4:   Select  $\widehat{\mathbf{v}}_0$  ▷ e.g., s.t.  $h(\mathbf{u}, \widehat{\mathbf{v}}_0) = \varphi_d(\mathbf{u})$ 
5:   while  $\neg$ converged do
6:     Sample  $\mathbf{U} = \{\mathbf{u}^k, k = 1, \dots, N\} \in \mathbb{R}^{N \times d}$  ▷  $\mathbf{u}^k \stackrel{i.i.d.}{\sim} h(\mathbf{u}, \widehat{\mathbf{v}}_{i-1})$ 
7:      $\mathbf{G} = G(\mathbf{U}) \in \mathbb{R}^{N \times 1}$ 
8:     if  $\text{CoV}[I[\mathbf{G}]/\Phi(-\mathbf{G}/\sigma_i)] \leq \delta_{\text{target}}$  then ▷ CoV of likelihood ratio of  $h_*$  and  $\eta_i(\mathbf{u})$ 
9:       Set  $m = i - 1$ 
10:       $\mathbf{W} = I[\mathbf{G}^k \leq 0]\varphi_d(\mathbf{u}^k)/h(\mathbf{u}^k, \widehat{\mathbf{v}}_m)$  ▷ Likelihood ratio of  $h_*$  and  $h(\mathbf{u}, \widehat{\mathbf{v}}_m)$ 
11:      Estimate the failure probability

```

$$\widehat{p}_{\text{iCE}} = \widehat{\mathbb{E}}(\mathbf{W})$$

```

12:      Compute the failure probability estimate's CoV

```

$$\widehat{\delta}_{\text{iCE}} = \sqrt{\frac{\widehat{\mathbb{V}}(\mathbf{W})}{N\widehat{\mathbb{E}}(\mathbf{W})^2}}$$

```

13:      Set converged = true
14:    else
15:      Compute  $\sigma_i$  from Eq. (7.13)
16:      Compute  $\widehat{\mathbf{v}}_i$  from Eq. (7.12)
17:       $i = i + 1$ 
18:    return  $\widehat{p}_{\text{iCE}}, \widehat{\delta}_{\text{iCE}}, m, \{h(\mathbf{u}, \widehat{\mathbf{v}}_i), i = 1, \dots, m\}$ 

```

The second point of departure of iCE from CE is given by the parametric density model choice. When working in standard-normal space, typical choices for $h(\mathbf{u}, \mathbf{v})$ are the d -dimensional single Gaussian density (SG) [24] or a d -dimensional Gaussian mixture (GM) [32, 24]. For the single Gaus-

sian, Eqs. (7.10) and (7.12) can be solved analytically and for the Gaussian mixture, the solution is identified through the expectation maximization (EM) algorithm. However, within importance sampling algorithms, both models perform poorly in higher-dimensional problems ($d \geq 20$). A detailed discussion of the issue can be found in Refs. [24, 2, 29]. In Ref. [52], a von Mises-Fisher model for the direction in \mathbf{U} -space is proposed to remedy these issues. For the iCE method, this model is extended by a Nakagami distribution for the radius of any point in standard-normal space, which yields the von Mises-Fisher-Nakagami-mixture model (vMFNM) for $h(\mathbf{u}, \mathbf{v})$, where $\mathbf{v} = [\mathbf{m}, \mathbf{\Omega}, \boldsymbol{\mu}, \boldsymbol{\kappa}, \boldsymbol{\alpha}]$ [42]. $\boldsymbol{\alpha} \in \mathbb{R}^K$ are the mixture weights of the K components, $\mathbf{m} \in \mathbb{R}^K$ and $\mathbf{\Omega} \in \mathbb{R}^K$ are the shape and spread parameters of the K Nakagami distributions and $\boldsymbol{\kappa} \in \mathbb{R}^K$ and $\boldsymbol{\mu} \in \mathbb{R}^{K \times d}$ are the concentration and mean direction parameters of the K von Mises-Fisher distributions. Eq. (7.12) can be solved through a weighted expectation-maximization algorithm. The number of components in the mixture K can be either prescribed through prior knowledge of the reliability problem (e.g., knowledge of the number of disjunct failure regions) or - in moderate dimensions - identified through a clustering algorithm such as DBSCAN [21]. For details, see [42, 24].

7.3 Conditional reliability analysis

7.3.1 Problem Statement

The interest is in computing the failure probability conditional on a d_B -dimensional subset of the input random vector $\boldsymbol{\Theta}$. Let this subset be $\boldsymbol{\Theta}_B : \Omega \rightarrow \mathbb{D}_B \subseteq \mathbb{R}^{d_B}$ with joint CDF F_B . Further let $\boldsymbol{\Theta}_A : \Omega \rightarrow \mathbb{D}_A \subseteq \mathbb{R}^{d_A}$ with joint CDF F_A be the complement of $\boldsymbol{\Theta}_B$ over $\boldsymbol{\Theta}$ such that we may reorder the inputs and write $\boldsymbol{\Theta} = [\boldsymbol{\Theta}_A, \boldsymbol{\Theta}_B]^T$. The failure probability conditional on $\boldsymbol{\Theta}_B$ is defined by the integral

$$P_F(\boldsymbol{\Theta}_B) = \mathbb{P}(F|\boldsymbol{\Theta}_B) = \int_{\mathbb{D}_A} \mathbb{I}[g(\boldsymbol{\theta}_A; \boldsymbol{\Theta}_B) \leq 0] f(\boldsymbol{\theta}_A|\boldsymbol{\Theta}_B) d\boldsymbol{\theta}_A = \mathbb{E}_{f(\boldsymbol{\theta}_A|\boldsymbol{\Theta}_B)} [\mathbb{I}(g(\boldsymbol{\Theta}) \leq 0)|\boldsymbol{\Theta}_B]. \quad (7.14)$$

Using the isoprobabilistic transformation T from Subsection 7.2.1 we recast Eq. (7.14) in standard-normal space:

$$P_F(\mathbf{U}_B) = \mathbb{P}(F|\mathbf{U}_B) = \int_{\mathbb{R}^{d_A}} \mathbb{I}[G(\mathbf{u}_A; \mathbf{U}_B) \leq 0] \varphi_{d_A}(\mathbf{u}_A) d\mathbf{u}_A = \mathbb{E} [\mathbb{I}(G(\mathbf{U}) \leq 0)|\mathbf{U}_B]. \quad (7.15)$$

Note that \mathbf{U}_A and \mathbf{U}_B are independent and thus we have $f(\mathbf{u}_A|\mathbf{u}_B) = f(\mathbf{u}_A) = \varphi_{d_A}(\mathbf{u}_A)$.

As mentioned before, possible applications include quantiles $\underline{P}_F/\bar{P}_F$, surrogate models $\hat{P}_F(\boldsymbol{\Theta}_B)$ or a density estimate $\hat{f}(p_F)$ of $P_F(\boldsymbol{\Theta}_B)$. In practice, these quantities are computed based on n_B samples of $P_F(\boldsymbol{\Theta}_B)$, $\{p^j, j = 1, \dots, n_B\}$, where

$$p^j = P_F(\mathbf{u}_B^j) = \int_{\mathbb{R}^{d_A}} \mathbb{I}(G(\mathbf{u}_A, \mathbf{u}_B^j) \leq 0) \varphi_{d_A}(\mathbf{u}_A) d\mathbf{u}_A, \quad \mathbf{u}_B^j \stackrel{i.i.d.}{\sim} \varphi_{d_B}, \quad j = 1, \dots, n_B. \quad (7.16)$$

The computational cost associated with this setting can be considerably higher than that of a conventional reliability analysis as n_B d_A -dimensional reliability problems of form Eq. (7.14) have to be solved instead of a single one. Fig. 7.1 illustrates the estimation of conditional failure probabilities with iCE and the associated parametric importance densities constructed in the process. Therein,

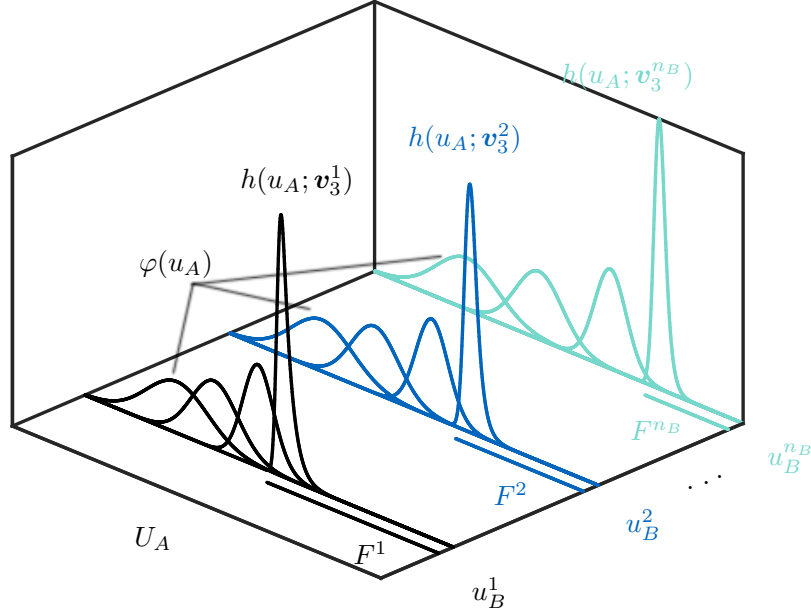


Figure 7.1: Illustration of iCE densities: $h(u_A; \mathbf{v}_i^j)$ is the importance density used at the i -th iCE step in the j -th conditional reliability problem in standard-normal space. Intermediate densities at equal iCE-steps i associated with different conditional reliability problems j tend to look identical in this illustration but are generally not. An exception is the nominal density $\varphi(u_A)$, which is identical for each problem.

\mathbf{v}_i^j is the parameter vector of the parametric importance density constructed in the i -th step of the j -th iCE run (i.e., the run solving the j -th reliability problem). The parameter vector \mathbf{v} now receives a superscript to identify the reliability problem to which it belongs while its subscript indicates the associated iCE step.

The main contribution of this paper is an algorithm that efficiently solves the sequence of conditional reliability problems given in Eq. (7.16). The connection of this problem sequence to a conceptually similar one arising in RBDO is discussed in the following subsection. Our framework is based on the iCE method for conventional reliability analysis. The basic idea is to reuse information from - or more precisely: densities constructed in - past problems to alleviate the computational cost in the current estimation. The two key tasks of such an algorithm are the identification of suitable biasing densities amongst solved reliability computations on the one hand, and the efficient integration of these densities in the estimation of the current conditional failure probability on the other hand. Identification and integration are addressed in Subsection 7.3.3 and Subsection 7.3.4, respectively.

7.3.2 Connection to RBDO

Reliability-based design optimization (RBDO) may be defined as the minimization of a deterministic cost function under constraints on the failure probability given the system design. To this end, the design is parametrized by means of a set of design variables. Then, a problem similar to Eq. (7.16) arises as the failure probability of the system has to be evaluated repeatedly and conditional on sev-

eral points in the design space. In such case, $\boldsymbol{\theta}_B^j$ would represent the design variables' values in the j -th iteration of the RBDO program. Due to this similarity, information reuse is also interesting for solving RBDO problems and has been put forward in this context in Ref. [10]. There, an influence hypersphere around each Θ_B -sample in \mathbb{D}_B is defined to identify suitable previously constructed importance densities. An important difference to conditional reliability estimation is the fact that values of Θ_B are not based on randomly sampling from $f(\boldsymbol{\theta}_B)$ but are inherently ordered as they are generated by an optimisation procedure. While this can significantly simplify the source identification task discussed in Subsection 7.3.3, it incurs the additional cost of computing gradients of the model with respect to Θ_B . ℓ_2 -distance-based information reuse in \mathbb{D}_B is a promising approach as long as the design space dimension remains moderate. [17] use a nearest neighbour search to identify such an ordering based on the ℓ_2 -distance amongst Θ_B -samples in \mathbb{D}_B . However, in our experience, such a heuristic for the proximity of reliability problems is not robust if either g is not sufficiently well-behaved (e.g., not sufficiently linear in $\boldsymbol{\theta}_B$) or the dimension of \mathbb{D}_B is large. In the latter case, the heuristic will suffer from the concentration of distance in high dimensions [6].

7.3.3 Source selection

We reuse information by identifying parametric importance densities constructed for previous conditional reliability problems in the sequence in Eq. (7.16) that are, in some sense, well-suited to estimate the current conditional failure probability. Thus, each parameter set $\{\mathbf{v}_i^j, i = 1, \dots, m_j, j = 1, \dots, n_B\}$ is stored in a candidate pool during the computation of the j -th problem, with m_j denoting the number of steps in the iCE method solving the j -th reliability problem. Within importance sampling, the fitness of an importance density for a given reliability problem can be characterized in terms of its proximity to the optimal importance density, e.g., in terms of an f -divergence measure. In brief, an f -divergence $D_f(p|q)$ measures dissimilarity between distributions with PDFs p and q as the expected value of the likelihood ratio p/q weighted with a function f : $D_f(p|q) = \mathbb{E}_q[f(p/q)]$. Different choices of f lead to different divergence measures such as the squared Hellinger distance ($f(s) = 2(1 - \sqrt{s})$) or the Kullback-Leibler-divergence ($f(s) = s \log s$).

Mode search according to Beaurepaire et al.

Estimating $D_f(p|q)$, with p the target density and q any density in the candidate pool, requires a considerable amount of LSF evaluations per candidate density for source identification only (in addition to the estimation cost). Instead, Ref. [5] proposes a heuristic to reduce the identification cost to a single LSF evaluation per candidate density in the context of RBDO. There, the failure probability is re-evaluated with sequential importance sampling conditional on various design parameter values provided by an optimization sequence. The heuristic is based on evaluating the current (ℓ -th) LSF $G(\mathbf{u}_A, \mathbf{u}_B^\ell)$ at the mode $\bar{\mathbf{u}}_i^j$ of each stored parametric density. The fittest importance density amongst all available candidates is identified as the one whose mode evaluation is closest to 0:

$$[I_\ell, J_\ell] = \arg \min_{\substack{i=1, \dots, m_j, \\ j=1, \dots, n_B}} |G(\bar{\mathbf{u}}_i^j, \mathbf{u}_B^\ell)|, \quad (7.17)$$

i.e., the importance density of the ℓ -th reliability problem is selected as the importance density constructed at the I_ℓ -th step of the J_ℓ -th conditional reliability problem.

Extension for mixtures and the CE-framework

In this work, we extend the idea presented in Ref. [5] by identifying multiple potentially suitable candidate densities and combining them into a mixture. Instead of identifying a single density amongst the members of the candidate pool, we evaluate the current LSF at the mode of each candidate density to identify a mixture density. The weight α_i^j of the candidate density constructed in the i -th iCE step of the j -th reliability problem is computed as the inverse absolute value of the LSF at the density mode whereby the weights are normalized so that they sum to unity. Mixture components whose weights fall below a threshold value (we choose the threshold as 0.01) are eliminated from the mixture to prevent dilution of the mixtures. The source identification procedure for mixtures

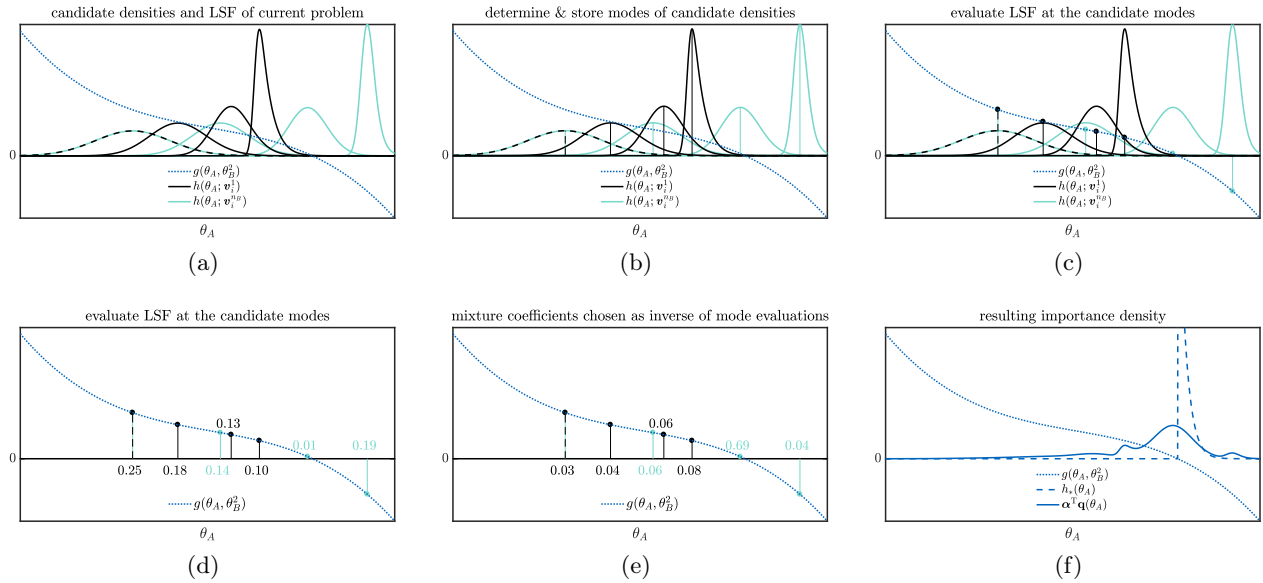


Figure 7.2: Step-wise illustration of the mode search algorithm (top left to bottom right). The perspective is obtained by a projection of Fig. 7.1 along the U_B -axis. (a) Candidate densities (importance densities from reliability problems 1 and 3) along with LSF of current reliability problem (problem 2) are depicted. (b) The modes of the candidate densities are identified. (c) + (d) The LSF of the current problem is evaluated at these modes. (e) The normalized reciprokes of the mode evaluations form the mixture weights. (f) The importance density mixture for the current problem is computed.

is detailed in Alg. 11, which returns the importance mixture $q_\alpha = \alpha^T \mathbf{q}$ for each reliability problem but the first. \mathbf{q} is the vector of all retained candidates in the current mixture and α is the vector of associated mixture weights. Fig. 7.2 illustrates the algorithm in one dimension ($d_A = 1$). The candidate density modes can be computed exactly without additional LSF evaluations due to their parametric nature.

Algorithm 11 Mode-based source identification

Input current conditional LSF $G(\mathbf{u}_A, \mathbf{u}_B^\ell)$, candidate modes $\bar{\mathbf{u}}_i^j$, candidate densities $h(\mathbf{u}_A; \mathbf{v}_i^j)$
Output number of mixture components M , mixture coefficients α , mixture densities \mathbf{q}

```

1: procedure SOURCE-ID( $G, \bar{\mathbf{u}}_i^j, h(\mathbf{u}_A; \mathbf{v}_i^j)$ )
2:   Compute mode evaluations  $\bar{g}_i^j = G(\bar{\mathbf{u}}_i^j, \mathbf{u}_B^\ell)$  ▷  $i \in 1, 2, \dots, m_j, j \in 1, 2, \dots, \ell - 1$ 
3:   Compute mixture coefficients  $a_i^j = \frac{1/\bar{g}_i^j}{\sum_{j=1}^{\ell-1} \sum_{i=1}^{m_j} 1/\bar{g}_i^j}$ 
4:   Initialize  $\alpha, \mathbf{q}$ 
5:   for all  $a_i^j$  do ▷ Gather tuples identifying important candidates
6:     if  $a_i^j > 0.01$  then
7:       Add:  $\alpha \leftarrow a_i^j, \mathbf{q} \leftarrow h(\mathbf{u}_A; \mathbf{v}_i^j)$ 
8:
9:   Renormalize mixture coefficients  $\alpha = \alpha / \|\alpha\|_1 \in [0.01, 1]^{M \times 1}$  ▷
    $M = \sum_{j=1}^{\ell-1} \sum_{i=1}^{m_j} \mathbb{I}(a_i^j > 0.01)$ 
10:  return  $M, \alpha, \mathbf{q}$ 

```

7.3.4 Information reuse for iCE: mixture-based and controlled importance sampling

Once the mixture q_α is identified, it can be used for importance sampling. The mixture-based importance sampling (M-IS) estimate of p^ℓ reads

$$\hat{p}_{\text{MIS}}^\ell = \sum_{k=1}^n \frac{\varphi_{d_A}(\mathbf{u}_A^k)}{q_\alpha(\mathbf{u}_A^k)} \mathbb{I}[G(\mathbf{u}_A^k, \mathbf{u}_B^\ell) \leq 0], \quad \mathbf{u}_A^k \stackrel{i.i.d.}{\sim} q_\alpha. \quad (7.18)$$

The M-IS estimate's CoV is found in the same way as that of the previously discussed standard IS estimate, i.e.,

$$\hat{\delta}_{\text{MIS}}^\ell = \frac{1}{\hat{p}_{\text{MIS}}^\ell} \sqrt{\frac{1}{n^2} \sum_{k=1}^n \left(\frac{\varphi_{d_A}(\mathbf{u}_A^k)}{q_\alpha(\mathbf{u}_A^k)} \mathbb{I}[G(\mathbf{u}_A^k, \mathbf{u}_B^\ell) \leq 0] - \hat{p}_{\text{MIS}}^\ell \right)^2}. \quad (7.19)$$

The M-IS is one of two IS estimates employed in this work. The second IS estimate results from the application of a variance reduction technique to the M-IS estimate and is known as controlled importance sampling (C-IS) [38]. The C-IS estimate can be viewed as a control variates analogue for probability densities. The classical control variates method is a variance reduction technique that may be applied to unbiased estimators $\hat{\mu}$ of $\mathbb{E}[f(\mathbf{X})]$, where \mathbf{X} is a random vector with CDF $F_{\mathbf{X}}$ [50]. The idea is to use a random vector of d_C control variates \mathbf{C} that is correlated with $\hat{\mu}$ and has known mean $\mathbf{c} = \mathbb{E}[\mathbf{C}]$ to construct a new unbiased estimate with lower variance compared to $\hat{\mu}$ as:

$$\hat{\mu}_\beta = \hat{\mu} + \beta^\top (\mathbf{C} - \mathbf{c}).$$

β is found by minimizing $\mathbb{V}[\hat{\mu}_\beta]$, which yields [50]

$$\beta_{\text{opt}} = \Sigma_C^{-1} \sigma_{\hat{\mu}\mathbf{C}},$$

where Σ_C is the covariance matrix of \mathbf{C} and $\sigma_{\hat{\mu}\mathbf{C}}$ is the d_C -dimensional vector of covariances between the components of \mathbf{C} and $\hat{\mu}$. The minimal variance reads

$$\mathbb{V}[\hat{\mu}_{\beta_{\text{opt}}}] = \mathbb{V}[\hat{\mu}] - \sigma_{\hat{\mu}\mathbf{C}}^\top \Sigma_C^{-1} \sigma_{\hat{\mu}\mathbf{C}},$$

where $\sigma_{\hat{\mu}C}^T \Sigma_C^{-1} \sigma_{\hat{\mu}C} \geq 0$, with the expression becoming 0 only if $\sigma_{\hat{\mu}C} = \mathbf{0}$, i.e., if μ and C are uncorrelated. That is, control variates are based on exploiting knowledge about a quantity that is correlated with the estimation target, where the larger the correlation, the larger the variance reduction.

