Optimal Uncertainty Quantification of a risk measurement from a computer code

Jerome Stenger

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*
Partenaire CIFRE : *EDF R&D*

Présentée et soutenue le *02/10/2020* **par :**

JÉRÔME STENGER

Optimal Uncertainty Quantification of a risk measurement from a computer code

---

JURY

JEAN BACCOU	Ingénieur Chercheur	Examinateur
FABRICE GAMBOA	Professeur des Universités	Directeur de thèse
ELISABETH GASSIAT	Professeur des Universités	Examinateur
BERTRAND IOOSS	Chercheur Senior	Encadrant
MERLIN KELLER	Chercheur Expert	Co-directeur de thèse
JEAN-BERNARD LASSERRE	Directeur de Recherche Émérite	Examinateur
JEAN-MICHEL MARIN	Professeur des Universités	Rapporteur
TIM SULLIVAN	Professeur Junior	Rapporteur

École doctorale et spécialité :
MITT : Domaine Mathématiques : Mathématiques appliquées
Unité de Recherche :
Institut de Mathématiques de Toulouse (UMR 5219)
Directeur de Thèse :
Fabrice GAMBOA
Rapporteurs :
Jean-Michel MARIN et *Tim SULLIVAN*

Remerciements

Ces remerciements concluent les trois dernières années de ma vie, trois années folles, riches de changements, de rencontres, d'épanouissement. Ces remerciements sont aussi également d'une grande importance puisque c'est l'unique section que beaucoup vont lire. Ce n'est pas sans émotion qu'une grande partie de ces lignes ont été écrites confiné dans mon appartement au cours d'une pandémie mondiale qui je l'espère changera pour le mieux et profondément l'humanité (edit : rien ne va changer).

First of all, I would like to address my heartfelt thanks and my deepest gratitude to Pr. Tim Sullivan and Pr. Jean-Michel Marin without whom this PhD thesis would have never been published.

Cette thèse n'aurait jamais été possible sans un encadrement d'une remarquable qualité. Une dream team faisant rêver n'importe quel thésard !

- Fabrice, merci d'avoir été mon directeur, j'aurais souhaité passer plus de temps à Toulouse, mais les choses sont d'une telle efficacité que quelques heures passées ensemble se transforment en plusieurs mois de travail pour débriefer toutes les idées échangées ! Merci d'avoir toujours trouvé le temps, ne serait-ce par un petit coup de fil, pour m'aider pendant les périodes creuses. Merci évidemment pour toute la confiance que tu m'as accordée et de m'avoir proposé une multitude d'opportunités pour poursuivre, si l'envie m'en prenait, une carrière académique. Enfin, merci de m'avoir fait voyager, littéralement bien sûr, mais aussi à travers tes propres aventures : « Jérôme je t'appelle rapidos du Japon, ça te dirait de venir au Mexique le mois prochain ? » !

- Merlin, c'est grâce à toi que cette thèse a pu démarrer. Je suis venu passer le premier jour mon entretien à reculons, mais j'ai tout de suite su en arrivant ici que la thèse allait bien se dérouler. Merci pour tout, pour ta gentillesse, ta bienveillance infinie, ta patience qui a su résister même à l'épreuve de mes fautes d'orthographe. Je n'aurais pas pu rêver d'un meilleur encadrant. Merci également pour tous les nouvelles pistes et idées que tu m'as apportées, preuve de ton investissement dans cette thèse finalement assez théorique et surtout non bayésienne !

- Bertrand, merci pour cette efficacité, ces relectures impressionnantes, ces conseils d'une impeccable précision, et ton exigence qui pousse à se dépasser. Merci également pour ta bienveillance et tes compliments réguliers. Tu pourras montrer ces lignes à tes chefs la prochaine fois que quelqu'un en doute ! Enfin, merci de m'avoir fait rire, parce que si les réunions c'est chiant, une réunion avec Bertrand ça l'est beaucoup moins !

Et bien évidemment, merci à tous les trois pour les moments passés ensemble, vous

m'avez confirmé qu'il est possible de bien travailler avec légèreté et bonne humeur ! Merci pour ces rires et ces bières échangés.

Sur le plan scientifique, je voudrais remercier les personnes que j'ai eu la chance de rencontrer et grâce à qui j'ai beaucoup appris, en particulier parmi les membres du GdR MASCOT-NUM, de la chaire OQUAIDO, ou les habitués de l'école d'été ETICS. Je pense aux jeunes qui m'ont beaucoup fait rire, Adrien Spagnol, Clément Benard ou Mathieu Carmassi, aux moins jeunes aux côtés desquels j'ai beaucoup appris, Amandine Marrel, Olivier Roustant, Josselin Garnier, Rodolphe Le Riche et j'en oublie beaucoup d'autres, et enfin remercier les beaucoup plus vieux comme Sebastien DaVeiga que j'ai un toujours pris beaucoup de plaisir à taquiner.

Ceux qui me connaissent pourront peut-être confirmer mes dires, le travail c'est bien mais à petite dose. Et pourtant, quel bonheur j'ai eu à venir travailler tous les jours dans cet environnement extraordinaire. Je ne pourrais pas remercier individuellement tous les collègues de Chatou car la liste serait trop longue. Mais c'est grâce à chacun d'entre vous que se maintient cette ambiance chaleureuse, amicale et studieuse qui motive tous les matins à prendre les transports et venir travailler. Un petit mot pour les plus remarquables, merci à Josselin Delmas pour ta capacité à dire autant de conneries à la minute et à Emmanuel Remy pour réussir à me faire rire en un seul regard. Merci aux autres collègues pour toutes les discussions autour d'un bon café. Je parle de collègues mais bon nombre d'entre vous représente bien plus. Les concernés se reconnaitront mais vous avez été nombreux à me proposer ou à me dépanner lors de mes galères d'appartement. Qu'il s'agisse d'un lit, d'un canapé, ou parfois d'une planche de bois (rue St-Denis pour ne viser personne). Je voudrais remercier les meilleurs des ingénieurs, Pauline, Elias, Louis, Laura, et aussi Claire toujours partante pour rigoler ; tu m'as beaucoup fait rire entre tes jeudi soir tisane/ puzzle puis tes weekends festifs. Merci aussi à Thibault pour m'avoir lâchement abandonné dans le suivi de cette thèse pour poursuivre ta carrière de direction. Tu as le management dans le sang, basé sur un système d'effort/récompense ; tu as attendu que je fasse mes preuves en bar avant de me considérer comme un pair. Je voudrais également remercier Vincent pour ta sensibilité et ton authenticité, et bien sûr pour toutes les histoires que tu partages après quelques verres. Je pense aussi à Clément, tu seras éternellement mon stagiaire, merci d'être bizarre. En parlant de bizarre, merci à Anne pour les fous rires et pour ton inspirante liberté d'être qui a profondément changé ma vie. Enfin, je n'oublie pas la team des thésards : Azénor, Sami, Antoine, Thomas, Alvaro, merci pour tous les moments partagés ensemble. Et une dédicace spéciale pour Pablo, tu m'as toujours bien fait rire dans ces moments d'échanges porte fermée, pour se plaindre, râler, et évacuer quelques tensions accumulées. Et puis Paul merci évidemment pour ces séances d'escalade et cette éphémère bromance, t'es un peu mon Parisien préféré.

Je voudrais également remercier Charlotte, car cette thèse je te la dois en grande partie parce que tu m'as soutenu et accompagné depuis le début. Tu m'as donné la force de démarrer et de poursuivre cette aventure tout au long de ces années vécues ensemble, bien que nos chemins se soient maintenant séparés. Je n'oublie pas tout ce que tu m'as apporté.

Bien évidemment, je remercie profondément ma famille qui a toujours été et sera toujours présente sans jamais rien attendre en retour. Merci à mes parents sans qui rien n'aurait été possible, merci de s'être assuré que je ne manque jamais de rien et pour m'avoir toujours soutenu dans mes projets et dans ma reconversion pour devenir enseignant. Merci à mes beaux-frères, à mes sœurs Lise, Floriane, Céline et Elodie, à mes neveux et à mes nièces pour les moments et les rires partagés. Ce n'est jamais

reposant mais ça fait toujours du bien de nous retrouver ensemble ! Enfin, une attention particulière à mon frère Sébastien, qui, malgré moi, continue d'être mon modèle à suivre, tant pour sa réussite professionnelle que pour ses qualités humaines.

Et puis enfin je voudrais remercie la team Carqueiranne, ma deuxième famille, parce que quoiqu'il arrive, quoique je fasse, quoi que l'avenir nous réserve, je sais que vous serez toujours à mes côtés. Merci à Thibs, Alice, Mull, Batman, Robin, Dyno, Chloé, au Blond, à Alicia, pour ces litres d'alcool ingérés transformés grâce à vous en larmes de rires. Vous m'avez aussi accompagné à votre manière dans cette thèse, certainement pas sur le plan scientifique, car vous êtes tous des incompétents, mais vous avez été à mes côtés dans mon cœur, je vous aime. Vous êtes mon sas de décompression, ma raison d'apprécier autant les vacances.

Enfin, je voudrais remercier ma meilleure amie, Marie, la preuve vivante qu'il n'existe pas de « trop » dans la gentillesse. Merci de réfléchir avec ton cœur, je suis persuadé que l'humanité aurait tout à apprendre de toi. Tu as été là pendant les périodes les plus difficiles qui ont agité cette thèse. Merci d'avoir ce don de trouver les mots qui soignent. Merci de me comprendre.

Contents

List of Figures

List of Tables

List of Algorithms

Acronyms

CDF cumulative distribution function. 10

cpu central processing unit. 23

DE differential evolution. 65

GMP generalized moment problem. 119

Gp Gaussian process. 8

HSIC Hilbert-Schmidt independence criterion. 13

IBLOCA intermediate break loss of coolant accident. 22

iid independent and identically distributed. 11

LARS least angle regression strategy. 120

LHS latin hypercube sampling. 23

MC Monte-Carlo. 12

NN neural network. 111

OUQ optimal uncertainty quantification. 1

PCT peak cladding temperature. 22

PDF probability density function. 11

PII primary influential inputs. 8

PLI perturbed-law based indices. 15

PoF probability of failure. 12

QoI quantity of interest. 12

SDP semi definite programming. 118

UQ uncertainty quantification. 8

Notations

$\mathcal{B}$	Borel σ-algebra on $\mathcal{X}$. 10
$C_b(\mathcal{X})$	Set of all bounded continuous real valued function on $\mathcal{X}$. 29
G	Computer model or its metamodel. 7
$\mathcal{M}_+(\mathbb{K})$	Set of all positive measure on $\mathbb{K}$. 119
$\mathbb{R}[X_1, \ldots, X_d]$	Multivariate polynomial ring at d variables. 119
$\mathcal{S}^c$	Complement of a set $\mathcal{S}$. 119
∂M	Border of the set M. 51
$\overline{\mathbf{co}}$	Closed convex hull. 30
$\mathbf{co}$	Convex hull. 35
$\mathbf{int}(M_{n-1})$	Border of the set M. 53
$\leqq$	Designed either an inequality $\leq$ or an equality $=$. 33
$M \succeq 0$	The matrix M is positive semi-definite. 123
$\mathbf{supp}(\mu)$	Support of a probability measure μ. 55
$\lvert\alpha\rvert$	Maximum of a tuple $\alpha \in \mathbb{N}^d$. 122
d	Dimension of the input vector of the model. 7
$\mathbb{E}_{\mu_1,\ldots,\mu_d}$	Expectation with respect to the product measure. 40
$\mathbb{E}_{-i}$	Expectation over all but the ith input parameter. 90
$\mathbb{E}$	Expectation. 13
$\mathcal{X}$	Input space of the model, subset of $\mathbb{R}^d$. 7
$\mathbb{1}$	Indicator function. 12

$\mathbb{P}$	Probability. 10
ϕ	Quantity of interest. 28
$\mathbb{R}$	Real line. 7
$\mathcal{P}(\mathcal{X})$	Set of probability measure on $\mathcal{X}$. 10
$X \sim f$	The random variable X is distributed according to μ. 11

General Introduction

The PhD originates from the close collaboration between EDF R&D and the university of Toulouse III - Paul Sabatier. From a general point of view, this thesis is related to important industrial challenges. This thesis places itself in the scope of robustness analysis, which is an emerging field in the domain of uncertainty quantification. It consists of analyzing the response of a computer model with uncertain inputs to the perturbation of one or several of its input distributions. It is especially required when evaluating industrial risks, as in nuclear safety.

The manuscript is composed of two parts, the first one constitutes the main subject of this PhD and is composed of eight chapters whose contents are detailed below. We strongly recommend reading the thesis in chronological order. Chapters are not self-contained, however we made sure to include as many cross references as necessary for easier readability. The second part constitutes the appendix. Keywords related to each chapter are detailed in Table 1.

Part I.

Chapter 1 aims at introducing fundamental concepts in uncertainty quantification. We present the probabilistic framework, oriented toward measure theory notions. We also introduce the concept of robustness analysis related to the idea of a second level uncertainty. A state of the art is then presented on that matter.

Chapter 2 presents different computer models under study in this thesis. This PhD originates from industrial challenges and real-world engineering problems. Two computer models are presented, the first one is a toy example, but it is complex enough to illustrate with interest the methodology. The second model is a real-word application related to nuclear safety, it will later serve for testing our framework and its limits.

Chapter 3 addresses the theoretical aspect of the thesis. We demonstrate a generalization of the optimal uncertainty quantification (OUQ) reduction theorem. We show in an abstract setting that the optimization of quasi-convex lower semicontinuous function over a convex subset of a locally convex topological vector space can be reduced to the generator of this subset, assuming its existence. Moreover, we study practical measure space, over which we later proceed the optimization. We deal with two spaces somehow related, the moment class and the unimodal moment class. The input joint distribution of the computer model is assumed to belong in a product of measure space that is also studied in this chapter.

Chapter 4 introduces the canonical parameterization of the moment class. We assume that the moment classes are defined by *classical* moment constraints. Then, we

describe a well-suited parameterization of the quantity of interest. We prove that there exists a bijection between the extreme points of the moment class and a pavement of the form $[0,1]^n$ where n depends in the number of moment constraints. We also detail the main algorithm of the thesis that serves for optimizing the QoI.

Chapter 5 presents numerous practical quantity of interest. For instance, a probability of failure, a quantile, a superquantile or even a Sobol index. We investigate both their mathematical interest to fit the OUQ framework, and their practical interest that we illustrate in the toy case introduced in Chapter 2. We also show in this section how the robust Bayesian framework is a specific case of our framework.

Chapter 6 aims at applying the whole methodology in the real-world engineering case introduced in Chapter 2. We further detail how to account for the metamodel-induced uncertainty, so that this chapter gives a guideline on how to proceed practically a robust analysis using the OUQ framework.

Chapter 7 addresses the complex problem of dependency between inputs. We expose several leads that have been studied with their benefits and their flaws. The first one deals with the canonical moment parameterization in high dimension. The second is based on semi definite programming of a relaxed equivalent optimization problem. And the last lead focuses in the copula theory for modeling inputs dependence. All the different solutions explored face the challenging issue of moving towards high dimensional problems.

Chapter 8 gives a general conclusion to this work. It discusses the limitation of the OUQ methodology applied in an industrial context. It also exposes perspectives of this work, some of them being still ongoing work.

Part II

Appendix A presents our contribution in sensitivity analysis. We study the perturbed law indices that consists of analysing the response of a computer model with uncertain inputs to the perturbation of one or several of its input distributions. This work focus in defining rigorously a coherent perturbation. The proposed methodology is based on the Fisher-Rao metric on manifolds of probability distribution and related tools from Lagrangian mechanic. This perturbation definition is then used to compute quantile-oriented robustness indices.

Appendix B analyzes an algorithm for the sequential estimation of a conditional quantile in the context of real stochastic codes with vector-valued inputs. Our algorithm is based on k-nearest neighbors smoothing within a Robbins-Monro estimator. We discuss the convergence of the algorithm under some conditions on the stochastic code. We provide non-asymptotic rates of convergence of the mean squared error and we discuss the tuning of the algorithm's parameters.

Appendix C addresses different versions of the differential evolution algorithm. This is the algorithm used for optimizing the quantity of interest related to the first part of this manuscript.

Appendix D constitutes a extended french summary of the thesis.

Table 1: Keywords of the content of this thesis

Keywords	**Chapter**
Uncertainty modeling / Probability theory / OUQ / Robustness analysis	Chapter 1
Computer model / Numerical experiment / Black-box	Chapter 2
Bauer maximum principle / Quasi convexity / Moment class / Unimodal moment class	Chapter 3
Canonical moments / Global optimization / QoI computation	Chapter 4
Probability of Failure / Quantile / Superquantile / Sobol Indices / Robust Bayesian analysis	Chapter 5
Thermal hydraulic / Computer model / Robust quantile	Chapter 6
Dependence / Generalized Moment Problem / Semi definite programming / Copula	Chapter 7

Publications and Communications

Published works related to the content of this thesis are listed herebelow.

The following publication is related to the Chapter 4 and constitutes one my main contribution to this work. It also presents the main application found in Chapters 2 and 6 of this manuscript:

Stenger, J., Gamboa, F., Keller, M., and Iooss, B. 2020. Optimal Uncertainty Quantification of a Risk Measurement from a Thermal-Hydraulic Code using Canonical Moments. In *International Journal for Uncertainty Quantification 10*, 1, 35-53.

This publication, written for the 4th conference on geometric science of information, presents similar results to the first one but emphasizes the geometrical interpretation of the canonical moments:

Stenger, J., Gamboa, F., Keller, M., and Iooss, B. 2019. Canonical Moments for Optimal Uncertainty Quantification on a Variety. In *Geometric Science of Information.* Springer, Cham, 571-578.

The content of this publication can be found in Chapters 3 and 5, it constitutes the second main contribution to this work:

Stenger, J., Gamboa, F., and Keller, M. 2020. Optimization of a Quasi-Convex Function Over Product Measure Sets. *Preprint. arXiv: 1907.07934.*

This publication is the result of the 6 months internship of Clément Gauchy that I supervised at EDF with the help of Roman Sueur and Bertrand Iooss. Its content differs from main topic of this thesis, for that reason it has been exposed in the Appendix A:

Gauchy, C., Stenger, J., Sueur, R., and Iooss, B. 2019. An Information Geometry Approach for Robustness Analysis in Uncertainty Quantification of Computer Codes. *Preprint. hal-02425477.*

I contributed to this last publication by correcting and improving an existing paper, it can be found in the Appendix B:

Labopin-Richard, T., Gamboa, F., Garivier, A. and Stenger, J. 2019. Conditional Quantile Sequential Estimation for Stochastic Codes. In *Journal of Statistical Theory and Practice 13*.

Part I

Contributions in Optimal Uncertainty Quantification

1

Introduction: Uncertainty Quantification of Computer Models

« The most exciting phrase to hear in science, the one that heralds new discoveries, is not "Eureka!" but "That's funny..." »

Isaac Asimov

Contents

1.1 Motivation - Industrial Context

In modern engineering studies, numerical modeling and simulation have become major tools. When confronted to physical phenomena, repeated practical experimentation quickly becomes too costly or too complex to set up. Modern engineering study thus replaces the experimentation with a numerical modeling of the process. Those numerical models are governed by complicated equations aiming at reproducing the physical phenomenon. For instance, in order to pursue complex engineering risk management analysis, the long, expensive, or dangerous if not impossible real physical experiment reproducing an accidental scenario cannot be considered. In nuclear engineering, it might preferable to have a numerical simulator predicting the thermal and hydraulic phenomenon happening in a accidental scenario than performing the true experiment. In another setting, engineers in financial planning and analysis aim at budgeting and forecasting a process embodied as a computer experiment. For instance it is necessary to determine whether a wind farm will be profitable before building the structure, thus numerical estimation of the power production should be conducted.

Most physical systems are subject to either lack of knowledge about the governing physical laws or incomplete and limited information about model parameters such as input variable properties, initial or boundary conditions. In that context, it is essential to evaluate the accuracy of the numerical model. Assessing the reliability of the engineering study imposes to take into account the sources of uncertainty that may be tainting the result of the computer code.

This chapter aims at presenting the overall methodology currently used by engineers that pursue an uncertainty quantification study. Section 1.2 presents some elements of typical computer codes under study. We depict in Section 1.3 the different sources of uncertainties affecting the result of the computer results. Then, Section 1.4 gives the general methodology to take into account these uncertainties in an engineering study, we introduce in this section some probabilistic elements and notations. We study some second level uncertainty in Section 1.5 which is the main topic of this thesis. Last section gives a picture of the aim of this PhD.

1.2 Computer Model

1.2.1 Black-box Function

In this thesis, few assumptions are made about the computer code under study. Because the physical phenomena it reproduces are complex, we consider the code acts as a black-box function. That means that given a set of input parameters, the only information available is the scalar output of the computer code. The mathematical formulation of the problem is represented hereunder

$$\begin{array}{rccc} G: & \mathcal{X} \subset \mathbb{R}^d & \to & \mathcal{Y} \subset \mathbb{R} \\ & \mathbf{x} & \mapsto & y = G(\mathbf{x}) \end{array} \tag{1.1}$$

where G denotes the computer model and $\mathbf{x} = (x_1, \ldots, x_d)$ is a set of input variables in $\mathbb{R}^d$. Each input represents a physical or an environmental parameter. As they represent physical parameters, the input parameters are most of the time bounded, so that the input space $\mathcal{X}$ can be considered compact.

In this thesis, the computer code G is considered deterministic in contrast to the stochastic computer experiments which return different result values for several calls

with a the exact same set of input variable. The model G is a more or less high fidelity representation of the physical process, it may itself belongs to a chain of multi-physics numerical codes. This implies that a single run of the code can be computationally costly, taking up to several days. Notice, that it is not because the code G is a black box function that it prevents the understanding of the physical process. It is still possible to analyse how an input impacts the result of the numerical experiment.