Ref. [38] introduces a second mixture q_{β} of control densities with coefficients β into the M-IS estimate and add a correction term to preserve its unbiasedness (corresponding to $\beta^T \mathbf{c}$ above). Ref. [38] states that, ideally, for the M-IS sampler these densities are the ones that constitute the importance mixture q_{β} (namely, \mathbf{q}). The C-IS estimate reads:

$$\hat{p}_{\text{CIS}}^{\ell} = \sum_{k=1}^n \frac{\varphi_{d_A}(\mathbf{u}_A^k) \mathbb{I}[G(\mathbf{u}_A^k, \mathbf{u}_B^{\ell}) \leq 0] - q_{\beta}(\mathbf{u}_A^k)}{q_{\alpha}(\mathbf{u}_A^k)} + \|\beta\|_1, \quad (7.20)$$

where $q_{\beta} = \beta^T \mathbf{q}$. The second summand is the correction term that preserves unbiasedness. Since q_{β} is a density and thus integrates to 1, we have $\mathbb{E}_{q_{\alpha}}[\beta_i q_i / q_{\alpha}] = \beta_i \forall i \in 1, 2, \dots, M$. Optimal variance reduction can be achieved by minimizing the variance of the C-IS estimate jointly over the additional free coefficients β and its associated estimate $\hat{p}_{\text{CIS}}^{\ell}(\beta)$. The estimate's variance can be computed based on the sample from q_{α} as

$$(\hat{\sigma}_{\text{CIS}}^{\ell})^2 = \frac{1}{n} \sum_{k=1}^n \left(\frac{\varphi_{d_A}(\mathbf{u}_A^k) \mathbb{I}[G(\mathbf{u}_A^k, \mathbf{u}_B^{\ell}) \leq 0] - q_{\beta}(\mathbf{u}_A^k)}{q_{\alpha}(\mathbf{u}_A^k)} + \|\beta\|_1 - \hat{p}_{\text{CIS}}^{\ell} \right)^2. \quad (7.21)$$

Following Ref. [38], minimizing Eq. (7.21) can be cast as a multiple linear regression of the model $Y(\mathbf{u}) = \mathbf{c}^T \mathbf{Z}(\mathbf{u})$ with the extended coefficient vector $\mathbf{c} = [\hat{p}_{\text{CIS}}^{\ell}, \beta^T]^T$ and

$$\begin{aligned} Y(\mathbf{u}) &= \varphi_{d_A}(\mathbf{u}) \mathbb{I}[G(\mathbf{u}, \mathbf{u}_B^{\ell}) \leq 0] / q_{\alpha}(\mathbf{u}) \\ \mathbf{Z}(\mathbf{u}) &= [1, q_1(\mathbf{u})/q_{\alpha}(\mathbf{u}) - 1, q_2(\mathbf{u})/q_{\alpha}(\mathbf{u}) - 1, \dots, q_M(\mathbf{u})/q_{\alpha}(\mathbf{u}) - 1]^T. \end{aligned}$$

The multiple linear regression program reads

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c} \in \mathbb{R}^{1 \times M+1}} \frac{1}{n} \sum_{k=1}^n \left[Y(\mathbf{u}_A^k) - \mathbf{c}^T \mathbf{Z}(\mathbf{u}_A^k) \right]^2, \quad (7.22)$$

where $\hat{\beta}_{\text{opt}} = \hat{\mathbf{c}}_{2:M+1}$ and $\hat{p}_{\text{CIS}}^{\ell}(\hat{\beta}_{\text{opt}}) = \hat{\mathbf{c}}_1$. For simplicity, $\hat{p}_{\text{CIS}}^{\ell}$ shall always denote the minimum variance estimate $\hat{p}_{\text{CIS}}^{\ell}(\hat{\beta}_{\text{opt}})$ and its CoV is denoted by $\hat{\delta}_{\text{CIS}}^{\ell}$. The latter can be computed directly from the standard error of multiple linear regression as

$$\hat{\delta}_{\text{CIS}}^{\ell} = \frac{1}{\hat{p}_{\text{CIS}}^{\ell}} \sqrt{\frac{1}{n(n-M-1)} \sum_{k=1}^n \left[Y(\mathbf{u}_A^k) - \hat{\mathbf{c}}^T \mathbf{Z}(\mathbf{u}_A^k) \right]^2}. \quad (7.23)$$

Eqs. (7.19) and (7.23) provide the means to determine the accuracy of the two p^{ℓ} -estimates. In Section 7.4, we test the efficiency and accuracy of both the M-IS and C-IS estimates against a standard iCE run starting from the nominal distribution p . Based on Alg. 10, this baseline estimate will have CoV $\hat{\delta}_{\text{iCE}}$. Thus, the goal is to achieve $\hat{\delta}_{\text{MIS/CIS}}^{\ell} \leq \delta_{\text{iCE}}$ at lower computational cost compared to the total cost of the iCE baseline. In iCE, δ_{target} is prescribed for the CoV of the weights of the optimal IS density with respect to its current smooth approximation h_i . This is

equivalent to requiring that $\sqrt{N}\widehat{\delta}_{\text{iCE}} \lesssim \delta_{\text{target}}$. The inequality is exact if $h_i(\mathbf{u})$ and $h(\mathbf{u}; \mathbf{v}_i)$ are equal. Hence, it is reasonable to enforce

$$\widehat{\delta}_{\text{MIS/CIS}}^\ell \leq \frac{\delta_{\text{target}}}{\sqrt{N}}, \quad 2 \leq \ell \leq n_B. \quad (7.24)$$

A straightforward way to ensure criterion Eq. (7.24) with as few samples as possible is to incrementally add samples drawn from the importance mixture to the estimate until convergence. In practice, we start with Δn samples and iteratively increase the number of samples by batches of Δn , where we set $\Delta n = N/100$. The maximum number of samples is set to the number of samples per level in the iCE procedure N . Convergence is likely to be achieved within N samples, if the identified importance mixture q_α is a good approximation of the optimal importance density h^* , i.e., if $\text{CoV}[h^*/q_\alpha] \leq \sqrt{N}\delta_{\text{target}}$.

7.3.5 Preconditioning iCE

If convergence is not achieved within N samples, it is still likely that $\text{CoV}[h^*/q_\alpha] \leq \text{CoV}[h^*/\varphi_{d_A}]$, such that replacing the nominal density in Alg. 10 with q_α leads to a reduced number of steps in the iCE sequence, m . To precondition iCE in this way, one may use the N samples drawn from q_α and evaluated for the M-IS/C-IS-estimate, effectively entering crefalg:iCE in line 8 with \mathbf{v}_0 corresponding to the parameter set of q_α . In Alg. 12, this preconditioned version of iCE is called as preconditioned-iCE($G(\mathbf{u}_A, \mathbf{u}_B^j), \delta_{\text{target}}, N, d_A, q_\alpha, \mathbf{U}, \mathbf{G}$), where the three additional arguments represent a set of samples \mathbf{U} and the corresponding LSF evaluations \mathbf{G} , drawn from q_α . If the preconditioned iCE step is performed, it is likely because the history of already computed conditional reliability problems does not contain problems that are sufficiently informative with respect to the current one. Therefore, once the current problem is solved with preconditioned iCE, the resulting importance density $h(\mathbf{u}_A; \mathbf{v}_{m_\ell}^\ell)$ is added to the pool of candidate densities. In this way, the pool grows adaptively and with each solved conditional reliability problem, it becomes more informative for the problems left in the sequence. The final procedure is outlined in Subsection 7.3.6 by Alg. 12.

7.3.6 The computational procedure

In Alg. 12, we summarise the overall procedure to estimate a sequence of conditional reliability problems using Alg. 10 & Alg. 11 as well as the samplers described in Eq. (7.18) or Eq. (7.20).

7.4 Numerical experiments

7.4.1 Parameter Study: sequential processing chain

In the first example, we consider a sequence of processing steps. Each step is modelled as a Poisson process and arrival of the first jump indicates finalisation of a step and triggers the subsequent step. The arrival time of the first jump in the i -th step is distributed exponentially with rate parameter

Algorithm 12 Importance sampling with information reuse

Input LSF $G(\mathbf{u})$, input space dimension d , target CoV δ_{target} , sample increment Δn , samples per level N , a set of B -samples $\mathbf{U}_B = \{\mathbf{u}_B^j, j = 1, \dots, n_B\} \in \mathbb{R}^{d_B \times n_B}$

Output conditional failure estimates $\hat{\mathbf{p}} \in [0, 1]^{n_B \times 1}$, estimate CoVs $\hat{\boldsymbol{\delta}} \in \mathbb{R}^{n_B \times 1}$, no. of levels $\mathbf{m} \in \mathbb{N}^{n_B \times 1}$

```

1: procedure IS-IR( $g, \delta_{\text{target}}, N, d, \mathbf{U}_B$ )
2:    $j \leftarrow 1$ 
3:    $g_j(\mathbf{u}_A) = g(\mathbf{u}_A, \mathbf{u}_B^j)$ 
4:    $[\hat{\mathbf{p}}^j, \hat{\boldsymbol{\delta}}^j, \mathbf{m}^j, \{h(\mathbf{u}_A, \hat{\mathbf{v}}_i), i = 1, \dots, m^j\}] = \text{iCE}(g_j, \delta_{\text{target}}, N, d_A)$  ▷ Alg. 10
5:   Add  $\mathbf{q} \leftarrow \{h(\mathbf{u}_A, \hat{\mathbf{v}}_i), i = 1, \dots, m^j\}$ 
6:   for  $j = 2, \dots, n_B$  do
7:      $[M, \boldsymbol{\alpha}, \mathbf{q}] = \text{Source-ID}(g, \bar{\mathbf{u}}_i^j, h(\mathbf{u}_A; \mathbf{v}_i^j))$  ▷ Alg. 11
8:      $q_\alpha = \boldsymbol{\alpha}^T \mathbf{q}$ 
9:      $n = \Delta n$ 
10:    Sample  $\mathbf{U}_A = \{\mathbf{u}_A^k, k = 1, \dots, n\} \in \mathbb{R}^{n \times d_A}$  ▷  $\mathbf{u}_A^k \stackrel{i.i.d.}{\sim} q_\alpha$ 
11:    Compute  $\mathbf{G} = G(\mathbf{U}_A, \mathbf{u}_B^j) \in \mathbb{R}^{n \times 1}$ 
12:    while  $n \leq N$  do
13:      With  $(\mathbf{U}_A, \mathbf{G})$ , compute  $\hat{p}_{\text{MIS}}^j / \hat{p}_{\text{CIS}}^j$  based on Eq. (7.18)/Eq. (7.20)
14:      With  $(\mathbf{U}_A, \mathbf{G})$ , compute  $\hat{\delta}_{\text{MIS}}^j / \hat{\delta}_{\text{CIS}}^j$  based on Eq. (7.19)/Eq. (7.23)
15:      if  $\hat{\delta}_{\text{MIS/CIS}}^j \leq \delta_{\text{target}} / \sqrt{N}$  then
16:        break
17:      else if  $n \leq N$  then
18:        Sample  $\mathbf{U}_{A,\text{new}} = \{\mathbf{u}_{A,\text{new}}^k, k = 1, \dots, \Delta n\} \in \mathbb{R}^{\Delta n \times d_A}$  ▷  $\mathbf{u}_{A,\text{new}}^k \stackrel{i.i.d.}{\sim} q_\alpha$ 
19:        Compute  $\mathbf{G}_{\text{new}} = g(\mathbf{U}_{A,\text{new}}, \mathbf{u}_B^j) \in \mathbb{R}^{\Delta n \times 1}$ 
20:        Add  $\mathbf{U}_A \leftarrow \mathbf{U}_{A,\text{new}}$ 
21:        Add  $\mathbf{G} \leftarrow \mathbf{G}_{\text{new}}$ 
22:      else if  $n = N$  and  $\hat{\delta}_{\text{MIS/CIS}}^j > \delta_{\text{target}} / \sqrt{N}$  then
23:         $g_j(\mathbf{u}_A) = g(\mathbf{u}_A, \mathbf{u}_B^j)$ 
24:         $[\hat{\mathbf{p}}^j, \hat{\boldsymbol{\delta}}^j, \mathbf{m}^j, h(\mathbf{u}_A, \hat{\mathbf{v}}_{m^j}^j)] = \text{preconditioned-iCE}(g_j, \delta_{\text{target}}, N, d_A, q_\alpha, \mathbf{U}_A, \mathbf{G})$ 
25:        Add  $\mathbf{q} \leftarrow h(\mathbf{u}_A, \hat{\mathbf{v}}_{m^j}^j)$ 
26:  return  $\hat{\mathbf{p}}, \hat{\boldsymbol{\delta}}, \mathbf{m}$ 

```

λ_i , which is an uncertain parameter. The goal is to estimate the probability of the total processing time exceeding a threshold. The LSF thus reads

$$g(\mathbf{t}, \boldsymbol{\lambda}) = T - \underbrace{\sum_{i=1}^{d_A} t_i}_{\sim \text{Hypoexp}(\boldsymbol{\lambda})}, \quad t_i \sim \text{Exp}(\lambda_i), \quad (7.25)$$

where Exp denotes the exponential distribution and Hypoexp denotes the hypoexponential distribution. The sum of d_A independent exponential random variables with rate parameters $\boldsymbol{\lambda} \in \mathbb{R}_+^{d_A \times 1}$ is hypoexponentially distributed [7]. The exponential and hypoexponential CDF read

$$F_{\text{Exp}}(x; \lambda) = 1 - e^{-\lambda x} \quad \forall x \geq 0,$$

$$F_{\text{Hyp}}(x; \boldsymbol{\lambda}) = 1 - \sum_{i=1}^{d_A} e^{-\lambda_i x} \prod_{j=1, j \neq i}^{d_A} \frac{\lambda_j}{\lambda_j - \lambda_i} \quad \forall x \geq 0.$$

With known rate parameter vector $\boldsymbol{\lambda}$, the failure probability is given as

$$\mathbb{P}(\text{F}|\boldsymbol{\lambda}) = \mathbb{P}(g(\mathbf{t}) \leq 0) = 1 - F_{\text{Hyp}}(T; \boldsymbol{\lambda}). \quad (7.26)$$

We let $\boldsymbol{\theta}_A = \mathbf{t} \in \mathbb{R}^{d_A \times 1}$, $\boldsymbol{\Theta}_B \sim \mathcal{U}_{d_B}(0, 1) \in \mathbb{R}^{d_B \times 1}$, where $d_B = n \cdot d_A$, $n \in \mathbb{N}$, and

$$\frac{1}{\lambda_i} = \frac{1}{n} \sum_{\substack{j=n \cdot i \\ -n+1}}^{n \cdot i} \theta_{B,j}, \quad i = 1, 2, \dots, d_A. \quad (7.27)$$

That is, each inverse rate parameter (scale parameter) is computed as the average of n different reducible variables. With these definitions and transforming the LSF to standard-normal space, Eq. (7.25) reads

$$G(\mathbf{u}_A, \mathbf{u}_B) = T - \sum_{i=1}^{d_A} F_{\text{Exp}}^{-1}(\Phi(u_{A,i}); \lambda_i), \quad \frac{1}{\lambda_i} = \frac{1}{n} \sum_{\substack{j=n \cdot i \\ -n+1}}^{n \cdot i} \Phi(u_{B,j}). \quad (7.28)$$

We define $\mathbf{A} \in \{0, 1\}^{d_A \times d_B}$ such that $\boldsymbol{\lambda}^{-1} = \mathbf{A}\boldsymbol{\Theta}_B/n$. \mathbf{A} has a banded structure with elements of the main diagonal and the first $n-1$ diagonals above the main diagonal being 1 and the rest 0. An analytical expression for $P_{\text{F}}(\mathbf{U}_B)$ follows from Eq. (7.26):

$$P_{\text{F}}(\mathbf{U}_B) = 1 - F_{\text{Hyp}}(T; (\mathbf{A}\Phi(\mathbf{U}_B)/n)^{-1}). \quad (7.29)$$

By inverting Eq. (7.29) one may compute T_p corresponding to the first-order approximation of a fixed unconditional failure level p as

$$T_p = F_{\text{Hyp}}^{-1}(1 - p; (\mathbf{A}\Phi(\mathbb{E}[\mathbf{U}_B])/n)^{-1}) = F_{\text{Hyp}}^{-1}(1 - p; \mathbf{2}), \quad (7.30)$$

where $\mathbf{2}$ denotes a d_A -dimensional vector of twos. In the following application, if not stated otherwise, $d_B = 100$ and $n = 5$ such that $d_A = 20$. The unconditional failure probability is chosen as $\mathbb{P}(\text{F}) = 10^{-5}$ and the number of $\boldsymbol{\Theta}_B$ -samples as $n_B = 100$. We use the relative root-mean-squared error (RMSE) to measure the accuracy of a failure probability estimate \hat{p} conditional on $\boldsymbol{\theta}_B$, which reads

$$e(\hat{p}|\boldsymbol{\theta}_B) = \frac{1}{P_{\text{F}}(\boldsymbol{\theta}_B)} \sqrt{\mathbb{E} \left[(\hat{p}(\boldsymbol{\theta}_B) - P_{\text{F}}(\boldsymbol{\theta}_B))^2 \right]}, \quad (7.31)$$

and plot the mean of $e(\hat{p}|\theta_B)$ taken over the n_B Θ_B -samples in Fig. 7.3. The expectation in Eq. (7.31) is computed with 100 repetitions of the analysis using the same Θ_B sample set. The proposed algorithms (M-IS & C-IS) are benchmarked against an iCE reference solution with equal number of samples per level $N = 1000$ and target CoV $\delta_{\text{target}} = 1.5$ for iCE, M-IS and C-IS. This corresponds to a target CoV of the conditional failure probability estimate of approximately 0.05. A series of parameter studies is devised to illustrate the behaviour of our algorithm depending on the failure magnitude of the problem ($\mathbb{P}(F)$ - through choosing different T_p thresholds according to Eq. (7.30)), its dimensionality in A -space (d_A) and B -space (d_B) and the number of conditional reliability problems present in the sequence (n_B). The estimator accuracy (mean and variance) of

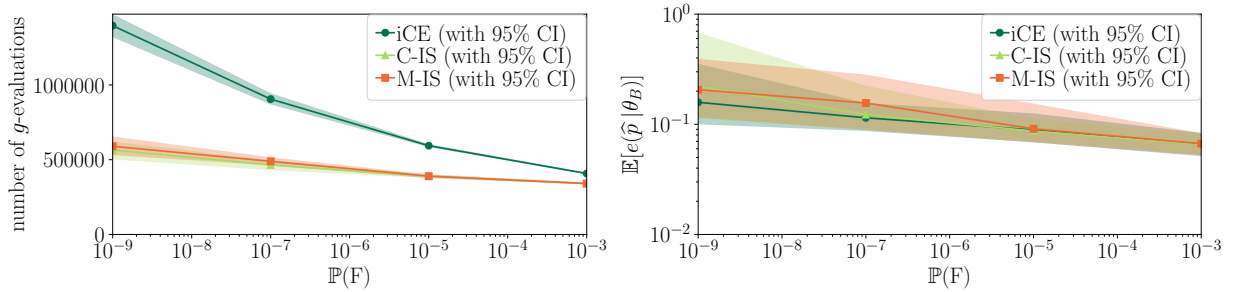


Figure 7.3: Failure probability estimation: computational cost (left) and RMSE mean including confidence intervals (right) with varying unconditional failure probability levels.

both M-IS and C-IS-based information reuse samplers matches that of the reference solution when varying $\mathbb{P}(F)$ over a wide range (10^{-3} and 10^{-9} , see Fig. 7.3). The computational savings of our information reuse samplers increase as $\mathbb{P}(F)$ decreases and reach $> 60\%$ saved effort at $\mathbb{P}(F) = 10^{-9}$. This is somewhat intuitive as the number of required steps in the iCE sequence grows with decreasing failure probability magnitude such that the savings potential rises. C-IS does not notably outperform M-IS, i.e., the control variate effect is negligible at all investigated $\mathbb{P}(F)$ -values. Fig. 7.4 shows

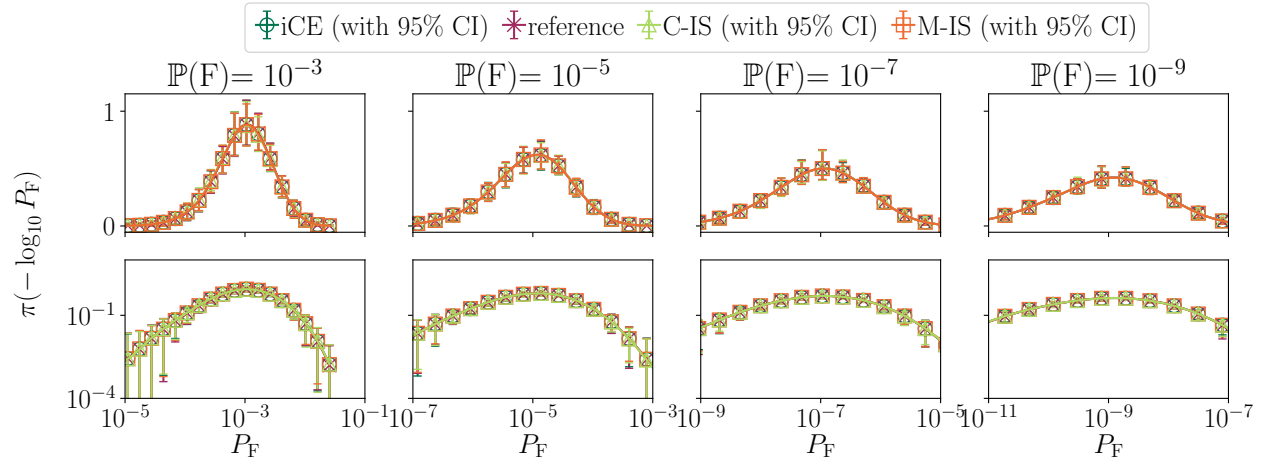


Figure 7.4: Conditional failure probability densities in linear (upper) and logarithmic (lower) scale with varying unconditional failure probability levels.

the density estimates corresponding to the $\mathbb{P}(F)$ -study. Both mean and confidence intervals (CI) of the density estimates produced with M-IS/C-IS match exactly with the reference solution.,The

confidence intervals here are again based on 100 repeated runs of iCE, M-IS and C-IS while drawing a *different* sample set $\{\theta_B^j, j = 1, \dots, n_B\}$ in each repetition (as opposed to the plots in Fig. 7.3, which are based on 100 repetitions given an identical sample set). That is, the CIs in Fig. 7.4 represent the aggregation of statistical uncertainty stemming from the failure estimation (iCE/M-IS/C-IS) *and* the density estimation based on n_B samples. This procedure (100 repetitions at identical type B-sample

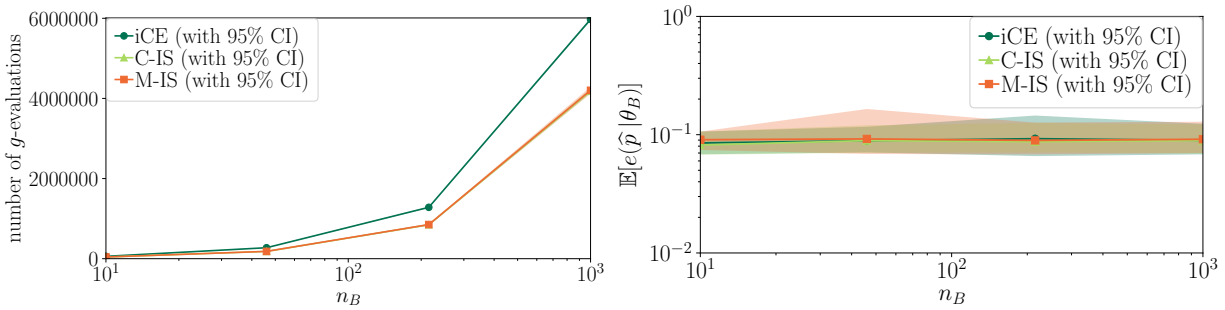


Figure 7.5: Computational cost (left) and mean relative RMSE of conditional failure probability estimates including confidence intervals (right) with varying number of type B-variable samples.

for cost/error analysis and 100 random draws of the type B -sample for computing P_F density CIs) is used for any following parameter study as well. The second parameter study considers the number of type B-samples n_B . There, the relative savings potential at $\mathbb{P}(F) = 10^{-5}$ amounts to 30% (this is also evident from the first parameter study in Fig. 7.3) at all investigated sample sizes n_B , while the absolute savings scale proportionally with the total computational effort or n_B (Fig. 7.5, left). Although the relative RMSE of all failure probability estimators is independent of n_B (Fig. 7.5), P_F density CIs are shrinking around the mean density estimates as n_B increases (Fig. 7.6) due to the aggregation of failure probability and density estimation uncertainty in these computations (the uncertainty in the density estimation). As n_B rises further, we expect an increase in relative savings due to an increasingly dense population of the type B-variable space with samples. This, in turn, will produce more correlation amongst the LSFs, which facilitates more efficient information reuse. n_B does not affect the relative RMSE of the failure probability estimates (Fig. 7.5, right). Finally,

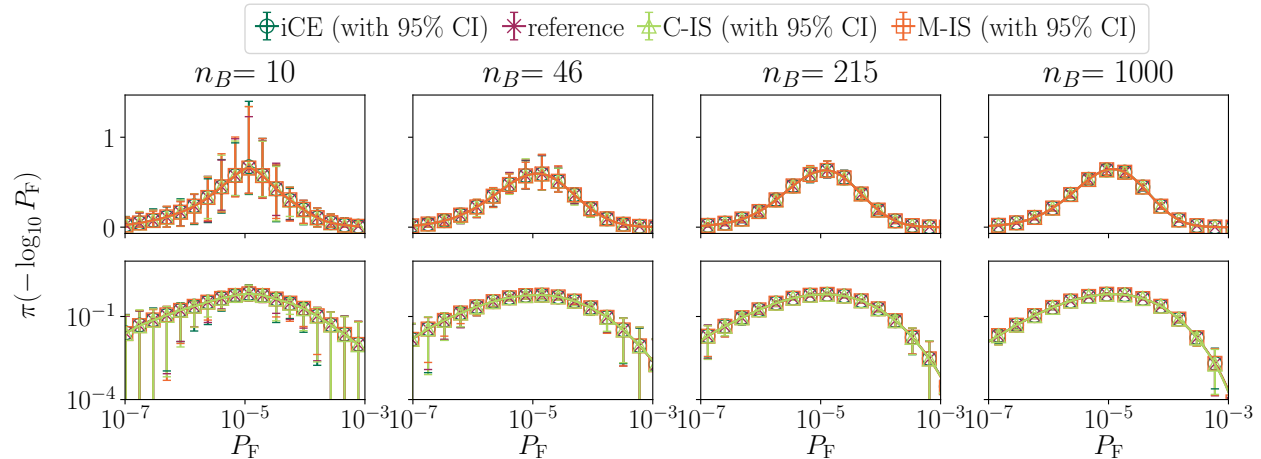


Figure 7.6: Conditional failure probability densities in linear (upper) and logarithmic (lower) scale with varying conditional sample size n_B .

we study the influence of the subspace dimensions d_A and d_B on the performance of our method. Fig. 7.7 shows the progression of computational cost and RMSE as d_B increases and $d_A = 10$, $\mathbb{P}(F) = 10^{-5}$ and $n_B = 100$ are fixed. The corresponding conditional failure probability densities - plotted in Fig. 7.8 - reveal a decreasing variability in P_F as d_B rises. This is to be expected as every rate parameter (type A-variable) in λ is averaged over an increasing number of type B variables. That is, the rate parameter variance - and thus also the variance of P_F - scale inversely with d_B . As d_B increases from 10 to 1000, the computational savings increase from 45% to $> 70\%$ while the error (RMSE) decreases slightly due to the decreasing variability of P_F . The same effect is observed when increasing d_A while keeping $d_B = 100$ (and again $d_B = 100$, $\mathbb{P}(F) = 10^{-5}$ and $n_B = 100$) constant: the variance of P_F increases with rising d_A (Fig. 7.10), which in turn causes a slight increase of the relative RMSE (Fig. 7.9). Fig. 7.9 also reveals an intermediate d_A regime, in which the computational savings are relatively low ($\approx 20\%$) due to a plateau in the computational cost of the reference solution between $d_A = 10$ and $d_A = 25$. In conclusion, this study demonstrates

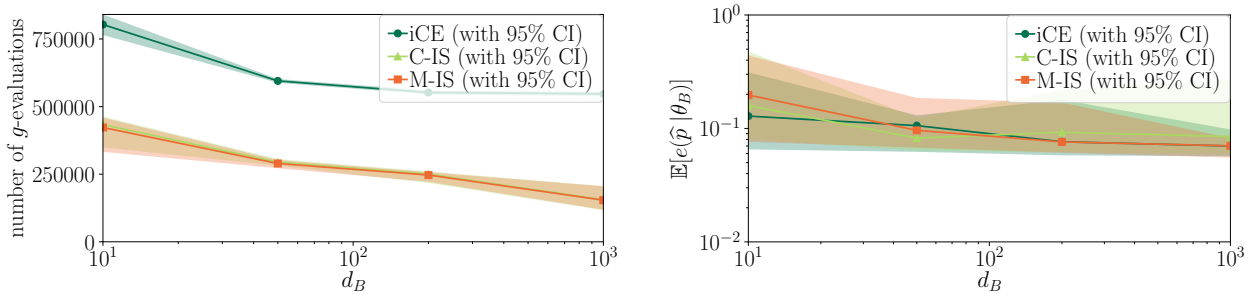


Figure 7.7: Failure probability estimation: computational cost (left) and RMSE mean including confidence intervals (right) with varying problem dimension d_B .

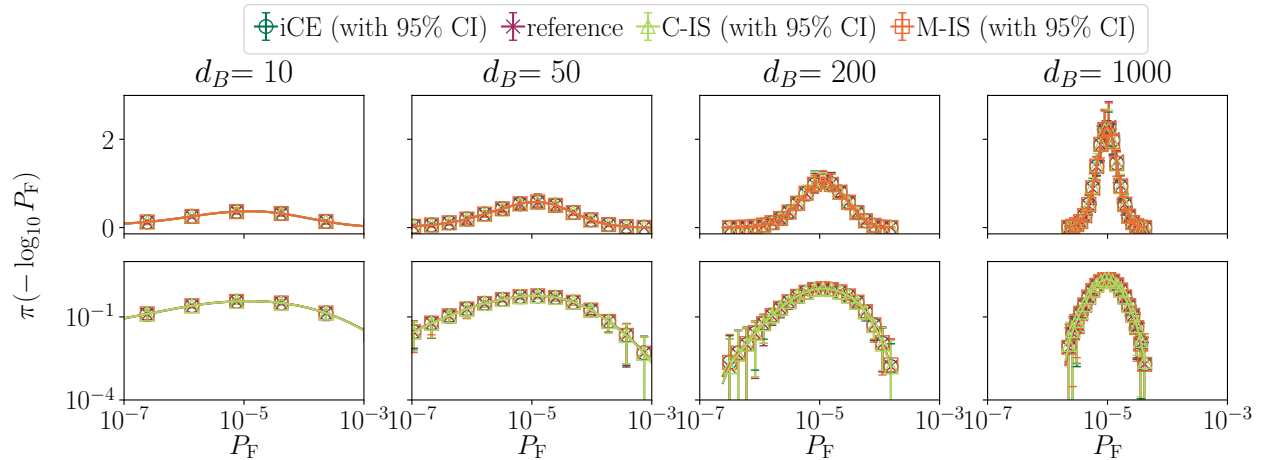


Figure 7.8: Conditional failure probability densities in linear (upper) and logarithmic (lower) scale with varying dimension d_B .

the robustness of the information reuse-based M-IS/C-IS estimators with respect to the problem dimensions in both type A & B - variables while facilitating computational savings of $\approx 25\% - 75\%$.

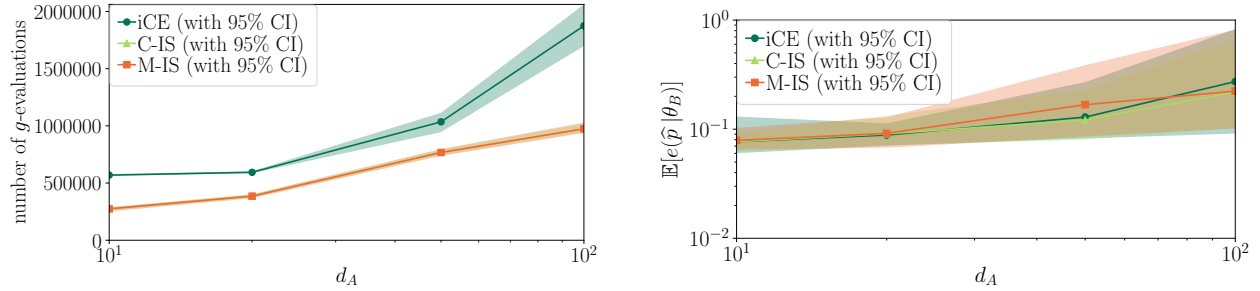


Figure 7.9: Failure probability estimation: computational cost (left) and RMSE mean including confidence intervals (right) with varying problem dimension d_A .

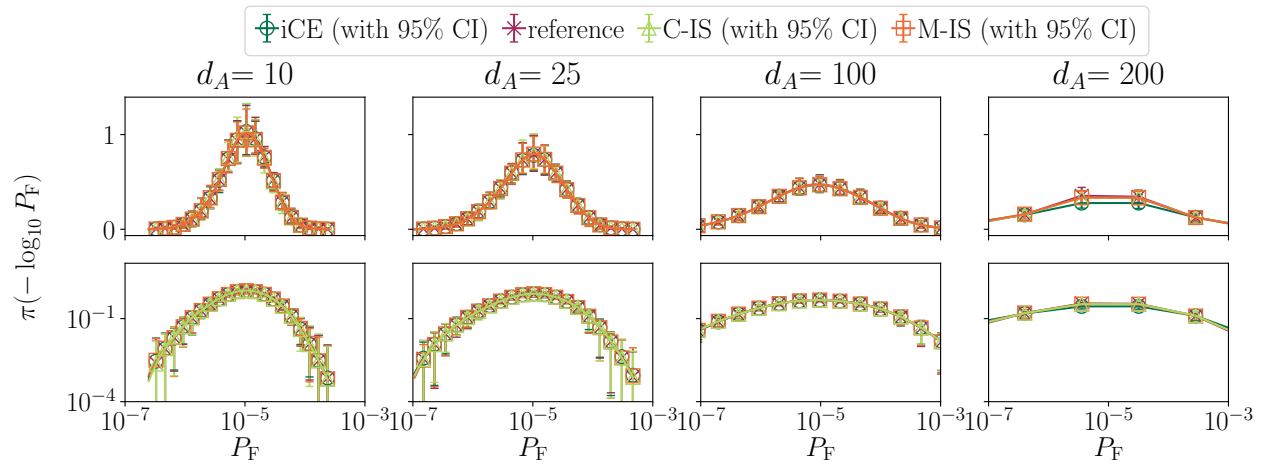


Figure 7.10: Conditional failure probability densities in linear (upper) and logarithmic (lower) scale with varying dimension d_A .

7.4.2 Case Study: monopile foundation in plastically behaving soil

Problem Setup

This case study is based on a finite element model for the interaction of a monopile foundation of an offshore wind turbine (Fig. 7.11) with stiff, plastic soil. Deterministic parameters of the monopile are its depth $L = 30$ m, diameter $D = 6$ m, wall thickness $t = 0.07$ m, Poisson ratio $\nu = 0.3$ and Young's modulus $E = 2.1 \cdot 10^5$ MPa. The uncertain inputs comprise the lateral load H as well as the undrained shear strength s of the soil and hyperparameters of both quantities. The engineering model consists of a nonlinear finite element code whose setup is described in detail in Ref. [11] and the probabilistic model considered there has been modified following Ref. [28]: s is considered both uncertain and increasing in mean with soil depth z . It is thus modelled by a random field with linear mean drift along the soil depth coordinate z . Given an underlying homogeneous Gaussian random field $\tilde{s}(z, \Theta)$

$$\{\tilde{s}(z) : 0 \leq z \leq L\} \sim \mathcal{N}(0, \sigma_{\tilde{s}}),$$

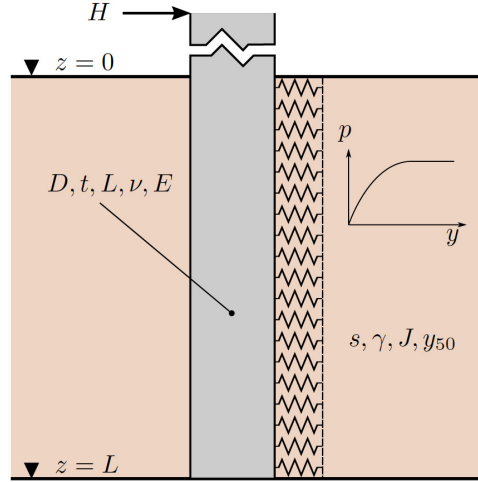


Figure 7.11: Wind turbine monopile foundation [11].

the non-homogeneous random field representing the shear strength of the soil can be expressed as

$$\begin{aligned} s(z, \Theta) &= s_0 + s_1 \sigma'(z) \exp \{ \tilde{s}(z, \Theta) \} \\ &= s_0 + s_1 \gamma z \exp \{ \tilde{s}(z, \Theta) \}, \end{aligned}$$

where γ is the soil unit weight, $\sigma'(z) = \gamma z$ is the effective vertical stress, s_0 is the undrained shear strength at ground level and s_1 is the drift parameter governing the mean increase of s with increasing soil depth. $\tilde{s}(z, \Theta)$ models the intra-site variability. That is, at a given site with known deterministic s_0 and s_1 , it describes the inherent variability of the undrained shear strength. In order to describe the inter-site variability, the parameters s_0 and s_1 are modelled probabilistically as well. The homogeneous RF \tilde{s} is equipped with an exponential correlation function:

$$\rho_{\tilde{s}\tilde{s}}(z', z'') = \exp \left\{ -\frac{2|z' - z''|}{\theta_{\tilde{s}}} \right\},$$

with vertical soil scale of fluctuation $\theta_{\tilde{s}} = 1.9m$ [45] and standard deviation $\sigma_{\tilde{s}} = 0.3$ [45, 46]. We assume the soil to be stiff and plastic according to the classification provided in Ref. [47]. There, the specific soil weight range is given with $17 - 19kN/m^3$, thus we set $\gamma = 18kN/m^3$. The mean cohesion range is given with $20 - 50kN/m^2$ by Ref. [47] and Ref. [9] lists the mean range of the undrained shear strength ratio s_u/σ' as $0.23 - 1.4$. We fit log-normal distributions for s_0 and s_1 by setting the 10 % and 90% quantiles of the distributions equal to the lower and upper bounds of these ranges. The resulting parameters are detailed in Tab. 7.1 along with uncertain parameters for the load H , namely μ_H and σ_H . The mean and CoV of the load Gumbel distribution in Tab. 7.1 are conditional on the parameters a_H (location parameter) and b_H (scale parameter):

$$\begin{aligned} \mu_{H|a_H, b_H} &= \mu_{a_H} + \gamma_E \mu_{b_H} \\ \delta_{H|a_H, b_H} &= \frac{\pi}{\sqrt{6}} \frac{\mu_{b_H}}{\mu_{H|a_H, b_H}}, \end{aligned}$$

where γ_E is the Euler-Mascheroni constant. \tilde{s} is simulated by means of the midpoint method. That is, the spatial domain $[0, L]$ is discretized with n spatial elements and \tilde{s} is represented by means of n random variables with joint distribution $\mathcal{N}(\mathbf{0}, \Sigma_{\tilde{s}\tilde{s}})$. The random variables represent the random field values at the element midpoints. Thus, the covariance matrix $\Sigma_{\tilde{s}\tilde{s}}$ is computed by evaluating

Table 7.1: Input variable definitions of the monopile foundation.

Input	Distribution	Mean μ	CoV δ
ξ [-]	Standard-Normal	$\mathbf{0}$	<i>n.d.</i> ($\Sigma_{\xi\xi} = \mathbf{I}_{n \times n}$)
s_0 [kPa]	Log-Normal	33.7094	0.3692
s_1 [kPa]	Log-Normal	0.7274	0.8019
H [kN]	Gumbel	$\mu_{P a_H, b_H}$	$\delta_{P a_H, b_H}$
a_H [kN]	Log-Normal	2274.97	0.2
b_H [kN]	Log-Normal	225.02	0.2

$\sigma_{\xi}^2 \rho_{\xi\xi}(z', z'')$ at the element midpoints. The number of elements is chosen such that 95% of the inherent RF variability is captured by the RF discretization, leading to $n = 82$ in this example. Therefore, the total input dimension is $d = 87$. As the sampling approaches are implemented in the standard normal space, the midpoint random variables are transformed to independent standard normal random variables, denoted as ξ , by means of the Cholesky decomposition of $\Sigma_{\xi\xi}$. The model output $Y = \mathcal{Y}(\theta)$ is the maximum occurring stress in the foundation. The LSF is given by

$$G(\mathbf{u}) = \sigma_{crit} - \mathcal{Y}(T^{-1}(\mathbf{u})),$$

with T the transformation from the original input probability space \mathbb{D}_A to standard-normal space and $\sigma_{crit} = 100$ MPa the stress threshold, which corresponds to a system failure probability of $\mathbb{P}(F) = 3.6 \cdot 10^{-4}$ (estimated with MC and $\hat{\delta}_{MC} = 0.1187$). Depending on the availability of measurements and inter-site data, the assignment of inputs to either Θ_A and Θ_B may vary. We illustrate two cases, the first of which features a high-dimensional reducible space \mathbb{D}_B as we set $\Theta_A = [a_H, b_H, H]$ and $\Theta_B = [s_0, s_1, \xi]$ while the second has high-dimensional \mathbb{D}_A , where $\Theta_A = [\xi, H]$ and $\Theta_B = [s_0, s_1, a_H, b_H]$.

Efficiency & Accuracy

Due to the large computational cost of a single evaluation of the monopile foundation model, we compare single runs of M-IS/C-IS over a given type B-sample to the iCE reference, which is repeated 26 times over that same sample to estimate a confidence interval. For both type B-variable configurations, we set the number of samples per iCE level to $N = 500$ and choose $\delta_{target} = 1.5$ and $n_B = 100$. The monopile example is well-suited to demonstrate the dependence of information reuse-based savings potential on the partitioning of inputs in type A and type B. Figs. 7.12 and 7.13 (both: right) show that the conditional failure probabilities are computed accurately with the information reuse estimators for both type B-variable configurations. However, while for the configuration $\Theta_B = [s_0, s_1, \xi]$, these results are obtained with $\approx 25\%$ of the reference computational cost, the second configuration $\Theta_B = [s_0, s_1, a_H, b_H]$ admits almost no savings (Figs. 7.12 and 7.13, left). This is mostly due to the random field hyperparameters a_H and b_H that are present in the second configuration and cause a large fraction of P_F -variability: the conditional failure probabilities span roughly 6 orders of magnitude in the second configuration (10^{-8} - 10^{-2} , Fig. 7.13, right) versus only 2 orders of magnitude in the first (10^{-4} - 10^{-2} , Figure Fig. 7.12, right). This high variability

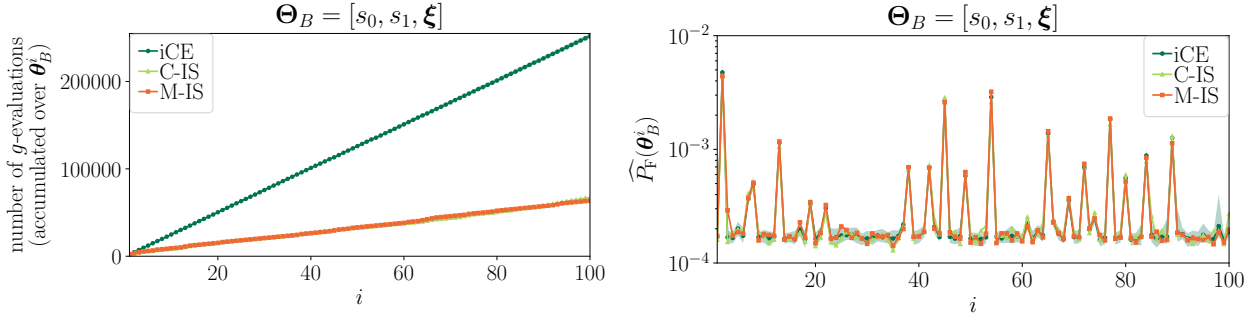


Figure 7.12: Monopile configuration 1: computational cost (left) and conditional failure probability estimates at Θ_B samples (right) with iCE reference solution mean and 95% CI.

delays the construction of a pool of candidate densities that are relevant for information reuse. As the number of Θ_B -samples increases we expect the computational savings to increase.

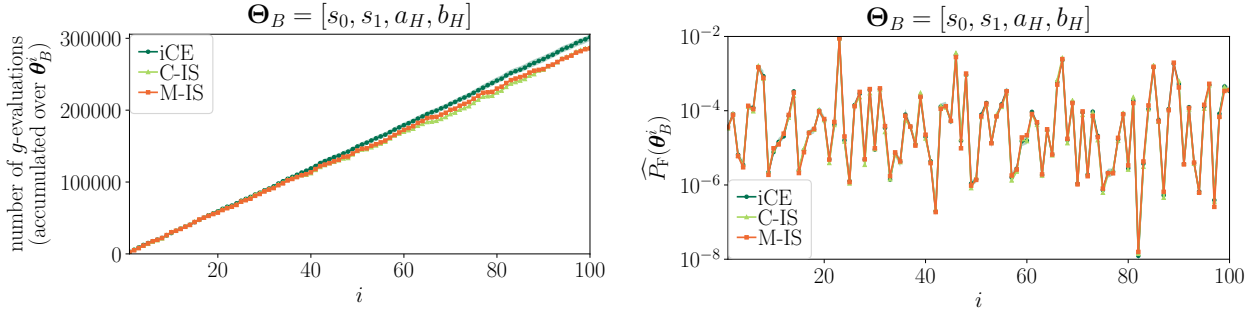


Figure 7.13: Monopile configuration 2: computational cost (left) and conditional failure probability estimates at Θ_B samples (right) with iCE reference solution mean and 95% CI.

Conditional densities and global sensitivity analysis

Based on $n_B = 500$ (first configuration) and $n_B = 100$ (second configuration) type B-variable samples, we compute confidence intervals on P_F (Tab. 7.2) and a kernel density estimate of its distribution (Figs. 7.14 and 7.15, left). These are intervals of the random variable P_F itself. That means they are not measures of estimation accuracies but rather a property of the conditional random variable. Estimation accuracy is measured by the confidence intervals of the estimator $\widehat{P}_F(\theta_B^i)$ as depicted in Figs. 7.12 and 7.13. The number of samples per iCE level is set to $N = 1000$. The intervals again demonstrate the vast increase of P_F -variability in between the first to the second type B-variable configuration. Moreover, we compute reliability-oriented variance-based sensitivity indices according to Ref. [16]: the n_B conditional failure probability samples are used to construct a surrogate (a partial least squares-based polynomial chaos expansion [41]) of $\log P_F(\Theta_B)$, $\log \widehat{P}_F(\Theta_B)$. Sobol' and total-effect indices of $\log \widehat{P}_F$ with respect to Θ_B can be computed by post-processing the surrogate model coefficients [18]. The sensitivity indices indicate that the random field drift gradient s_1 contributes by far the largest variability to P_F in the first configuration while in the second, the

load hyperparameters a_H and b_H dominate the random field parameters (Figs. 7.14 and 7.15, right). We observe, that Sobol' and total-effect indices are much closer to one another under the second configuration indicating an absence of interaction effects. Under the first configuration, such higher-order effects play a more prominent role, which is due to the fact that in this configuration all type B-variables are random field-related. In other words, any pair of parameters belonging to the same probabilistic model is likely to exhibit stronger dependencies than a pair belonging to two different probabilistic models (here: to the random field model and the random load model). In the latter case, the pair of parameters can only interact through the FE-model, where however first-order effects seem to be dominant. Asymptotically, the mean estimates of the conditional failure probability provided in Tab. 7.2 will coincide with the unconditional failure probability of the system.

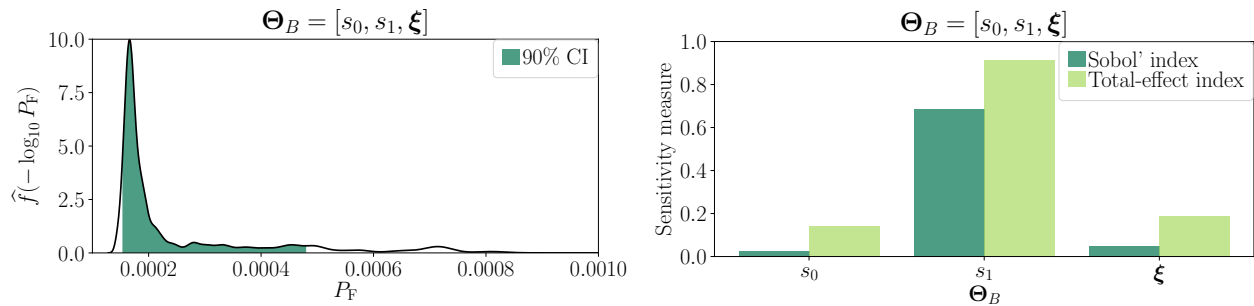


Figure 7.14: Monopile configuration 1: conditional failure PDF with confidence intervals (left) and variance-based failure sensitivity indices (right) with $n_B = 500$.

Table 7.2: P_F -confidence intervals for two type B-variable configurations.

	$\Theta_B = [s_0, s_1, \xi]$		$\Theta_B = [s_0, s_1, a_H, b_H]$	
Mean	$2.537 \cdot 10^{-4}$		$2.828 \cdot 10^{-4}$	
interval width	lower CI	upper CI	lower CI	upper CI
90%	$1.530 \cdot 10^{-4}$	$4.812 \cdot 10^{-4}$	$7.053 \cdot 10^{-7}$	$1.457 \cdot 10^{-3}$
95%	$1.476 \cdot 10^{-4}$	$7.060 \cdot 10^{-4}$	$3.910 \cdot 10^{-7}$	$2.325 \cdot 10^{-3}$
99%	$1.444 \cdot 10^{-4}$	$2.789 \cdot 10^{-3}$	$1.308 \cdot 10^{-8}$	$8.658 \cdot 10^{-3}$

The computed quantities help answer questions such as: ‘Is there something to gain from gathering additional information on any of the type B inputs?’ And ‘If so, which parameters should we learn and update by collecting additional information on them?’ Based on the confidence intervals, the answer to the first question may be based on predefined maximally admissible bounds on the interval widths or upper semi-widths (for the failure probability, naturally, the upper tail of the distribution is the decisive one). Under the second configuration, the large target variability motivates collecting additional information for uncertainty reduction, whereas, in the first configuration the target variability is already quite low. Then, in order to decide which type B-variables require an update one may resort to the variable ranking provided by the Sobol' and total-effect indices of $\log \widehat{P}_F$: for the second configuration, the load dispersion parameter b_H contributes most of the variance in $\log \widehat{P}_F$ and thus should receive priority in the acquisition of additional data.