Using a deterministic simulation model does not remove the study from being affected by uncertainties. The art of quantifying the impact of the uncertainties on the result of the computer experiment is known in the literature as uncertainty quantification (UQ) (De Rocquigny et al., 2008; Sullivan, 2015).

1.2.2 Metamodels

In complex engineering problem, such as the ones encountered in nuclear safety management analysis, the computer codes are usually expensive to run. Further study and in particular UQ study becomes a difficult task to process with such models. In order to circumvent this problem, a widely accepted method consists in replacing the cpu time expensive numerical experiment by an inexpensive mathematical function called a metamodel, a surrogate model, a code emulator, or a response surface (De Lozzo, 2015; Kleijnen and Sargent, 2000). Optimally, the metamodel is a high-fidelity substitution to the model. In the following, we briefly present different famous methods for building an emulator. Practical construction of a metamodel is described in Chapter 2.

First of all, the metamodel is built from a finite set of computer simulations that must be as representative as possible of the code in the variation domain of its uncertain inputs. Generally, space-filling designs are created with a fixed budget in order to ensure a full coverage of the input space (Fang et al., 2005). When the dimension of the input vector $\mathbf{x}$ is high, screening strategies may be used in order to select the primary influential inputs (PII) on the model output variability and rank them by decreasing influence. In that case, the surrogate model is built from the learning sample to fit the simulator output, considering only the PII as the explanatory inputs (Iooss and Marrel, 2019). The remaining inputs remain fixed to a default value.

Different solutions for building metamodels are found in the literature (splines, neural networks, polynomial chaos expansion, etc.), Gaussian process (Gp) regression (Sacks et al., 1989; Le Gratiet et al., 2017; Rasmussen and Williams, 2005), also known as Kriging, has been of a particular interest. It makes the assumption that the response of the computer code is the realization of a Gaussian process. The Gp is conditioned on the code observations from the learning sample. It provides the basis for handy statistical inference, as the conditioning of the Gp remains Gaussian. In that, we have at our disposal simple analytical formula of the predictor and mean squared error of the prediction.

1.3 Sources of Uncertainties

In order to confidently use a computer code, one needs to quantify its ability to predict the response of the real system. Therefore, we need to identify the possible uncertainties tainting the result of the numerical simulation. From the reality to the construction of the model, many different sources of error may affect the predictivity quality of the result. Moreover, classifying the sources of uncertainty depend on the semantics,

and there exists different schools on that subject. One can refer to Kiureghian and Ditlevsen (2009) for an review on that matter.

Considering a source of uncertainty, one would like to know whether some further analyses could yield a reduction of its associated error. In other words, would it be possible to get complementary information in a reasonable budget so that we improve our confidence in the prediction of the model. Therefore, from an engineering point of view, we distinguish two types of uncertainty:

- *Aleatory uncertainty.* It refers to the intrinsic randomness of the real system. For instance, the wind loads passing through a wind turbine or a coin flip. It is considered that given a realistic budget, this kind of uncertainty remains irreducible. It affects the input vector $\mathbf{x}$ of the model (see Eq. (1.1)), meaning this uncertainty affects the physical variable of the system and not the computer code directly.

- *Epistemic uncertainty.* This relates to the lack of knowledge of the analyst. This uncertainty can be reduced by acquiring more information (i.e. data, measurements, expert judgements). Epistemic uncertainty affects either the input vector $\mathbf{x}$ or the model G.

Some consider that aleatory and epistemic uncertainty cannot be distinguished as their difference relies on the budget considered to acquire new information (JCGM, 2008). In Fu (2012), the author studies the Strickler coefficient which measure the friction of a river bed. It is both tainted with epistemic uncertainty due to the small sample size used for the calibration and aleatory due to the random change of flow and river bed properties during the seasons. This classification gives a pragmatic way for engineers to distinguish which source of uncertainty is worth allocating budget to improve the confidence in the prediction. For instance, the wind load of a wind turbine could be faithfully captured given an enormous amount of sensor data, and the result of a coin flip is fully determined by the physical initial conditions of the experiment. Nevertheless, the sources of uncertainty can be differently classified in the following way:

- *Variability.* It refers to the natural variability of a real system. For instance, the variability of manufactured mechanical parts.

- *Model error.* This relates to the error made when building a model. For instance, it encompasses the numerical approximation errors, but also the error due to simplified physical equations modeling the real phenomena. At last, a supplementary source of error appears when the expensive model is replaced with a metamodel (see Section 1.2.2).

- *Input uncertainty.* Many sources of uncertainty affect the knowledge of the input. It arises from a lack of data, in addition the data itself can be imprecise or possibly sparse depending on the quality of the measurements. Some information on the input are also given by expert judgment that is naturally biased.

In this thesis, we are going to focus more specifically on the input uncertainty. A widely accepted method for quantifying the impact of a lack of knowledge of the input over the output is to use a probabilistic framework.

1.4 Uncertainty Propagation

The uncertainties tainting the input variable can be captured by means of different modeling (intervals, random set, fuzzy probabilities, info-gap, among others... see Ferson and Ginzburg (1996); Ferson and Oberkampf (2009); Paté-Cornell (1996); Qiu et al. (2008); Ben-Haim (2006)). The most commonly used is the probabilistic framework. In this thesis, we make extensive use of measure theory so that some prerequisites are necessary. However, we present in the following some generalities, further details will be presented when needed in the thesis. For an extensive overview of uncertainty quantification, one can refer to Sullivan (2015); De Rocquigny et al. (2008)

1.4.1 Probabilistic Modeling

Let $\mathcal{X}$ be a topological space, in most cases $\mathcal{X}$ will refer to a subset of $\mathbb{R}^d$. We denote $\mathcal{P}(\mathcal{X})$ the set of all probability measures on $(\mathcal{X}, \mathcal{B})$ where $\mathcal{B}$ is the Borel σ-algebra on $\mathcal{X}$. An element $\mu \in \mathcal{P}(\mathcal{X})$ induces a probability space $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mu)$. A random vector $\mathbf{X} = (X_1, \ldots, X_d)$ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is a (measurable) function such that

$$\mathbf{X}: \quad \mathcal{B} \to \mathbb{R}^d \tag{1.2}$$

$$\omega \mapsto X(\omega) \ . \tag{1.3}$$

If $\mathbf{X}$ is a random vector, then for every Borel set B of $\mathbb{R}^d$, we have $\mathbf{X}^{-1}(B) \in \mathcal{B}$. We can define a probability measure $\mu_{\mathbf{X}}$ satisfying

$$\mu_{\mathbf{X}}(B) = \mathbb{P}(\mathbf{X} \in B) = \mathbb{P}(\mathbf{X}^{-1}(B)).$$

Moreover, given a probability measure μ on $\mathbb{R}^d$, we can associate a random vector $\mathbf{X}_\mu$ satisfying for all $B \in \mathcal{B}(\mathbb{R}^d)$

$$\mathbb{P}(\mathbf{X}_\mu \in B) = \mu(B).$$

So that we may speak indifferently of a measure or a random variable. A probability measure μ on $\mathbb{R}^d$ is often given in terms of a cumulative distribution function (CDF) denoted by F_μ and defined for every vector $\mathbf{h} = (h_1, \ldots, h_d) \in \mathbb{R}^d$ by

$$F_\mu(\mathbf{h}) = \mathbb{P}_\mu(\mathbf{X} \leq \mathbf{h}) = \mathbb{P}_\mu(X_1 \leq h_1, \ldots, X_d \leq h_d) = \mu\left(\prod_{1 \leq i \leq d}]-\infty, h_i]\right) \ . \tag{1.4}$$

We highlight that the traditional split of probability courses into two segments, that is the study of discrete distributions and the study of "continuous" distributions, is unnecessary. Indeed, in a measure theoretic treatment, absolute continuity is defined for one measure with respect to another. In all generality the Lebesgue decomposition theorem states that any probability measure μ can be decomposed with respect to a reference measure ν:

$$\mu = \mu_{ac} + \mu_s \ ,$$

where μ_{ac} is absolutely continuous with respect to ν and where μ_s and ν are singular. In $\mathbb{R}^d$, the decomposition can be refined. Hence, a measure μ can always be decomposed with respect to the Lebesgue measure on $\mathbb{R}^d$ (Hewitt and Ross, 1979, Theorem 19.20) as.

$$\mu = \mu_{ac} + \mu_d + \mu_s \ .$$

Where μ_{ac} is absolutely continuous with respect to the Lebesgue measure, μ_d is a discrete measure that is supported on a countable set and μ_s is singularly continuous, i.e., it is supported on a Lebesgue null-set, and the probability of each point in that set is zero (see the famous Cantor distribution as example).

On one hand if the measure μ is absolutely continuous with respect to the Lebesgue measure, i.e. $\mu = \mu_{ac}$ then we define its probability density function (PDF) f_μ as the Radon-Nikodym derivative of the measure μ with respect to the Lebesgue measure. On the other hand if the measure is discrete then $\mu = \mu_d$, the density function is defined as the Radon-Nikodym derivative with respect to the counting measure on the support of μ.

In all this work, we will switch freely between measure, random variable, distribution function or density according to our convenience. However, a certain preference goes for the measure formalism. Indeed, consider computing the probability of a measurable set $B \in \mathbb{R}^d$ with respect to a measure μ. The notation

$$\mathbb{P}(B) = \int_{\mathbb{R}^d} \mathbb{1}_B(\mathbf{x}) f(\mathbf{x}) \, \mathrm{d}\mathbf{x} \ ,$$

is usually preferred for absolutely continuous measure with pdf f. However, the more general Stieltjes form of integral, namely,

$$\mathbb{P}(B) = \int_{\mathbb{R}^d} \mathbb{1}_B(\mathbf{x}) \, \mathrm{d}\mu(\mathbf{x}) \ ,$$

also known as Fourier-Stieltjes transform, is defined in all cases.

Independence will often be supposed between the different input variables so that the d-dimensional input probability distribution F_μ can be written as a product of its marginal distribution $\prod_{1 \leq i \leq d} F_{\mu_i}$ and $f_\mu = \prod_{1 \leq i \leq d} f_{\mu_i}$. In other words, at independence the input space $\mathcal{X}$ is a pavement written $\prod_{1 \leq i \leq d} \mathcal{X}_i$ and the measure $\mu \in \mathcal{P}(\mathcal{X})$ is decomposed $\mu = \otimes_{1 \leq i \leq d} \mu_i$ where $\mu_i \in \mathcal{P}(\mathcal{X}_i)$ is the ith canonical projection of μ.

When the assumption of independence is not verified, the multivariate distribution will be defined with the marginals distribution and a copula that represent the dependence structure between each variables. This follows the Sklar's Theorem characterization of multivariate distributions (Nelsen, 2006).

1.4.2 Choice of Input Distributions

Engineers gather many sources of information to better precise their understanding of the input variables. However, because the information can be incomplete or even contradictory, each input is modeled as a random variable. The probabilistic modeling translates the uncertainty and the variability of the input of the model.

We precise some usual sources of information. The first one is some expert judgment, it can help to identify some specific constraints that the input distributions should satisfy. This source is intrinsically biased, however human experience is often precious to fill incomplete data. Indeed, the second main source of information is the data available from possible measurements. The analyst would like to find a probabilistic density that fits best those data. To do so, the dataset of size n is considered as an independent and identically distributed (iid) sample $\mathscr{X} = \{\mathbf{x}^{(i)}\}_{1 \leq i \leq n}$ of a probability distribution μ that needs to be reconstructed. The purpose is therefore to find the distribution associated with the underlying input random vector $\mathbf{X}$. The notation $X \sim f$ means that the random vector of the model $\mathbf{X}$ is distributed according to μ. Two different ways for inferring the probability can be employed.

- An assumption is made that the underlying probability distribution belongs to parametric classes of measures, for instance Gaussian, uniform, lognormal family etc. Determining the probability distribution that fits the data consists in estimating the best parameters of the distribution family. Different tools can be applied, such as Bayesian calibration or direct parameters estimation (Rohde, 2014). The Bayesian calibration and more specifically the prior distribution construction relies on the expert judgment. The direct parameters estimation can be realized using the maximum likelihood method.

- When no assumption is made, the underlying probability distribution is non-parametrically estimated, for instance by means of kernel density estimation or using an histogram approximation (Silverman, 1986).

The parametric approach is often preferred. The choice of the parametric family can be motivated using statistical tests such as the Kolmogorov-Smirnov or the Cramér-von Mises ones, to cite only the most famous (Nikolaidis et al., 2004). In this thesis, we also assume the existence, from a purely formal point of view, of a *true* distribution that represents the real variation of the input variable. This distribution is considered unreachable and only provides a conceptual model. In this way, our probabilistic modeling of the input aims at best approaching the *true* distribution.

1.4.3 Quantity of Interest

Because the input variables $\{X_i\}_{1\leq i\leq d}$ are now modelled as random variables, the output of the code Y is also a random variable. More precisely, suppose that every input X_i is independently distributed according to μ_i, for $1 \leq i \leq d$, and the input vector $\mathbf{X} = (X_1, \ldots, X_d)$ is distributed according to $\mu = (\mu_1, \ldots, \mu_d)$, then the output distribution function F_μ of the computer code satisfies the following equation

$$F_\mu(y) = \int_{\mathbb{R}^d} \mathbb{1}_{\{G(x_1,\ldots,x_d)\leq y\}} \, \mathrm{d}\mu_1(x_1) \ldots \mathrm{d}\mu_d(x_d) \; . \tag{1.5}$$

This result is known in the literature as the transport theorem (Barbé and Ledoux, 2007). Because this formula is generally not analytically tractable, we estimate in practice the output distribution using stochastic method, the most established and widely used being the Monte-Carlo (MC) method. It consists in generating a sample of the input random vector $\mathscr{X} = \{\mathbf{x}_i\}_{1\leq i\leq n}$, then to propagate the sample through the computer code G, so that we obtain a sample of the output distribution $\mathscr{Y} = \{G(\mathbf{x}_i)\}_{1\leq i\leq n}$ (De Rocquigny et al., 2008).

The output of the computer code is usually called the variable of interest. However, the statistician is not directly interested in the variable of interest but in a statistical indicator of the output called quantity of interest (QoI) (De Rocquigny et al., 2008). For instance, consider a code that computes the height of a river. In a reliability study the engineer would rather quantify the probability that the river height does not exceed a given threshold, for instance to evaluate the required height of a protection dike. In this example, the variable of interest is the height of the river, and the quantity of interest is the probability not to exceed a given threshold (a probability of failure (PoF)). Depending on the context many quantities of interest can be studied:

- it could be as stated a probability of failure $\mathbb{P}_\mu(G(\mathbf{X}) \leq h)$ with respect to a given threshold $h \in \mathbb{R}$. More generally, one can reconstruct the entire CDF, F_μ to characterize the whole variability of the output;

- to estimate a safety criterion, engineers are often interested in a conservative measure, such as a quantile $Q_p^L(\mu) = \inf\{h \in \mathbb{R} : F_\mu(h) \geq p\}$ (Oakley, 2004; Iooss and Marrel, 2019);

- one can estimate some moments, for instance $\mathbb{E}_\mu[G(\mathbf{X})]$ or $\mathrm{Var}_\mu[G(\mathbf{X})]$, in order to synthesize the statistical behavior of the output Y. Those are just specific cases of what is called *generalized* moments (Lasserre, 2010), that are defined with respect to a measurable quantity of the input measure $\mathbf{X} \sim \mu$

$$\mathbb{E}_\mu[\phi] = \int_{\mathbb{R}^d} \phi(\mathbf{x}) \, \mathrm{d}\mu(\mathbf{x}) \; ; \tag{1.6}$$

- uncertainty quantification is often, and should be, run in tandem with sensitivity analysis. It is the study of how the uncertainty in the output of the mathematical model can be divided and allocated to different sources of uncertainty in its inputs. Sensitivity analysis aims at determining and ordering by importance the relevance of each input in the variation of the output. This can take into account either the whole range of variation of the input then we talk about *global* sensitivity analysis, or it can be focused on the output variability over a small variation of the input around a nominal value then it is know as *local* sensitivity analysis. Three kinds of methods are distinguished: screening (coarse sorting of the most influential inputs among a large number), measures of importance (quantitative sensitivity indices) and deep exploration of the model behaviour (measuring the effects of inputs throughout their all variation range). Sensitivity analysis returns as a result some real and easily interpretable indices. The most famous are the Sobol' indices (Sobol', 1993), that decomposes the variance of the output of the model into fractions that can be attributed to each input (under independence assumption). Shapley effects (Owen, 2014; Song et al., 2016), Hilbert-Schmidt independence criterion (HSIC) indices (Da Veiga, 2013) are a non exhaustive list of examples that will be further specified in the following of the thesis. Each one of these indices can constitute our quantity of interest.

1.4.4 Summary

The above methodology can be summarized in Figure 1.1. As stated, Step A specifies the problem by setting a computer model or its replacement with a metamodel, its output and a quantity of interest on the output. The randomness of the output arises from Step B, which assigns probabilistic distributions to the inputs. In step B' comes a calibration and validation in order to decrease the error between the model and the observed data, typically by decreasing the uncertainty of the inputs and tuning the model parameters. A screening strategy in Step C' helps simplifying the problem by setting some non influential input parameter to fixed value, leaving only the PII as random variables. This feedback loop updates the metamodel and the input distributions. The Step C, that goes from the random inputs to the QoI constitutes what is known as uncertainty propagation.

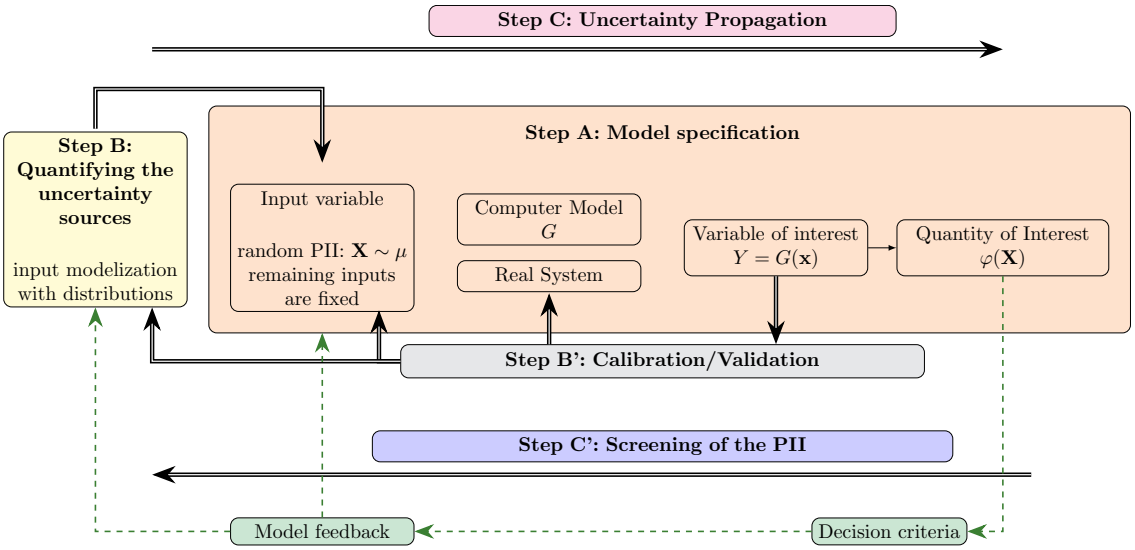


Figure 1.1: This scheme represents the usual UQ methodology employed in engineering study.

1.5 Second Level Uncertainty Quantification

The uncertainties in the input variables are taken into account with the widely accepted methodology presented Section 1.4. However, the probability distributions characterizing the uncertain input may themselves be uncertain. Indeed, contradictory expert opinion may yield difficulty to select a probability model, and the lack of information in the input variables is inevitably affecting the choice of the distribution. For instance, consider an input variable with few number of observations and no expert knowledge. Statistical tests (for instance Pearson' chi squared test, G-test (Chernoff and Lehmann, 1954)) could accept many different probability density models, but only with low power and poor confidence in the model parameters.

This uncertainty affecting the choice of a probabilistic model can be seen as a second level uncertainty. This lack of knowledge can affect the modeling in different forms:

- the type of the input distribution is uncertain, the choice of a parametric model can be unadapted (uniform, triangular, normal, lognormal). The parametric modeling is itself questionable. Though it helps providing a simple numerical framework to approximate the *true* distribution, it is only a limited interpretation of reality. An unique distribution is in practice challenging to identify;

- the other source of uncertainty affecting the probability distributions comes from the lack of confidence in the parameters. The parameters are uncertain because they are estimated through limited data.

In both cases, the input distribution is imprecisely defined, these uncertainties show limits in the probabilistic modeling approach. Because the UQ methodology is based on this probabilistic modeling, a second level uncertainty in the distribution affects the robustness of the whole UQ study. In the presence of uncertainty the modification of the input distribution can significantly change the value of the quantity of interest. Quantifying the impact of this second level uncertainty is a challenging problem and the topic of this PhD.