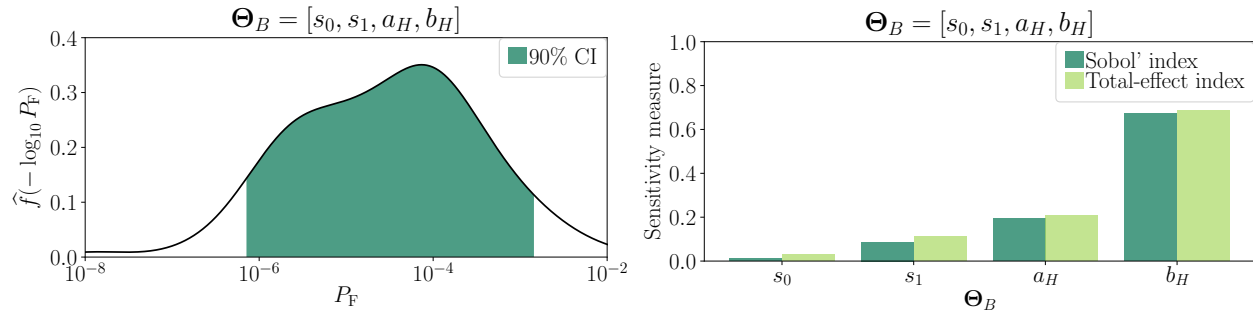


Figure 7.15: Monopile configuration 2: conditional failure PDF with confidence intervals (left) and variance-based failure sensitivity indices (right) with $n_B = 100$.

7.5 Conclusion

This paper reviews conditional reliability analysis, i.e., the estimation of the probability of failure conditional on a subset of the uncertain inputs. Such conditional probabilities of failure are functions of the input they are conditioned upon and are useful in a multitude of contexts such as sensitivity analysis, quantification and communication of lack of knowledge and decision analysis. The estimation of conditional failure probabilities is a computationally intensive task. We present a method to efficiently perform conditional reliability estimation by reusing information throughout the computation. Information reuse is realized through importance densities from previous computation steps for importance sampling estimates of conditional failure probabilities. We propose a strategy for the selection of these densities and test two importance sampling estimators that efficiently incorporate them to reduce the estimator variance, namely: mixture importance sampling and controlled importance sampling.

In two numerical examples, we find that both mixture and controlled importance sampling perform similarly and provide up to 76% computational savings compared to a baseline method without information reuse. A parameter study reveals the robustness of the proposed method both with respect to the magnitude of the probability of failure (down to $\mathbb{P}(F) = 10^{-9}$) and the dimensionality of both type A- and type B- variable spaces (scenarios with up to 200 type A-variables and 1000 type B-variables are investigated). The second example showcases the application of our method to compute the distribution and global sensitivity indices of two differently conditioned probabilities of failure for a wind turbine foundation model with 87 inputs. We find that the potential for computational savings offered by information reuse depends on the variability contributed to P_F by the type B-variables. Increasing variability reduces correlation amongst the limit-state functions in the reliability problem sequence and therefore the re-usability of importance densities.

In order to further increase the potential for computational savings of information reuse, surrogate models may be used to replace the expensive engineering model at various points in the algorithm. For example, the sequence of importance densities occurring in iCE could be based on locally reconstructed surrogate models. Similar approaches have been used in the context of conventional reliability analysis [40] and Bayesian updating [53].

7.6 Acknowledgment

We acknowledge support of the German Research Foundation (DFG) through Grant STR 1140/6-1 under SPP 1886. The third author acknowledges support of the AEOLUS center under the U.S. Department of Energy Applied Mathematics MMICC award DE-SC0019303. The first author would like to thank Anirban Chaudhuri and Boris Kramer who have been very helpful in discussing information reuse concepts as well as importance sampling and its variants.

References

- [1] S. Au and J. Beck. “A new adaptive importance sampling scheme for reliability calculations”. In: *Structural Safety* 21.2 (1999), pp. 135–158.
- [2] S. Au and J. Beck. “Important sampling in high dimensions”. In: *Structural Safety* 25.2 (2003), pp. 139–163.
- [3] S.-K. Au and J. L. Beck. “Estimation of small failure probabilities in high dimensions by subset simulation”. In: *Probabilistic Engineering Mechanics* 16.4 (2001), pp. 263–277.
- [4] S. Bansal and S. H. Cheung. “On the evaluation of multiple failure probability curves in reliability analysis with multiple performance functions”. In: *Reliability Engineering & System Safety* 167 (2017). Special Section: Applications of Probabilistic Graphical Models in Dependability, Diagnosis and Prognosis, pp. 583–594.
- [5] P. Beaurepaire et al. “Reliability-based optimization using bridge importance sampling”. In: *Probabilistic Engineering Mechanics* 34 (2013), pp. 48–57.
- [6] K. Beyer et al. “When Is “Nearest Neighbor” Meaningful?” In: *Database Theory — ICDT’99*. Ed. by C. Beeri and P. Buneman. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 217–235.
- [7] G. Bolch et al. *Queuing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*. Jan. 2006.
- [8] C. G. Bucher. “Adaptive sampling — an iterative fast Monte Carlo procedure”. In: *Structural Safety* 5.2 (1988), pp. 119–126.
- [9] Z. Cao, Y. Wang, and D. Li. “Quantification of prior knowledge in geotechnical site characterization”. In: *Engineering Geology* 203 (2016), pp. 107–116.
- [10] A. Chaudhuri, B. Kramer, and K. E. Willcox. “Information Reuse for Importance Sampling in Reliability-Based Design Optimization”. In: *Reliability Engineering & System Safety* (2020), p. 106853.
- [11] I. Depina et al. “Coupling the cross-entropy with the line sampling method for risk-based design optimization”. In: *Structural and Multidisciplinary Optimization* 55.5 (2017), pp. 1589–1612.
- [12] A. Der Kiureghian. “First-and second-order reliability methods”. In: *Engineering Design Reliability Handbook*. Ed. by E. Nikolaidis, D. M. Ghiocel, and S. Singhal. Boca Raton, FL: CRC Press, 2005. Chap. 14.

- [13] A. Der Kiureghian and O. Ditlevsen. “Aleatory or epistemic? Does it matter?” In: *Structural Safety* 31.2 (2009), pp. 105–112.
- [14] O. Ditlevsen and H. O. Madsen. *Structural reliability methods*. John Wiley & Sons Ltd, 1996.
- [15] V. Dubourg, B. Sudret, and F. Deheeger. “Metamodel-based importance sampling for structural reliability analysis”. In: *Probabilistic Engineering Mechanics* 33 (2013), pp. 47–57.
- [16] M. Ehre, I. Papaioannou, and D. Straub. “A framework for global reliability sensitivity analysis in the presence of multi-uncertainty”. In: *Reliability Engineering & System Safety* 195 (2020), p. 106726.
- [17] M. Ehre, I. Papaioannou, and D. Straub. “Efficient Conditional Reliability Updating with Sequential Importance Sampling”. In: *PAMM* 18.1 (2018), e201800282.
- [18] M. Ehre, I. Papaioannou, and D. Straub. “Global sensitivity analysis in high dimensions with PLS-PCE”. In: *Reliability Engineering & System Safety* 198 (2020), p. 106861.
- [19] M. Ehre et al. “Conditional reliability analysis in high dimensions based on controlled mixture importance sampling and information reuse”. In: *Computer Methods in Applied Mechanics and Engineering* 381 (2021), p. 113826.
- [20] S. Engelund and R. Rackwitz. “A benchmark study on importance sampling techniques in structural reliability”. In: *Structural Safety* 12.4 (1993), pp. 255–276.
- [21] M. Ester et al. “A density-based algorithm for discovering clusters in large spatial databases with noise”. In: AAAI Press, 1996, pp. 226–231.
- [22] M. H. Faber. “On the treatment of uncertainties and probabilities in engineering decision analysis”. In: *Journal of Offshore Mechanics and Arctic Engineering* 127.3 (2005), pp. 243–248.
- [23] S. Ferson et al. “Summary from the epistemic uncertainty workshop: consensus amid diversity”. In: *Reliability Engineering & System Safety* 85.1 (2004), pp. 355–369.
- [24] S. Geyer, I. Papaioannou, and D. Straub. “Cross entropy-based importance sampling using Gaussian densities revisited”. In: *Structural Safety* 76 (2019), pp. 15–27.
- [25] J. Helton and W. Oberkampf. “Alternative representations of epistemic uncertainty”. In: *Reliability Engineering & System Safety* 85.1 (2004). Alternative Representations of Epistemic Uncertainty, pp. 1–10.
- [26] M. Hohenbichler and R. Rackwitz. “Improvement Of Second-Order Reliability Estimates by Importance Sampling”. In: *Journal of Engineering Mechanics* 114.12 (1988), pp. 2195–2199.
- [27] W.-C. Hsu and J. Ching. “Evaluating small failure probabilities of multiple limit states by parallel Subset Simulation”. In: *Probabilistic Engineering Mechanics* 25.3 (2010), pp. 291–304.
- [28] S.-H. Jiang, I. Papaioannou, and D. Straub. “Bayesian updating of slope reliability in spatially variable soils with in-situ measurements”. In: *Engineering Geology* 239 (2018), pp. 310–320.
- [29] L. Katafygiotis and K. Zuev. “Geometric insight into the challenges of solving high-dimensional reliability problems”. In: *Probabilistic Engineering Mechanics* 23.2 (2008). 5th International Conference on Computational Stochastic Mechanics, pp. 208–218.
- [30] P. Koutsourelakis, H. Pradlwarter, and G. Schuëller. “Reliability of structures in high dimensions, part I: algorithms and applications”. In: *Probabilistic Engineering Mechanics* 19.4 (2004), pp. 409–417.

- [31] D. P. Kroese, R. Y. Rubinstein, and P. W. Glynn. “Chapter 2 - The Cross-Entropy Method for Estimation”. In: *Handbook of Statistics: Machine Learning: Theory and Applications*. Vol. 31. Handbook of Statistics. Elsevier, 2013, pp. 19–34.
- [32] N. Kurtz and J. Song. “Cross-entropy-based adaptive importance sampling using Gaussian mixture”. In: *Structural Safety* 42 (2013), pp. 35–44.
- [33] M. Lemaire, A. Chateauneuf, and J.-C. Mitteau. *Structural reliability*. Wiley-ISTE, 2009.
- [34] H.-S. Li, Y.-Z. Ma, and Z. Cao. “A generalized Subset Simulation approach for estimating small failure probabilities of multiple stochastic responses”. In: *Computers & Structures* 153 (2015), pp. 239–251.
- [35] L. W. Ng and K. E. Willcox. “Monte Carlo information-reuse approach to aircraft conceptual design optimization under uncertainty”. In: *Journal of Aircraft* 53.2 (2015), pp. 427–438.
- [36] A. O’Hagan and J. E. Oakley. “Probability is perfect, but we can’t elicit it perfectly”. In: *Reliability Engineering & System Safety* 85.1 (2004), pp. 239–248.
- [37] J. E. Oakley. “Decision-Theoretic Sensitivity Analysis for Complex Computer Models”. In: *Technometrics* 51.2 (2009), pp. 121–129.
- [38] A. Owen and Y. Zhou. “Safe and effective importance sampling”. In: *Journal of the American Statistical Association* 95.449 (2000), pp. 135–143.
- [39] I. Papaioannou, K. Breitung, and D. Straub. “Reliability sensitivity estimation with sequential importance sampling”. In: *Structural Safety* 75 (2018), pp. 24–34.
- [40] I. Papaioannou, M. Ehre, and D. Straub. “Efficient PCE representations for reliability analysis in high dimensions”. In: *Proceedings of the 19th working conference of the IFIP Working Group 7.5 on Reliability and Optimization of Structural Systems*. Ed. by J. Song. ETH Zürich, 2018.
- [41] I. Papaioannou, M. Ehre, and D. Straub. “PLS-based adaptation for efficient PCE representation in high dimensions”. In: *Journal of Computational Physics* 387 (2019), pp. 186–204.
- [42] I. Papaioannou, S. Geyer, and D. Straub. “Improved cross entropy-based importance sampling with a flexible mixture model”. In: *Reliability Engineering & System Safety* 191 (2019), p. 106564.
- [43] I. Papaioannou, C. Papadimitriou, and D. Straub. “Sequential importance sampling for structural reliability analysis”. In: *Structural Safety* 62 (2016), pp. 66–75.
- [44] B. Peherstorfer, B. Kramer, and K. Willcox. “Multifidelity Preconditioning of the Cross-Entropy Method for Rare Event Simulation and Failure Probability Estimation”. In: *SIAM/ASA Journal on Uncertainty Quantification* 6.2 (2018), pp. 737–761.
- [45] K.-K. Phoon and F. H. Kulhawy. “Characterization of geotechnical variability”. In: *Canadian Geotechnical Journal* 36.4 (1999), pp. 612–624.
- [46] K.-K. Phoon and F. H. Kulhawy. “Evaluation of geotechnical property variability”. In: *Canadian Geotechnical Journal* 36.4 (1999), pp. 625–639.
- [47] R. Rackwitz. “Reviewing probabilistic soils modelling”. In: *Computers and Geotechnics* 26.3 (2000), pp. 199–223.
- [48] M. Rosenblatt. “Remarks on a multivariate transformation”. In: *The Annals of Mathematical Statistics* 23.3 (1952), pp. 470–472.
- [49] R. Y. Rubinstein. “Optimization of computer simulation models with rare events”. In: *European Journal of Operational Research* 99.1 (1997), pp. 89–112.

- [50] R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo Method*. 3rd. Wiley Publishing, 2017.
- [51] E. Ullmann and I. Papaioannou. “Multilevel Estimation of Rare Events”. In: *SIAM/ASA Journal on Uncertainty Quantification* 3.1 (2015), pp. 922–953.
- [52] Z. Wang and J. Song. “Cross-entropy-based adaptive importance sampling using von Mises-Fisher mixture for high dimensional reliability analysis”. In: *Structural Safety* 59 (Mar. 2016), pp. 42–52.
- [53] J. Zhang and A. Taflanidis. “Bayesian posterior sampling using a metamodel-based sequential approach”. In: *Proceedings of the 19th working conference of the IFIP Working Group 7.5 on Reliability and Optimization of Structural Systems*. Ed. by J. Song. ETH Zürich, 2018.

A framework for global reliability sensitivity analysis in the presence of multi-uncertainty

Original Publication

M. Ehre, I. Papaioannou, and D. Straub. “A framework for global reliability sensitivity analysis in the presence of multi-uncertainty”. In: *Reliability Engineering & System Safety* 195 (2020), p. 106726.

Author’s contribution

Max Ehre, Iason Papaioannou and Daniel Straub developed the concept for the reliability sensitivity framework. Max Ehre wrote the underlying code and carried out all numerical experiments. Max Ehre drafted the manuscript under the guidance of Iason Papaioannou and Daniel Straub.

Abstract

In reliability analysis with numerical models, one is often interested in the sensitivity of the probability of failure estimate to changes in the model input. In the context of multi-uncertainty, one wishes to separate the effect of different types of uncertainties. A common distinction is between aleatory (irreducible) and epistemic (reducible) uncertainty, but more generally one can consider any classification of the uncertain model inputs in two subgroups, type A and type B. We propose a new sensitivity measure for the probability of failure conditional on type B inputs. On this basis, we outline a framework for multi-uncertainty-driven reliability sensitivity analysis. A bi-level surrogate modelling strategy is designed to efficiently compute the new conditional reliability sensitivity measures. In the first level, a surrogate is constructed for the model response to circumvent possibly expensive evaluations of the numerical model. By solving a sequence of reliability problems condi-

tional on samples of type B random variables, we construct a level 2-surrogate for the logarithm of the conditional probability of failure, using polynomial bases which allow to directly evaluate the variance-based sensitivities. The new sensitivity measure and its computation are demonstrated through two engineering examples.

8.1 Introduction

Reliability analysis is concerned with the evaluation of the probability of failure of an engineering system. The system can be described probabilistically in terms of the input random vector Θ with joint cumulative distribution function (CDF) F_{Θ} and a deterministic model \mathcal{Y} mapping each Θ to an output $Y = \mathcal{Y}(\Theta)$. The performance of the system can be assessed in terms of its limit-state function g (LSF). The limit-state function defines the failure modes of a system and by convention takes values below 0 in the failure domain, which is a subset of the entire input variable space Ω_{Θ} . The system probability of failure is given by [15]

$$\mathbb{P}(F) = \mathbb{E}_{\Theta} [\mathbf{I}(g(\Theta) \leq 0)] = \int_{\Omega_{\Theta}} \mathbf{I}[g(\theta) \leq 0] \pi_{\Theta}(\theta) d\theta, \quad (8.1)$$

where the indicator function \mathbf{I} equals 1 on the failure domain $\{\theta \in \Omega_{\theta} : g(\theta) \leq 0\}$ and 0 on its complement and π_{Θ} is the joint probability density function (PDF) of Θ .

Standard Monte Carlo methods fail to efficiently solve Eq. (8.1) if $\mathbb{P}(F)$ is very small, which is typically the case for failure probabilities. Methods which are specifically designed to solve Eq. (8.1) for very small $\mathbb{P}(F)$ are referred to as structural reliability methods (SRM). SRM can be categorized into approximation (e.g., FORM, SORM [50, 13]) and sampling methods (e.g., importance sampling [7, 17, 23], line sampling [31], subset simulation [1] and sequential importance sampling [46, 3]), both of which can be combined with surrogate modelling (e.g., [18, 57]).

To account for different types of uncertainties (i.e., multi-uncertainty), one may consider a segmentation of the random input vector in two disjunct subsets $\Theta = [\Theta_A, \Theta_B]^T$. The variables Θ_B can be interpreted as *epistemic* (sometimes also termed *reducible*) and the variables in Θ_A as *aleatory* (*irreducible*). Irrespective of the interpretation of Θ_B , we aim at explicitly quantifying its influence on $\mathbb{P}(F)$. This is, for instance, useful to identify those inputs in Θ_B whose uncertainty should be reduced by means of additional data in order to increase the accuracy of the reliability analysis.

Fig. 8.1 illustrates our framework for reliability analysis and updating in the presence of multi-uncertainty. To obtain information on the potential influence of each component of Θ_B on the reliability analysis, we consider the probability of failure conditional on Θ_B [14]:

$$\begin{aligned} P_F(\theta_B) &= \mathbb{P}(F | \Theta_B = \theta_B) \\ &= \mathbb{E}_{\Theta_A} [\mathbf{I}(g(\Theta_A, \theta_B) \leq 0) | \Theta_B = \theta_B] \\ &= \int_{\Omega_{\Theta_A}} \mathbf{I}[g(\theta_A, \theta_B) \leq 0] \pi_{\Theta_A | \Theta_B}(\theta_A | \theta_B) d\theta_A. \end{aligned} \quad (8.2)$$

Note, that $P_F(\theta_B)$ is a scalar function of the outcome of a random vector and hence is itself a random variable. A similar way of handling uncertainty separation is presented under the term

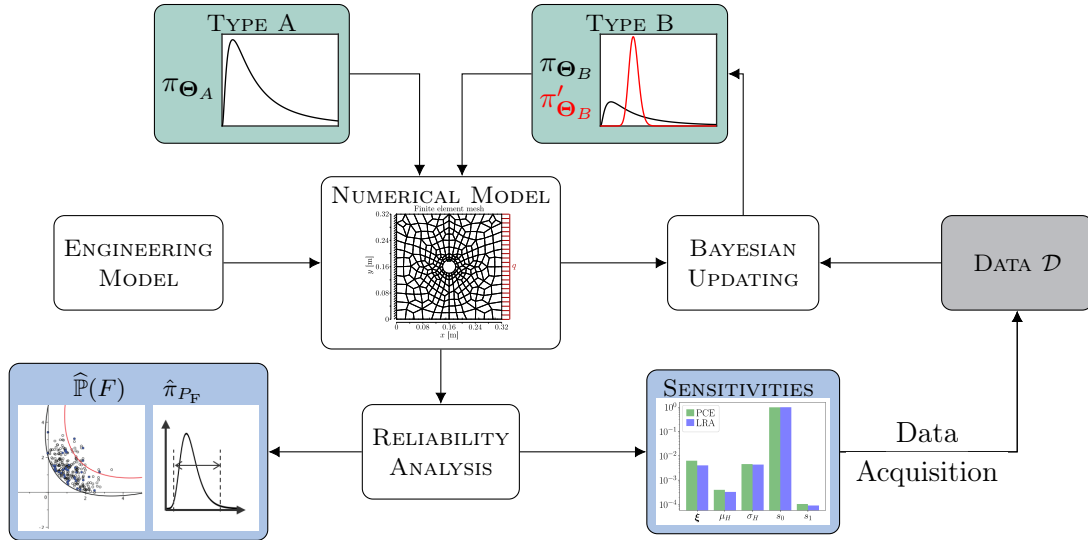


Figure 8.1: Framework for reliability sensitivity analysis in the presence of multi-uncertainty. The model \mathcal{Y} is subject to irreducible (Θ_A) and reducible (Θ_B) probabilistic inputs. Estimates of the distribution of the probability of failure conditional on Θ_B , $\hat{\pi}_{P_F}$, may be obtained through repeated reliability analyses fixed at different θ_B . Thus, $\hat{\pi}_{P_F}$ is subject to change when updating components of Θ_B . The proposed reliability sensitivities quantify the obtainable variability reduction in P_F resulting from uncertainty reduction (e.g., through Bayesian updating) in the components of Θ_B . In this way, they allow for an informed selection of data acquisition measures.

"second-order probability" in the program package Dakota [[dakota_manual](#)] and is also discussed by other authors [14, 10].

In the literature, a variety of metrics for quantifying sensitivities with respect to different quantities of interest can be found. They can be grouped according to scope (local [42] vs. global importance [54, 6]), moment-dependence (e.g., variance-based [54, 24, 26, 52] vs. distribution-based [6, 35]) and considered inputs (deterministic parameters [43] vs. random variables [54, 24, 26, 52]). A comprehensive overview over global sensitivity measures is given in [25] and [51].

Most often, these measures are computed for the outcome of the numerical model. In contrast, reliability-oriented sensitivity analysis is concerned with determining the sensitivity of the output of a reliability analysis to the model inputs. Many approaches dedicated to reliability-oriented sensitivities are concerned with determining local derivative-based sensitivity measures for deterministic model parameters in the limit-state function [43] or deterministic distribution parameters of the input [10, 59, 55, 16]. Alternatively, [2] computes failure probability histograms conditional on design parameters (LSF or distribution parameters) by artificially considering them as random variables. These histograms can be used to compute global sensitivity measures for the parameters. When using approximative reliability methods like FORM/SORM, reliability sensitivities are obtained from the analysis in form of the component (or α -) values of the design point (or most probable point of failure) [13, 36]. These can be interpreted as global, variance-based sensitivity indices of the first-(FORM) and second-(SORM) order Taylor expansions of g around the design point [36] in standard-normal space. Recently, [28] extended this idea to account for a mixture of several significant points and regions along the hypersurface $g = 0$ by means of a Gaussian mixture (GM) model through defining a participation factor for each of the regional design point contributions of the GM

components. [35] compute variance-based sensitivities for the indicator function $I(g \leq 0)$ and show, that this is equivalent to computing the moment-independent sensitivity measure proposed in [6] for the probability of failure. They use a surrogate model to relax the computational cost induced by the sampling-based sensitivity computation approach. Along the same lines, [58] also use a surrogate (Kriging) modelling approach when computing variance-based sensitivities for the probability of failure conditional on uncertain distribution parameters of the input variables. [40] compute variance-based sensitivity indices for a probability of failure conditional on distribution parameters using importance sampling. These indices bear some similarity to the definition of sensitivity indices used in the context of stochastic black boxes or stochastic emulators by, e.g., [38]. Here, a fraction of the model uncertainty corresponding to Θ_A is considered model-inherent or latent and thus Sobol' indices are defined for a conditional mean, which retains only a dependence in Θ_B and marginalizes over Θ_A . [38] compute Sobol' indices for the response of such an inherently stochastic model by constructing surrogates jointly for the model response mean and variance. Combining this idea with that of [35] - i.e. to analyze the indicator function - yields the reliability-oriented Sobol' indices discussed in [58, 40]. Alternative sensitivity measures for rare event probabilities include the use of quantiles [39] or perturbation of input densities [33] to globally quantify influence of model inputs on rare event probabilities.

In this contribution, we introduce a reliability-oriented sensitivity measure that is based on the variance decomposition of the logarithm of the conditional probability of failure, defined in Eq. (8.2). The proposed measure enables the quantification of the influence of the type B (reducible inputs), which may consist of either input variables or uncertain distribution parameters thereof, on the magnitude of the probability of failure (Fig. 8.1). To efficiently compute the proposed reliability sensitivities, we introduce a 2-level surrogate based approach. In the first level, a surrogate of the model response is constructed and is further used to compute the probability of failure conditional on samples of the type B-variables. The resulting conditional probability estimates are further used to construct a polynomial basis surrogate model of the logarithm of the conditional probability of failure. The coefficients of the level-2 surrogate can be post-processed to compute an estimate of the sought sensitivity indices.