Several works have been dedicated to studying this second level UQ, while this topic has recently gained considerable interest for engineering application, one of its

earliest treatment dates back to Knight (1921). A non exhaustive state of the art of these works are given hereunder:

- the first work dates back to the late 1980s and is called robust Bayesian analysis (Berger, 1990; Ruggeri et al., 2005), it aims at quantifying the range of variation of the quantity of interest under modification of the prior distribution in a Bayesian framework. The prior distribution is supposed to belong to some set of probability measures that is representative of the uncertainty affecting the prior. Over this set of probability distribution, the robust Bayesian analysis computes lower and upper bounds of the quantity of interest. If the range of the quantity of interest over the set of prior is small, then the result is said to be robust in the sense that the priors choice has small impact on the quantity of interest;

- more recently, the OUQ framework has been introduced by Owhadi et al. (2013). They consider uncertainty in the input distribution of an UQ study. The quantity of interest is assumed to be an affine function of the input distribution, for instance a PoF. The approach is quite similar to the robust Bayesian analysis as the input probability measure is assumed to belong to a set of distributions satisfying *generalized* moment constraints (see Eq. (1.6)). Once again this set should be compatible with the information and uncertainty of the input variables. They also compute upper or lower bounds on their quantity of interest, it is called optimal in the sense that it is the sharpest bound that is consistent with the information encapsulated in the set of input distributions. Because this PhD has been motivated by Owhadi et al. (2013), we extensively use in the following the notation of their framework;

- in sensitivity analysis, second level UQ has become popular, the aim being to quantify how the uncertainty on the input distribution can affect the sensitivity indices. For instance, in Meynaoui et al. (2019) the authors investigate the impact of a second level uncertainty on the HSIC dependence measures. To that effect, they define a second level sensitivity indices whose purpose is to characterize the sensitivity of the HSIC indices to the uncertainty affecting the input probability distribution. In Hart and Gremaud (2019) is studied the robustness of the Sobol' indices to a perturbation of the input probability distribution. To do so, they study the maximal local variation of the Sobol' indices by computing its Fréchet derivative with respect to the input distribution. In Chabridon (2018) is presented a set of dedicated reliability-oriented Sobol' indices taking the second level of input uncertainty into account in a parametric framework. The authors separates the aleatory (irreducible) and epistemic (reducible) uncertainties via a disaggregated version of the input random variables. At last, the perturbed-law based indices (PLI) (Lemaître et al., 2015; Sueur et al., 2017; Gauchy et al., 2019) (see also Appendix A) is a sensitivity index that is dedicated to measure the impact of an imprecise specification of input distributions on the reliability analysis results. They study the relative variation of the quantity of interest under a perturbation of the input distribution. A somehow similar idea can be found in Pesenti et al. (2019) where the perturbation of the input density (called a stress) yields the computation of the derivative of the quantity of interest. In addition the dependency of each input is taken into account directly and indirectly in their so called cascade sensitivity indices;

- in financial statistics, the terminology differs but the mathematical modeling

is similar to ours. The input probabilistic densities describe risk factors that are aggregated through a loss function G. The scalar random output of this model represents a portfolio loss. The quantity of interest is computed through a risk measure that only depends on the distribution of the portfolio loss, the measure of risk should satisfy some statistical properties also known as coherency axioms (Artzner et al., 1999). Historical risk measures are the value at risk (VaR) equivalent to a quantile and the conditional value at risk (CVaR) also know as the expected shortfall (ES) or superquantile (Rockafellar and Uryasev, 2002). Second level uncertainty also arises in financial statistics as the portfolio loss distribution may be uncertain. In Cont et al. (2010), the authors investigate whether a risk measure is sensitive to the small perturbation of the empirical distribution computed from the available data. For this purpose, they define the robustness of a risk measure as its continuity with respect to the Prokhorov metric on the set of integrable random variable;

- the theory of imprecise probability (Walley, 2000) allows for partial probability specifications. It permits modeling separately aleatory and epistemic sources of uncertainty (Schöbi, 2017), which is not the case when a unique measure μ models both sources of uncertainty in an input X. A large amount of theory exists in the literature: evidence theory (Dempster, 2008; Shafer, 1976), probability-boxes (Ferson and Ginzburg, 1996), fuzzy variables (Möller and Beer, 2004), possibility theory (Dubois and Prade, 1988), info-gap theory (Ben-Haim, 2006) constitute a non-exhaustive list of examples. All consider that a unique probability distribution can be hard to identify. The most commonly used are probably the evidence theory and the probability-boxes. Focusing in the evidence theory, it consists in determining bounds on some specific events. When these informations are gathered, it yields lower and upper distribution called *necessity measure* and *possibility measure* which correspond to the envelope of a set of probability distributions. In a similar manner, probability boxes directly set lower and upper bounds on the CDF in between which the true but unknown distribution should lie. At all events, probability boxes are a special case of evidence theory, we refer to Schöbi (2017) for a nice and concise overview on imprecise probability.

This non exhaustive list of examples outlines the raise of awareness and the need to take into account the lack of confidence in the input probability distributions in an engineering study.

1.6 Objectives and Conclusion

This chapter has presented the context in which the PhD takes place. Starting from a computer model, we introduced the way the uncertainty on the input variables is taken into account through a probabilistic modelization. However, this methodology is fundamentally limited by the uncertainty tainting the identification of a good probabilistic model.

In this PhD Thesis we propose a way to take into account this second-level uncertainty. We extend the work of Owhadi et al. (2013) and use the OUQ framework developed to that end. Although the semantic differs, most of the examples presented in Section 1.5 correspond to the same optimization problem. For instance, we will prove that the robust Bayesian framework is a specific case of our generalized OUQ

setting. This does not mean that all such frameworks should be united as one, because the purpose of their robustness study and their interpretation are different. Nevertheless, the underlying optimization problem can be summarized in the following way: "because the specification of an unique input distribution is challenging, we instead consider a set of probability measure". This set should encompass all the uncertainty tainting the input probability measure, and should be compatible with the information we possess about the input variables. Over this set, there exist bounds (possibly infinite) of the quantity of interest. Engineers are interested in those bounds. Indeed, in the context of engineering safety management analysis computer models are used to measure safety margins, those margins depend on the choice of the input distribution. The maximal margins over the measure space guaranty the safety of the structure even in the worst case scenario. And the seek for robustness yields the decision for the maximal safety margins. This general optimization problem is very complex since measure spaces are in all generality non parametric and infinite dimensional. Moreover, in the OUQ framework of Owhadi et al. (2013) the function optimized (the QoI) is required to be linear in the input distribution which is restrictive.

In this Thesis we generalize this OUQ setting. We consider a measure space called (unimodal) moment class, which enforce moment constraints to (unimodal) probability measures. This space possesses nice topological properties that will be extensively study in Chapter 3. We make no assumption on the underlying computer model, which makes the framework very appealing for industrial application. The numerical simulator is seen as a scalar black box function that might be expensive. In the later case, we present in Chapter 6 a methodology to replace the costly computer model by a surrogate model while containing its uncertainty. We intend to optimize and assess bounds of a general QoI over the handy moment space. A large variety of quantity of interest are presented in Chapter 5, all intensively illustrated on examples.

As this thesis is related to important applicative challenges, we intend to give the interpretation of the QoI bounds in terms of engineering analysis. In particular, we study two cases related to risk management analysis. We introduce in the next chapter the computer models that will be used to evaluate performance of our algorithms and practical implementation of our methodology.

2

Application Cases

« What surprises me most is "Man", because he sacrifices his health in order to make money. Then he sacrifices money to recuperate his health. And then he is so anxious about the future that he doesn't enjoy the present; the result being he doesn't live in the present or the future; He lives as if he's never going to die, and then he dies having never really lived. »

Tenzin Gyatso

Contents

Introduction

In this Chapter we detail the computer models that will serve as illustrations in the following of the manuscript. The computer experiments under study fit the model described in Chapter 1. They are deterministic functions taking for input physical or environmental parameters that are uncertain. We put in action in this chapter the uncertainty propagation methodology presented in Section 1.4, in order to illustrate and better understand the benefits of this approach.

Two different models are detailed in the two first sections, the first one is a simple computer code that calculates the water height of a river subject to a flood event. It has already been extensively used in several works, see for instance Pasanisi et al. (2012); Iooss and Lemaître (2015); Stenger et al. (2020). The second one is a real engineering computer experiment called CATHARE2 (V2.5_3mod3.1). It simulates the time evolution of physical quantities during a thermal hydraulic transient in a water pressurized nuclear power plant. Contrary to the toy case which possesses an analytical formula, this computer model is expensive to run and is considered to be a black box function. In the last section, we emphasize the limits of the classical uncertainty propagation approach, consequently motivating the following of this thesis.

2.1 Flood Toy Model

The model of interest concerns a flooded river simulation, which is useful in assessing the risk of submergence of a dike protecting industrial sites nearby a river. To this purpose, we consider a hydraulic model implementing a simplified version of the one dimensional hydro-dynamical equations of Saint-Venant. Under the assumptions of uniform and constant flowrate and large rectangular sections, the maximal annual water height H is computed from four parameters Q, K_s, Z_m, and Z_v with the following analytical formula:

$$H = \left(\frac{Q}{300 K_s \sqrt{\frac{Z_m - Z_v}{5000}}} \right)^{3/5} . \tag{2.1}$$

The parameter Z_v represents the river downstream level (in meter), and Z_m the river

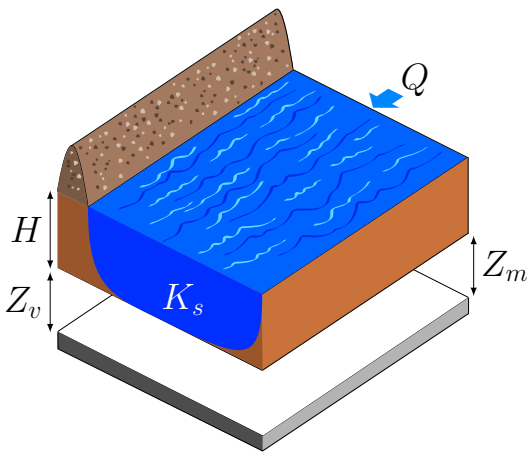
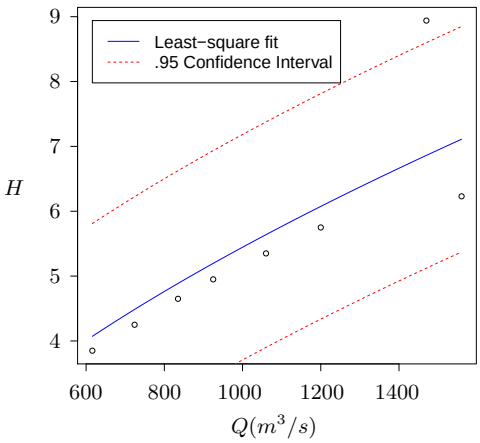


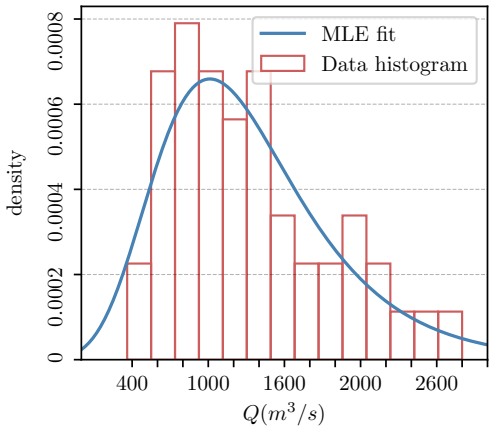
Figure 2.1: This scheme represents the height H of a river computed from the four parameters Q, K_s, Z_v, and Z_m using Eq. (2.1).

upstream level (in meter), Q is the maximal annual flowrate in (m^3/s) and K_s is the Manning-Strickler coefficient $(m^{1/3}/s)$ characterizing surface roughness and sinuosity of the riverbed. The constants 300 and 5000 are not adimensional and represent width and length (in m) of the river. The four parameters are depicted in Figure 2.1. The randomness of the variables is due to their spatio-temporal variability, our ignorance of their value, or some inaccuracies of their estimation. We suppose in the following that all four inputs are stochastically independent. In order to assess their distribution we have at our disposal a sample of 47 maximal annual flowrates measured on this river, and a sample of 8 couples flowrate/height established at the same localization.

Because Q represents the maximal annual flowrate, it has been associated to a Gumbel distribution $\mathcal{G}u(a, b)$ with mode a and scale factor b. This is justified by extreme value theory (Coles, 2001) as the Gumbel distribution belongs to the family of generalized extreme value distributions. The parameters of the Gumbel distribution have been estimated via a *plug-in* approach using the maximum likelihood estimator on the sample of 47 maximal annual flowrates (Pasanisi et al., 2012). The Manning-Strickler coefficient is tainted with epistemic uncertainty and is assumed to follow a normal distribution $\mathcal{N}(m, \sigma)$ with mean m and variance σ^2. The distribution parameters are estimated in a *plug-in* approach using a least-squared regression of Eq. (2.1) on the 8 couples flowrate/height. These data driven parameter estimations are depicted in Fig. 2.2. The distribution of Z_v and Z_m are supposed to be triangularly distributed with parameters determined by expert opinion. Finally, the distributions are truncated in order to circumvent non physically viable values. Those information are summarized in Table 2.1.



(a) Least squared adjustment of the normal distribution over the 8 couples flowrate/height.



(b) Maximum likelihood estimation of the Gumbel parameters over the 47 maximal annual flowrates.

Figure 2.2: Data driven estimation of the parameters of the flowrate distribution and the Mannin-Strieckler distribution.

In order to assess the dike's height for protecting the coast in case of a flood, one would like to compute a high quantile, here of order 95%. A direct *plug-in* estimation of the quantile is realized with a Monte-Carlo simulation of $N = 10^6$ code evaluations (Iooss and Lemaître, 2015). This is easily computed on any ordinary computer as the model is a simple analytical formula. The CDF of the output of the code is represented in Fig. 2.3, it gives a quantile of order 0.95 equals to 4.22 meters.

Table 2.1: Initial distribution of the 4 inputs of the hydraulic model.

Variable	Description	Distribution	Truncation
Q	Annual maximum flow rate	$\mathcal{G}u(1013, 558)$	$[160, 3580]$
K_s	Manning-Strickler coefficient	$\mathcal{N}(30, 7.5)$	$[12.55, 47.45]$
Z_v	Depth measure of the river downstream	$\mathcal{T}(49, 50, 51)$	–
Z_m	Depth measure of the river upstream	$\mathcal{T}(54, 54.5, 55)$	–

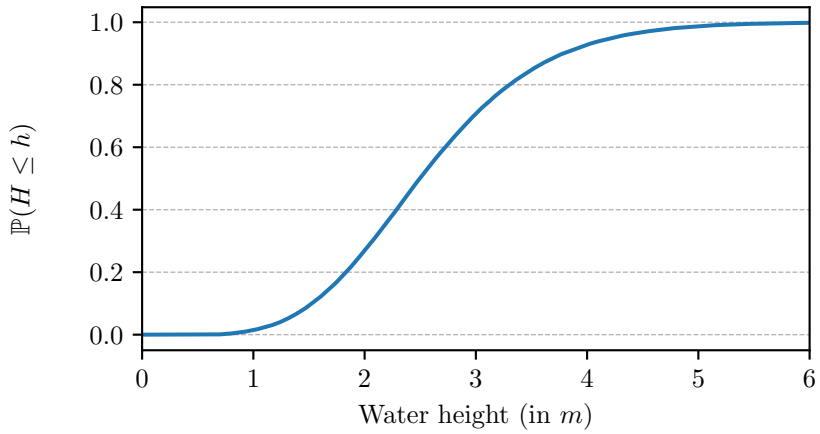


Figure 2.3: Distribution of the flood model with input distributions detailed in Table 2.1. Computation realized with a Monte-Carlo sample of size $n = 10^6$ simulations.

In order to better understand how the computer model behave, one would like to know which parameter is most influential on the river height. In global sensitivity analysis, Sobol' indices are the most popular sensitivity measures because they are easily interpreted: each Sobol' index represents a share of the output variance and the indices sum to 1 (Saltelli and Tarantola, 2002). To this purpose, we compute the first order and total Sobol' indices of the inputs of the flood model [Eq. (2.1)]. The asymptotically efficient pick-freeze estimator (Prieur and Tarantola, 2015) is used with an elementary Monte Carlo matrix of size 10^6, which gives a total cost of $N = 6 \times 10^6$ model runs and a standard deviation of the indices' estimation error smaller than 10^{-3}. As shown in Table 2.2 we observe that the variable Q is clearly more influential than the variable K_s whereas Z_v and Z_m appear to have almost no influence on the output.

Table 2.2: Sobol' indices estimates of the flood model inputs.

Sobol' indices	Q	K_s	Z_v	Z_m
First-order indices	0.764	0.206	0.007	0.002
Total indices	0.785	0.226	0.007	0.002

While this model has the advantage of being cheap and simple to study, a low dimensional and analytical computer experiment is not really representative of a real engineering study. In the next section, we study a code currently used at EDF for nuclear safety management analysis.

2.2 Thermal-Hydraulic model based on CATHARE code

The nuclear industry faces major issues as the ageing of facilities and the strengthening of regulatory authorities' requirements (Bucalossi et al., 2010). For instance, the operators have to study the loss of coolant accident resulting from a break of the primary loop of a pressurized water nuclear reactor. This scenario can be simulated using system thermal-hydraulic computer codes which include tens of physical parameters (Mazgaj et al., 2016; Sanchez-Saez et al., 2018).

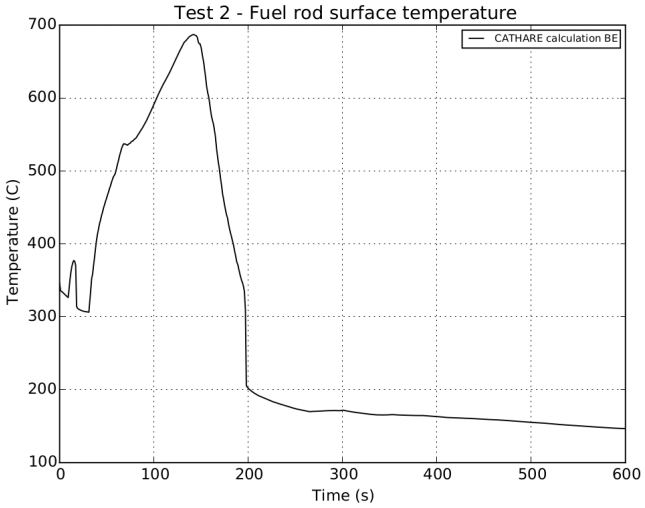


Figure 2.4: The thermal transient during an IBLOCA scenario simulated by the code CATHARE. The variable of interest (the output of our model) is the PCT of the fuel rod.

This industrial application concerns the study of the peak cladding temperature (PCT) of fuel rods in case of an intermediate break loss of coolant accident (IBLOCA) in the primary loop of a nuclear pressurized water reactor. This application has been studied in Iooss and Marrel (2019) and this section mainly gathers the main elements of their analysis. According to operation rules, this temperature must remain below a threshold to prevent any deterioration of the reactor state. The thermal-hydraulic transient caused by this accidental scenario is simulated with the code CATHARE2 (V2.5_3mod3.1) (Geffraye et al., 2011), providing a time depending temperature profile of the nuclear core assemblies (Mazgaj et al., 2016) (see Fig 2.4).

The thermal hydraulic model involves boundary and initial conditions, as well as many physical parameters (heat transfer coefficients, friction coefficients, critical flow rates, etc.) whose exact values are known with limited precision as they are calculated by the way of other quantities measured via small-scale physical experiments as shown in Fig. 2.5, or observed during periodic inspections (as for instance the characteristics of hydraulic pumps). The probability distributions of the inputs can be obtained from data, expert knowledge or recovered by solving inverse problems on an experimental database (Baccou et al., 2019).

The input uncertainties are propagated inside this model and the UQ objective consists of estimating a high-order quantile of the PCT (which is the model output). This quantile is interpreted as a pessimistic estimate of the PCT. Like any scientific approach, this methodology is based on assumptions, which regulatory authorities ask to evaluate the impact in exhibited results. Indeed, it is required that the operator

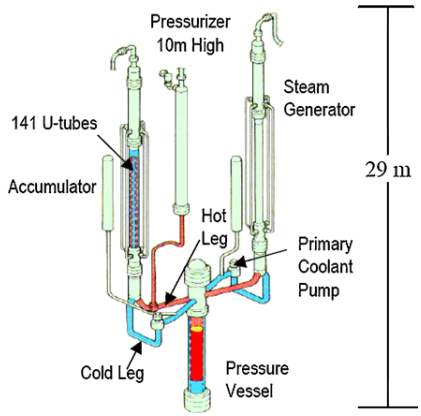


Figure 2.5: Large scale test facility located in Japan. This mock-up of a water pressured nuclear reactor has a 1/1 height scale and 1/48 volume scale. The pressure vessel is electrically heated and it is possible to carry out physical experiment of an IBLOCA scenario.

of the nuclear power plants conducts studies in such a way to ensure that actual risks are overestimated. By this "conservatism principle" they are bound to choose the most pessimistic assumption each time a modeling decision has to be made. In deterministic studies, this simply consists in taking the most penalizing values for each of the input variables. This way, the resulting computation is supposed to simulate a worst case scenario for the examined risk. It is however not straightforward to implement such a principle when the numerical code is complex with interactions between inputs and non-monotonic effects of inputs. Indeed, the most penalizing value depends on the mutual interaction between variables. It is even harder to extend this rationale to a UQ framework aiming at representing all potential scenarios with related occurrence plausibility.

The CATHARE code is expensive to compute and one run on an ordinary computer takes around 20 minutes. Nevertheless, considering both physical phenomena and dimensions of the system, our case study is a simplified model with respect to a realistic modeling of a reactor. With expensive computer codes, uncertainty propagation, risk measurement such as high quantile inference, or system robustness analysis are difficult tasks to carry out. In order to circumvent this problem, we propose to replace the expensive model with a Gaussian process regressor. In this use-case, the code takes $d = 27$ scalar input variables all considered uncertain. In order to build the metamodel, the number of simulations chosen is a tradeoff between the central processing unit (cpu) time required for each simulation and the number of input parameters. For uncertainty propagation and metamodel building purpose, it is a common rule to chose at least 10 times the dimension d of the input vector (Loeppky et al., 2009). We have at our disposal a sample of $n = 1000$ simulations performed using a space filling latin hypercube sampling (LHS) in dimension 27 (Fang et al., 2005).