The structure of the paper is as follows: In Section 8.2, the basics of global sensitivity analysis are discussed and the new reliability sensitivity index is introduced. Section 8.3 provides an introduction to polynomial basis surrogate modelling (polynomial chaos expansions (PCEs) and k approximations (LRAs)) and their relation to global sensitivity measures. Section 8.4 outlines a two-level framework for the surrogate-driven estimation of the novel reliability sensitivities, which is subsequently tested on two numerical examples in Section 8.5. The studied examples consist of a 12-dimensional elastic truss structure and a 87-dimensional wind turbine monopile foundation in plastically behaving soil. Section 8.6 contains a discussion of the presented method and an outlook on possible extensions and improvements of the introduced framework.

8.2 Global Reliability Sensitivity Analysis

The sensitivity measures for reliability proposed here belong to the realm of variance-based sensitivity methods. After a brief introduction of the underlying idea of variance-based sensitivity indices, their application within reliability analysis is discussed. We discuss the computational cost of these newly

introduced indices when estimated with classical Monte Carlo-based (MC) approaches and thereby motivate a surrogate-driven computational approach, which is discussed in the next section.

8.2.1 Variance-Based Sensitivity Analysis

Consider a d -dimensional standard uniform random vector $\mathbf{U} \sim \mathcal{U}_d(0, 1)$, representing the input uncertainty of some model, and an output quantity of interest (QOI) $Q = f(\mathbf{U})$, with $f : [0, 1]^d \rightarrow \mathbb{R}$. Variance-based sensitivity indices rely on estimating the contribution of a single (first-order index) or a combination of the inputs \mathbf{U} (n -th order index, n being the number of variables considered in the combination) to the output variance of the QOI. These fractions can be found by projecting the QOI on an orthogonal functional basis, which leads to a unique representation of f , namely the Sobol'-Hoeffding decomposition [54]

$$f(\mathbf{U}) = f_0 + \sum_{i=1}^d f_i(U_i) + \sum_{i=1}^d \sum_{j=i+1}^d f_{ij}(U_i, U_j) + \cdots + f_{12\dots d}(\mathbf{U}). \quad (8.3)$$

Here, all basis functions have zero mean except for $f_0 = \mathbb{E}[Q]$. If Q depends on pairwise independent inputs with arbitrary distribution with known marginal CDFs F_{Θ_i} , an isoprobabilistic transformation can be used to generalize Eq. (8.3). This transformation is given by $T : \Theta_i \rightarrow F_{\Theta_i}(\Theta_i)$, $1 \leq i \leq d$, and the transformed decomposition is obtained by setting $\mathbf{U} = T(\Theta)$ in Eq. (8.3). By virtue of the orthogonality property, the variance of f is merely the sum of the partial variances of all the basis functions. The partial variance corresponding to a set of inputs indicated by the index set $\mathcal{A} \subseteq \{1, 2, \dots, d\}$ is found as the variance of $f_{\mathcal{A}}$. The corresponding Sobol' index is then defined by the ratio

$$S_{Q,\mathcal{A}} = \frac{\mathbb{V}[f_{\mathcal{A}}]}{\mathbb{V}[f]}. \quad (8.4)$$

The Sobol' index measures the variance fraction that originates from a particular combination of variables exclusively without considering interaction of $\Theta_{\mathcal{A}}$ with $\Theta_{\sim\mathcal{A}}$, where \sim denotes the complement of \mathcal{A} over the total set $\{1, 2, \dots, d\}$. [24] defined the total-effect indices, which measure the partial variances of all basis functions that include $\Theta_{\mathcal{A}}$:

$$S_{Q,\mathcal{A}}^T = \frac{\sum_{\mathcal{I} \supseteq \mathcal{A}} \mathbb{V}[f_{\mathcal{I}}]}{\mathbb{V}[f]}. \quad (8.5)$$

Alternatively, both Sobol' and total-effect indices can be cast in terms of variances of conditional expectations as [51]:

$$\begin{aligned} S_{Q,\mathcal{A}} &= \frac{\mathbb{V}_{\Theta_{\mathcal{A}}} [\mathbb{E}_{\Theta_{\sim\mathcal{A}}} [Q | \Theta_{\mathcal{A}}]]}{\mathbb{V}[Q]}, \\ S_{Q,\mathcal{A}}^T &= \frac{\mathbb{E}_{\Theta_{\sim\mathcal{A}}} [\mathbb{V}_{\Theta_{\mathcal{A}}} [Q | \Theta_{\sim\mathcal{A}}]]}{\mathbb{V}[Q]} \\ &= 1 - \frac{\mathbb{V}_{\Theta_{\sim\mathcal{A}}} [\mathbb{E}_{\Theta_{\mathcal{A}}} [Q | \Theta_{\sim\mathcal{A}}]]}{\mathbb{V}[Q]}. \end{aligned} \quad (8.6)$$

8.2.2 The proposed reliability sensitivity indices

In order to define variance-based sensitivities related to the reliability of a system one has to choose Q appropriately. In reliability analysis, the QOI is the failure event F and the associated probability of failure. Since F is defined via the indicator function of the failure domain, [35] propose to compute importance rankings through the variance decomposition of the indicator function $I(g \leq 0)$. They do so by means of a surrogate modelling technique to avoid the slow convergence of standard MC-estimators in cases where the QOI is a rare event.

Within the multi-uncertainty setting, we propose to perform the variance decomposition of the conditional probability of failure defined in Eq. (8.2). More precisely, we propose using

$$Q = \log P_F(\Theta_B), \quad (8.7)$$

as the QOI, where \log indicates the logarithm to the base 10. Q is a measure for the magnitude of the conditional probability of failure. In this way, we focus the sensitivity analysis on possibly substantial/magnitude-altering changes in the estimate of $P(\Theta_B)$. Note that in the mono-uncertainty case, in which Θ_A is empty and all uncertainty is reducible, Q takes values 0 or $-\infty$. This obstructs the computation of the proposed index in such case. The concepts introduced in the following may be generalized to any QOI that can be cast in terms of the expectation of a functional (e.g., any moment of Y). Consider now an arbitrary subset of Θ_B , which is denoted by $\Theta_{\mathcal{B}}$, and its complement $\Theta_{\sim\mathcal{B}}$ such that $\Theta_B = \{\Theta_{\mathcal{B}}, \Theta_{\sim\mathcal{B}}\}$. The novel sensitivity indices for the variable subset $\Theta_{\mathcal{B}}$ are given by

$$S_{\log P_F, \mathcal{B}} = \frac{\mathbb{V}_{\Theta_{\mathcal{B}}} [\mathbb{E}_{\Theta_{\sim\mathcal{B}}} [\log P_F | \Theta_{\mathcal{B}}]]}{\mathbb{V} [\log P_F]}, \quad (8.8)$$

$$S_{\log P_F, \mathcal{B}}^T = 1 - \frac{\mathbb{V}_{\Theta_{\sim\mathcal{B}}} [\mathbb{E}_{\Theta_{\mathcal{B}}} [\log P_F | \Theta_{\sim\mathcal{B}}]]}{\mathbb{V} [\log P_F]}. \quad (8.9)$$

Substituting Eq. (8.2) in the above, one arrives at the following expressions for the novel sensitivity indices:

$$S_{\log P_F, \mathcal{B}} = \frac{\mathbb{V}_{\Theta_{\mathcal{B}}} [\mathbb{E}_{\Theta_{\sim\mathcal{B}}} [\log \{\mathbb{E}_{\Theta_A} [I(g \leq 0) | \Theta_B]\} | \Theta_{\mathcal{B}}]]}{\mathbb{V}_{\Theta_B} [\log \{\mathbb{E}_{\Theta_A} [I(g \leq 0) | \Theta_B]\}]} \quad (8.10)$$

$$S_{\log P_F, \mathcal{B}}^T = 1 - \frac{\mathbb{V}_{\Theta_{\sim\mathcal{B}}} [\mathbb{E}_{\Theta_{\mathcal{B}}} [\log \{\mathbb{E}_{\Theta_A} [I(g \leq 0) | \Theta_B]\} | \Theta_{\sim\mathcal{B}}]]}{\mathbb{V}_{\Theta_B} [\log \{\mathbb{E}_{\Theta_A} [I(g \leq 0) | \Theta_B]\}]} \quad (8.11)$$

While these expressions appear cumbersome, they exhibit key features of the new indices:

1. The variance decomposition of the total variance contributed by Θ_B rather than Θ is performed, which is reflected by the normalizing constants in Eqs. (8.10) and (8.11).
2. Due to the expectation \mathbb{E}_{Θ_A} , the employed QOI is smooth on Ω_{Θ_B} . In particular, it is non-binary as opposed to the QOI underlying the indices proposed by [35].
3. The log-transformation of the conditional probability of failure focuses the sensitivity analysis on the probability of failure magnitude.

A key motivation for defining the sensitivity metric based on $\log P_F$ is that the probability of failure and the resulting risk estimates are typically interpreted in terms of their orders of magnitude [8,

34]. Additionally, the distribution of P_F is heavily right-skewed and its support covers multiple orders of magnitude, as opposed to the distribution of $\log P_F$ which is supported on a single order of magnitude. This renders the estimation of the associated sensitivity indices more robust for the log-transformed QOI.¹

Given that the two measures will attribute different importance to the same variable, it is worthwhile discussing which QOI is the more appropriate choice depending on the analysis goal. If the conditional P_F -distributions associated with two inputs have identical mean and variance, they will be attributed equal importance under $Q = P_F$. Under $Q = \log P_F$ this is not necessarily the case, as the variance of $\log P_F$ also depends on higher-order moments of P_F , most prominently its skewness. In particular, under $Q = \log P_F$, more heavily right-skewed distributions will have a higher importance when computing sensitivities. In other words, the variability over small failure probability magnitudes is given more importance in this case. Such a behaviour is desirable when the goal of the sensitivity analysis is to identify variables which contribute most to the inaccuracy of the failure probability estimate. This is useful to, e.g., decide on which of the inputs to collect more data or perform more measurements in order to increase the accuracy of a reliability analysis or a reliability-based design. If, on the other hand, the analysis aims at identifying those variables which are responsible for a large predictive failure probability, e.g. in the context of retrofitting and fortification, one should use $Q = P_F$ rather than $Q = \log P_F$.

The proposed indices reveal connections to other global reliability sensitivity measures presented in the literature. In particular, when decomposing the conditional probability of failure directly instead of its log-transformation, the Sobol' indices of the indicator function defined in [35] are recovered up to a constant factor (see Section 8.A). Computing Sobol' indices of P_F , that is without the log-transformation, has been proposed in [40] for the case where Θ_B consists of distributional parameters only. [58] also discuss these indices and propose to efficiently compute them using Gaussian process surrogates.

8.2.3 Monte-Carlo estimators

This section reviews common problems when tackling reliability sensitivities with sampling methods. [26] and [52] provide Monte Carlo-estimators for the expressions in Eq. (8.6). Based on a set of n_s d -dimensional Θ -samples, $n_s \cdot (d + 2)/2$ model evaluations are necessary to compute them, where n_s is the samples size required by the analysis. A reasonable estimate for n_s is the number of samples required for the MC-estimation of $\mathbb{E}[Q]$ at a prescribed accuracy. These estimators may be intractable if a model evaluation is computationally expensive, d is large or Q is given by a failure event with small associated probability of failure. In the last case, intractability arises from the required amount of samples (i.e., evaluations of the g -function) to accurately estimate $\mathbb{E}[Q]$. It is in the order of $100/\mathbb{P}(F)$ when the allowed coefficient of variation of the estimator is 10%. Therefore, when $\mathbb{P}(F)$ is very small, n_s becomes prohibitively large.

For the proposed sensitivity indices, the computational burden would amount to a multiple of what is needed for the computation of sensitivity indices of the indicator function of g . This is due to the need to solve Eq. (8.2) $n_s \cdot (d+2)/2$ times, which may in turn require many g -evaluations per solution.

¹Another such transformation, which in a sense regularizes the outcome space, would be given by considering the generalized reliability index $\beta = -\Phi^{-1}(P_F)$ [15] rather than P_F .

Conversely, computing the indices associated with $Q = \mathbb{I}(g \leq 0)$ requires a single g -evaluation at each sample to determine whether $g \leq 0$. Therefore, the computational effort scales approximately as the average number of g -calls necessary to solve Eq. (8.2) at constant n_s (although, at convergence, n_s would likely be considerably smaller compared to the case where $Q = \mathbb{I}(g \leq 0)$). However, the smoothness in our choice of Q is key to an entirely surrogate-driven sensitivity computation, which facilitates the use of only a small fraction of the samples required in the sampling-based procedure. Two types of surrogate models have been tested and are detailed in the subsequent section.

8.3 Polynomial Basis Surrogate Modelling

Let Θ be a random vector on the outcome space \mathbb{R}^d with joint CDF F_{Θ} whose elements are mutually independent and $Y = \mathcal{Y}(\Theta)$, with $\mathcal{Y} : \mathbb{R}^d \rightarrow \mathbb{R}$. If Y has finite mean-square, i.e., $\mathbb{E}_{\Theta}[\mathcal{Y}(\Theta)^2] < \infty$, then the function \mathcal{Y} belongs in a Hilbert space \mathcal{H} on which an inner product of any two functions $g, h \in \mathcal{H}$ is defined as

$$\langle g(\boldsymbol{\theta}), h(\boldsymbol{\theta}) \rangle_{\mathcal{H}} = \mathbb{E}_{\Theta}[g(\Theta)h(\Theta)] = \int_{\mathbb{R}^d} g(\boldsymbol{\theta})h(\boldsymbol{\theta})\pi_{\Theta}(\boldsymbol{\theta})d\boldsymbol{\theta}, \quad (8.12)$$

where $\pi_{\Theta}(\boldsymbol{\theta})$ is the joint PDF of Θ . g and h are orthogonal if

$$\langle g(\boldsymbol{\theta}), h(\boldsymbol{\theta}) \rangle_{\mathcal{H}} = \mathbb{E}_{\Theta}[g(\Theta)h(\Theta)] = 0. \quad (8.13)$$

Note, that if g and h can be written as products of univariate functions of the components of Θ , the following holds:

$$\langle g(\boldsymbol{\theta}), h(\boldsymbol{\theta}) \rangle_{\mathcal{H}} = \prod_{i=1}^d \mathbb{E}_{\Theta_i}[g_i(\Theta_i)h_i(\Theta_i)]. \quad (8.14)$$

Given a complete and orthonormal basis of \mathcal{H} , $\{h_i(\boldsymbol{\theta}), i \in \mathbb{N}\}$, Y may be expressed as a linear combination of the basis functions:

$$Y = \mathcal{Y}(\Theta) = \sum_{i=0}^{\infty} a_i h_i(\Theta). \quad (8.15)$$

Then, since $\mathcal{Y} \in \mathcal{H}$, the approximation

$$\hat{Y} = \hat{\mathcal{Y}}(\Theta) = \sum_{i=0}^m a_i h_i(\Theta) \quad (8.16)$$

asymptotically ($m \rightarrow \infty$) converges to Y in the mean-square sense. For $d = 1$, a possible choice of basis functions related to certain standard distribution types of π_{Θ} are polynomial families $\{\psi_i(\theta), i = 0, \dots, m\}$, which are identified by means of the Askey scheme [60]. This lays the foundation for both PCEs and LRAs. They differ with respect to how the multi-dimensional base polynomials are defined and how the expansion coefficients a_i are determined. For $d > 1$, due to Eq. (8.14), multi-dimensional basis polynomials Ψ_k can be easily constructed as products of the one-dimensional canonical polynomials $\psi_j^{(i)}$. Note, that in the multi-dimensional case, the additional superscript i is used as the various inputs may have different marginal PDFs and thus the $\psi^{(i)}$ would represent different polynomial families for different i .

8.3.1 Polynomial Chaos Expansions

Given the polynomial family of the i -th input θ_i up to m_i -th order $\{\psi_j^{(i)}(\theta_i), j = 0, \dots, m_i\}$, the j -th multi-dimensional basis function reads

$$\Psi_j(\boldsymbol{\theta}) = \prod_{i=1}^d \psi_{\alpha_{ji}}^{(i)}(\theta_i), \quad (8.17)$$

where $\boldsymbol{\alpha}$ contains all combinations of d -dimensional index sets each assigning a polynomial order to each input θ_i such that the total polynomial order $|\boldsymbol{\alpha}_j| = \sum_{i=1}^d \alpha_{ji} \leq p, 0 \leq j \leq P-1$. The number of basis functions P is given by

$$P = \binom{d+p}{p} \quad (8.18)$$

and the polynomial chaos expansion (PCE) reads

$$\hat{\mathcal{Y}}^{PCE}(\boldsymbol{\Theta}) = \sum_{j=0}^{P-1} a_j \prod_{i=1}^d \psi_{\alpha_{ji}}^{(i)}(\Theta_i). \quad (8.19)$$

The coefficients \mathbf{a} are identified through a projection of \mathcal{Y} onto the space spanned by $\{\Psi_j, j = 0, \dots, P-1\}$. In this work, we evaluate \mathbf{a} using an ordinary least-squares (OLS) approach, which approximates the projection of \mathcal{Y} onto the PCE basis [4]. Consider a set of n_s samples \mathbf{X} of the input random vector $\boldsymbol{\Theta}$ called the experimental design and corresponding \mathcal{Y} -evaluations \mathbf{Y} , where $\mathbf{X} \in \mathbb{R}^{n_s \times d}$ and $\mathbf{Y} \in \mathbb{R}^{n_s \times 1}$. The collection $\mathcal{E} = \{\mathbf{X}, \mathbf{Y}\}$ is referred to as the training set. The data matrix $\boldsymbol{\Psi}$ collects the evaluation of each of the multi-dimensional basis polynomials at each point in \mathbf{X}

$$\boldsymbol{\Psi} = \Psi_j(\mathbf{x}^{(i)}), 1 \leq i \leq n_s, 1 \leq j \leq P, \quad (8.20)$$

where $\mathbf{x}^{(i)}$ is the i -th row of \mathbf{X} . The vector of all P PCE-coefficients \mathbf{a} is then obtained by regressing \mathbf{Y} on $\boldsymbol{\Psi}$ which gives

$$\mathbf{a} = (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathbf{Y}. \quad (8.21)$$

Eq. (8.18) indicates a fast growth of the OLS problem size with increasing dimension d . This motivates the use of sparse PCE methods, which are also applied in this work. Sparse PCE reduces P by penalizing the number of terms in the PCE through solving a regularized least-squares problem [5]. In this way, the method elicits a minimal number of basis functions such as to best explain the output variance.

8.3.2 Canonical Decomposition

Low-rank approximations (LRA) have been introduced originally to represent high-dimensional tensors by means of lower-dimensional tensors [21]. A specific format of such approximations are canonical decompositions, in which tensors are approximated by means of a linear combination of products of one-dimensional tensors [22]. The idea extends to continuous spaces where a multivariate function is approximated by a linear combination of products of univariate functions:

$$\hat{\mathcal{Y}}^{LRA}(\boldsymbol{\Theta}) = \sum_{j=1}^r a_j \prod_{i=1}^d \sum_{k=0}^{m_i} z_{ijk} \psi_k^{(i)}(\Theta_i), \quad 0 \leq k \leq m_i, \quad 1 \leq i \leq d. \quad (8.22)$$

Therein, an additional set of coefficients \mathbf{z} appears, which can be efficiently determined by solving reduced, univariate least squares problems over the directions $i = 1, \dots, d$ repeatedly (while keeping all remaining directions constant in each step; this is often referred to as alternating least squares). In a second step, the coefficients \mathbf{a} are determined via OLS. A detailed description of the procedure is given in [11] and [30]. The number of coefficients in the canonical decomposition is $\mathcal{O}((m+1) \cdot d \cdot r)$ assuming a constant polynomial degree m in all dimensions. This linearity in d is remarkable and explains the advantage this format offers over classical PCE where the coefficients grow factorially with the dimension as described in Eq. (8.18).

8.3.3 Surrogate-Based sensitivity indices

Both PCEs and LRAs can be used to infer first-order and total sensitivity indices directly from the computed model coefficients. Rather than searching estimates of the expressions in Eq. (8.6), the similarity of the underlying orthogonal Sobol' decomposition in Eq. (8.3) with Eq. (8.19) and Eq. (8.22) is exploited and the expressions Eq. (8.4) and Eq. (8.5) can be computed directly. [56] showed that the Sobol' decomposition of the PCE is readily obtained by collecting any multi-dimensional orthogonal polynomials depending on identical variable subsets $\Theta_{\mathcal{A}}$ into $f_{\mathcal{A}}(\Theta_{\mathcal{A}})$. Therefore, computing the partial variance of the PCE model associated with a subset of variables $\Theta_{\mathcal{A}}$ amounts to summing the squared coefficients of the respective multi-dimensional basis polynomials in which the elements of $\Theta_{\mathcal{A}}$ occur (exclusively for Sobol' indices and collectively for total indices). The same concept can be applied to LRAs even though the compressed format (product) renders the evaluation somewhat more tedious. For a given subset of the input variables denoted by the index set \mathcal{A} , $\mathcal{I}^{\mathcal{A}} \in \{0, 1\}^d$ is a boolean index vector s.t. $\mathcal{I}_i^{\mathcal{A}} = 0$ if $i \notin \mathcal{A}$ and $\mathcal{I}_i^{\mathcal{A}} = 1$ if $i \in \mathcal{A}$. In the same way, for the j -th row of $\boldsymbol{\alpha}_j$, $\mathcal{I}_i^{\boldsymbol{\alpha}_j} = 0$ if $\alpha_{ij} = 0$ and $\mathcal{I}_i^{\boldsymbol{\alpha}_j} = 1$ if $\alpha_{ij} > 0$. Then, the PCE-based sensitivity indices read [56]:

$$\hat{S}_{Q,\mathcal{A}} = \sum_{\substack{\mathcal{I}^{\mathcal{A}} = \mathcal{I}^{\boldsymbol{\alpha}_j}, \\ 1 \leq j \leq P-1}} a_j^2 \bigg/ \sum_{1 \leq j \leq P-1} a_j^2, \quad \hat{S}_{Q,\mathcal{A}}^T = \sum_{\substack{(\mathcal{I}^{\mathcal{A}})^T \mathcal{I}^{\boldsymbol{\alpha}_j} > 0, \\ 1 \leq j \leq P-1}} a_j^2 \bigg/ \sum_{1 \leq j \leq P-1} a_j^2 \quad (8.23)$$

and the LRA-based indices [29]:

$$\hat{S}_{Q,\mathcal{A}} = \frac{\sum_{j=1}^r \sum_{j'=1}^r a_j a_{j'} \left(\prod_{i \notin \mathcal{A}} z_{ij,0} z_{ij',0} \right) \left(\prod_{i' \in \mathcal{A}} \left(\sum_{k=1}^{m_i} z_{i'jk} z_{i'j'k} \right) \right)}{\sum_{j=1}^r \sum_{j'=1}^r a_j a_{j'} \left(\left(\prod_{i=1}^d \left(\sum_{k=1}^{m_i} z_{ijk} z_{ij'k} \right) \right) - \left(\prod_{i=1}^d z_{ij,0} z_{ij',0} \right) \right)} - \frac{\sum_{j=1}^r \sum_{j'=1}^r a_j a_{j'} \left(\prod_{i=1}^d \left(\sum_{k=1}^{m_i} z_{ijk} z_{ij'k} \right) \right)}{\sum_{j=1}^r \sum_{j'=1}^r a_j a_{j'} \left(\left(\prod_{i=1}^d \left(\sum_{k=1}^{m_i} z_{ijk} z_{ij'k} \right) \right) - \left(\prod_{i=1}^d z_{ij,0} z_{ij',0} \right) \right)}, \quad (8.24)$$

$$\hat{S}_{Q,\mathcal{A}}^T = 1 - \frac{\sum_{j=1}^r \sum_{j'=1}^r a_j a_{j'} \left(\prod_{i \in \mathcal{A}} z_{ij,0} z_{ij',0} \right) \left(\prod_{i' \notin \mathcal{A}} \left(\sum_{k=1}^{m_i} z_{i'jk} z_{i'j'k} \right) \right)}{\sum_{j=1}^r \sum_{j'=1}^r a_j a_{j'} \left(\left(\prod_{i=1}^d \left(\sum_{k=1}^{m_i} z_{ijk} z_{ij'k} \right) \right) - \left(\prod_{i=1}^d z_{ij,0} z_{ij',0} \right) \right)} - \frac{\sum_{j=1}^r \sum_{j'=1}^r a_j a_{j'} \left(\prod_{i=1}^d \left(\sum_{k=1}^{m_i} z_{ijk} z_{ij'k} \right) \right)}{\sum_{j=1}^r \sum_{j'=1}^r a_j a_{j'} \left(\left(\prod_{i=1}^d \left(\sum_{k=1}^{m_i} z_{ijk} z_{ij'k} \right) \right) - \left(\prod_{i=1}^d z_{ij,0} z_{ij',0} \right) \right)}. \quad (8.25)$$

In summary, Eqs. (8.23) and (8.25) provide a way of computing the variance fraction in a QOI \mathcal{Q} caused by a subset of surrogate model inputs identified by \mathcal{A} . One can compute the effect of the variables in \mathcal{A} exclusively (and call this the Sobol' index \hat{S}) or including any interactions of these variables with the remainder of the inputs (and call this the total-effect index \hat{S}^T).

8.4 Conditional Surrogate-Based Reliability Sensitivities

The computation of sensitivity indices via polynomial surrogates requires the QOI to be sufficiently smooth. In particular, any attempts to obtain surrogate-based indices of the indicator function of the failure domain $I(g \leq 0)$ directly in such a manner must fail due to the discontinuity in $I(g \leq 0)$. However, the log-transformed conditional probability of failure is continuous in the space of Θ_B so that one may compute the proposed sensitivity indices with polynomial surrogates. To this end, we devise a two-level surrogate modelling procedure. Building a surrogate of $\log P_F$ (level 2) requires an experimental design that consists of samples of Θ_B and the associated probabilities of failure given each of these samples. That is, one has to solve n_2 reliability problems, where n_2 is the experimental design size for the final surrogate. An auxiliary (level 1) surrogate is built for the actual model \mathcal{Y} , based on which the reliability computations can be conducted. This approach is referred to as global in the following. Alternatively, a local surrogate model may be constructed at each sample of Θ_B for solving its associated reliability problem. The local surrogate has lower-dimensional input compared to the global surrogate because the input space is reduced from Ω_{Θ} to Ω_{Θ_A} , and is therefore cheaper to determine; however, it has to be recomputed for each Θ_B -sample. This approach will be referred to as local. Prior to construction, an isoprobabilistic transformation to an independent standard-normal space is used, such that both PCEs and LRAs can be consistently constructed using the orthogonal polynomial basis with respect to the standard normal probability measure, which is the Hermite polynomial basis [60].

8.4.1 Level 1

On the first level, the goal is to construct a surrogate model for the original model $\mathcal{Y}(\Theta)$, describing the engineering system.

Global Approach

In the global approach, the model is evaluated at the level 1-experimental design, which yields the level 1-training set $\mathcal{E}_1 = \{\mathbf{X}_1, \mathbf{Y}_1\}$, where $\mathbf{X}_1 \in \mathbb{R}^{n_1 \times d}$ is drawn from π_{Θ} and $\mathbf{Y}_1 = \mathcal{Y}(\mathbf{X}_1) \in \mathbb{R}^{n_1 \times 1}$ and n_1 is the number of points in \mathcal{E}_1 . The overall number of original model evaluations is thereby limited to n_1 because any subsequent computations, namely the reliability analyses, will be run with the level 1-surrogate. In this level, any kind of surrogate modelling technique can be utilized to run the reliability analysis. However, the quality of the reliability sensitivity estimates mostly depends on the quality of the surrogate and the applied SRM in level 1. Thus, tuning the method in this component will yield the most substantial improvements in estimating the reliability indices. [18] introduced a Kriging-driven Monte-Carlo sampling approach, which enriches the Kriging experimental design according to a learning function that favors large model uncertainty close to the limit-state face $g = 0$ and [37] applied a similar idea to PCEs. A recently introduced surrogate-driven sequential sampling approach for reliability analysis explores the failure domain sequentially and reconstructs a surrogate model at each intermediate step in the sampling procedure [44]. All of these approaches have been shown to substantially improve the unconditional (predictive) reliability estimate. However, using such an active learning-strategy for the level 1-surrogate with a learning criterion based on the unconditional LSF could lead to inaccurate representations of the LSFs conditional on the Θ_B -samples. If a set of Θ_B -samples is drawn ahead of the level 1-surrogate construction, one may consider other, more suitable learning criteria.

Local Approach

For nonlinear models, Ω_{Θ_B} may contain regions in which the conditional probability of failure becomes either very small or very large. Global surrogate methods may fail to reconstruct the model accurately in such regions and thereby introduce an error in the estimator of the proposed sensitivity indices. In such case, one may instead use cheap local surrogates which are reconstructed at each Θ_B -sample. At the i -th sample $\theta_B^{(i)}$, the local surrogate $\hat{\mathcal{Y}}|_{\theta_B^{(i)}}(\Theta_A)$ is constructed based on a local training set $\mathcal{E}_1^{(i)} = \{\mathbf{X}_1^{(i)}, \mathbf{Y}_1^{(i)}\}$, where $\mathbf{X}_1^{(i)} \sim \pi_{\Theta_A}$ and $\mathbf{Y}_1^{(i)} = \mathcal{Y}(\mathbf{X}_1^{(i)}, \theta_B^{(i)})$. The local approach is particularly suitable if $d_B = \dim(\Theta_B)$ is large ($d_B/d \rightarrow 1$), i.e., most variables are reducible, since then, the resulting conditional reliability problems are low-dimensional (they are posed on Ω_{Θ_A} , which has dimension $d_A = \dim(\Theta_A) = d - d_B$). In such case, the local surrogates depend on a low-dimensional input such that they can be computed using small experimental designs. Within the local approach, standard active-learning approaches can be used at each Θ_B -sample. A more detailed discussion of computational cost depending on the variable splitting in Θ_A and Θ_B is given in Subsection 8.4.4.

Effective surrogate model dimension for input mixtures

When considering both random inputs and uncertain parameters of these inputs at the same time, one can reduce the effective dimension of the experimental design over which to construct the level 1-surrogate. Consider a random vector Θ , where $\Theta_2 = M$, $\Theta_3 = \Sigma$ are uncertain parameters of Θ_1 which is a random variable. The probabilistic model of Θ_1 , through the conditional CDF $F_{\Theta_1|M,\Sigma}$, establishes an exact relationship between Θ_1 , M and Σ . If $F_{\Theta_1|M,\Sigma}$ is invertible in Θ_1 , the following hierarchical sampling strategy yields a sample from Θ_1 based on a sample $\mathbf{u} = [u_1, u_2, u_3]^T$ from the standard-uniform distribution.

$$\begin{aligned} m &= F_M^{-1}(u_1), \\ \sigma &= F_\Sigma^{-1}(u_2), \\ \theta_1 &= F_{\Theta_1|M,\Sigma}^{-1}(u_3|m, \sigma) = F_{\Theta_1|M,\Sigma}^{-1}(u_3|F_M^{-1}(u_1), F_\Sigma^{-1}(u_2)). \end{aligned} \quad (8.26)$$

Since the surrogate is constructed in standard-normal space, an isoprobabilistic transformation is used to compute the corresponding standard-normal sample v_1 :

$$v_1 = \Phi^{-1}[F_{\Theta_1}(\theta_1)], \quad (8.27)$$

where $\Phi(\cdot)$ is the standard-normal CDF. Eqs. (8.27) and (8.26) facilitate the construction of a surrogate model for the marginalized input space, which does not contain M and Σ anymore. This can also be understood as eliminating M and Σ by constructing the predictive distribution of Θ_1 . In this way, the surrogate accuracy is improved through replacing what would have been an approximation of the interaction amongst Θ_1 , M and Σ with their exact relationship. Instead of marginalizing the input space, one may also work in the higher-dimensional space and obtain larger experimental designs at no additional computational cost by sampling $[\Theta_1, M, \Sigma]$ from iso- Θ_1 -surfaces, as done in [53]. Then, the PCE will still approximate the exact relationship amongst Θ_1 , M and Σ , though at considerably better accuracy through the larger training set. The predictive CDF of Θ_1 , F_{Θ_1} , can be computed via numerical integration, e.g., Monte-Carlo integration:

$$\begin{aligned} F_{\Theta_1}(\theta_1) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{\Theta_1|M,\Sigma}(\theta_1|m, \sigma) \pi_M(m) \pi_\Sigma(\sigma) dm d\sigma \\ &= \mathbb{E}[F_{\Theta_1|M,\Sigma}(\theta_1|M, \Sigma)] \\ &\approx \frac{1}{N} \sum_{i=1}^N F_{\Theta_1|M,\Sigma}(\theta_1|m^i, \sigma^i), \quad m^i \sim \pi_M, \quad \sigma^i \sim \pi_\Sigma. \end{aligned} \quad (8.28)$$

8.4.2 Level 2

The level 2-surrogate $\widehat{\log P_F}$ is based on the training set $\mathcal{E}_2 = \{\mathbf{X}_2, \mathbf{Y}_2\}$, where $\mathbf{X}_2 \in \mathbb{R}^{n_2 \times d_B}$ is drawn from π_{Θ_B} , $\mathbf{Y}_2 = \log P_F(\mathbf{X}_2) \in \mathbb{R}^{n_2 \times 1}$ and n_2 is the number of points in \mathcal{E}_2 . In level 2, we focus on polynomial basis surrogates, such that no additional (not even surrogate-model-based) sampling is required to obtain the reliability sensitivities (Section 8.3). The level 2-surrogate model additionally provides a cheap and approximate method for updating probability-of-failure-densities. That is, upon obtaining posterior Θ_B -samples through Bayesian updating, the associated posterior failure

density (and its mean, the predictive reliability estimate) can be computed through the analytical function $\widehat{\log P_F}(\Theta_B)$ rather than solving a reliability problem for each of the posterior samples. The polynomial basis in $\widehat{\log P_F}$ is orthogonal with respect to the prior input joint density. Thus, coefficient-based postprocessing (moments, sensitivities) of the posterior conditional probability of failure, should be handled with great care.

8.4.3 The Framework

The analysis proceeds in the following way:

1. Elicit a variable subset of interest Θ_B . Obtain n_2 Θ_B -samples, $\{\theta_B^{(i)}\}_{1 \leq i \leq n_2}$ (e.g., based on latin hypercube sampling or a low-discrepancy sequence).
2. In case Θ_B comprises a mixture of inputs and uncertain parameters thereof, marginalize the input space according to Eqs. (8.27), (8.26) and (8.28).
3. Select the global or local strategy for the level 1-surrogate.

a) Global

Sample n_1 Θ -samples and evaluate the model \mathcal{Y} at these samples. Based on these, build the global level 1-surrogate over Ω_{Θ} , $\widehat{\mathcal{Y}}$.

For each Θ_B -sample, use a structural reliability method and the global level 1-surrogate-based limit-state function \widehat{g} to compute

$$\widehat{P}_F^{(i)} = \mathbb{E}_{\Theta_A} [\mathbb{I}(\widehat{g}(\Theta_A, \theta_B^{(i)}) \leq 0) | \theta_B^{(i)}] = \int_{\Omega_{\Theta_A}} \mathbb{I}(\widehat{g}(\theta_A, \theta_B^{(i)}) \leq 0) \pi_{\Theta_A | \Theta_B}(\theta_A | \theta_B^{(i)}) d\theta_A.$$

b) Local

For each Θ_B -sample, sample n_1 Θ_A -samples and evaluate the model \mathcal{Y} at these samples. Based on these, build the local level 1-surrogate over Ω_{Θ_A} conditional on $\theta_B^{(i)}$, $\widehat{\mathcal{Y}} | \theta_B^{(i)}$ and its associated local limit-state function $\widehat{g}_i(\theta_A)$ to compute $\widehat{\mathcal{Y}} | \theta_B^{(i)}$

$$\widehat{P}_F^{(i)} = \mathbb{E}_{\Theta_A} [\mathbb{I}(\widehat{g}_i(\theta_A) \leq 0)] = \int_{\Omega_{\Theta_A}} \mathbb{I}(\widehat{g}_i(\theta_A) \leq 0) \pi_{\Theta_A | \Theta_B}(\theta_A | \theta_B^{(i)}) d\theta_A.$$

4. From the set $\{\theta_B^{(i)}, \log(\widehat{P}_F^{(i)})\}_{i=1, \dots, n_2}$, build the level 2-surrogate $\widehat{\log P_F}(\Theta_B)$.
5. Obtain variance-based sensitivity indices of $\widehat{\log P_F}$ by means of the model coefficients.

The procedure outlined above is also sketched in Fig. 8.2 (global) & Fig. 8.3 (local).

8.4.4 Computational cost

Fig. 8.4 shows the behaviour of computational cost in both levels when using either a global or a local or no surrogate strategy at all in level 1. The computational cost is measured in terms of

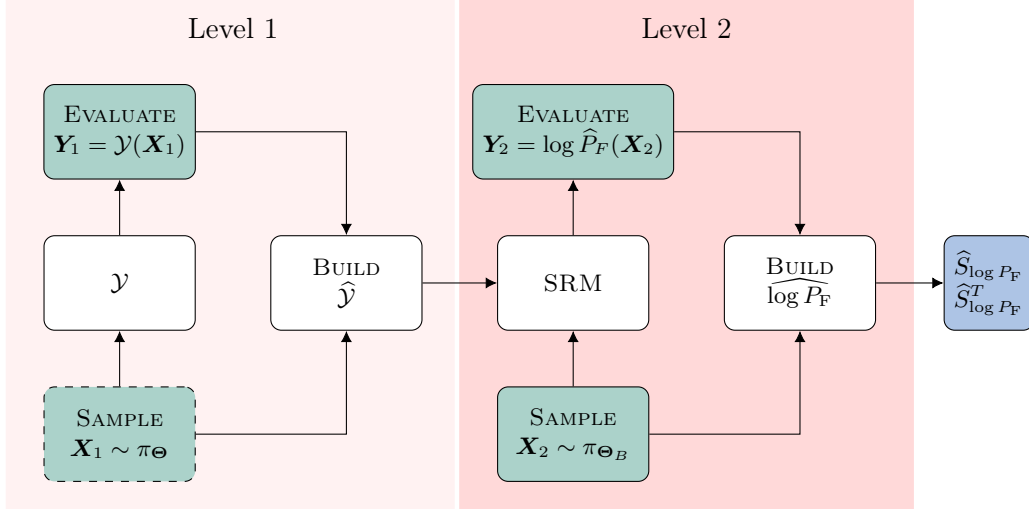


Figure 8.2: Flow diagram of 2-level surrogate-based conditional reliability sensitivities with global surrogate in level 1, where the dashed boundary marks the starting point.

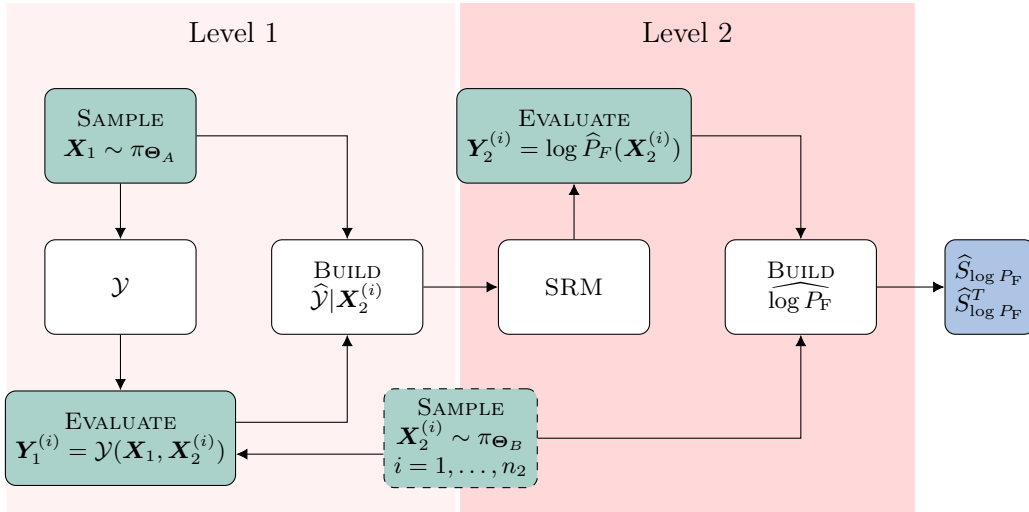


Figure 8.3: Flow diagram of 2-level surrogate-based conditional reliability sensitivities with local surrogates in level 1, where the dashed boundary marks the starting point.

number of \mathcal{Y} -evaluations. To proceed, we make the assumption that the number of samples required to learn a surrogate model in input dimension d is $n = 2 \cdot (d + 1)$. Note that this is a rather crude assumption and is accurate only for mildly nonlinear models \mathcal{Y} . The number of samples required in level 2, $n_2 = 2(d_B + 1)$, is the same in all three versions, though for the global surrogate model, this does not influence the computational cost. At fixed model dimension d , the global surrogate modelling costs are $n_{1,global} = n_{global} = 2(d + 1)$, i.e., they are constant, depend only on level 1 and the total dimension and not on d_A or d_B . For the local surrogate strategy, the cost in level 1 is given by $n_{1,local} = 2(d_A + 1)$ per Θ_B -sample. The total cost thus reads $n_{local} = n_{1,local} \cdot n_2$. When using no level 1-surrogate at all, n_2 reliability problems are solved using the original model \mathcal{Y} . In order to estimate the computational cost in the surrogate-free approach, the number of required \mathcal{Y} -evaluations to solve a d_A -dimensional reliability problem, $n_{1,\mathcal{Y}}$, has to be estimated. Therein, d_A influences which method should be used to solve the problem. Tab. 8.1 provides a qualitative rela-

tionship between the dimensionality and computational cost of a d_A -dimensional reliability problem ($n_{1,\mathcal{Y}}$ in Fig. 8.4). The total cost in this approach is computed as $n_{\mathcal{Y}} = n_{1,\mathcal{Y}} \cdot n_2$.

Table 8.1: Exemplatory cost of reliability analyses at different d_A using different suited SRM.

Problem Dimension	SRM	no. of required samples N
1	Bisection	4
2	FORM	10
10	Importance Sampling	100
100	Subset Simulation	10000

Fig. 8.4 (right-hand side) shows, that both the local surrogate and surrogate-free strategy yield the highest overall computational cost when d_B and d_A have similar size, i.e., neither reducible nor irreducible uncertainties dominate the model input. This is due to two counter-acting effects: on the one hand, as the effective local surrogate model dimension equals d_A , the local strategy requires less samples per surrogate when d_A is small. Similarly, if not using a surrogate at all in level 1, the reliability analysis performed with the original model will require less model evaluations due to the decreasing d_A . On the other hand, decreasing d_A implies an increasing reducible space dimension d_B . Thus, in order to build accurate level 2-surrogates, n_2 needs to increase, which implies a multiplication of the overall number of required local surrogates/original-model-based reliability analyses. Here, the global surrogate model seems to be the most efficient. This, however, is only true when \mathcal{Y} is mildly nonlinear. For models that exhibit stronger nonlinear behaviour, preserving the accuracy of the global model may require more model evaluations than the local strategy. The required surrogate and SRM accuracies increase as the predictive failure probability decreases. Analyses with predictive failure probabilities $\leq 10^{-7}$ (potentially with conditional failure probabilities several orders of magnitude below the predictive level) will benefit from failure-oriented enrichment in the surrogate construction and/or choosing a suitable SRM. In many application cases, the local strategy could be operated efficiently as the number of reducible inputs is considerably lower than the number of irreducible inputs $d_B \ll d_A$.

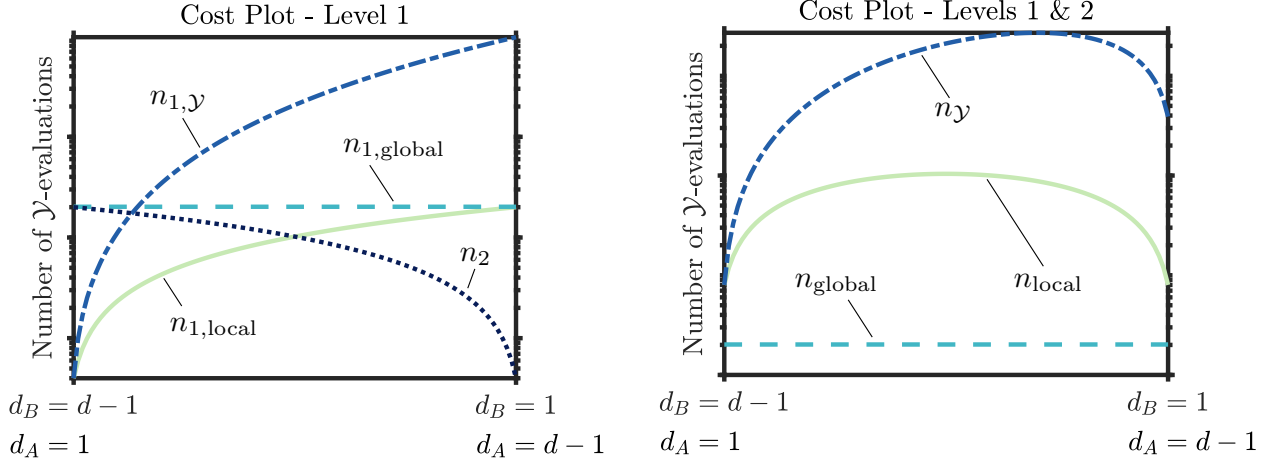


Figure 8.4: Comparison of global (*global*) vs. local (*local*) vs. no (\mathcal{Y}) surrogate strategy in level 1 in terms of computational cost (number of \mathcal{Y} -calls) over reducible space dimension d_B . Left: cost per level, right: total cost.

8.5 Numerical examples

The novel reliability sensitivity indices are investigated and demonstrated through two applications, namely an elastic truss of moderate dimensionality (12 dimensions) and a monopile foundation in plastic soil involving a random field model (87 dimensions). Both examples feature both independent input variables as well as uncertain parameters thereof. The level 1-surrogate is constructed with LRAs which yield consistently smaller global and conditional (on failure samples) global errors. This is in accordance with the findings of [30].

8.5.1 Elastic truss

The truss structure (Fig. 8.5) consists of 13 rods, where horizontal and diagonal rods have log-normally distributed cross-sections A_1 , A_2 and Young's moduli E_1 , E_2 , respectively. The truss sustains 6 vertical point loads $P_1 - P_6$ which are modelled as Gumbel-distributed [29, 32]. [20] presents results for an analysis of the original elastic truss with the proposed sensitivity framework. Here an extension of the truss model featuring hyperparameters is discussed. Namely, the parameters a_P , b_P of the load Gumbel distribution are assumed uncertain and log-normally distributed. The mean and coefficient of variation of the load Gumbel distribution in Tab. 8.2 are conditional on the parameters a_P (location parameter) and b_P (scale parameter):

$$\mu_{P|a_P,b_P} = \mu_{a_P} + \gamma_E \mu_{b_P} \quad (8.29)$$

$$\delta_{P|a_P,b_P} = \frac{\pi}{\sqrt{6}} \frac{\mu_{b_P}}{\mu_{P|a_P,b_P}}, \quad (8.30)$$

where γ_E is Euler's constant. It is further assumed that all point loads share the same distribution parameters. Tab. 8.2 summarizes the probabilistic input models.

The limit-state function is defined by means of a threshold for the maximum vertical truss deflection, i.e.,

$$g(\Theta) = u_{lim} - u_{max}(\Theta),$$

where $u_{lim} = \{14\text{cm}, 16\text{cm}, 18\text{cm}\}$ are considered, which correspond to system failure probabilities of $\mathbb{P}(F) = \{6.2 \cdot 10^{-3}, 1 \cdot 10^{-3}, 1.7 \cdot 10^{-4}\}$ (estimated with Monte Carlo, estimator coefficient of variation $\leq 5\%$). In view of the Bayesian interpretation of the variable sets Θ_A (irreducible) and Θ_B (data

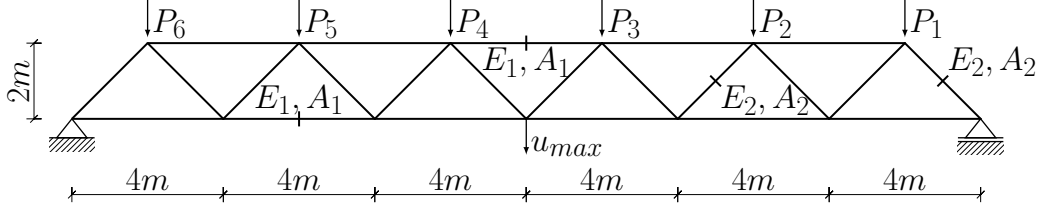


Figure 8.5: 2-D truss example.

available, reducible), we choose the latter to comprise quantities which are typically reducible by means of measurements (such as material properties) or archive data (such as load hyperparameters). Thus, we select $\Theta_B = [E_1, A_1, E_2, A_2, a_P, b_P]$ and $\Theta_A = [P_1, \dots, P_6]$. For this example, the global

Table 8.2: Input variable definitions of the elastic truss.