A screening based on the HSIC dependence measure (Da Veiga, 2013) was performed on the $n = 1000$ learning simulations. The HSIC quantifies how dependent an input of the computer model X_i is to the output $Y = G(\mathbf{X})$. To do so, HSIC evaluates how close the joint distribution $\mu_{X_i,Y}$ is to the product $\mu_{X_i}\mu_Y$. It is defined as the squared Hilbert-Schmidt norm of the associated cross-covariance operator; this criterion equals zero if and only if the input is independent from the output. Notice that the screening based on the HSIC takes into account the whole output variability. From the estimated HSIC, De Lozzo and Marrel (2016) proposes several approaches to construct some statistical hypothesis test. In our application the hypothesis "$\mathcal{H}_0^{(k)}$: the

input X_i is independent from the output Y" is tested against its alternative "$\mathcal{H}_1^{(i)}$: X_i and Y are dependent". The significance level of this test, defined as the probability of rejecting H_0 when it is true, is set to 0.1. Nine inputs are found influential by rejecting the test. These selected inputs, designated as PII, are given in Table 2.3. The PII's distributions derive from the CIRCE method (Damblin and Gaillard, 2020;

Table 2.3: Nine most influential inputs of the CATHARE model (Iooss and Marrel, 2019).

Variable	Bounds	Initial distribution (truncated on bounds)	Physical input
$n°10$	$[0.1, 10]$	$\mathcal{LN}(0, 0.76)$	Interfacial friction in the hot legs
$n°22$	$[0, 12.8]$	$\mathcal{N}(6.4, 4.27)$	Wall-liquid friction undersaturated
$n°25$	$[11.1, 16.57]$	$\mathcal{N}(13.79, 2.05)$	Wall-liquid friction saturated
$n°2$	$[-44.9, 63.5]$	$\mathcal{U}(-44.9, 63.5)$	Heat transfer in the core
$n°12$	$[0.1, 10]$	$\mathcal{LN}(0, 0.76)$	Interfacial friction in the steam generators inlet plena
$n°9$	$[0.1, 10]$	$\mathcal{LN}(0, 0.76)$	Interfacial friction in the steam generators outlet plena
$n°14$	$[0.235, 3.45]$	$\mathcal{LN}(-0.1, 0.45)$	Interfacial friction in the core
$n°15$	$[0.1, 3]$	$\mathcal{LN}(-0.6, 0.57)$	Interfacial friction in the upper plenum
$n°13$	$[0.1, 10]$	$\mathcal{LN}(0, 0.76)$	Interfacial friction in the downcomer

De Crécy, 2001), where in a Bayesian setting the posterior distributions are computed with respect to an experimental database.

The Gp is thereupon build on the PII reduced space, conditioned from the available $n = 1000$ simulations. The remaining 18 inputs are fixed to their nominal values. We recall that a Gp is uniquely characterized by its mean and its autocovariance function (Rasmussen and Williams, 2005). We usually consider in computer experiments an anisotropic stationary covariance, here chosen as a Matérn 5/2 (Rasmussen and Williams, 2005, Chapter 4). The metamodel accuracy is evaluated using the predictivity coefficient Q^2 (Le Gratiet et al., 2017):

$$Q^2 = 1 - \frac{\sum_{i=1}^{n_{test}} (y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n_{test}} (y_i - \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} y_i)^2} ,$$

where $(y_i)_{1 \leq i \leq n_{test}}$ are the corresponding observed outputs of $(\mathbf{x}_i)_{1 \leq i \leq n_{test}}$ an input test sample, and $(\widehat{y}_i)_{1 \leq i \leq n_{test}}$ are the metamodel predictions. We use a leave-one-out strategy in order to perform the validation on the learning sample and obtain $Q_2 = 0.92$. We recall that the closer the predictivity coefficient is to one, the better is the prediction quality.

Once the predictive metamodel has been built, it can be used to perform uncertainty propagation and in particular, estimates quantiles. The most trivial approach

to evaluate a quantile with a Gp metamodel, known as the *plug-in* approach, simply consists in computing the quantile of the predictor of the metamodel. Because the expectation of the Gp is a cheap and deterministic function of the input, the *plug-in* approach provides a fast but deterministic expression of the quantile with no confidence intervals. Indeed, as it exploits MC methods, one knows the quantile estimation error, but not the metamodel-induced uncertainty. Moreover, for high quantiles, this method tends to substantially underestimate the true quantile (Cannamela et al., 2008).

To address this issue, Oakley (2004) proposes to take into account the covariance structure of the Gp metamodel. The quantile definition is therefore applied to the global Gp and yields a random variable, whose expectation can be considered as the quantile estimator and its variance an indicator of its prediction accuracy. This *full-Gp* approach thus naturally provides confidence intervals. In practice, the estimation of the quantile with the *full-Gp* approach (Iooss and Marrel, 2019) is based on stochastic simulations conditional on the learning sample of the Gp metamodel. Therefore, it requires simulations of several trajectories of the conditioned Gaussian process and their associated quantile. The 95%-quantile estimated with these two methodologies are explicited in Table 2.4

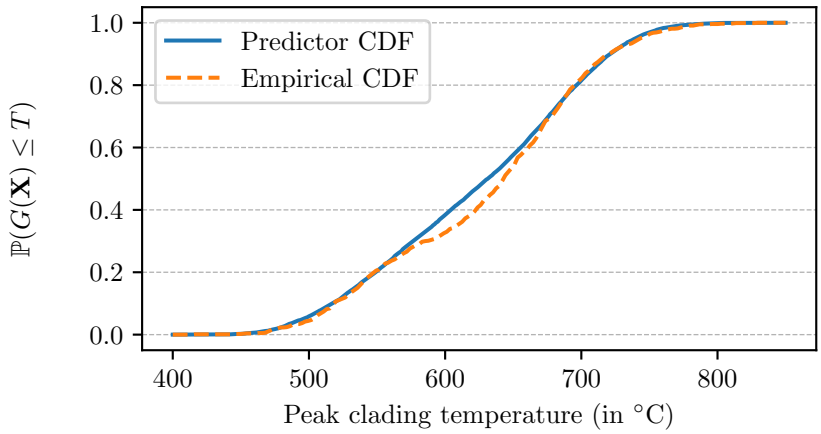


Figure 2.6: Distribution of the CATHARE model, the empirical distribution corresponds to the $n = 1000$ simulations, the predictor is the one of the Gp metamodel computed from a MC sample of size $n = 10^6$ simulations.

Table 2.4: 95%-quantile estimates of the CATHARE model and their 90% confidence interval (CI). The empirical is based on the $n = 1000$ learning sample, the *plug-in* and *full-Gp* use respectively the predictor of the metamodel and its complete covariance structure.

	Empirical	*plug-in*	*full-Gp*
Estimate	746.80	738.28	743.49
90%-CI	[743.7, 747.41]	-	[742.19, 745.10]

Conclusion

The role of this chapter was to present two different use-cases that will serve to illustrate our methodology. The flood model will be extensively used as it is computationally

cheap, low dimension, and quite easily interpreted because of its linearity. This toy model is an analytical deterministic function of four inputs, its simplicity will help understanding the behavior of our algorithms in many situations. The second model named CATHARE is more challenging to assess. Although it has been simplified in this work *via* its surrogate model, it remains representative of the phenomena encountered in a water pressurized nuclear reactor, but with a reduced simulation cost. The CATHARE model stays challenging enough regarding the optimization problem that we further look into. It is representative of a real engineering problem. Indeed, it may be represented as an input-output black-box computer experiment only known from a limited number of numerical simulations.

In both use-cases, the inputs represent physical parameters affected by uncertainties. These input variables are gathered under a random vector with given PDF detailed respectively in Tables 2.1 and 2.3. These PDF are set from the collection of information such as expert analysis or experimental database. It allows the computation of certain QoI, for instance a quantile of the output used to establish safety margins. The quality of the estimation depends on the allocated number of simulations, or on the metamodel predictivity accuracy. But most importantly, this QoI depends on the input distribution which is tainted by uncertainty as detailed in Chapter 1. The input PDF in Tables 2.1 and 2.3 are *one* possible modeling of the physical parameters considered by engineers to be reasonably adequate for representing the physical reality with respect to the lack of knowledge. But many different PDF suits the available information and could serve as input model. Hence, this exact modeling is nothing but a choice made by the analyst in order to pursue the study.

Consequently, in the next chapters we investigate an OUQ strategy to account for the uncertainty affecting the input distribution. Because it is hard to identify a unique PDF, we consider instead a class of probability measures that embodies all the uncertainty in the input distribution. The QoI, such as a quantile, is then maximized over this set. This optimum is interpreted by the analyst as robust with respect to the input unsure specification. For instance, if assessing the quantile of the flood model, OUQ of the quantile guaranties a safety margin robust with respect to the input PDF imprecision. This way, the dike's height is chosen independently of the input distribution choice as long as it belongs in the probability measure space that must be representative of the uncertainty. This non parametric optimization problem and its solution are detailed in the next Chapter.

3

Optimization Problem and Reduction

« *We're in a giant car heading towards a brick wall and everyone's arguing over where they're going to sit.* »

David Suzuki

Contents

Introduction

In this chapter we present the mathematical foundation of this manuscript. In Chapter 1 we introduced the general context that motivated this work, whose purpose is to propose a methodology to quantify the impact of an imprecise input probability distribution on a quantity of interest that derives from a computer model. In Section 1.5 we presented some work investigating a second level of uncertainty. In particular, the OUQ framework in Owhadi et al. (2013) establishes the basis of this thesis. We will extensively use the notations they introduced, as this section can be seen as a generalization of their OUQ framework.

This approach can be summarized in the following way: we consider a quantity of interest on the output of a computer model. This quantity of interest is a scalar quantity that should be statistically relevant for engineering studies. Some examples were presented in Section 1.4.3, such as a PoF, a quantile, or a sensitivity index. In the context of probabilistic modeling of uncertainties, the quantity of interest is seen as a real function taking for argument the input probability distribution. It is mathematically defined in the following way:

$$\begin{array}{rccc} \phi: & \mathcal{P}(\mathcal{X}) & \to & \mathbb{R} \\ & \mu & \mapsto & \phi(\mu)\ . \end{array} \tag{3.1}$$

In a second level UQ approach, we consider that the choice of the input distribution μ is tainted with uncertainty. In order to take into account how this lack of knowledge impacts the QoI, we are interested in a worst case scenario. That means that we investigate the input distribution that penalizes most the QoI. To that extent, we compute bounds for ϕ over a set of probability measures $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ that is chosen to be as representative as possible of the uncertainty affecting the input distribution. Therefore, we aim to solve the following optimization problem:

$$\sup_{\mu \in \mathcal{A}} \phi(\mu)\ . \tag{3.2}$$

This optimization problem is generally non tractable as the set of probability measures $\mathcal{A}$ is infinite dimensional. Fortunately, under some conditions specified in the following, it is possible to replace the optimization problem 3.2 by an equivalent optimization of the QoI ϕ over a finite dimensional subset Δ of $\mathcal{A}$. This result is known as the reduction Theorem.

In the following, this mathematical approach is taken as general as possible. We show in Section 3.1 how our problem is a particular application of primitive topological results. Primitive in the sense that the proofs rely on quite simple arguments, and that few assumptions are made on the measure set $\mathcal{A}$ and the function ϕ. The generality of the theoretical results is what makes the framework appealing as it can be adapted to a large number of different applications, some of them being extensively developed in this thesis. Then, in Section 3.2, we specify some measure space and detail their topological structure and their extreme points. Those measure spaces, called *moment classes*, are exceptional for both their interesting mathematical properties and their practical engineering applications that we will illustrate in Chapters 4 and 5. Section 3.3 is dedicated to presenting the construction of a product measure space, so that we progressively leave the abstract framework to connect it to the engineering modeling of computer models introduced in Chapters 1 and 2. Indeed, the input joint distribution $\mu = (\mu_1, \ldots, \mu_d)$ of a computer model will be assumed to belong to a product of measure spaces, in order to enforce the mutual independence of the inputs $\{\mu_i\}_i$.

In all generality, this chapter focuses on the optimization space, its construction and its properties, to optimize a QoI that is purposely left abstract. In Chapter 5, we will specify several interesting QoIs suited for practical applications.

3.1 Mathematical Results

3.1.1 Generalization of the reduction Theorem

In the following, we assume that $\mathcal{A}$ is a convex subset of a locally convex topological vector space Ω. We emphasize that, for the practical applications considered in this thesis, Ω will be the set of all probability measures $\mathcal{P}(\mathcal{X})$ on a Polish space $\mathcal{X}$. Therefore, the local convexity of Ω is always verified. Indeed, $\mathcal{P}(\mathcal{X})$ can be considered as a subset of the closed unit ball of the topological dual of $C_b(\mathcal{X})$, the set of all bounded continuous real valued functions on $\mathcal{X}$. Hence, $\mathcal{P}(\mathcal{X})$ inherits its topology which is known as the topology of the weak* convergence. The weak* topology is always locally convex since it is induced by seminorms (Aliprantis and Border, 2007, Chapitre 15).

The optimization of convex functions has been widely studied since several decades. While it often consists in minimizing a convex functional, it appears that convex function are also attractive for maximum search. On this matter, the Bauer maximum principle (Choquet et al., 1969) states that the maximum of a convex upper semicontinuous function defined on a compact convex subset of a locally convex topological vector space is reached on the extreme points. Our main result is basically an extension of the Bauer maximum principle in three principal directions:

- We focus on quasi-convex lower semicontinuous functions instead of convex upper semicontinuous functions. A quasi-convex function is defined on a convex subset $\mathcal{A}$ of a topological vector space and satisfies the following inequality, for all $x, y \in \mathcal{A}$ and $\lambda \in [0,1]$:
$$f(\lambda x + (1-\lambda)y) \leq \max\{f(x); f(y)\} \ .$$
Most of the properties of convex functions extend to quasi-convex functions so that they are a well tailored generalization of convexity for the purpose of optimization. We refer to Greenberg and Pierskalla (1971) for an excellent review on quasi-convex functions. An equivalent definition of quasi-convexity is that the lower level sets $\{x \in \mathcal{A} : f(x) \leq \alpha\}$ are convex for all $\alpha \in \mathbb{R}$. Adding the lower semicontinuity hypothesis, we require the function to have all its lower levels set both convex and closed. It is remarkable that we assume the lower semicontinuity of the function in order to maximize. In contrast, the upper semicontinuity required in the Bauer maximum principle is a more standard assumption for function maximization. We intend to clarify this hypothesis in the following.

- Let Δ be a given subset $\Delta \subsetneq \mathcal{A}$, we assume that for every element x in $\mathcal{A}$, there exists a probability measure $\boldsymbol{\nu}$ supported on Δ, such that
$$f(x) = \int_\Delta f(s)\, \mathrm{d}\boldsymbol{\nu}(s) \ , \text{ for any linear function } f \text{ of the topological dual of } \Omega. \tag{3.3}$$
We say that x is the barycenter of $\boldsymbol{\nu}$ when Eq. (3.3) is satisfied. Further, when Eq. (3.3) stands for all $x \in \mathcal{A}$, then Δ is called the generator of $\mathcal{A}$. No assumption of uniqueness of this representation is needed in this work, however if uniqueness stands then $\mathcal{A}$ is a simplex. This is also known as the integral (or barycentrical)

representation property of an element $x \in \mathcal{A}$. We assume in the following that Δ is the generator of $\mathcal{A}$. We emphasize that such an hypothesis on $\mathcal{A}$ includes the compact case. Indeed, from the Krein-Milman theorem, it holds that every point of a metrizable compact convex set is the barycenter of a probability measure carried by the extreme points (Choquet et al., 1969, Theorem 27.6). So that the extreme points generate any compact convex set. Nonetheless, here the existence of a generator is a strong assumption, not necessarily easy to check. In the following, we specify two different measure spaces non necessarily compact for which the integral representation holds.

- Finally, we study a product structure, meaning that the optimization set is constructed from the product of d convex subsets $\mathcal{A}_i$ of a locally convex topological vector space Ω_i, for $1 \leq i \leq d$. We aim at optimizing a quasi-convex lower semi-continuous function over a product space $\mathcal{A} \subset \prod_{i=1}^{d} \mathcal{A}_i$, where every marginal set $\mathcal{A}_i$ is supposed to be generated by a subset $\Delta_i \subset \mathcal{A}_i$. We will show that the product space $\prod_{i=1}^{d} \mathcal{A}_i$ inherits the integral representation property and that $\prod_{i=1}^{d} \Delta_i$ is exactly the generator of $\prod_{i=1}^{d} \mathcal{A}_i$. Hence, the "generator" property transfers to the product structure. However, the optimization is realized on some specific subset $\mathcal{A} \subset \prod_{i=1}^{d} \mathcal{A}_i$ that no longer satisfies the barycentrical property. We will study the extreme points of $\mathcal{A}$ when this subset possesses a very particular structure.

Our main theorem is therefore an extension of the Bauer maximum principle in three directions: the quasi-convexity of the optimization function replaces the convexity. The tensorization generalizes the structure of the optimization space, and the existence of an integral representation on marginal sets generalizes the compact case. By doing so, the framework we build is quite general and it unifies under one theory several second level uncertainty frameworks such as the robust Bayesian analysis and OUQ presented in Section 1.5. We also intend to present new applications of this framework illustrated in some practical examples in Chapter 5.

Let us begin with some simple yet important results and their proof before introducing the main reduction Theorem.

3.1.2 Preliminary results

Those first two lemmas are of great importance and gather the main arguments of our demonstration. The demonstration of the reduction Theorem 3.3 is a direct application of those Lemmas. We recall that the objective is to show that the optimum of a quasi-convex lower semicontinuous function on convex set $\mathcal{A}$ of a locally convex topological vector space can be found on the generator Δ of $\mathcal{A}$. In the following, if no reference is shown next to the result, it indicates that the proof are our own original work.

Lemma 3.1. *Let $\mathcal{A}$ be a convex subset of a locally convex topological vector space Ω. If any point $x \in \mathcal{A}$ is the barycenter of some probability measure $\boldsymbol{\nu}$ supported on $\Delta \subset \mathcal{A}$, then $\mathcal{A} \subset \overline{co}(\Delta)$, where $\overline{co}(\Delta)$ denotes the closed convex hull of Δ.*

Proof. Let $K = \overline{\text{co}}(\Delta)$. We suppose that there exists $x_0 \in \mathcal{A} \backslash K$. By applying the Hahn-Banach separation theorem, there exists a continuous linear map $l : \Omega \to \mathbb{R}$, such that $\sup_{x \in K} l(x) < C < l(x_0)$, for some real C. The lower level set

$$Z = \{x \in \mathcal{A} \mid l(x) \leq C\}$$

obviously contains Δ. Let $\boldsymbol{\nu}_0$ be the representative measure of x_0, supported on Δ so that $\boldsymbol{\nu}_0(Z) = 1$. Then,

$$l(x_0) = \int_Z l \, \mathrm{d}\boldsymbol{\nu}_0 \leq C < l(x_0) \ ,$$

leading to a contradiction. □

The next Lemma expresses the supremum of a quasi-convex function on a closed convex hull of some subset (Bereanu, 1972).

Lemma 3.2. *Let $\mathcal{A}$ be a convex set of a locally convex topological vector space. Let $\phi : \mathcal{A} \to \mathbb{R}$ be a quasi-convex lower semicontinuous function. If Y is an arbitrary subset of $\mathcal{A}$ and $\overline{\mathrm{co}}(Y)$ denotes its closed convex hull, then*

$$\sup_{\overline{\mathrm{co}}(Y)} \phi(x) = \sup_{Y} \phi(x) \ ,$$

Proof. If $\sup_Y \phi(x) = \infty$, there is nothing to prove. So we assume that $a := \sup_Y \phi(x)$ is finite. Let $Z_a = \{x \in \mathcal{A} \mid \phi(x) \leq a\}$.
Obviously, we have $Y \subset Z_a$. But Z_a is convex as ϕ is quasi-convex. Further, it is closed as ϕ is lower semicontinuous. Therefore, we have $\overline{\mathrm{co}}(Y) \subset Z_a$ because of the minimal property of the closed convex hull. Hence,

$$\sup_{\overline{\mathrm{co}}(Y)} \phi(x) \leq \sup_{Z_a} \phi(x) \leq a = \sup_{Y} \phi(x) \ ,$$

The converse is obvious. □

Remark. It is remarkable that we assume the lower semicontinuity of the function to maximize. In contrast, the upper semicontinuity hypothesis in the Bauer maximum principle is a more standard assumption for function maximization. The proof of Lemma 3.2 clarifies the role of the assumption. Indeed, it is used to enforce the closedness of the set $Z_a = \{x \in \mathcal{A} \mid \phi(x) \leq \sup_Y \phi(x)\}$. This argument differs from Choquet's demonstration of the Bauer maximum principle (Choquet et al., 1969, p.102). Therein, the authors study the closedness of the almost similar set $\{x \in \mathcal{A} \mid \phi(x) = \sup_Y \phi(x)\}$ for an upper semicontinuous function ϕ on the compact set $\mathcal{A}$. Doing so, the assumptions of compactness and upper semicontinuity in the Bauer maximum principle ensure that the optimum of the function ϕ is reached, which is not necessarily the case in our frame.

From Lemma 3.1 and 3.2 we establish the next Theorem. It is an analogous to the Bauer maximum principle, where the compactness assumption is replaced by the assumption of integral representation. The integral representation is always satisfied on compact sets, thanks to the Choquet representation theorem (Choquet et al., 1969, p.153). So that the next theorem is analogous to the Bauer maximum principle when compactness is assumed. This means that our assumption of integral representation is, in a way, more general:

Theorem 3.3. *Let $\mathcal{A}$ be a convex subset of a locally convex topological vector space Ω. We assume that every point $x \in \mathcal{A}$ is the barycenter of a probability measure $\boldsymbol{\nu}$ supported on $\Delta \subset \mathcal{A}$. Let $\phi : \mathcal{A} \to \mathbb{R}$ be a quasi-convex lower semicontinuous function. Then,*

$$\sup_{x \in \mathcal{A}} \phi(x) = \sup_{x \in \Delta} \phi(x) \ .$$

Proof. From Lemma 3.1

$$\mathcal{A} \subset \overline{\text{co}}(\Delta) \ .$$

Then applying Lemma 3.2 on the lower semicontinuous quasi-convex function ϕ, we obtain

$$\sup_{\mathcal{A}} \phi(x) \leq \sup_{\overline{\text{co}}(\Delta)} \phi(x) = \sup_{\Delta} \phi(x) \ .$$

The converse inequality is obvious. □

Theorem 3.3 gives a way to optimize a quasi-convex lower semicontinuous functional on an abstract space by reducing the problem to the extreme points. In the following we intend to construct products of convex sets $\{\mathcal{A}_i\}_i$ in order to relate to our practical application cases, in which the multidimensional input is endowed with a joint probability distribution. The next proposition highlights that the existence of an integral representation on each marginal space $\mathcal{A}_i$ also implies the existence of an integral representation on the product space $\prod_i \mathcal{A}_i$.

Proposition 3.4. *Let $\mathcal{A}_i$ be convex subsets of locally convex topological vector spaces Ω_i with generators Δ_i, such that the integral representation property holds on every marginal space $\mathcal{A}_i$. Then any element of $\prod_{i=1}^d \mathcal{A}_i$ is also the barycenter of a probability measure supported by $\prod_{i=1}^d \Delta_i$.*

Proof. Let $\mathbf{x}$ be in $\prod_{i=1}^d \mathcal{A}_i$, so that $\mathbf{x} = (x_1, \dots, x_d)$, with $x_i \in \mathcal{A}_i$. Because of the integral representation property of $\mathcal{A}_i$, there exists a probability measure $\boldsymbol{\nu_i}$ supported by Δ_i such that x_i is the barycenter of $\boldsymbol{\nu_i}$, i.e. $x_i = \int_{\Delta_i} s_i \, \mathrm{d}\boldsymbol{\nu}(s_i)$. We denote $\boldsymbol{\nu} = (\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_d)$, then for any function $\phi \in C_b(\prod_{i=1}^d \mathcal{A}_i)$, we have $\phi(\mathbf{x}) = \phi(x_1, \dots, x_d)$, but for every $x_2, \dots, x_d \in \prod_{i=2}^d \mathcal{A}_i$, the function $x_1 \mapsto \phi(x_1, \dots, x_d)$ belongs in $C_b(\mathcal{A}_1)$ so that

$$\begin{aligned} \phi(x_1, \dots, x_d) &= \phi\left(\int_{\Delta_1} s_1 \mathrm{d}\boldsymbol{\nu_1}(s_1), \ x_2, \dots, x_d \right) \ , \\ &= \int_{\Delta_1} \phi\left(s_1, x_2, \dots, x_d\right) \mathrm{d}\boldsymbol{\nu_1}(s_1) \ . \end{aligned}$$

Repeating this argument for each input, it holds from Fubini's Theorem

$$\begin{aligned} \phi(x_1, \dots, x_d) &= \int_{\Delta_1} \cdots \int_{\Delta_d} \phi(s_1, \dots, s_d) \, \mathrm{d}\boldsymbol{\nu}_1(s_1) \dots \mathrm{d}\boldsymbol{\nu}_d(s_d) \ , \\ &= \int_{\prod_{i=1}^d \Delta_i} \phi(\mathbf{s}) \, \mathrm{d}\boldsymbol{\nu}(\mathbf{s}) \ , \end{aligned}$$

as $\boldsymbol{\nu}_i$ is supported by Δ_i. Hence, $\mathbf{x}$ is the barycenter of $\boldsymbol{\nu}$ supported on $\prod_{i=1}^d \Delta_i$. □

The proposition above warrants the existence of the integral representation on the product set $\prod_{i=1}^d \mathcal{A}_i$ whenever any marginal set $\mathcal{A}_i$ possesses itself an integral representation property. In this situation there is no mathematical difference between the marginal set and the product set. Indeed $\prod_{i=1}^d \mathcal{A}_i$ is itself a convex set of a locally convex topological vector space with the barycentrical property. Therefore, the previous results and more specifically Theorem 3.3 holds for the product set. However, in the following of this chapter, we will focus on a subset $\mathcal{A}$ of the product set $\prod_{i=1}^d \mathcal{A}_i$ with different topological properties. Section 3.3 introduces in detail the construction of the subset $\mathcal{A}$, but first we leave behind the abstract topological notation and start introducing some practical measure spaces. These measure spaces are directly related to our target applications, where the quasi-convex semicontinuous function represents a QoI, defined as the output of a computer model, to be optimized over a measure space representing all the admissible input probability distributions.

3.2 Extreme Points Characterization

3.2.1 Moment Class

Assume now that $\mathcal{X}$ is a Polish space. We study a convex subspace of $\mathcal{P}(\mathcal{X})$, called the moment class. All measures in the moment class $\mathcal{A}^*$ satisfy *generalized* moment constraints [see Eq. (1.6)]. That is, a measure $\mu \in \mathcal{A}^*$ verifies $\mathbb{E}_\mu[\varphi_i] \leqq 0$, for measurable functions $\varphi_1, \ldots, \varphi_n \in C_b(\mathcal{X})$. Because $\mathcal{X}$ is Suslin, all measures $\mu \in \mathcal{P}(\mathcal{X})$ are regular. In fact, in all this section Polish space can be replaced by Suslin space. Hence, the following Theorem 3.5 due to Winkler holds.

Theorem 3.5 (Extreme points of moment classes (Winkler, 1988, p.586))**.** *Consider the space $\mathcal{P}(\mathcal{X})$ of measures defined on the σ-algebra of Borel sets on a Suslin space $\mathcal{X}$, and measurable functions $\varphi_1, \ldots, \varphi_n$ on $\mathcal{X}$. Then, for any measure μ in the moment class $\mathcal{A}^* = \{\mu \in \mathcal{P}(\mathcal{X}) \mid \mathbb{E}_\mu[\varphi_i] \leqq 0, 1 \leq i \leq n\}$, there exists a probability measure $\boldsymbol{\nu}$ supported on $\Delta^*(n)$ such that μ is the barycenter of $\boldsymbol{\nu}$, where*

$$\Delta^*(n) = \left\{ \mu \in \mathcal{A}^* \ : \ \mu = \sum_{i=1}^{n+1} \omega_i \delta_{x_i},\ \omega_i \geq 0,\ x_i \in \mathcal{X} \right\} .$$

is the set of discrete probability measures of $\mathcal{A}^*$ supported on at most $n+1$ points.

The case of equality in the constraints defining $\mathcal{A}^*$ is covered by this result (Winkler, 1988, p.586). Theorem 3.5 states that the extreme points of a class of measures with n *generalized* moment constraints are discrete measures supported on at most $n+1$ points of $\mathcal{X}$. In particular, in case that the system of constraint functions φ_i is degenerated, then the cardinal of the support is strictly lower than $n+1$.

The heuristic behind this theorem derives from the decomposition of any distribution into a "infinite convex combination" of Dirac masses, which can thus be seen as the elementary units of the probability set. In the simple example hereunder, we give a didactic illustration to demonstrate how a probability optimized over a measure space reaches its optimum in a discrete distribution. The reader can admit for now that a probability is a function that suits the reduction Theorem 3.3.

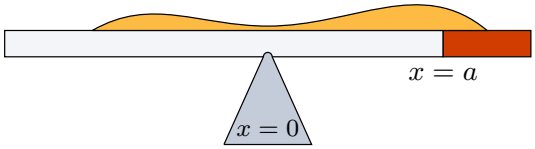
Illustration. This illustration is directly taken from Owhadi et al. (2013). Consider a seesaw with 1kg of sand on it, which can be arranged in any way so that it stays balanced around $x = 0$. One wishes to know how much mass can be poured on the region $x \geq a$, see Fig. 3.1a.

A mathematically equivalent problem can be stated as follows: we consider a probability measure μ (with mass 1), with mean $\mathbb{E}_\mu[X] = 0$. The optimization problem consists in maximizing the probability

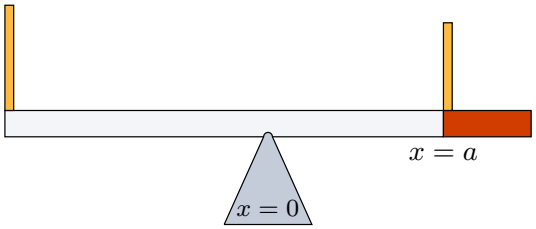
$$\sup_{\mu \in \mathcal{A}^*} \mathbb{P}_\mu(X \geq a) \ , \tag{3.4}$$

where $\mathcal{A}^* = \{\mu \in \mathcal{P}([l,u]) \mid \mathbb{E}_\mu[X] = 0\}$. We admit for now that this function satisfies the assumptions of the OUQ reduction Theorem 3.3. Therefore, it holds that the optimal measure is discrete, and supported on at most two Dirac masses. Let us examine with factual considerations how to maximize quantity Eq. (3.4).

First, consider the case $a \geq 0$, see the representation in Fig. 3.1. Then, one needs to maximize the arm-lever on the left-hand side in order to minimize the quantity of



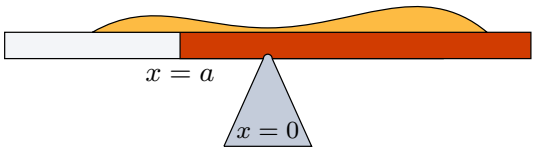
(a) A seesaw with 1kg of sand, balanced around $x = 0$.



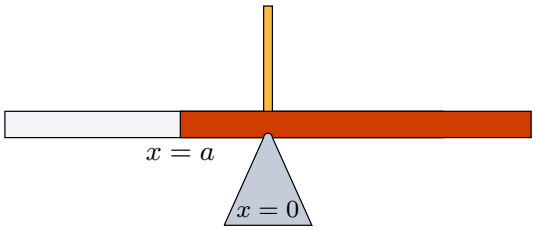
(b) Optimal solution for maximizing the quantity of sand in the region $x \geq a$.

Figure 3.1: In this problem and in the case $a \geq 0$, the optimal solution is supported on two Dirac masses.

sand used, therefore pour the sand on the extreme left of the seesaw. On the other hand to put the maximal amount of sand in the region $x \geq a$, one needs to minimize the arm-lever on the right-hand size by putting the remaining sand on $x = a$. The solution is as expected a discrete measure supported on two Dirac masses located in $x = a$ and in the left extremity of the seesaw.



(a) A seesaw with 1kg of sand, balanced around $x = 0$.



(b) Optimal solution for maximizing the quantity of sand in the region $x \geq a$.

Figure 3.2: In this problem and in the case $a \leq 0$, the optimal solution is supported on one Dirac mass.

Now, consider the case $a \leq 0$, see the representation Fig. 3.2. Then, obviously pouring all the sand on $x = 0$ is solution to the problem, see Fig. 3.2b. Therefore, the optimal measure is supported on exactly one Dirac mass.

In both situations $a \geq 0$ and $a \leq 0$ the optimal measure is discrete, and supported on at most two Dirac masses as stated by the reduction Theorem 3.5. Indeed, the measure set $\mathcal{A}^*$ is defined by $n = 1$ constraint on the mean. Also, notice, that the optimal solution is physically unrealistic.

The moment class is very interesting as we have a full characterization of its extreme points. We emphasize that the moment class is not parametric but its extreme points are, since they are discrete with finite support. Hence, it provides a convenient basis for the numerical optimization of a quasi-convex lower semicontinuous function. In the following we investigate a second measure space, defined as a subset of the moment class, and with very similar properties as we are going to demonstrate.

3.2.2 Unimodal Moment Class

In this section, $\mathcal{X}$ denotes an interval of the real line $\mathbb{R}$. Let μ be a probability distribution on $\mathcal{X}$, and let F_μ be its distribution function. The measure μ is said to

be unimodal with mode at $\theta \in \mathbb{R}$, whenever F_μ is convex on $]-\infty, \theta[$ and concave on $]\theta, +\infty[$. We denote $\mathcal{H}_\theta(\mathcal{X})$ the set of all probability measures on $\mathcal{X}$ which are unimodal at θ. The set $\mathcal{H}_\theta(\mathcal{X})$ is closed but not necessarily compact ($\mathcal{H}_\theta(\mathbb{R})$ is not compact, see (Bertin et al., 1997, p.19)). Clearly, given its distribution function, any uniform probability measure on an interval of the form $\text{co}(\{\theta, z\})$, $z \in \mathcal{X}$ (co is the convex hull) including the Dirac mass in θ, is unimodal at θ. We define the set $\mathfrak{U}_\theta(\mathcal{X})$ of these uniform probability measures

$$\mathfrak{U}_\theta(\mathcal{X}) = \{\mathcal{U}(\theta, z) \text{ is uniformly distributed on } \text{co}(\{\theta, z\}), z \in \mathcal{X}\} \ .$$

This set is closed in $\mathcal{P}(\mathcal{X})$ (Bertin et al., 1997, p.19). In this section, we are interested in the convex subset $\mathcal{A}^\dagger$ of unimodal measures satisfying *generalized* moment constraints $\mathbb{E}_\mu[\varphi_i] \leqq 0$, for measurable functions $\varphi_1, \ldots, \varphi_n$. This subspace is called an unimodal moment class and an equivalent of Theorem 3.5 holds:

Theorem 3.6 (Extreme points of unimodal class)**.** *Consider the space $\mathcal{H}_\theta(\mathcal{X})$ of unimodal measures on an interval $\mathcal{X} \subset \mathbb{R}$ with mode θ, and measurable functions $\varphi_1, \ldots, \varphi_n$ on $\mathcal{X}$. Then, for any measure μ in the unimodal moment class $\mathcal{A}^\dagger = \{\mu \in \mathcal{H}_\theta(\mathcal{X}) \mid \mathbb{E}_\mu[\varphi_i] \leqq 0, 1 \leq i \leq n\}$, there exists a probability measure $\boldsymbol{\nu}$ supported on $\Delta^\dagger(n)$ such that μ is the barycenter of $\boldsymbol{\nu}$. Here*

$$\Delta^\dagger(n) = \left\{ \mu \in \mathcal{A}^\dagger \mid \mu = \sum_{i=1}^{n+1} \omega_i \mathcal{U}(\theta, z_i),\ \omega_i \geq 0,\ z_i \in \mathcal{X} \right\} \ .$$

Elements of $\Delta^\dagger(n)$ are mixtures of at most $n+1$ uniform distributions supported on $\text{co}(\theta, z)$ for some $z \in \mathcal{X}$.

Proof. The proof is quite technical and as it is not of particular interest, its details are kept to the bare minimum. We mainly gather different results to prove our point, all precisely stated so that the interested reader might refer to it. Let $\mathcal{X}$ be an interval of $\mathbb{R}$, and $\mathcal{H}_\theta(\mathcal{X})$ the set of all probability measures on x which are unimodal at $\theta \in \mathcal{X}$. From (Bertin et al., 1997, p.19) we now that $\mathcal{H}_\theta(\mathcal{X})$ is a simplex, meaning that every probability measure in $\mathcal{H}_\theta$ is the barycenter of a uniform probability measure supported on $\mathfrak{U}_\theta(\mathcal{X})$. This implies that the uniform distribution in $\mathfrak{U}_\theta(\mathcal{X})$ are the elementary units of the set of unimodal distribution $\mathcal{H}_\theta$. Choquet (Choquet et al., 1969, p.160) used another definition of simplex. A convex subset K of a locally convex topological vector space is called a Choquet simplex if and only if the cone $\tilde{K} = \{(\lambda x, \lambda) : x \in K, \lambda > 0\}$ is a lattice cone in its own order (that is, the vector space $\text{span}(\tilde{K})$ is a lattice when its positive cone is taken to be $\tilde{K}$). The important point is that these two definitions are connected, and from (Winkler, 1985, p.47) it holds that each simplex is a Choquet simplex. Hence, $\mathcal{H}_\theta(\mathcal{X})$ is also a Choquet simplex. Moreover, in finite dimensional compact sets these two definitions are equivalent. We now define

$$\begin{aligned} K &= \{\mu \in \mathcal{H}_\theta(\mathcal{X}) : \varphi_1, \ldots, \varphi_n \text{ are } \mu\text{-integrable}\} \ , \\ S(\mu) &= \left(\int \varphi_1 \, d\mu, \ldots, \int \varphi_n \, d\mu \right) \ , \\ W &= S[K] \cap \prod_{i=1}^{n}]-\infty, 0] \text{ or } W = S[K] \cap \{(0, \ldots, 0)\} \ . \end{aligned}$$

It is already known that the extreme point set of $\mathcal{H}_\Theta(\mathcal{X})$ is precisely $\mathfrak{U}_\theta(\mathcal{X})$ as shown in (Bertin et al., 1997, p.19). Indeed, a classical result due to Khintchine Khintchine

(1938) states that uniform distributions constitute the "elementary units" of the set of all unimodal probability measures. However, we wish to known how the extreme points of $\mathfrak{U}_\theta(\mathcal{X})$ also characterize the extreme points of a convex subset of $\mathcal{H}_\Theta(\mathcal{X})$ defined as $\mathcal{A}^\dagger = S^{-1}[W] = \{\mu \in \mathcal{H}_\Theta(\mathcal{X}) \mid \mathbb{E}_\mu[\varphi_i] \leq 0,\ 1 \leq i \leq n\}$. Consequently, we wish to apply (Winkler, 1988, Proposition 2.1) that states that the set $\mathcal{A}^\dagger$ satisfies

$$\operatorname{ex}\{\mathcal{A}^\dagger\} \subset \Delta^\dagger(n) = \left\{\mu \in \mathcal{A}^\dagger \mid \mu = \sum_{i=1}^{n+1} \omega_i \mathcal{U}(\theta, z_i), \omega_i \geq 0, z_i \in \mathcal{X}\right\}. \tag{3.5}$$

However, in order to apply (Winkler, 1988, Proposition 2.1), it remains to be checked that K is linearly compact (meaning each of its line meets K in a compact interval). By the main theorem in Kendall (1962), it is sufficient to show that $\mathbb{R}_+ \cdot K$ is a lattice cone in its own order (condition (2^0) in the main theorem is an equivalent formulation of linear compactness as shown in the same reference on p.369). Of course the cone $\mathbb{R}_+ \cdot \mathcal{H}_\theta(\mathcal{X})$ is a lattice cone in its own order because it is a Choquet simplex. Now, choose $\mu \in \mathbb{R}_+ \cdot K$ and $\nu \in \mathbb{R}_+ \cdot \mathcal{H}_\theta(\mathcal{X})$ such that $(\mu - \nu) \in \mathbb{R}_+ \cdot K$, then $\nu \in \mathbb{R}_+ \cdot K$ since

$$\int |\varphi_i|\, d\nu \leq \int |\varphi_i|\, d\mu \quad \text{for every } i = 1, \ldots, n.$$

Hence, $\mathbb{R}_+ \cdot K$ is a hereditary subcone of $\mathbb{R}_+ \cdot \mathcal{H}_\theta(\mathcal{X})$ and consequently a lattice cone in its own order. This proves that K is linearly compact and that (Winkler, 1988, Proposition 2.1) applies. It follows that the set $\mathcal{A}^\dagger = S^{-1}[W] = \{\mu \in \mathcal{H}_\theta(\mathcal{X}) \mid \mathbb{E}_\mu[\varphi_i] \leq 0\}$ satisfies Eq. (3.5).

Now that the extreme points of $\mathcal{A}^\dagger$ are classified and observing that this set is closed with respect to the weak topology; Corollary 1 in Weizsäcker and Winkler (1979) concludes that every measure in $\mathcal{A}^\dagger$ has an integral representation supported on $\Delta^\dagger(n)$. □

As for Theorem 3.5, this theorem derives from the decomposition of any unimodal distribution into an "infinite convex combination" of uniform distributions (Khintchine, 1938), so that uniform distributions can be seen as the elementary units of the unimodal class. Indeed, historically the unimodal class was studied by Khintchine who revealed the fundamental relationship between the set of unimodal probability distributions and uniform probability measures. It was later demonstrated in Bertin and Theodorescu (1984) that the Khinchin Theorem may be considered as a non-compact form of the Krein-Milman Theorem (Choquet et al., 1969, p.105). In Sivaganesan and Berger (1989), one can find a first application of the optimization of a functional on a unimodal moment class in the context of robust Bayesian analysis, however no rigorous proof of the topological structure of the unimodal moment class was stated.

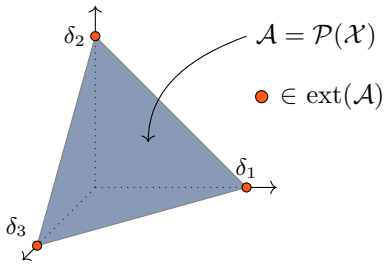
One could also be interested by the set of all symmetric unimodal probability measures. Let assume that $\mathcal{X}$ denotes an interval of the real line $\mathbb{R}$ and $\mathscr{S}_\theta(\mathcal{X})$ the set of all symmetric unimodal probability measures on $\mathcal{X}$ which are symmetric unimodal at θ. Then, the set of all uniform probability measures on an interval of the form $[\theta - z, \theta + z]$, $z \geq 0$ generates $\mathscr{S}_\theta(\mathcal{X})$. This result, almost identical to Theorem 3.6, just ensures that the extremes measures (the uniform distributions) are symmetric over the mode. The proper demonstration of this result would be similar to the one of Theorem 3.6, by noticing that a uniform probability measure on an interval of the form $[\theta - z, \theta + z]$, $z \geq 0$ constitutes the basic components of the symmetric uniform distribution set $\mathscr{S}_\theta(\mathcal{X})$. It reads as follows

Corollary 3.7. *Consider the space $\mathscr{S}_\theta(\mathcal{X})$ of symmetric unimodal measures on an interval $\mathcal{X} \subset \mathbb{R}$ with mode θ, and measurable functions $\varphi_1, \ldots, \varphi_n$ on $\mathcal{X}$. Then, for any measure μ in the symmetric unimodal moment class $\mathcal{A}^\diamond = \{\mu \in \mathscr{S}_\theta(\mathcal{X}) \mid \mathbb{E}_\mu[\varphi_i] \leqq 0, 1 \leq i \leq n\}$, there exists a probability measure $\boldsymbol{\nu}$ supported on $\Delta^\diamond(n)$ such that μ is the barycenter of $\boldsymbol{\nu}$, where*

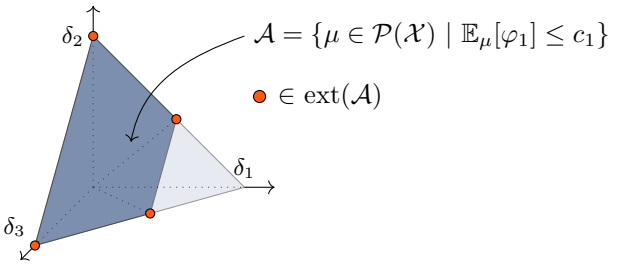
$$\Delta^\diamond(n) = \left\{ \mu \in \mathcal{A}^\dagger \mid \mu = \sum_{i=1}^{n+1} \omega_i \mathcal{U}(\theta - z_i, \theta + z_i),\ \omega_i \geq 0,\ z_i \geq 0 \right\} .$$

Elements of $\Delta^\diamond(n)$ are mixtures of at most $n+1$ uniform distributions that are symmetric over the mode.