Quantity	Distribution	Mean μ	CoV δ
A_1 [m^2]	Log-Normal	$2 \cdot 10^{-3}$	0.1
A_2 [m^2]	Log-Normal	$1 \cdot 10^{-3}$	0.1
E_1, E_2 [Pa]	Log-Normal	$2.1 \cdot 10^{11}$	0.1
$P_1 - P_6$ [N]	Gumbel	$\mu_{P a_P, b_P}$	$\delta_{P a_P, b_P}$
a_P [N]	Log-Normal	46624	0.2
b_P [N]	Log-Normal	3375	0.2

surrogate modelling strategy is employed, as the truss behaves only mildly nonlinear. Both level 1- and level 2-experimental designs are obtained via latin hypercube sampling. $n_1 = 200$ samples are used to construct the level 1-LRA, while $n_2 = 1000$ points are used to evaluate the level 2-surrogate. The analysis is repeated 20 times redrawing random level 1-experimental designs which yields the estimator statistics provided in Figs. 8.6 and 8.7. These are computed for the difference of the surrogate-based estimator from the direct Monte-Carlo (DMC) reference solution, i.e.,

$$\epsilon_Q = Q - Q_{DMC}. \quad (8.31)$$

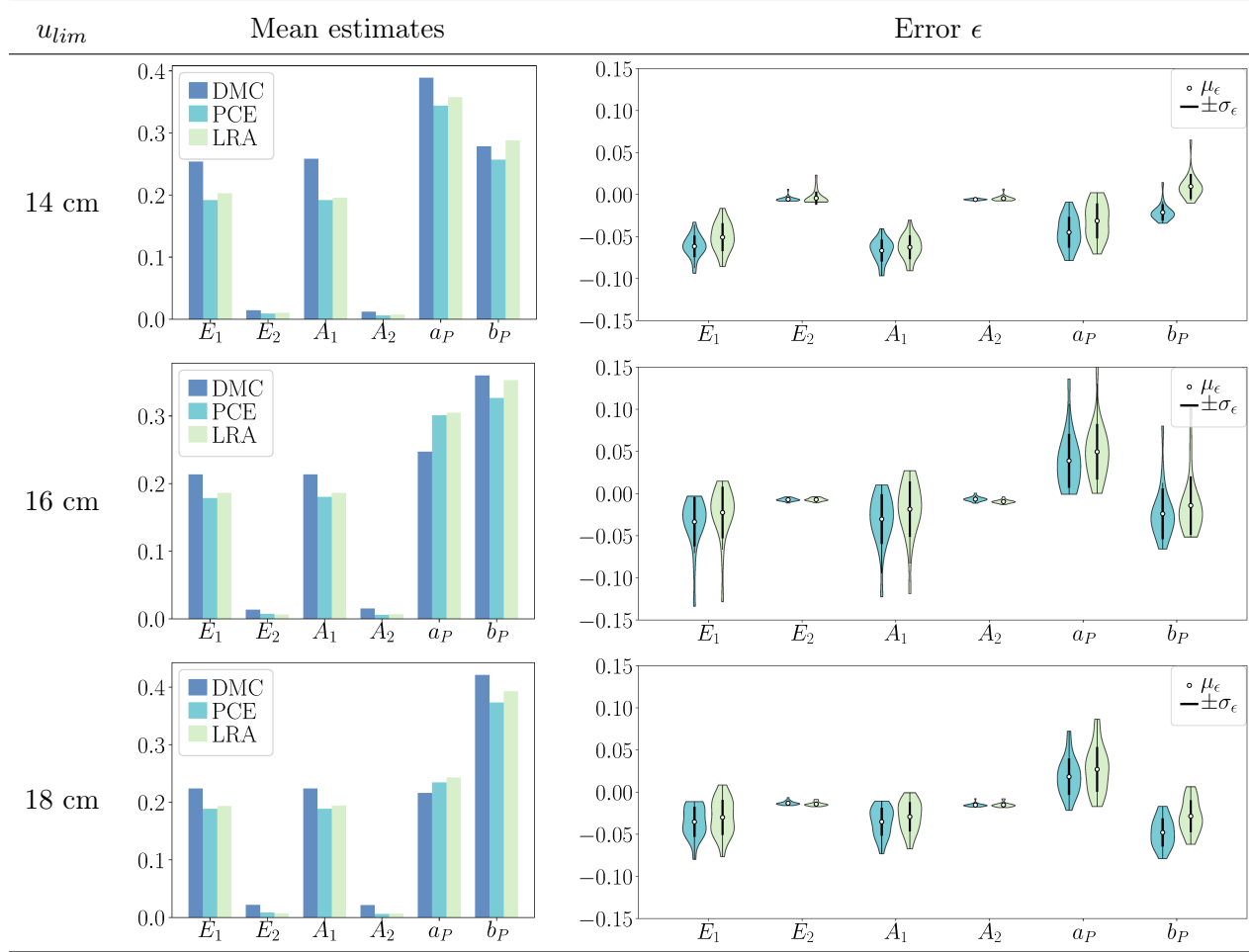
All reliability analyses have been performed with the first-order reliability method (FORM). For the reference solution, $n_{DMC} = 2 \cdot 10^5$ samples have been used implying the solution of $8 \cdot 10^5$ reliability problems of dimension 6 (since the type B -variable space has dimension 6, see Subsection 8.2.3).

While both PCE and LRA-based approaches capture the variable importance ranking correctly, the LRA-based approach performs slightly but consistently better in the mean compared to the PCE-based approach. The least important variables E_2, A_2 are estimated with the smallest error



Figure 8.6: $\log P_F$ first-order Sobol' indices: mean estimates and errors ($n_1 = 200, n_2 = 1000$).

mean and variance, which, however, is due to their small true magnitude. Intuitively, estimation accuracy and index magnitude should depend on one another reciprocally; the larger the index the more accurate it is estimated based on a given set of samples. Indeed, relative to their respective magnitudes, the indices of the most important indices a_P and b_P are estimated more accurately.

Figure 8.7: $\log P_F$ Total-effect indices: mean estimates and errors ($n_1 = 200, n_2 = 1000$).

8.5.2 Monopile Foundation

In this example, we consider a finite element model for the interaction of a monopile foundation of an offshore wind turbine (Fig. 8.8) with stiff, plastic soil. Deterministic parameters of the monopile are its depth $L = 30$ m, diameter $D = 6$ m, wall thickness $t = 0.07$ m, Poisson ratio $\nu = 0.3$ and Young's modulus $E = 2.1 \cdot 10^5$ MPa. The uncertain inputs comprise the lateral load H as well as the undrained shear strength s of the soil and hyperparameters of both quantities. The engineering model setup follows [12] and the probabilistic model considered there has been modified following [27]. s is considered both uncertain and increasing in mean with soil depth z . It is thus modelled by a random field with linear mean drift along the soil depth coordinate z . Given an underlying stationary Gaussian random field $\tilde{s}(z, \Theta)$

$$\{\tilde{s}(z) : 0 \leq z \leq L\} \sim \mathcal{N}(0, \sigma_{\tilde{s}}),$$

the non-stationary random field representing the shear strength of the soil can be expressed as

$$\begin{aligned} s(z, \Theta) &= s_0 + s_1 \sigma'(z) \exp\{\tilde{s}(z, \Theta)\} \\ &= s_0 + s_1 \gamma z \exp\{\tilde{s}(z, \Theta)\}, \end{aligned}$$

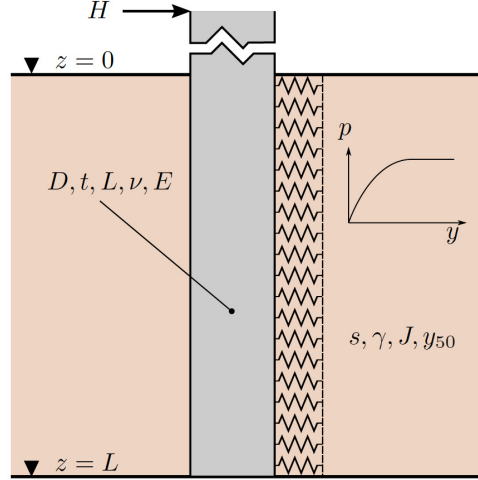


Figure 8.8: Wind turbine monopile foundation [12].

where γ is the soil unit weight, $\sigma'(z) = \gamma z$ is the effective vertical stress, s_0 is the undrained shear strength at ground level and s_1 is the drift parameter governing the mean increase of s with increasing soil depth. $\tilde{s}(z, \Theta)$ models the intra-site variability. That is, at a given site with known deterministic s_0 and s_1 , it describes the inherent variability of the undrained shear strength. In order to describe the inter-site uncertainty in s , the parameters s_0 and s_1 are modeled probabilistically as well. The stationary RF \tilde{s} is taken to be correlated with exponential-type

$$\rho_{\tilde{s}\tilde{s}}(z', z'') = \exp \left\{ -\frac{2|z' - z''|}{\theta_{\tilde{s}}} \right\},$$

with vertical soil scale of fluctuation $\theta_{\tilde{s}} = 1.9m$ [47] and standard deviation $\sigma_{\tilde{s}} = 0.3$ [47, 48]. We assume the soil to be stiff and plastic according to the classification provided in [49]. There, the specific soil weight range is given with $17 - 19kN/m^3$, whence we set $\gamma = 18kN/m^3$. The mean cohesion range is given with $20 - 50kN/m^2$ by [49] while [9] lists the mean range of the undrained shear strength ratio s_u/σ' as $0.23 - 1.4$. We fit log-normal distributions for s_0 and s_1 by setting the 10 % and 90% quantiles of the distributions equal to the lower and upper bounds of these ranges. The resulting parameters are detailed in Tab. 8.3 along with uncertain parameters for the load H , namely μ_H and σ_H . The mean and coefficient of variation of the load Gumbel distribution in Tab. 8.3 are conditional on the parameters a_H (location parameter) and b_H (scale parameter) according to Eqs. (8.29) and (8.30). \tilde{s} is simulated by means of the midpoint method. That is, the spatial domain $[0, L]$ is discretized with n spatial elements and \tilde{s} is represented by means of n random variables with joint distribution $\mathcal{N}(\mathbf{0}, \Sigma_{\tilde{s}\tilde{s}})$. The random variables represent the random field values at the element midpoints. Thus, the covariance matrix $\Sigma_{\tilde{s}\tilde{s}}$ is computed by evaluating $\sigma_{\tilde{s}}^2 \rho_{\tilde{s}\tilde{s}}(z', z'')$ at the element midpoints. The number of elements is chosen such that 95% of the inherent RF variability is captured by the RF discretization, leading to $n = 82$ in this example. Therefore, the total input dimension is $d = 87$. As the surrogate modeling approach requires independent inputs, the midpoint random variables are transformed to independent standard normal random variables, denoted as ξ , by means of the Nataf transform. The model output $Y = \mathcal{Y}(\Theta)$ is the maximum occurring stress in the foundation. The limit-state function is given by

$$g(\Theta) = \sigma_{crit} - \mathcal{Y}(\Theta),$$

where three different stress thresholds $\sigma_{crit} = \{80 \text{ MPa}, 100 \text{ MPa}, 120 \text{ MPa}\}$ are considered with corresponding system failure probabilities $\mathbb{P}(F) = \{3.0 \cdot 10^{-3}, 3.6 \cdot 10^{-4}, 8.0 \cdot 10^{-5}\}$ (estimated with

Table 8.3: Input variable definitions of the monopile foundation.

Input	Distribution	Mean μ	CoV δ
ξ [-]	Standard-Normal	$\mathbf{0}$	<i>n.d.</i> ($\Sigma_{\xi\xi} = \mathbf{I}_{n \times n}$)
s_0 [kPa]	Log-Normal	33.7094	0.3692
s_1 [kPa]	Log-Normal	0.7274	0.8019
H [kN]	Gumbel	$\mu_{H a_H, b_H}$	$\delta_{H a_H, b_H}$
a_H [kN]	Log-Normal	2274.97	0.2
b_H [kN]	Log-Normal	225.02	0.2

MC and $\text{CoV} = \{0.0409, 0.1187, 0.2500\}$). For this example, we consider a scenario in which uncertainties about all the input hyperparameters as well as the inherent variability of the shear strength can be reduced through additional data, i.e., $\Theta_B = [\xi, a_H, b_H, s_0, s_1]$. This leaves the inherent load variability as the only remaining aleatory input, i.e., $\Theta_A = H$. This example has proven extremely challenging for common global surrogate models (polynomial basis surrogates, adaptive kriging surrogates). Therefore, the local surrogate modelling strategy is chosen, which, in this case, is an efficient choice since the limit-state function to be approximated at each Θ_B -sample is one-dimensional. As little as four training points are necessary per Θ_B -sample. Based on the local surrogate, the corresponding one-dimensional reliability problem can be solved using a bisection- or Newton-procedure due to the monotonicity of the limit-state-function the same SRM is applied for the reference solution with the original model \mathcal{Y}). Accurate estimates of the conditional reliability sensitivities are achieved with an overall 2000 evaluations of the limit-state-function for the second example with the introduced framework. This is a conservative choice and satisfying accuracy may be achieved with a considerably lower number of model evaluations as is evident from Figs. 8.11 and 8.12.

The algorithm is run 20 times redrawing level 2-experimental designs to compute errors ϵ arising from exploring Ω_{Θ_B} randomly. These errors are measured against a DMC-based reference solution according to Eq. (8.31). The reference solution is based on $n_s = 20000$ independent samples which yields a total of $1.64 \cdot 10^6$ reliability problems since $d_B = 86$. A single reference solution (fixed σ_{crit}) obtained with a Matlab implementation took approximately 48 hours to complete on 10 Intel Xeon E5-2697 v3 14-core nodes, which emphasizes the need for surrogate modelling-based estimators. In order to assess the random field influence, the sum of all first-order indices of the elements of ξ are reported for both the reference solution and the surrogate-based approach. This is motivated by the impossibility of efficiently estimating a sensitivity index of order 86 with the MC-methods of [26, 52]. However, the surrogate-based estimators indicate negligible interaction of the elements of ξ (all Sobol' and total-effect indices are virtually identical) thus justifying this approximation.

The sensitivities depicted in Figs. 8.9 and 8.10 summarize the effect of the inherent random field variability in the order- n -indices of the random vector ξ . The PCE and LRA level 2-surrogates yield similar estimates for the reliability sensitivities (Figs. 8.9 and 8.10): The load dispersion parameter b_H is identified as the most influential input to the monopile reliability analysis across all investigated scenarios. The dominance of b_H becomes more pronounced at higher critical stress levels. PCE-based sensitivity estimate means are in slightly better agreement with the reference

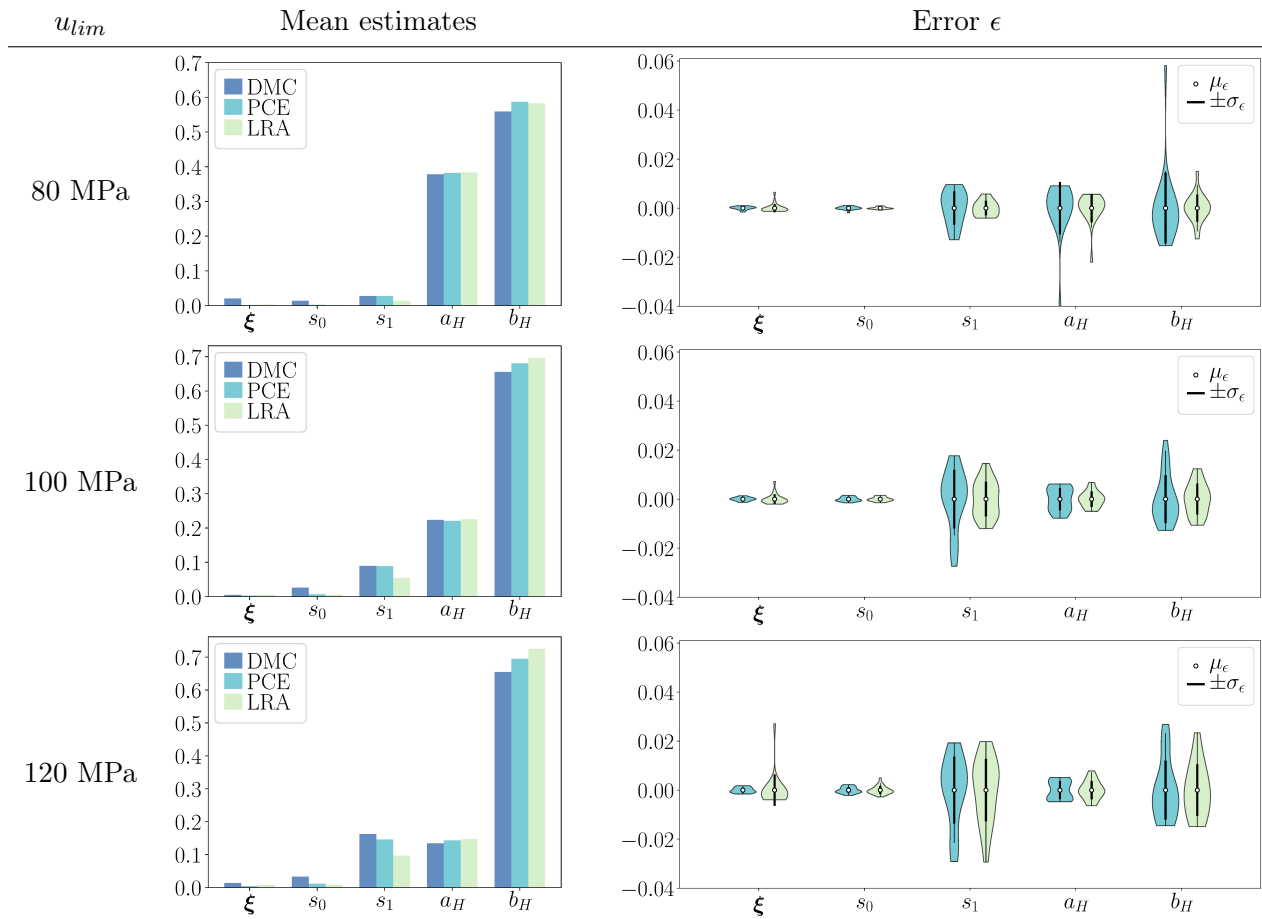


Figure 8.9: $\log P_F$ first-order Sobol' indices: mean estimates and errors ($n_1 = 4$ (local), $n_2 = 2000$).

solution than LRA-based estimates. Moreover, PCE-based estimates exhibit consistently smaller variability resulting from the random level 2-experimental design. Error means and variabilities are of comparable magnitude for all computed indices in this example. However, as discussed for the truss example, indices of larger magnitude should be estimated more accurately at a given amount of information. This is true when considering the estimation error relative to the index magnitudes. Then, estimates of the most influential variables a_H and b_H exhibit comparably small variability due to the random level 2-design of experiments while the less important variables' estimators prove more sensitive in this respect. In Figs. 8.11 and 8.12, the evolution of the error means and variances at different level 2-experimental design sizes n_2 are depicted. The error converges in mean and standard deviation as n_2 increases. The error mean apparently is bounded from below which is likely due to a bias in the reference solution which arises as the MC-estimates are not fully converged at $n_c = 20000$ independent samples. Ultimately, if a decision-maker were to choose whether to acquire data on either of the uncertain model inputs and on which input in particular, they should choose b_H according to the introduced reliability sensitivity framework.

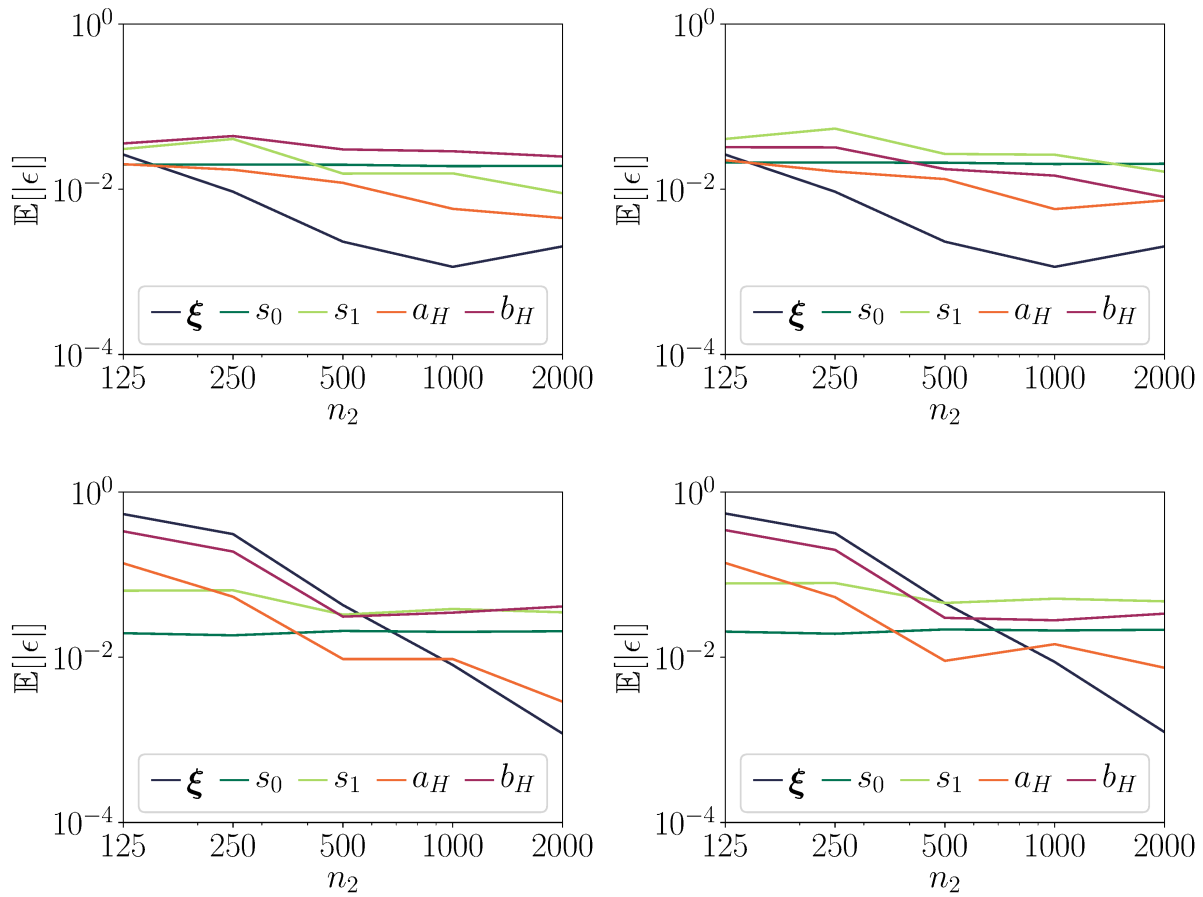


Figure 8.11: Sobol' (left column) and total-effect indices (right column) error mean vs. level 2 experimental design size n_2 for PCE- (upper row) and LRA-based (lower row) at $\sigma_{crit} = 100$ MPa.

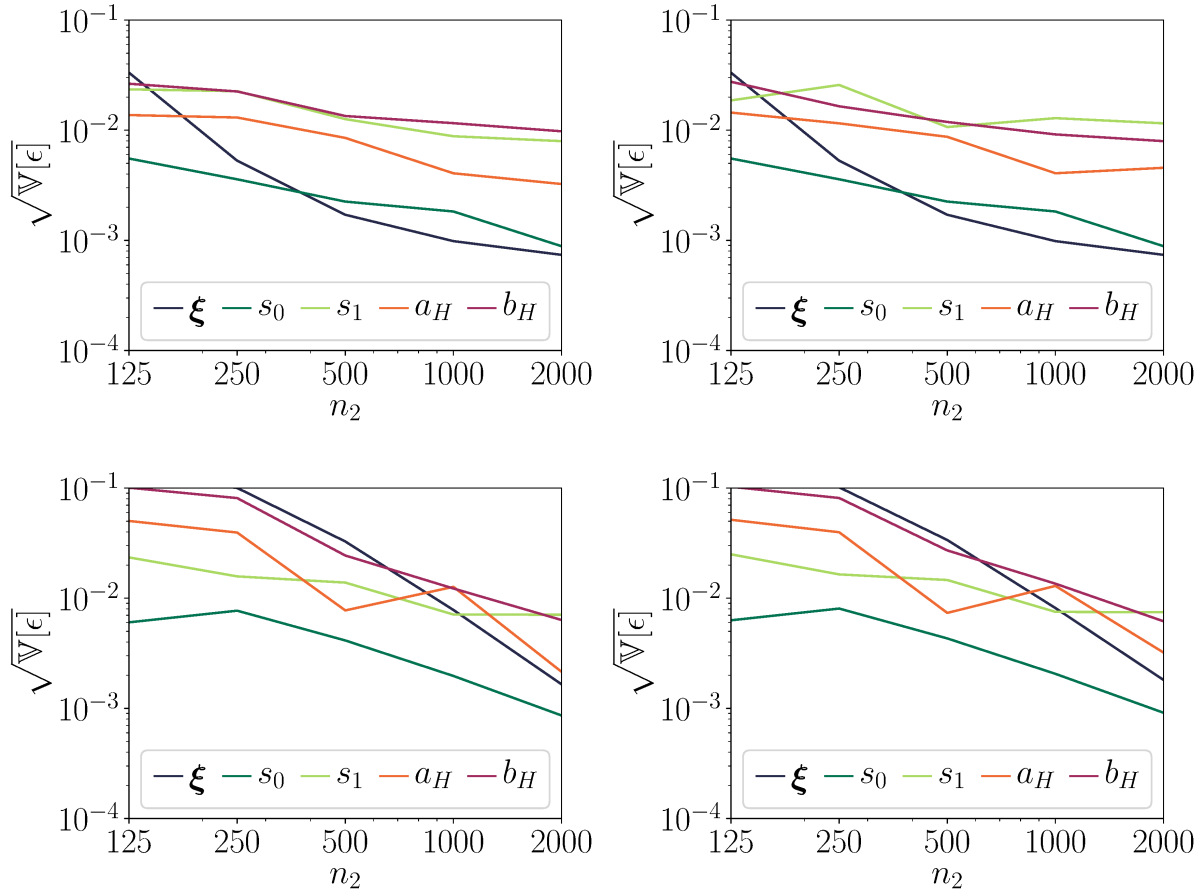


Figure 8.12: Sobol' (left column) and total-effect indices (right column) error standard deviation vs. level 2 experimental design size n_2 for PCE- (upper row) and LRA-based (lower row) at $\sigma_{crit} = 100$ MPa.