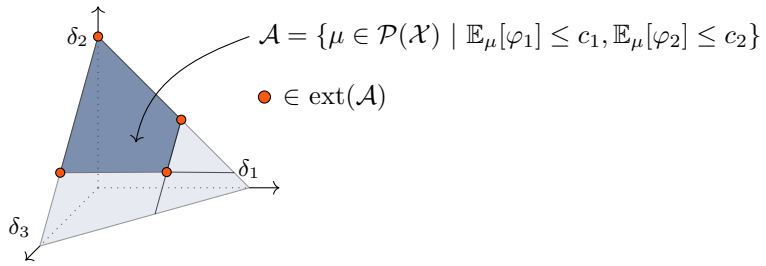
In the following, we propose a simple illustration to visualize why and how the extreme points of an (unimodal) moment class are convex combinations of either Dirac masses or uniform distributions. We emphasize that Dirac masses are the extreme points of the set of all probability measures $\mathcal{P}(\mathcal{X})$ on $\mathcal{X}$, which is basically the moment class without moment constraints. Likewise, uniform distributions are the extreme points of the set of all unimodal measures $\mathcal{H}_\theta(\mathcal{X})$, which is also the unimodal moment class without moment constraint. So that, the heuristic is that adding moment constraints to these sets transforms the initial extreme points into finite mixtures of them.



(a) Without constraints the extreme points of $\mathcal{P}(\mathcal{X})$ are Dirac masses.



(b) With one arbitrary constraint, the extreme points of $\mathcal{P}(\mathcal{X})$ are convex combinations of at most two Dirac masses.



(c) With two arbitrary constraints, the extreme points of $\mathcal{P}(\mathcal{X})$ are convex combinations of at most three Dirac masses.

Figure 3.3: Illustration of the extreme points of a probability space on a finite set, the extreme points being finite convex combinations of Dirac masses, the cardinality of which is the number of constraints plus one.

Illustration. This illustration is taken from (Sullivan, 2015, p.304). As stated in the proof, the optimization space is a simplex. We intend to give a geometrically simple interpretation of these extreme points. Consider $\mathcal{X} = \{1; 2; 3\}$, a discrete finite space composed of three elements. Therefore, it is well known that the space $\mathcal{P}(\mathcal{X})$

of all probability measures on $\mathcal{X}$ is isomorphic to the simplex of $\mathbb{R}^3$. Indeed, every point has three coordinates in its barycentrical representation on the extreme points $\{(1,0,0),(0,1,0),(0,0,1)\}$ that sum to one, such that it corresponds uniquely to the weights of a mixture distributed on $\{\delta_1;\delta_2;\delta_3\}$. We refer to Fig. 3.3 for a visual representation. The extreme points of $\mathcal{P}(\mathcal{X})$ are the Dirac masses supported on $\{1;2;3\}$, see Fig. 3.3a. When adding new constraints, the extreme points change but each new extreme point can be written as a convex combination of Dirac masses, see Fig. 3.3b and Fig. 3.3c.

The sets $\mathcal{A}^\dagger$ and $\mathcal{A}^*$ are very interesting in our applications. Indeed, measure spaces are non-obvious sets and it is generally not straightforward to exhibit their extreme points. Moreover, the *generalized* moment constraints are very generic. They provide a handy way of building measure spaces that encompass all the uncertainty affecting the input distribution of a computer model. Indeed, on one hand when no information is available in an input distribution then no constraint should be enforced. On the other hand, if the input distribution is well-known it is possible to enforce constraints on a large number of moments. Therefore, tuning the number of constraints allows to adjust the size of the optimization space, in order to obtain bounds on the QoI that are as tight as possible, while still reflecting its uncertainty in a realistic way.

We recall that, in our applications, the input of the numerical experiment is modeled as a random vector. Because we assume the mutual independence between each of its components, we now study the product of measure spaces. In what follows, we investigate the reduction theorem on such a structure.

3.3 Product Structure

3.3.1 Construction of product measure spaces

This section gives all the notations and details for the construction of the optimization set that will be used in the next chapters.

The moment class and the unimodal moment class, introduced in Section 3.2 have similar properties, so that they are gathered under the same notation. Indeed, we enforce *generalized* moment constraints in both cases. The difference lies in the unimodality of the measures of $\mathcal{A}^\dagger$, while $\mathcal{A}^*$ can contain any Radon measure. The difference between Theorems 3.5 and 3.6 is that the nature of the extreme points are somewhat different. Indeed, the generator of the unimodal moment class $\mathcal{A}^\dagger$ is the set of finite convex combination of uniform distributions, while the generator of the moment class $\mathcal{A}^*$ is the set of finite convex combination of Dirac masses. The notation are already cumbersome, so that we voluntarily do not develop further in this thesis the set of symmetric unimodal moment class to avoid overburdening. However, we will make some comment on the nature of the moment class that can be easily interchanged inside the construction of the product space.

To begin with, we first detail the construction of the product measure space. For $p \in \{1,\dots,d\}$, let

$$\mathcal{X} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_p \tag{3.6}$$

be a product of p Polish spaces $\mathcal{X}_1,\dots,\mathcal{X}_p$, and $d-p$ real intervals $\mathcal{X}_{p+1},\dots,\mathcal{X}_d$, so that it corresponds respectively to p moment classes (Section 3.2.1) and $d-p$ unimodal moment classes (Section 3.2.2). Precisely, given some real numbers $\theta_i \in \mathcal{X}_i$,

for $p < i \leq d$ and some measurable functions $\varphi_i^{(j)} : \mathcal{X}_i \to \mathbb{R}$ for $1 \leq j \leq N_i$ and $1 \leq i \leq d$, we construct d measure spaces with the integral representation property as follows:

$$\begin{aligned} \mathcal{A}_i = \mathcal{A}_i^* &= \left\{ \mu_i \in \mathcal{P}(\mathcal{X}_i) \mid \mathbb{E}_{\mu_i}[\varphi_i^{(j)}] \leqq 0 \text{ for } j = 1, \dots, N_i \right\} \text{ for } 1 \leq i \leq p \ , \\ \mathcal{A}_i = \mathcal{A}_i^\dagger &= \left\{ \mu_i \in \mathcal{H}_{\theta_i}(\mathcal{X}_i) \mid \mathbb{E}_{\mu_i}[\varphi_i^{(j)}] \leqq 0 \text{ for } j = 1, \dots, N_i \right\} \text{ for } p < i \leq d \ . \end{aligned} \tag{3.7}$$

Therefore, we leave behind the asterisk or the dagger notation as $\mathcal{A}_i$ designates either a moment class on a Polish space, or an unimodal moment class on an interval as presented in Section 3.2. We denote by $\Delta_{N_i} \subset \mathcal{A}_i$, the generator of the space $\mathcal{A}_i$, as defined in Section 3.2. Summarizing, we have

$$\begin{aligned} \Delta_i(N_i) = \Delta_i^*(N_i) &= \left\{ \mu_i \in \mathcal{A}_i \mid \mu_i = \sum_{k=1}^{N_i+1} \omega_i^{(k)} \delta_{x_i^{(k)}},\ x_i^{(k)} \in \mathcal{X}_i \right\} \text{ for } 1 \leq i \leq p \ , \\ \Delta_i(N_i) = \Delta_i^\dagger(N_i) &= \left\{ \mu_i \in \mathcal{A}_i \mid \mu_i = \sum_{k=1}^{N_i+1} \omega_i^{(k)} \mathcal{U}(\theta_i, z_i^{(k)}),\ z_i^{(k)} \in \mathcal{X}_i, \right\} \text{ for } p < i \leq d \ . \end{aligned} \tag{3.8}$$

where every weight $\omega_i^{(k)}$ is in $[0, 1]$ because the measures are probability measures. With these definitions and as discussed in the previous section, any measure $\mu_i \in \mathcal{A}_i$ is the barycenter of a probability measure supported on $\Delta_i(N_i)$, the set of convex combinations of at most $N_i + 1$ Dirac masses or uniform distributions. We recall that from Proposition 3.4 this construction implies that any element of $\prod_{i=1}^d \mathcal{A}_i$ is the barycenter of a measure supported on $\prod_{i=1}^d \Delta_i(N_i)$.

Remark 1. The product spaces $\prod_{i=1}^d \mathcal{A}_i$ and $\prod_{i=1}^d \Delta_i(N_i)$ are equipped with the product σ-algebra, not to be confused with the Borel σ-algebra of the product.

Remark 2. One could replace any unimodal moment class $\mathcal{A}_i^\dagger$ by a symmetric unimodal moment class $\mathcal{A}_i^\diamond$. Consequently, in the extreme points product set, one should also replaced the nature of the mixture of uniform by swapping $\Delta_i^\dagger(N_i)$ in $\Delta_i^\diamond(N_i)$. Those two moment classes can be interchanged anytime in this thesis. However, for clarity reasons we do not develop the symmetric unimodal moment class further.

The following definition highlights the meaning of quasi-convexity and lower semi-continuity of a function on a product space.

Definition 1. A function $\phi : \prod_{i=1}^d \mathcal{A}_i \to \mathbb{R}$ is said to be marginally quasi-convex (respectively *lsc*) if for all $\{\mu_k \in \mathcal{A}_k,\ k \neq i\}$, the function $\mu_i \mapsto \phi(\mu_1, \dots, \mu_d)$ is quasi convex (respectively *lsc*) for the topology of $\mathcal{A}_i$.

Notice that, if ϕ is globally quasi-convex (respectively *lsc*) for the product topology then it is marginally quasi-convex (respectively marginally *lsc*). Indeed, if ϕ is globally *lsc*, then $\{\mu \in \prod_{i=1}^d \mathcal{A}_i \mid \phi(\mu_1, \dots, \mu_d) > \alpha\}$ is open for all α and as the canonical projections are open maps, $\{\mu_i \in \mathcal{A}_i \mid \phi(\mu_1, \dots, \mu_d) > \alpha\}$ is also open. The claim on quasi-convexity is obvious. Having defined properly the product spaces, we present in the following the main result of this thesis, that is the reduction Theorem of quasi-convex lower semi continuous function on a product of measure spaces.

3.3.2 Reduction Theorem

Let us consider a subset $\mathcal{A}$ of the product space $\prod_{i=1}^d \mathcal{A}_i$. Because $\mathcal{A}$ is a subset, we do not have any information on its topological structure, even though the extreme points

of $\prod_{i=1}^d \mathcal{A}_i$ are known. Indeed, we recall that from Proposition 3.4, any element of the product space $\prod_{i=1}^d \mathcal{A}_i$ is the barycenter of a measure supported on $\prod_{i=1}^d \Delta_i(N_i)$, that implies that the topological properties of the product space $\prod_{i=1}^d \mathcal{A}_i$ are similar to the ones of any marginal space $\mathcal{A}_i$ for $1 \leq i \leq d$. Hence, in all generality the reduction theorem 3.3 can be directly applied to the product space $\prod_{i=1}^d \mathcal{A}_i$.

Following the construction of the product space detailed in Section 3.3.1, any measure $\mu_i \in \mathcal{A}_i$ belongs to either a moment space or an unimodal moment space. Hence, μ_i always satisfies N_i *generalized* moment constraints, and the product measure $\mu = (\mu_1, \dots, \mu_d) \in \prod_{i=1}^d \mathcal{A}_i$ realizes only marginal moment constraints. We add one layer of complexity by enforcing additional *generalized* moment constraints directly on the product measure $\mu \in \prod_{i=1}^d \mathcal{A}_i$, such that, we have

$$\mathbb{E}_\mu[\varphi^{(j)}] \leq 0 \text{ for } N \text{ measurable functions } \varphi^{(j)} : \mathcal{X} \to \mathbb{R},\ 1 \leq j \leq N\ .$$

We emphasize that the notation $\mathbb{E}_\mu = \mathbb{E}_{\mu_1,\dots,\mu_d}$ is quite unconventional as it corresponds to the expectation with respect to the product measure. The subset of interest $\mathcal{A} \subset \prod_{i=1}^d \mathcal{A}_i$, which happens to be the optimization space, concatenates the moment constraints on both the marginal distributions and the product measure. It writes

$$\mathcal{A} = \left\{ \mu \;\middle|\; \begin{array}{l} \mu = (\mu_1, \dots, \mu_d) \in \prod_{i=1}^d \mathcal{P}(\mathcal{X}_i), \text{ satisfying the constraints:} \\ \text{for } \varphi^{(j)} : \mathcal{X} \to \mathbb{R} \text{ and } \varphi_i^{(j)} : \mathcal{X}_i \to \mathbb{R}, \\ \quad \bullet\ \mathbb{E}_\mu[\varphi^{(j)}] \leqq 0 \text{ for } j = 1, \dots, N, \\ \quad \bullet\ \mathbb{E}_{\mu_i}[\varphi_i^{(j)}] \leqq 0 \text{ for } j = 1, \dots, N_i \text{ and } i = 1, \dots, d \end{array} \right\}, \tag{3.9}$$

or equivalently

$$\mathcal{A} = \left\{ \mu = (\mu_1, \dots, \mu_d) \in \prod_{i=1}^d \mathcal{A}_i \mid \mathbb{E}_\mu[\varphi^{(j)}] \leqq 0,\ 1 \leq j \leq N \right\}, \tag{3.10}$$

as the marginal constraints are embeded in the marginal measure spaces $\mathcal{A}_i$. To resume the interpretation of subset $\mathcal{A}$, any measure μ in $\mathcal{A}$ is a product measure written $(\mu_1, \dots, \mu_d)$ supported on $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_d$. We enforce numerous constraints, the N first *generalized* moment constraints are set on the product measure μ. Moreover, N_i other *generalized* moment constraints are carried by all the marginal measure μ_i, for $1 \leq i \leq d$.

The following reduction theorem is one of our main result, and will be extensively used all along this thesis. It shows that the set $\mathcal{A}$ is very interesting as it remains possible to characterize its extreme points. They are product of mixture of either Dirac masses or uniform distributions, but with cardinality depending both on the marginal constraints and the joint constraints. The reduction Theorem reads:

Theorem 3.8 (Reduction Theorem (Stenger et al., 2019a))**.** *Let $\mathcal{A}, \mathcal{A}_i$ and $\Delta_i(N_i)$ be defined respectively in Eq. (3.10), Eq. (3.7) and Eq. (3.8). Let $\phi : \prod_{i=1}^d \mathcal{A}_i \to \mathbb{R}$ be a marginally quasi-convex lower semicontinuous function. Then,*

$$\sup_{\mu \in \mathcal{A}} \phi(\mu) = \sup_{\substack{(\mu_1,\dots,\mu_d) \in \prod \mathcal{A}_i \\ \mathbb{E}_{\mu_1,\dots,\mu_d}[\varphi^{(j)}] \leqq 0 \\ 1 \leq j \leq N}} \phi(\mu) \quad = \quad \sup_{\substack{(\mu_1,\dots,\mu_d) \in \prod \Delta_i(N_i+N) \\ \mathbb{E}_{\mu_1,\dots,\mu_d}[\varphi^{(j)}] \leqq 0 \\ 1 \leq j \leq N}} \phi(\mu)$$

Proof. We recall that $\mathcal{A} := \prod_{i=1}^d \mathcal{A}_i$ is a product of measure spaces, where $\mathcal{A}_i$ is either a moment space or an unimodal moment space. Therefore, each measure $\mu_i \in \mathcal{A}_i$ satisfies

N_i moment constraints. More precisely, for the measurable functions $\varphi_i^{(j)} : X_i \to \mathbb{R}$, we have $\mathbb{E}_{\mu_i}[\varphi_i^{(j)}] \leqq 0$ for $1 \leq j \leq N_i$ and $1 \leq i \leq d$.

Moreover, in Theorem 3.8, we also enforce constraints on the product measure $\mu = (\mu_1, \dots, \mu_d) \in \prod_{i=1}^d \mathcal{A}_i$, such that, for each measurable function $\varphi^{(j)} : \mathcal{X} \to \mathbb{R}$, $1 \leq j \leq N$, we have $\mathbb{E}_\mu[\varphi^{(j)}] \leqq 0$.

Let ϕ be a marginally quasi-convex lower semicontinuous function. Then,

$$\sup_{\substack{(\mu_1,\dots,\mu_d)\in\prod_{i=1}^d \mathcal{A}_i \\ \mathbb{E}_{\mu_1,\dots,\mu_d}[\varphi^{(j)}]\leqq 0 \\ 1\leq j\leq N}} \phi(\mu) = \sup_{\mu_1\in\mathcal{A}_1} \dots \sup_{\mu_{d-1}\in\mathcal{A}_{d-1}} \sup_{\substack{\mu_d\in\mathcal{A}_d \\ \mathbb{E}_{\mu_1,\dots,\mu_d}[\varphi^{(j)}]\leqq 0 \\ 1\leq j\leq N}} \phi(\mu_1, \dots, \mu_d) \ .$$

Now, for fixed $\mu_1, \dots \mu_{d-1}$, we may rewrite the jth constraint on the joint distribution as

$$\mathbb{E}_{\mu_1,\dots,\mu_d}[\varphi^{(j)}] = \mathbb{E}_{\mu_d}\left[\mathbb{E}_{\mu_1,\dots,\mu_{d-1}}[\varphi^{(j)}]\right] \leqq 0 \ , \quad \text{for } 1 \leq j \leq N \ .$$

This is a moment constraint on the measure μ_d. This means that there are in total N_d+N moment constraints enforced on μ_d. Therefore, μ_d has an integral representation supported on the set of convex combination of $N_d + N + 1$ extreme points (which are either Dirac masses, or uniform distributions). Hence, for fixed $\mu_1, \dots, \mu_{d-1}$, and because the function $\mu_d \longmapsto \phi(\mu_1, \dots, \mu_d)$ is quasi-convex and lower semicontinuous. We have from Theorem 3.3

$$\sup_{\substack{\mu_d\in\mathcal{A}_d \\ \mathbb{E}_{\mu_1,\dots,\mu_d}[\varphi^{(j)}]\leqq 0 \\ 1\leq j\leq N}} \phi(\mu_1, \dots, \mu_d) = \sup_{\substack{\mu_d\in\Delta_d(N_d+N) \\ \mathbb{E}_{\mu_1,\dots,\mu_d}[\varphi^{(j)}]\leqq 0 \\ 1\leq j\leq N}} \phi(\mu_1, \dots, \mu_d).$$

So that,

$$\sup_{\substack{(\mu_1,\dots,\mu_d)\in\prod_{i=1}^d \mathcal{A}_i \\ \mathbb{E}_{\mu_1,\dots,\mu_d}[\varphi^{(j)}]\leqq 0 \\ 1\leq j\leq N}} \phi(\mu) = \sup_{\substack{(\mu_1,\dots,\mu_d)\in\mathcal{A}_1\times\dots\times\mathcal{A}_{d-1}\times\Delta_d(N_d+N) \\ \mathbb{E}_{\mu_1,\dots,\mu_d}[\varphi^{(j)}]\leqq 0 \\ 1\leq j\leq N}} \phi(\mu) \ .$$

Consequently, the last component of μ can be replaced by some element of $\Delta_d(N_d+N)$. By repeating this argument to the other components, the result follows. □

The reduction theorem in this form states that the supremum of the quasi-convex function over the subset $\mathcal{A}$ of the product space can be computed once again considering only the d-fold product of finite convex combinations of extreme points of the marginal spaces. That is, consisting in finite convex combinations of either Dirac masses or uniform distributions. However, the difference lies in the cardinality of the finite mixtures. Indeed, if N_i constraints are enforced on every marginal measure μ_i and N constraints are enforced on the product measure μ, then the optimum is to be found marginally on convex combination of at most N_i+N+1 Dirac masses or uniform distributions for respectively the moment classes and unimodal moment classes.

Notice also that Theorem 3.8 extends the work of Owhadi et al. (2013). Indeed, in this paper the authors were the first to propose the reduction Theorem on a product space. Nevertheless, the optimization considered therein is restricted only to product of moment classes and did not include unimodal moment classes. Moreover, the optimized functional in Owhadi et al. (2013) is an affine function of the measure. This is a very particular case of our framework. Nevertheless, affine functions of a probability measure are useful, some of their properties are now discussed.

3.3.3 Relaxation of the lower semicontinuity assumption

From the very general and abstract Theorem 3.3, we developed a more specific reduction theorem 3.8 by specifying the optimization space. We focused on some specific measure spaces, namely product of moment classes and unimodal moment classes (Section 3.2). The optimization function was supposed in both theorems to be quasi-convex lower semicontinuous. We now consider some particular cases. By assuming certain regularity properties of the optimization function, we show that the lower semicontinuity hypothesis can sometimes be bypassed.

3.3.3.1 Measure affine functions

So far the function to be optimized in Theorem 3.8 was assumed to be both lower semicontinuous and quasi-convex. It appears that quasi-convexity covers a large class of functionals fitting numerous application cases. However, the lower semicontinuity assumption is not always satisfied in practical applications. So that, it is very interesting to relax this hypothesis.

In this section, we study some specific classes of functions that are called measure affine (Winkler, 1988). These functions and their optimization on product measure spaces have been already studied in Owhadi et al. (2013). We present hereunder proofs of reduction theorems for some specific functionals. The arguments differ from the ones in Section 3.1.2. We use hereunder the same abstract topological object from Section 3.1.2 but the reader may always picture $\mathcal{A}$ to be a measure space as in Section 3.3.1.

Definition 2. Let $\mathcal{A}$ be a convex subset of a locally convex topological vector space with generator Δ. A function ϕ is called measure affine whenever ϕ is integrable with respect to any probability measure $\boldsymbol{\nu}$ on Δ with barycenter $\mu \in \mathcal{A}$ and ϕ fulfills the following barycentrical formula

$$\phi(\mu) = \int_{\Delta} \phi(s)\, d\boldsymbol{\nu}(s) \ .$$

Remark. A measure affine function always satisfies for $\mu, \pi \in \mathcal{A}$ and $\lambda \in [0,1]$:

$$\phi(\lambda\mu + (1-\lambda)\pi) = \lambda\phi(\mu) + (1-\lambda)\phi(\pi) \ .$$

Hence, it is both quasi-convex and quasi-concave (such functions are also called quasi-affine).

We have already emphasized that the quasi convexity assumption is usually verified in practical applications. However, the optimization function often lacks lower semicontinuity. In the following, we show that the optimum of a measure affine function can still be computed only on the generator of the optimization set, independently of the regularity of the function ϕ. Theorem 3.9 is similar to Theorem 3.3, but with the assumptions of quasi-convexity and lower semicontinuity replaced by the measure affine property.

Theorem 3.9 (Reduction Theorem for Measure Affine Functions)**.** *Let $\mathcal{A}$ be a convex subset of a locally convex topological vector space with generator Δ. For any measure affine functional ϕ on $\mathcal{A}$ we have*

$$\sup_{\mu \in \mathcal{A}} \phi(\mu) = \sup_{\mu \in \Delta} \phi(\mu) \ ,$$

and,

$$\inf_{\mu\in\mathcal{A}} \phi(\mu) = \inf_{\mu\in\Delta} \phi(\mu) \ .$$

Proof. The proof is given for the supremum, but it is similar for the infimum. Given $\mu \in \mathcal{A}$, the integral representation property states that there exists a probability measure $\boldsymbol{\nu}$ supported on Δ such that μ is the barycenter of $\boldsymbol{\nu}$. Therefore,

$$\phi(\mu) = \int_{\Delta} \phi(s) d\boldsymbol{\nu}(s) \leq \sup_{s\in\Delta} \phi(s)$$

for any $\mu \in \mathcal{A}$. Hence, $\sup_{\mu\in\mathcal{A}} \phi(\mu) \leq \sup_{\mu\in\Delta} \phi(\mu)$, the converse inequality is obvious as $\Delta \subset \mathcal{A}$. □

Obviously, it is possible to extend Theorem 3.9 by enforcing moment constraints on the product measure as in Theorem 3.8. The proof of Theorem 3.10 is available in Owhadi et al. (2013) and relies on the same recursive argument as the one used for Theorem 3.8. In the same reference, the authors do not include the unimodal moment class, but it does not affect the scheme of proof.

Corollary 3.10 ((Owhadi et al., 2013, p.301))**.** *Let $\mathcal{A}, \mathcal{A}_i$ and $\Delta_i(N_i)$ be defined respectively in Eq. (3.10), Eq. (3.7) and Eq. (3.8). Let $\phi : \prod_{i=1}^{d} \mathcal{A}_i \to \mathbb{R}$ be measure affine. Then,*

$$\sup_{\mu\in\mathcal{A}} \phi(\mu) = \sup_{\substack{(\mu_1,\dots,\mu_d)\in\prod \mathcal{A}_i \\ \mathbb{E}_{\mu_1,\dots,\mu_d}[\varphi^{(j)}]\leqq 0 \\ 1\leq j\leq N}} \phi(\mu) \quad = \quad \sup_{\substack{(\mu_1,\dots,\mu_d)\in\prod \Delta_i(N_i+N) \\ \mathbb{E}_{\mu_1,\dots,\mu_d}[\varphi^{(j)}]\leqq 0 \\ 1\leq j\leq N}} \phi(\mu)$$

3.3.3.2 Ratio of measure affine functions

Theorem 3.9 and Theorem 3.10 states that the supremum of a measure affine functional can be searched only on the generator of the measure space $\mathcal{A}$. We examine some transformations of measure affine functions for which this theorem stays true and for which lower semicontinuity remains a non-necessary condition. While in the previous section we investigated measure affine functions, this result can be generalized to a larger class of functions. Hereunder, we show that Theorem 3.9 still holds for ratios of measure affine functions. This kind of functions is particularly interesting as they appear in many practical quantities of interest, as we will develop later in Chapter 5. In particular the integration of the robust Bayesian framework into the OUQ framework developed in Section 5.6 falls directly within the realm of the following proposition.

Proposition 3.11. *Let $\mathcal{A}$ be a convex set of measures with generator Δ. Let ϕ and ψ be two measure affine functions, with ϕ taking positive values $\psi > 0$. Then*

$$\sup_{\mu\in\mathcal{A}} \frac{\phi}{\psi} = \sup_{\mu\in\Delta} \frac{\phi}{\psi} \ ,$$

and,

$$\inf_{\mu\in\mathcal{A}} \frac{\phi}{\psi} = \inf_{\mu\in\Delta} \frac{\phi}{\psi} \ .$$

Proof. The proof is given for the supremum, but it is similar for the infimum. It is inspired from an existing result in (Sivaganesan and Berger, 1989, p.887). Given, $\mu \in$

$\mathcal{A}$, the integral representation property states that there exists a probability measure $\boldsymbol{\nu}$ supported on Δ with barycenter μ. Therefore,

$$\begin{aligned}\phi(\mu) &= \int_{\Delta} \phi(s)\, d\boldsymbol{\nu}(s) \ , \\ &= \int_{\Delta} \frac{\phi(s)}{\psi(s)} \psi(s)\, d\boldsymbol{\nu}(s) \ , \\ &\leq \sup_{\Delta} \frac{\phi}{\psi} \int_{\Delta} \psi(s)\, d\boldsymbol{\nu}(s) \ , \\ &= \sup_{\Delta} \frac{\phi}{\psi}\ \psi(\mu) \ .\end{aligned}$$

So that, $\phi(\mu)/\psi(\mu) \leq \sup_{\Delta} \phi/\psi$ for all $\mu \in \mathcal{A}$, hence $\sup_{\mathcal{A}} \phi/\psi \leq \sup_{\Delta} \phi/\psi$. The other inequality is obvious as $\Delta \subset \mathcal{A}$. □

We emphasize that the ratio of a convex function by a positive concave function is quasi-convex (Clausing, 1983, p.51). Thus, in the previous proposition the ratio ϕ/ψ is always quasi-convex.

Conclusion

In practical applications, the input parameter distribution is tainted by uncertainty. When selecting a measure space for optimizing a QoI, it should be as representative as possible of the uncertainty affecting the input distribution. While being compatible with the input's available information, the optimization set should encompass all the uncertainty without being too large, in order to avoid overconservative bounds on the QoI. The moment class and the unimodal moment class are very handy to work with. Indeed, the *generalized* moment constraints are flexible, and their number is not limited. The analyst can always adapt the number and form of constraints to tune the size of the optimization space.

In this chapter the theoretical framework of this thesis has been introduced. Our work actually corresponds to an extended version of the Bauer maximum principle. Proofs have been detailed, most of them are surprisingly short and elegant. The moment class and the unimodal moment class are the key elements for practical optimization. Indeed, although these classes are non parametric, their extreme points which are finite convex combination of Dirac masses or uniform distributions are. It is remarkable to be able to characterize the extreme points of such abstract spaces. We constructed a product of this measures spaces that characterizes the mutual independence of any marginal in the joint distribution. We also showed that extreme points of the product measure space reads as the product of the marginal extreme points. However, we studied a more interesting subspace of the product structure by enforcing new moment constraints in the joint distribution. The extreme points of the measure space have also been studied in a general version of the reduction Theorem.

We have at our disposal two versions of the reduction Theorem on the product measure space constructed in Section 3.3.1. The first one, Theorem 3.8, investigates the optimization of a quasi-convex lower semicontinuous function. The second one Corollary 3.10 focuses in measure affine function (or more generally ratio of measure affine function). This allows to drop out the lower semicontinuity assumption, which for some QoI can be hard to check. The reduction theorem presented here simplifies

the numerical optimization of the QoI. By restraining the optimization to the extreme points, the problem Eq. (3.2) becomes parametric and tractable.

The optimization problem is parameterized by the extreme points. To be specific, the QoI depends only on the weights and positions of the Dirac masses and uniform distributions composing the convex combinations. It holds that the maximum of the QoI is reached on these extreme points. Furthermore, the optimization is pursued globally. It is therefore required to explore the whole set of extreme points to reach the optimum. In this work we propose to tackle this problem using a differential evolutionary solver. This algorithm is known for its efficiency for solving OUQ problem (McKerns et al., 2012), but any global optimizer would also do. In order to pursue the optimization, the DE solver must efficiently move its population over the set of extreme points. Otherwise, many population vectors in the evolutionary algorithm are rejected reducing the overall performance of the algorithm.

A good parameterization of the optimization function is essential. At this state the QoI can be computed from the weights and positions of discrete measures. Then, to proceed the optimization one wishes to evolve weights and positions of convex combination of Dirac masses or uniform distribution, all while guarantying that the constraints are fulfilled. As we will see in the next chapter this problem is complex. A new parameterization of the problem is proposed under some restrictive conditions. It is based on the canonical moments instead of the weights and positions of the extreme measures. As we will later see it boosts the numerical performance of the optimization algorithm.

4

Canonical Moments Parameterization of the Extreme Set

« As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality. »

Albert Einstein

Contents

Introduction

The two optimization spaces studied in Chapter 3, namely the moment class and unimodal moment class, possess many interesting properties useful to address the problem of an imprecise input probability distributions. From an engineering point of view, enforcing *generalized* moment constraints on these classes is very handy, and can cover a broad scope of different situations. Moreover, their topological structure and specifically their extreme points are fully characterized and parametric. This is very convenient for the numerical optimization of a quasi-convex lower semicontinuous function, as stated in the reduction Theorem Eq. (3.8).

The QoI has no mathematical properties that one could take advantage of for improving the optimization efficiently. This function is not even necessarily continuous. Therefore, the optimization of the QoI requires using a global optimizer. For that reason, it is mandatory to probe efficiently the optimization space, in order to find the global optimum of the QoI. The current parameterization of QoI is not fulfilling as exposed in Section 4.1. Indeed, this function is computed based on the support points and weights of the extremal measures that are, we recall, convex combination of Dirac masses or uniform distributions, depending on their affiliation to either a moment class or an unimodal moment class. But these weights and positions, corresponding to the coordinates of the optimization vector in the DE solver, are evolving. The evolutionary algorithm struggles to guaranty that from an old generation of individuals, the new population members conform with weights and positions of discrete measures that satisfy the moment constraints. This difficulty to evolve the individuals in the algorithm while ensuring the moment constraints costs us to loose a high number of rejected population members, and reduces the overall performance of the algorithm. We emphasize that any global optimizer could replace the DE solver, however the problem remains as long as this parameterization stands. This issue being that we are dealing with a constrained optimization

Therefore, in this chapter we investigate a new parameterization of the moment class. More precisely, the computation of the QoI previously based on the weights and positions of discrete measures is entirely redesigned. In Section 4.1, the optimization space is specified and some additional assumptions are established. In particular, we put aside the *generalized* moment constraints to focus on *classical* moment constraints. This is necessary for pursuing the study. Section 4.2 exposes the canonical moments theory and its implementation in our optimization problem, first in the simple case of unidimensional measures. Canonical moments are the key elements of the new parameterization. Indeed, as we face moment constraints, problems involving moments and generation of measures have been studied for long under the name of: the Hamburger moment problem, the Stieltjes moment problem, or the Hausdorff moment problem (Shohat and Tamarkin, 1950; Gassiat, 1990). This shows how hard dealing with moment of a probability measure can be. Canonical moments help understanding these problems; it can be seen as a "normalization", in a sense to be precised, of the moments sequence of a measure. Informally, it corresponds to the relative positions of the moment sequence of a distribution in the space of the moment sequences of all probability measures. While Section 4.3 exposes the canonical moment parameterization of the moment class in an unidimensional frame, the last section extends to a product of (unimodal) moment class and introduces quite general algorithms for optimizing any QoI over moment classes using the canonical moments parameterization.

4.1 Specification of the Optimization Problem

In Chapter 3, the optimization space was constructed as the product of moment classes and/or unimodal moment classes. So far, the measures in these classes satisfy *generalized* moment constraints, which are very general but can be complex to handle. In the following, we make several assumptions in order to simplify the problem:

- For real-life computer models with several parameters, the joint input distribution belongs to the product space of moment classes and/or unimodal moment classes (see Section 3.3.1). The product measure translates the mutual independence of each input. Using this independence, the parameterization here under study applies marginally to the moment class and the unimodal moment class. Hence, for clarity's sake, this chapter, except in the last section, focuses on a scalar measure corresponding to a single input, instead of the whole random vector in the product space. Consequently, in the following we drop the notation "$\_i$" that indexes the input number. We intend to reconstruct in Section 4.3 the whole product space, but the parameterization of each marginal moment class goes well across the product structure.

- We focus from now on our work on *classical* moment constraints. That means that every measurable function $\varphi^{(j)}$ for $1 \leq j \leq N$ in Eq. (3.7) characterizing the constraints, now writes $\varphi^{(j)}(X) = X^j$. Hence, it corresponds to the first moment (the mean), the second moment, etc. It can happen that the available information (from expert opinion for instance) are summarized by the mean and variance of the input variable. Moreover, when the input distribution is fully elicited, it is always possible to extract its moments and convert them as constraints. Of course, this assumption is very restrictive, and usually engineers do not have knowledge of the moments but more often have access to some quantile which is a problem we cannot deal easily at this point. Nevertheless, we will show how restraining the problem to *classical* moment constraints allows to gain considerable performance by reparameterizing the optimization space using canonical moments. Indeed, *classical* moment constraints are required to reveal how canonical moments are inherent in the problem.

- So far, we have handled either inequality or equality constraints. We focus only on equality constraints for now, so that we actually study measures with a finite number of known moments. We will deal with the case of inequality constraints in the last section of this Chapter, when developing the algorithms.

- We wish to reparameterize either the moment class or the unimodal moment class. As a first step, we focus only on the moment class. We detail how to deal with the unimodal moment class in Section 4.3.1. Because the structure of these two spaces are very similar, only few changes in the methodology are necessary for the second case, relying on a simple trick.

- The input of the computer model often represents a physical parameter, hence it is a common assumption to assume that it is bounded. Consequently, the support of the input distribution is restricted to an interval $[a, b]$.

To summarize all the above assumptions, we consider in this chapter a bounded scalar measure μ belonging to the moment class:

$$\mathcal{A}_c^* = \left\{ \mu \in \mathcal{P}([a,b]) \mid \mathbb{E}_\mu[X^j] = c_j \text{ for } 1 \leq j \leq N \right\} , \tag{4.1}$$

which satisfies exclusively *classical* moment equality constraints. The extreme points of this class are known from Theorem 3.5 in the previous Chapter. The extreme set is given by:

$$\Delta_c^* = \left\{ \mu \in \mathcal{A}^* \mid \sum_{i=1}^{N+1} \omega_i \delta_{x_i},\ \omega_i \geq 0,\ x_i \in [a,b] \right\} . \tag{4.2}$$

which is the set of discrete measures on $[a,b]$ supported on at most $N+1$ points, and satisfying the N constraints. The optimization problem restricted to the extreme points of the optimization space reveals that the weights and positions in Eq. (4.2) of the input distributions provide a way of computing the QoI, and thus parameterize the optimization problem. However, in order to compute the optimal QoI from a global solver, one must be able to explore the whole set of admissible measures Δ_c^*, which can be challenging. The global optimization of the QoI requires to generate efficiently elements of the extreme set of the moment class Δ_c^*. Indeed, it is known that a global solver's efficiency is related to its ability to explore the optimization space (Price et al., 2005). As we focus our attention on the moment class Δ_c^*, one can easily see how difficult computing discrete measures with fixed moments can be. We give hereunder an example.

Example. Consider the set of all measures in $[0,1]$ with mean 0.5 and second moment 0.3:

$$\mathcal{A} = \left\{ \mu \in \mathcal{P}([0,1]) \mid \mathbb{E}_\mu[X] = 0.5,\ \mathbb{E}_\mu[X^2] = 0.3 \right\} . \tag{4.3}$$

This is a moment class defined by two constraints, the extreme points are therefore discrete measures supported by at most three points:

$$\Delta = \left\{ \mu = \sum_{i=1}^{3} \omega_i \delta_{x_i} \mid \sum_{i=1}^{3} \omega_i = 1,\ \sum_{i=1}^{3} \omega_i x_i = 0.5,\ \sum_{i=1}^{3} \omega_i x_i^2 = 0.3 \text{ with } \omega_i \geq 0,\ x_i \in [0,1] \right\} .$$

Now, assume that the QoI is a quasi-convex lower semicontinuous function of $\mu \in \mathcal{A}$, for instance a quantile of the distribution as we will see in Chapter 5. The optimization is thus reduced to the extreme set Δ by applying the reduction Theorem, which yields a certain parameterization of the problem. Indeed, the computation of the objective function depends on only six parameters: the three weights and the three positions of a discrete measure in Δ. The optimization is performed globally, so that we explore and generate randomly many elements of Δ. But randomly generating weights $\{\omega_i\}_{1\leq i\leq 3}$ and positions $\{x_i\}_{1\leq i\leq 3}$ in $[0,1]^6$ is extremely ineffective. The positions and weights should be solution of the following system:

$$\begin{cases} \omega_1 + \omega_2 + \omega_3 = 1 \\ \omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3 = 0.5 \\ \omega_1 x_1^2 + \omega_2 x_2^2 + \omega_3 x_3^2 = 0.3 \end{cases}$$

Hence, global optimization of the QoI equivalently requires the ability to explore all the solutions of this system.

Two ways to handle the problem appear. The first one consists in optimizing the QoI under constraints, a path followed by McKerns et al. (2012) with the Mystic Framework. We advocate a second option, preferable for computational performance, which consists in reformulating the objective function so that the constraints are inherently respected. This requires to identify a new parameterization adapted for the problem. Here, canonical moments (Dette and Studden, 1997) provide a surprisingly well-tailored approach.

4.2 Theory of Canonical Moments

4.2.1 Classical Moment Limits

The concept of canonical moments was first introduced in Skibinsky (1967). The main contribution of Skibinsky work lies in the original study of the geometrical aspects of general moment spaces (Skibinsky, 1977, 1986). In a number of further papers, Skibinsky proves numerous other interesting properties of canonical moments. More recently, Dette and Studden (1997) have shown the intrinsic relation between a measure μ and its canonical moments. These authors highlight the benefits of canonical moments in many areas of statistics, probability and analysis, such as design of experiments, or the Hausdorff moment problem (Hausdorff, 1923). In this last problem, one investigates whether an element of $[0,1]^n$ can correspond to the moment sequence of a measure supported on $[0,1]$. Many results in this section stem from the book of Dette and Studden (1997), which gives a very nice overview of canonical moments.

In this work, we deal with a problem involving moments. We study in Eq. (4.1) a measure space defined by moment constraints. Problems involving moments are difficult to deal with, in particular because raw moments have no intrinsic relations to their corresponding measure. For instance, let consider a uniform distribution on $[0,1]$, whose sequence of moments is $(\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots)$, and another uniform distribution supported this time on $[0,2]$ whose moment sequence is $(1, \frac{4}{3}, 2, \dots)$. Those two sequences have at first glance nothing in common, and from this last sequence of moment it is challenging to determine the underlying measure. As simple as this fact appears, there is no obvious relation between the moment sequence of a measure and its structure (by structure we mean the way the probability measure distribute its mass). This issue raises fundamental question, indeed there should be a simple link between a moment sequence and its associated distribution, in particular because a measure supported on a bounded set is uniquely defined by its moment sequence. This is a consequence of the Weierstrass approximation theorem, which states that polynomials are dense under the uniform norm in the space of continuous functions on $[0,1]$ (Feller, 1971, p. 223). One of the motivation for introducing canonical moments, was to find out how from a moment sequence one can determine characteristics of the underlying measure. Because raw moments are not suited to handle this question, canonical moments answer to the question "is it possible to *normalize* the moment sequence", in a way that remains to be defined.

Let us first define the moment space

$$M(a,b) := \{\mathbf{c}(\mu) : \mu \in \mathcal{P}([a,b])\} \ , \quad \text{with } a,b \in \mathbb{R}, a < b \ , \tag{4.4}$$

where $\mathbf{c}(\mu)$ denotes the sequence of all moments of a probability measure μ in $[a,b]$. This moment sequence always exists, as the support of the measure is bounded. The nth moment space $M_n(a,b)$ is defined by projecting $M(a,b)$ onto its first n coordinates, $M_n(a,b) = \{\mathbf{c}_n(\mu) = (c_1, \dots, c_n) \mid \mu \in \mathcal{P}([a,b])\}$. The moment space $M_n(a,b)$ is a subset of $\mathbb{R}^n$; we depict as an illustration $M_2(0,1)$ in Fig. 4.1. It represents the first two moments of any probability measure in $[0,1]$. Notice that the extreme points of this last moment space are located on the graph $\{(x,x^2), x \in [0,1]\}$ and that they correspond to the moment sequence of Dirac masses δ_x located on $x \in [0,1]$ with first two moments (x,x^2). More generally, the border (not to be mistaken with the extreme points) of the moment space M_n corresponds to the moment sequences of discrete measures supported on at most $n/2+1$ points (Dette and Studden, 1997, Theorem 1.2.5).