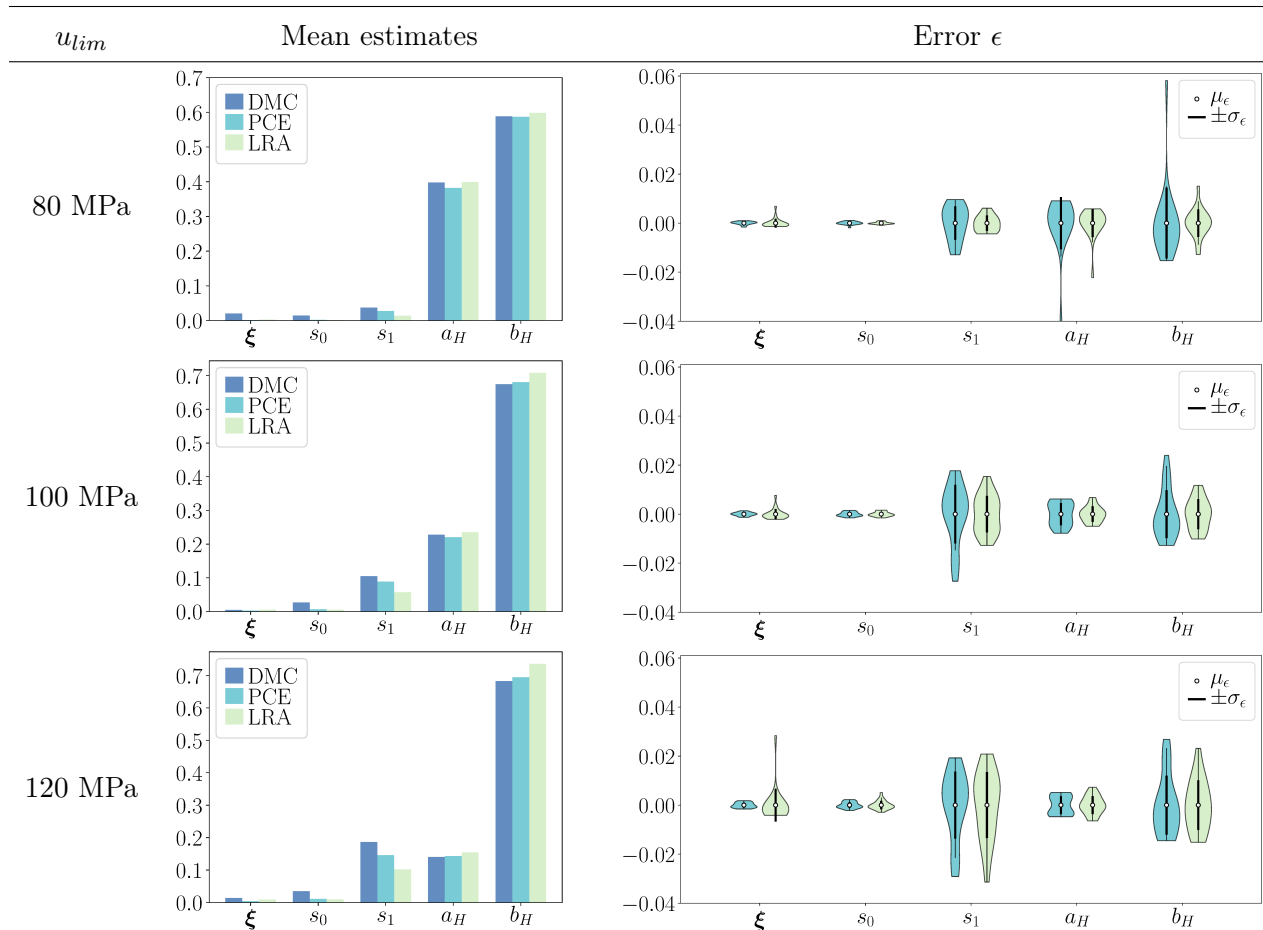


Figure 8.10: $\log P_F$ Total-effect Sobol' indices: mean estimates and errors ($n_1 = 4$ (local), $n_2 = 2000$).

8.6 Concluding Remarks

In this paper we describe a framework for reliability sensitivity analysis whose core is a novel, variance-based sensitivity index tailored to reliability analysis in the presence of multi-uncertainty. Multi-uncertainty refers to a separation of uncertain model inputs in different categories, e.g., aleatory and epistemic. The interest then lies in expressing the sensitivity of the probability of failure-estimate to the epistemic inputs. We devise a flexible two-level surrogate modelling approach which allows for a cost-efficient estimation of the proposed index. The approach relies on either building a global surrogate model over the entirety of input variables once and for all (global strategy) or on repeatedly recomputing cheaper, lower-dimensional surrogates (local strategy). An analysis of computational cost for both local and global surrogate modelling strategy has been carried out in dependence on how the model input is divided in reducible and irreducible.

We demonstrate the novel approach by means of two examples. In the elastic truss example, the sensitivity indices were estimated using a global low-rank approximation in level 1 at a total cost of 200 original model evaluations. The LRAs are more suitable for global surrogate-driven reliability computations (level 1) compared to the PCEs, as discussed in [30]. For the geotechnical example, namely a monopile foundation, a local surrogate modelling approach based on PCEs has been

adopted in level 1 yielding a total cost of 8000 original model evaluations. In level 2, the sparse PCEs perform slightly better than the LRA. Sensitivity estimates for both examples were validated with sampling-based reference estimates showing that the important input's indices are estimated accurately by the method. While in this paper only two possible choices for the level 1 surrogate are discussed, virtually any surrogate modelling technique could be applied here.

8.6.1 Discussion

The contribution of the proposed sensitivity index is threefold. First, it focusses the sensitivity analysis on a variable subset that is of interest in the presence of multi-uncertainty and data assimilation applications. Second, it represents a direct sensitivity measure for the probability of failure magnitude as opposed to indices based on the indicator function that rank influence on the failure hypersurface shape. Finally, the new index facilitates the entirely surrogate-driven computation of reliability sensitivities by smoothening the indicator function discontinuity through an integral formulation.

In accordance with intuition, estimators of sensitivity indices with small magnitude exhibit smaller errors compared to those of large magnitude. Relative to the magnitude, the opposite is true: at fixed amount of information (here: the level 1-training set size), the sensitivity indices with larger magnitude are estimated more reliably. Moreover, the variable importance ranking is accurately captured in all numerical examples.

The problem dimension, which can be handled by the approach is guided by the surrogate modelling techniques chosen in both levels. Using an arbitrary surrogate model that is capable of addressing high-dimensional problems in level 1 and LRAs or the recently introduced PLS-driven PCEs [45] in level 2, it is applicable up to several thousand input variables.

At a fixed accuracy (with respect to \mathcal{Y} -model output), the cost of the global surrogate modelling strategy remains constant irrespective of the fraction of reducible and irreducible uncertainty in the model input. The same is not true for the local strategy where two counteracting effects lead to the local strategy being most efficient if either of the uncertainty types dominates the model input. The local modelling strategy is more flexible and expected to perform well even for strongly nonlinear models, when the global modelling strategy may deteriorate. In such case, the required number of samples for a sufficiently accurate global surrogate model may overcompensate the strategy-based savings the global modelling strategy offers and render the local strategy the better choice. Therefore, the model under consideration should govern the decision which of either approach is used. Generally, we recommend to apply the local surrogate modelling strategy when the model under consideration exhibits significant nonlinear behaviour.

8.6.2 Outlook

The experimental design for the level 1-surrogate dominates the computational cost of the approach. The accuracy of the overall method is controlled by the capability of the level 1-surrogate to capture the tail behaviour of the model response and that of the level 2-surrogate to capture the second moment of its derived response (the conditional probability of failure). Thus, the level 1-surrogate represents the most crucial element of the framework both with respect to accuracy and computational cost. Next to the choice of surrogate, the level 1-performance would likely benefit from a more

guided selection of experimental design points in level 1, which addresses the method's requirement of surrogates to remain accurate close to $g = 0$ rather than minimize global error measures.

Moreover, the introduced sensitivity measure may be connected to decision-oriented sensitivity analysis and in particular the concept of expected value of partial perfect information (EVPPI). It is known that the classical first-order variance-based sensitivity measure coincides with the EVPPI for a special decision context (with a quadratic loss function) [41]. Based on this, the proposed framework may be adapted to facilitate computation of more relevant loss functions associated within decision analysis, which in turn promotes its applicability for decision support.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the priority programme SPP 1886 by grant STR 1140/6-1.

Appendices

8.A Connecting the proposed Sobol' index to its indicator function-based counterpart

Here, we discuss the connection of the novel sensitivity indices to the the measures proposed by [35]. In particular, we show that the Sobol' index of the conditional probability of failure (without logarithmic transformation) is identical to the Sobol' index of the indicator function $I(g \leq 0)$ up to a normalizing constant. In this case, the QOI reads

$$Q = P_F(\boldsymbol{\theta}_B) = \mathbb{E}_{\boldsymbol{\Theta}_A} [\mathbb{I}(g(\boldsymbol{\Theta}_A, \boldsymbol{\Theta}_B) \leq 0) | \boldsymbol{\Theta}_B = \boldsymbol{\theta}_B],$$

while in [35], $Q = \mathbb{I}(g(\boldsymbol{\Theta}) \leq 0)$ is used. Consider now an arbitrary subset of $\boldsymbol{\Theta}_B$ which is denoted by $\boldsymbol{\Theta}_B$ and its complement $\boldsymbol{\Theta}_{\sim B}$ such that $\boldsymbol{\Theta}_B = \{\boldsymbol{\Theta}_B, \boldsymbol{\Theta}_{\sim B}\}$. The Sobol' index of $Q = P_F$ for the variable subset $\boldsymbol{\Theta}_B$ is given by

$$S_{P_F, B} = \frac{\mathbb{V}_{\boldsymbol{\Theta}_B} [\mathbb{E}_{\boldsymbol{\Theta}_{\sim B}} [\mathbb{E}_{\boldsymbol{\Theta}_A} [\mathbb{I}(g \leq 0) | \boldsymbol{\Theta}_B] | \boldsymbol{\Theta}_B]]}{\mathbb{V}_{\boldsymbol{\Theta}_B} [\mathbb{E}_{\boldsymbol{\Theta}_A} [\mathbb{I}(g \leq 0) | \boldsymbol{\Theta}_B]]},$$

while the Sobol' for $Q = \mathbb{I}(g(\boldsymbol{\Theta}) \leq 0)$ reads

$$S_{\mathbb{I}(g \leq 0), B} = \frac{\mathbb{V}_{\boldsymbol{\Theta}_B} [\mathbb{E}_{\boldsymbol{\Theta}_A, \sim B} [\mathbb{I}(g \leq 0) | \boldsymbol{\Theta}_B]]}{\mathbb{V}_{\boldsymbol{\Theta}} [\mathbb{I}(g \leq 0)]}.$$

Here, $\Theta_{A,\sim B}$ denotes the union of Θ_A and $\Theta_{\sim B}$. Then, we have

$$\begin{aligned}
 S_{P_F, B} &= \frac{1}{\mathbb{V}[P_F]} \mathbb{V}_{\Theta_B} \left[\int_{\Omega_{\Theta_{\sim B}}} \pi(\boldsymbol{\theta}_{\sim B} | \Theta_B) \int_{\Omega_{\Theta_A}} \mathbb{I}(g(\boldsymbol{\theta}_A, \boldsymbol{\theta}_{\sim B}, \Theta_B) \leq 0) \pi(\boldsymbol{\theta}_A | \boldsymbol{\theta}_{\sim B}, \Theta_B) d\boldsymbol{\theta}_A d\boldsymbol{\theta}_{\sim B} \right] \\
 &= \frac{1}{\mathbb{V}[P_F]} \mathbb{V}_{\Theta_B} \left[\int_{\Omega_{\Theta_{A,\sim B}}} \mathbb{I}(g(\boldsymbol{\theta}_{A,\sim B}, \Theta_B) \leq 0) \pi(\boldsymbol{\theta}_{A,\sim B} | \Theta_B) d\boldsymbol{\theta}_{A,\sim B} \right] \\
 &= \frac{\mathbb{V}[I(g \leq 0)]}{\mathbb{V}[P_F]} \frac{\mathbb{V}_{\Theta_B} [\mathbb{E}_{\Theta_{A,\sim B}} [I(g \leq 0) | \Theta_B]]}{\mathbb{V}_{\Theta} [I(g \leq 0)]} \\
 &= \frac{\mathbb{V}[I(g \leq 0)]}{\mathbb{V}[P_F]} S_{I(g \leq 0), B}.
 \end{aligned}$$

The rescaling constant connecting the multi-uncertainty Sobol' index to its indicator function-based counterpart is the ratio of variance fractions contributed by all the input variables and the type B-variables only. This result is somewhat intuitive as the multi-uncertainty Sobol' index is defined with respect to the variance contributed by the type B-variables only.

References

- [1] S.-K. Au and J. L. Beck. "Estimation of small failure probabilities in high dimensions by subset simulation". In: *Probabilistic Engineering Mechanics* 16.4 (2001), pp. 263–277.
- [2] S. Au. "Reliability-based design sensitivity by efficient simulation". In: *Computers & structures* 83.14 (2005), pp. 1048–1061.
- [3] P. Beaurepaire et al. "Reliability-based optimization using bridge importance sampling". In: *Probabilistic Engineering Mechanics* 34 (2013), pp. 48–57.
- [4] M. Berveiller, B. Sudret, and M. Lemaire. "Stochastic finite element: a non intrusive approach by regression". In: *European Journal of Computational Mechanics* 15.1-3 (2006), pp. 81–92.
- [5] G. Blatman and B. Sudret. "Adaptive sparse polynomial chaos expansion based on least-angle regression". In: *Journal of Computational Physics* 230.6 (2011), pp. 2345–2367.
- [6] E. Borgonovo. "A new uncertainty importance measure". In: *Reliability Engineering & System Safety* 92.6 (2007), pp. 771–784.
- [7] C. G. Bucher. "Adaptive sampling — an iterative fast Monte Carlo procedure". In: *Structural Safety* 5.2 (1988), pp. 119–126.
- [8] S. K. C. Calman and G. Royston. "Personal paper: Risk language and dialects". In: *BMJ* 315.7113 (1997), pp. 939–942.
- [9] Z. Cao, Y. Wang, and D. Li. "Quantification of prior knowledge in geotechnical site characterization". In: *Engineering Geology* 203 (2016), pp. 107–116.
- [10] V. Chabridon et al. "Reliability-based sensitivity estimators of rare event probability in the presence of distribution parameter uncertainty". In: *Reliability Engineering & System Safety* (2018).
- [11] M. Chevreuil et al. "A Least-Squares Method for Sparse Low Rank Approximation of Multivariate Functions". In: *SIAM/ASA Journal on Uncertainty Quantification* 3.1 (2015), pp. 897–921.

- [12] I. Depina et al. “Coupling the cross-entropy with the line sampling method for risk-based design optimization”. In: *Structural and Multidisciplinary Optimization* 55.5 (2017), pp. 1589–1612.
- [13] A. Der Kiureghian. “First-and second-order reliability methods”. In: *Engineering Design Reliability Handbook*. Ed. by E. Nikolaidis, D. M. Ghiocel, and S. Singhal. Boca Raton, FL: CRC Press, 2005. Chap. 14.
- [14] A. Der Kiureghian and O. Ditlevsen. “Aleatory or epistemic? Does it matter?” In: *Structural Safety* 31.2 (2009), pp. 105–112.
- [15] O. Ditlevsen and H. O. Madsen. *Structural reliability methods*. John Wiley & Sons Ltd, 1996.
- [16] V. Dubourg and B. Sudret. “Meta-model-based importance sampling for reliability sensitivity analysis”. In: *Structural Safety* 49 (2014). Special Issue In Honor of Professor Wilson H. Tang, pp. 27–36.
- [17] V. Dubourg, B. Sudret, and F. Deheeger. “Metamodel-based importance sampling for structural reliability analysis”. In: *Probabilistic Engineering Mechanics* 33 (2013), pp. 47–57.
- [18] B. Echard, N. Gayton, and M. Lemaire. “AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation”. In: *Structural Safety* 33.2 (2011), pp. 145–154.
- [19] M. Ehre, I. Papaioannou, and D. Straub. “A framework for global reliability sensitivity analysis in the presence of multi-uncertainty”. In: *Reliability Engineering & System Safety* 195 (2020), p. 106726.
- [20] M. Ehre, iason Papaioannou, and D. Straub. “Efficient estimation of variance-based reliability sensitivities in the presence of multi-uncertainty”. In: *Proceedings of the 19th working conference of the IFIP Working Group 7.5 on Reliability and Optimization of Structural Systems*. Ed. by J. Song. ETH Zürich, 2018.
- [21] L. Grasedyck, D. Kressner, and C. Tobler. “A literature survey of low-rank tensor approximation techniques”. In: *GAMM-Mitteilungen* 36.1 (2013), pp. 53–78.
- [22] F. L. Hitchcock. “The Expression of a Tensor or a Polyadic as a Sum of Products”. In: *Journal of Mathematics and Physics* 6.1-4 (1927), pp. 164–189.
- [23] M. Hohenbichler and R. Rackwitz. “Improvement Of Second-Order Reliability Estimates by Importance Sampling”. In: *Journal of Engineering Mechanics* 114.12 (1988), pp. 2195–2199.
- [24] T. Homma and A. Saltelli. “Importance measures in global sensitivity analysis of nonlinear models”. In: *Reliability Engineering & System Safety* 52.1 (1996), pp. 1–17.
- [25] B. Iooss and P. Lemaître. “A review on global sensitivity analysis methods”. In: *Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications*. Ed. by C. Meloni and G. Dellino. Springer, 2015.
- [26] M. J. Jansen. “Analysis of variance designs for model output”. In: *Computer Physics Communications* 117.1 (1999), pp. 35–43.
- [27] S.-H. Jiang, I. Papaioannou, and D. Straub. “Bayesian updating of slope reliability in spatially variable soils with in-situ measurements”. In: *Engineering Geology* 239 (2018), pp. 310–320.
- [28] T. Kim and J. Song. “Generalized Reliability Importance Measure (GRIM) using Gaussian mixture”. In: *Reliability Engineering & System Safety* 173.C (2018), pp. 105–115.
- [29] K. Konakli and B. Sudret. “Global sensitivity analysis using low-rank tensor approximations”. In: *Reliability Engineering & System Safety* 156.Supplement C (2016), pp. 64–83.

- [30] K. Konakli and B. Sudret. “Polynomial meta-models with canonical low-rank approximations: Numerical insights and comparison to sparse polynomial chaos expansions”. In: *Journal of Computational Physics* 321 (2016), pp. 1144–1169.
- [31] P. Koutsourelakis, H. Pradlwarter, and G. Schuëller. “Reliability of structures in high dimensions, part I: algorithms and applications”. In: *Probabilistic Engineering Mechanics* 19.4 (2004), pp. 409–417.
- [32] S. H. Lee and B. M. Kwak. “Response surface augmented moment method for efficient reliability analysis”. In: *Structural Safety* 28.3 (2006), pp. 261–272.
- [33] P. Lemaître et al. “Density modification-based reliability sensitivity analysis”. In: *Journal of Statistical Computation and Simulation* 85.6 (2015), pp. 1200–1223.
- [34] E. Levine. “Improving risk matrices: the advantages of logarithmically scaled axes”. In: *Journal of Risk Research* 15.2 (2012), pp. 209–222.
- [35] L. Li et al. “Moment-independent importance measure of basic variable and its state dependent parameter solution”. In: *Structural Safety* 38 (2012), pp. 40–47.
- [36] M. Hohenbichler and R. Rackwitz. “Sensitivity and importance measures in structural reliability”. In: *Civil Engineering Systems* 3.4 (1986), pp. 203–209.
- [37] S. Marelli and B. Sudret. “An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis”. In: *Structural Safety* 75 (2018), pp. 67–74.
- [38] A. Marrel et al. “Global sensitivity analysis of stochastic computer models with joint meta-models”. In: *Statistics and Computing* 22.3 (2012), pp. 833–847.
- [39] V. Maume-Deschamps and I. Niang. “Estimation of quantile oriented sensitivity indices”. In: *Statistics & Probability Letters* 134 (2018), pp. 122–127.
- [40] J. Morio. “Influence of input PDF parameters of a model on a failure probability estimation”. In: *Simulation Modelling Practice and Theory* 19.10 (2011), pp. 2244–2255.
- [41] J. E. Oakley. “Decision-Theoretic Sensitivity Analysis for Complex Computer Models”. In: *Technometrics* 51.2 (2009), pp. 121–129.
- [42] I. Papaioannou, K. Breitung, and D. Straub. “Reliability sensitivity analysis with Monte Carlo methods”. In: *Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures*. CRC Press, 2014, pp. 5335–5342.
- [43] I. Papaioannou, K. Breitung, and D. Straub. “Reliability sensitivity estimation with sequential importance sampling”. In: *Structural Safety* 75 (2018), pp. 24–34.
- [44] I. Papaioannou, M. Ehre, and D. Straub. “Efficient PCE representations for reliability analysis in high dimensions”. In: *Proceedings of the 19th working conference of the IFIP Working Group 7.5 on Reliability and Optimization of Structural Systems*. Ed. by J. Song. ETH Zürich, 2018.
- [45] I. Papaioannou, M. Ehre, and D. Straub. “PLS-based adaptation for efficient PCE representation in high dimensions”. In: *Journal of Computational Physics* 387 (2019), pp. 186–204.
- [46] I. Papaioannou, C. Papadimitriou, and D. Straub. “Sequential importance sampling for structural reliability analysis”. In: *Structural Safety* 62 (2016), pp. 66–75.
- [47] K.-K. Phoon and F. H. Kulhawy. “Characterization of geotechnical variability”. In: *Canadian Geotechnical Journal* 36.4 (1999), pp. 612–624.

- [48] K.-K. Phoon and F. H. Kulhawy. “Evaluation of geotechnical property variability”. In: *Canadian Geotechnical Journal* 36.4 (1999), pp. 625–639.
- [49] R. Rackwitz. “Reviewing probabilistic soils modelling”. In: *Computers and Geotechnics* 26.3 (2000), pp. 199–223.
- [50] R. Rackwitz. “Reliability analysis—a review and some perspectives”. In: *Structural Safety* 23.4 (2001), pp. 365–395.
- [51] A. Saltelli, K. Chan, and E. Scott. *Sensitivity Analysis*. John Wiley & Sons, Inc., 2000.
- [52] A. Saltelli et al. “Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index”. In: *Computer Physics Communications* 181 (2010), pp. 259–270.
- [53] R. Schöbi and B. Sudret. “Global sensitivity analysis in the context of imprecise probabilities (p-boxes) using sparse polynomial chaos expansions”. In: *Reliability Engineering & System Safety* 187 (2019). Sensitivity Analysis of Model Output, pp. 129–141.
- [54] I. Sobol’. “Sensitivity Estimates for Nonlinear Mathematical Models”. In: *Math. Modeling & Comp. Exp* 1 (1993), pp. 407–414.
- [55] S. Song, Z. Lu, and H. Qiao. “Subset simulation for structural reliability sensitivity analysis”. In: *Reliability Engineering & System Safety* 94.2 (2009), pp. 658–665.
- [56] B. Sudret. “Global sensitivity analysis using polynomial chaos expansions”. In: *Reliability Engineering & System Safety* 93.7 (2008), pp. 964–979.
- [57] B. Sudret. “Meta-models for Structural Reliability and Uncertainty Quantification”. In: *Proc. 5th Asian-Pacific Symp. Struct. Reliab. (APSSRA 2012), Singapore* (Mar. 2012).
- [58] P. Wang, Z. Lu, and Z. Tang. “An application of the Kriging method in global sensitivity analysis with parameter uncertainty”. In: *Applied Mathematical Modelling* 37.9 (2013), pp. 6543–6555.
- [59] Y.-T. Wu. “Computational methods for efficient structural reliability and reliability sensitivity analysis”. In: *AIAA journal* 32.8 (1994), pp. 1717–1723.
- [60] D. Xiu and G. E. Karniadakis. “The Wiener–Askey polynomial chaos for stochastic differential equations”. In: *SIAM Journal on Scientific Computing* 24.2 (2002), pp. 619–644.