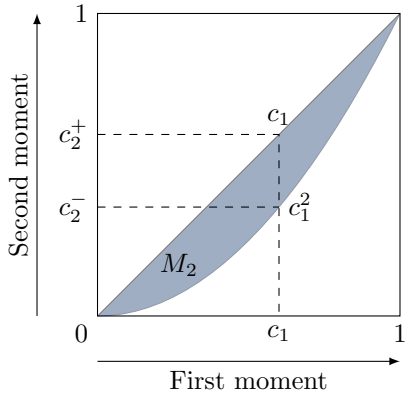


Figure 4.1: The moment space $M_2(0,1)$ and definition of c_2^+ and c_2^-. This shows that for a random variable X in $[0,1]$, if $\mathbb{E}[X] = c_1$, then necessarily $\mathbb{E}[X^2] \geq c_1^2$ and $\mathbb{E}[X^2] \leq c_1$.

As one can observe in Fig. 4.1, when the first moment c_1 is set, the value of the second moment is bounded between c_1 and c_1^2. Therefore, a natural way of normalizing the moments appears. It consists in computing the relative position of the moment sequence into the moment space. To that extend, we define the extreme values:

$$c_n^+ = \max\{c \in \mathbb{R} : (c_1, \dots, c_{n-1}, c) \in M_n(a,b)\} \ ,$$
$$c_n^- = \min\{c \in \mathbb{R} : (c_1, \dots, c_{n-1}, c) \in M_n(a,b)\} \ ,$$

which represent respectively the maximum and minimum values of the nth moment that a measure can have, when its moments up to order $n-1$ are equal to $\mathbf{c}_{n-1}$. Let us now define the nth canonical moment as the normalization of the nth moment between these two bounds:

Definition 3. The nth canonical moment of a measure with moment sequence $\mathbf{c}$ is defined recursively as

$$p_n = p_n(\mathbf{c}) = \frac{c_n - c_n^-}{c_n^+ - c_n^-} \ .$$

Several properties should be pointed out. First, canonical moments are defined up to degree $N = N(\mathbf{c}) = \min\{n \in \mathbb{N} \mid \mathbf{c}_n \in \partial M_n(a,b)\}$. Further, when N is finite p_N is either 0 or 1. Indeed, we know from (Dette and Studden, 1997, Theorem 1.2.5) that if a sequence of moment $\mathbf{c}_n$ belongs to the border of the moment space $\partial M_n(a,b)$, it implies that the underlying measure μ is uniquely determined, so that straightaway, $c_n^+ = c_n^-$ and the canonical moments are no further defined. We highlight that all the canonical moments belong to $[0,1]$. The following result is also important

Theorem 4.1 ((Dette and Studden, 1997, Theorem 1.3.2))**.** *Canonical moments are invariant under affine transformation of the support of the underlying measure μ, $y = a + (b-a)x$, with $b > a$ of $[0,1]$ onto $[a,b]$.*

Hence, we may restrict ourselves in the following to the case $a = 0, b = 1$. When applying the affine transformation $y = a + (b-a)x$ to the support of a measure μ in $[0,1]$, canonical moments are unchanged but not *classical* moments. The sequence of moments $\mathbf{c}' = (c_1', c_2', \dots)$ of the transformed measure μ' on $[a,b]$ are computed as follows:

$$c_n' = \frac{1}{(b-a)^n} \sum_{k=0}^{n} \binom{n}{k} (-b)^{(n-k)} c_k, \text{ for } n = 1, \dots, N \ . \tag{4.5}$$

The importance of Theorem 4.1 might seem limited, but it actually emphasizes the intrinsic relation between canonical moments and the structure of the related distribution. Indeed, we considered previously as an example the moment sequence of uniform distributions that were pretty much without interest. However, the canonical moment sequence of a uniform distribution on any bounded interval is unique and is given by

$$p_{2k-1} = \frac{1}{2},\ p_{2k} = \frac{k}{2k+1},\ k \geq 1\ . \tag{4.6}$$

This highlights that from a sequence of canonical moments one can recover information on how the probability mass spreads over $[0, 1]$. This is the canonical moments' greatest asset compared to the classical moments. The uniform distribution served as example but our interest lays here on discrete measures, which are very specific in their way of spreading the probability mass. The next sections are dedicated to the methodology of recovering discrete distributions from a given sequence of canonical moments.

We conclude this short introduction on canonical moments by presenting some interesting information about the moment space.

Remark (Volume of the nth moment space M_n (Dette and Studden, 1997, p.27)). The volume of the nth moment space M_n is given by

$$\text{Vol } M_n = \prod_{k=1}^{n} B(k, k)\ ,$$

where $B(k, k)$ is the Beta function defined by

$$B(p, q) = \int_0^1 x^{p-1}(1-x)^{q-1}\,\mathrm{d}x = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}\ ,$$

with Γ the Gamma function (Johnson et al., 1995). For instance, we obtain Vol $M_2 = 1/6$, Vol $M_3 = 1/180$, etc. For $n \to \infty$

$$\text{Vol } M_n \approx const \cdot 2^{-n^2}\ , \tag{4.7}$$

which demonstrates that the nth moment space forms a very small part of the nth cube $[0, 1]^n$. It helps understanding the difficulty to explore the optimization space Δ_c^* as the number of moment constraints increases. Indeed, if n moments are known, elements of Δ_c^* are discrete probability measures on $n + 1$ points. During the optimization of a QoI over this space, if one randomly generates weights and positions in $[0, 1]$, the probability to obtain a probability measure (that is with a moment sequence inside the moment space) quickly decreases to zero. Actually, M_n concentrates on a single points, more precisely the k first components of a sequence $c_n \in M_n$ converge exponentially fast in probability to the k first moments of the arcsine distribution (Gamboa and Lozada-Chang, 2004). The arcsine is interesting as all its canonical moments equal $1/2$.

4.2.2 Computation of the Canonical Moments

The description of canonical moments in Section 4.2.1 was geometrically intuitive. Given a sequence of moments $\mathbf{c} = (c_k)_{1 \leq k \leq N}$, that corresponds in our practical application to the moment constraints enforced on the input measure in $\mathcal{A}$ [see Eq. (4.1)], one wishes to compute the corresponding sequence of canonical moments $\mathbf{p} = (p_1, \ldots p_N)$.

In order to compute the canonical moments $\mathbf{p}$ from the moment sequence $\mathbf{c}$, a procedure is needed to calculate c_k^+ and c_k^-. This is accomplished with a sequence of determinants involving the moments c_k, which provides an explicit expression for p_k in terms of $c_1, c_2, \ldots, c_k$.

Let us consider a fixed vector $\mathbf{c}_n$; the problem of knowing whether there exists a probability measure μ in $[0,1]$ with n first moments $\mathbf{c}_n$ is called the Hausdorff moment problem. So that if $\mathbf{c_n}$ is in M_n, it is called a Hausdorff moment sequence and there exists a measure μ representing $\mathbf{c}_n$. There exists a characterization of Hausdorff moment sequences. More precisely, it is obvious that if $\mathbf{c}_n \in M_n$ then

$$\sum_{i=0} a_i x^i \geq 0 \text{ for } x \in [0,1] \text{ implies } \sum_{i=0}^{n} a_i c_i \geq 0 \ ,$$

The reverse implication is less obvious but also holds. So that a sequence $\mathbf{c}_n$ is in M_n, if and only if $\sum_{i=0}^n a_i c_i \geq 0$ whenever $(a_0, \ldots, a_n)$ are the coefficients of a nonnegative polynomial of degree n (Dette and Studden, 1997, Theorem 1.4.1). From a sequence $\mathbf{c}_n \in M_n$ and a polynomial $P(x) = \sum_{k=0}^n a_k x^k$, let define the function $L : P \mapsto L(P) := \sum_{k=0}^n a_k c_k$. Then, it follows that

$$L(P^2) = \sum_{k=0}^{n} \sum_{i=0}^{n} a_k a_i c_{i+k} \geq 0 \ ,$$

for all $(a_0, \ldots, a_n) \in \mathbb{R}^{n+1}$ as the polynomial P^2 is nonnegative. Thus the matrix $(c_{i+j})_{0 \leq i,j \leq m}$ is nonnegative definite. The same reasoning on different polynomials yields that the matrix $(c_{i+j+1} - c_{i+j+2})_{0 \leq i,j \leq m-1}$ is also nonnegative definite (see (Dette and Studden, 1997, p. 22) for details). It motivates the introduction of the Hankel determinants

$$\underline{H}_{2m} := \begin{vmatrix} c_0 & \cdots & c_m \\ \vdots & & \vdots \\ c_m & \cdots & c_{2m} \end{vmatrix} \qquad \overline{H}_{2m+1} := \begin{vmatrix} c_0 - c_1 & \cdots & c_m - c_{m+1} \\ \vdots & & \vdots \\ c_m - c_{m+1} & \cdots & c_{2m} - c_{2m+1} \end{vmatrix} \ , \tag{4.8}$$

$$\underline{H}_{2m+1} := \begin{vmatrix} c_1 & \cdots & c_{m+1} \\ \vdots & & \vdots \\ c_{m+1} & \cdots & c_{2m+1} \end{vmatrix} \qquad \overline{H}_{2m} := \begin{vmatrix} c_1 - c_2 & \cdots & c_m - c_{m+1} \\ \vdots & & \vdots \\ c_m - c_{m+1} & \cdots & c_{2m-1} - c_{2m} \end{vmatrix} \ .$$

Therefore, a Hausdorff moment sequence $\mathbf{c}_n$ is characterized by the property that $\underline{H}_k$ and $\overline{H}_k$ are nonnegative for all $1 \leq k \leq n$. Moreover, the moment sequence $\mathbf{c}_n$ is in the interior of M_n if and only if the Hankel determinants $\underline{H}_k$ and $\overline{H}_k$ are positive for $1 \leq k \leq n$. It is also worthwhile mentioning that $\underline{H}_n = 0$ or $\overline{H}_n = 0$ implies $\overline{H}_k = \underline{H}_k = 0$ for all $k \geq n+1$. Indeed, we recall that a symmetric matrix is nonnegative definite if and only if the leading principal minors are nonnegative. This follows because a vanishing Hankel determinant of index n implies $\mathbf{c}_n \in \partial\, M_n$. In this case, there exists a unique measure μ with moment sequence $\mathbf{c}$.

An examination of the Hankel determinants in Eq. (4.8) reveals that if $\mathbf{c}_{n-1} \in \text{int}(M_{n-1})$ then $\underline{H}_n$ is an increasing function of c_n. Therefore, for a fixed $\mathbf{c}_{n-1} \in \text{int}(M_{n-1})$, the nth moment c_n has a lower bound c_n^- determined by varying c_n in $\underline{H}_n$ until $\underline{H}_n = 0$. Similarly, the upper bound c_n^+ is determined by varying c_n in $\overline{H}_n$ up to $\overline{H}_n = 0$. This explains the upper and lower bar notation. The following result provides an explicit representation of the canonical moments in terms of Hankel determinants. The proof of this result is copied from Dette and Studden (1997). We display it because it helps understanding the general argumentation of this section.

Theorem 4.2 ((Dette and Studden, 1997, p. 23)). *If* $\mathbf{c}_{n-1} \in int(M_{n-1})$ *then*

$$c_n - c_n^- = \frac{\underline{H}_n}{\underline{H}_{n-2}} \ , \quad c_n^+ - c_n = \frac{\overline{H}_n}{\overline{H}_{n-2}} \ ,$$

and

$$p_n = \frac{\underline{H}_n \overline{H}_{n-2}}{\underline{H}_n \overline{H}_{n-2} + \overline{H}_n \underline{H}_{n-2}} \ , \tag{4.9}$$

where $\underline{H}_{-1} = \overline{H}_{-1} = \underline{H}_0 = \overline{H}_0 = 1$.

Proof. The value of p_n follows from the expressions for $c_n - c_n^-$ and $c_n^+ - c_n$ and from its definition in Eq. (4.6). To obtain the expression for $c_n - c_n^-$ we note again that $\underline{H}_n$ would be zero if we replace c_n by c_n^-. Then, writing $c_n = c_n^- + (c_n - c_n^-)$ for the last element in $\underline{H}_n$ gives $\underline{H}_n = (c_n - c_n^-)\underline{H}_{n-2}$. The value of $c_n^+ - c_n$ could be obtained in a similar way. □

The expression in the denominator of p_n in Eq. (4.9) can be written in a different form by noticing a basic relationship between Hankel determinants $\underline{H}_n \overline{H}_n = \underline{H}_{n-1}\overline{H}_{n+1} + \overline{H}_{n-1}\underline{H}_{n+1}$. This yields the following expression of the canonical moments:

$$c_n^+ - c_n^- = \frac{\underline{H}_{n-1}\overline{H}_{n-1}}{\overline{H}_{n-2}\underline{H}_{n-2}} \ , \quad p_n = \frac{\underline{H}_n \overline{H}_{n-2}}{\overline{H}_{n-1}\underline{H}_{n-1}} \ . \tag{4.10}$$

One can derive from these equations analytical formula for the canonical moments. For low moment orders, we get the simple expressions:

$$p_1 = c_1 \ , \quad p_2 = \frac{c_2 - c_1^2}{c_1(1 - c_1)} \ . \tag{4.11}$$

We also introduce the following quantities that will be of importance in the following:

$$\zeta_0 = 1 \ , \quad \zeta_1 = p_1 \ , \quad \zeta_k = p_k(1 - p_{k-1}) \qquad k \geq 2 \ , \tag{4.12}$$

which yields the relation

$$\zeta_k = \frac{\underline{H}_k \underline{H}_{k-3}}{\underline{H}_{k-1}\underline{H}_{k-2}} \ . \tag{4.13}$$

While Eq. (4.10) provides an explicit expression for the canonical moments, the general case involves computing a large amount of Hankel determinants. These computations are numerically expensive. However, one can exploit the special structure of the Hankel determinant to drastically decrease the computational budget. This leads to a recursive algorithm named the *Q-D* algorithm, which consists in constructing a table for moments $(c_1, c_2, \dots)$ of some measure μ, that leads to the quantities $(\zeta_1, \zeta_2, \dots)$ and hence canonical moments. This algorithm was first described in (Henrici, 1993, p.608). To sketch the *Q-D* algorithm, further Hankel determinants are needed, namely for $k \geq 1$:

$$H_k^{(n)} := \begin{vmatrix} c_n & c_{n+1} & \cdots & c_{n+k-1} \\ c_{n+1} & c_{n+2} & \cdots & c_{n+k} \\ \vdots & \vdots & & \vdots \\ c_{n+k-1} & c_{n+k} & \cdots & c_{n+2k-2} \end{vmatrix} \ , \tag{4.14}$$

and $H_0^{(n)} := 1$. Note that $H_k^{(n)}$ starts with c_n in the upper left corner and is of size of order k. The determinants $H_k^{(0)}$ and $H_k^{(1)}$ have been encountered as

$$H_k^{(0)} = \underline{H}_{2k-2} \quad \text{and} \quad H_k^{(1)} = \underline{H}_{2k-1} \ . \tag{4.15}$$

Now define $e_0^{(n)} := 0$ and for $n \geq 0$ and $m \geq 1$

$$\begin{aligned} q_m^{(n)} &= \frac{H_m^{(n+1)} H_{m-1}^{(n)}}{H_m^{(n)} H_{m-1}^{(n+1)}} \ , \\ e_m^{(n)} &= \frac{H_{m+1}^{(n)} H_{m-1}^{(n+1)}}{H_m^{(n)} H_m^{(n+1)}} \ . \end{aligned} \tag{4.16}$$

The *Q-D* algorithm is written for sake of clarity in a triangular array as depicted in Fig. 4.2. Observing Eq. (4.14) and Eq. (4.16) it follows that the values of $q_1^{(n)}$, $n \geq 0$ in the first non-zero column are

$$q_1^{(n)} = \frac{c_{n+1}}{c_n} \ ,$$

and by Eq. (4.13) and Eq. (4.15), the values of $q_m^{(0)}$ and $e_m^{(0)}$ on the first diagonal are

$$q_m^{(0)} = \zeta_{2m-1} \ , \quad e_m^{(0)} = \zeta_{2m} \ . \tag{4.17}$$

The canonical moments are then derived from Eq. (4.12). The algorithm recursively constructs the table in Fig. 4.2 using the following procedure which is numerically very effective.

Theorem 4.3 (*Q-D* algorithm (Dette and Studden, 1997, p. 30))**.** *If $q_m^{(n)}$ and $e_m^{(n)}$ are defined as in Eq. (4.16) then for $m \geq 1$ and $n \geq 0$*

$$\begin{aligned} e_m^{(n)} &= (q_m^{(n+1)} - q_m^{(n)}) + e_{m-1}^{(n+1)} \ , \\ q_{m+1}^{(n)} &= \frac{e_m^{(n+1)}}{e_m^{(n)}} q_m^{(n+1)} \ . \end{aligned} \tag{4.18}$$

4.2.3 From Canonical Moments to Discrete Measures

From a given sequence of canonical moments one needs to reconstruct support points and weights of a discrete measure. We introduce to this purpose the Stieljes transform, which connects the canonical moments with a corresponding measure. The Stieljes transform of a scalar measure μ is defined as

$$S(z) = z(z, \mu) = \int_a^b \frac{\mathrm{d}\mu(x)}{z - x} \ , \quad z \in \mathbb{C} \backslash \{\mathrm{supp}(\mu)\} \ . \tag{4.19}$$

The transform $S(z, \mu)$ is an analytical function of z in $\mathbb{C} \backslash \mathrm{supp}(\mu)$. We recall that a measure of the optimization set Δ_c^* in Eq. (4.2) is discrete. As it is the set of extreme points of moment class $\mathcal{A}_c^*$. If $\mu = \sum_{i=1}^{N+1} \omega_i \delta_{x_i}$ has a finite support then

$$S(z) = \int_a^b \frac{d\mu(x)}{z - x} = \sum_{i=1}^{N+1} \frac{\omega_i}{z - x_i} \ ,$$

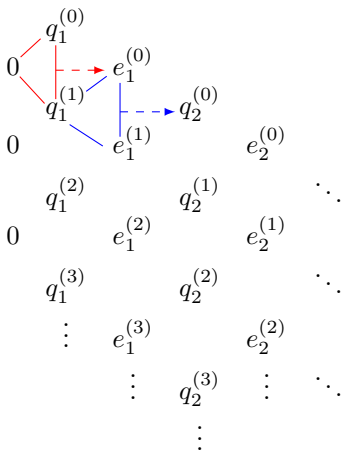


Figure 4.2: Visualization of the *Q-D* algorithm. The recursive computation of $e_m^{(n)}$ and $q_m^{(n)}$ relies on Eq. (4.18). The canonical moments are obtained from Eq. (4.17) and Eq. (4.12).

where the support points of the scalar measure μ are distinct and denoted by x_1, $\ldots, x_{N+1}$, with corresponding weights $\omega_1, \ldots, \omega_{N+1}$. Alternatively, the weights are given by $\omega_i = \lim_{z \to x_i}(z - x_i)S(z)$. Putting the expression Eq. (4.19) under the same denominator, the Stieljes transform can be rewritten as a ratio of two polynomials with no common zeros.

$$S(z) = \frac{Q^{(N)}(z)}{P_*^{(N+1)}(z)} \ , \tag{4.20}$$

where the zeros of $P_*^{(N+1)}(z) = \prod_{i=1}^{N+1}(z - x_i)$ are the support points of the measure μ. The associated weights can be recovered as

$$\omega_i = \frac{Q^{(N)}(x_i)}{\frac{d}{dx}P_*^{(N+1)}(x)|_{x=x_i}} \ . \tag{4.21}$$

We focus for now on recovering the support points $(x_i)_{1 \le i \le N+1}$ of a measure μ from its canonical moment sequence. We will show further than the weights are uniquely determined from the positions x_i and the constrained moment sequence $\mathbf{c}_N$. An important consequence is that the space Δ_c^* is parameterized only by the support points, as it will be clarified in the following. The Stieljes transform of a measure μ can also be written as a continuous fraction depending on the canonical moments of μ (Dette and Studden, 1997, Chapitre 4) . Some basic definitions and properties of continuous fraction are given in the following Lemma.

Lemma 4.4. *A finite continued fraction is an expression of the form*

$$b_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \ldots}} = b_0 + \frac{a_1}{\left\lfloor b_1 \right.} + \frac{a_2}{\left\lfloor b_2 \right.} + \cdots + \frac{a_n}{\left\lfloor b_n \right.} = \frac{A_n}{B_n} \ .$$

The quantities A_n and B_n are called the nth partial numerator and denominator. There exist basic recursive relations for the quantities A_n and B_n given by

$$\begin{aligned} A_n &= b_n A_{n-1} + a_n A_{n-2} \ , \\ B_n &= b_n B_{n-1} + a_n B_{n-2} \ , \end{aligned}$$

for $n \geq 1$ with initial conditions

$$\begin{aligned} A_{-1} &= 1 \quad , \quad A_0 = b_0 \ , \\ B_{-1} &= 0 \quad , \quad B_0 = 1 \ . \end{aligned}$$

The following important result gives the continuous fraction expansion of the Stieljes transform. Its proof is technical and we will omit it. Recall that the purpose is to generate the support points of a measure from its sequence of canonical moments. The reader can observe that for a given measure μ, we have a first characterization of the Stieljes transform relying on the support points in Eq. (4.20). Further, the next Theorem provides a second characterization of the Stieljes transform based on the canonical moments of μ. The two characterizations of the same object give rise to attractive results.

Theorem 4.5 ((Dette and Studden, 1997, Theorem 3.3.1)). *Let μ be a probability measure on the interval $[a, b]$ and $z \in \mathbb{C}\backslash[a, b]$, then the Stieltjes transform of μ has the continued fraction expansion:*

$$\begin{aligned} S(z) &= \cfrac{1}{z-a} \ - \ \cfrac{\zeta_1(b-a)}{1} \ - \ \cfrac{\zeta_2(b-a)}{z-a} \ - \ \dots \\ &= \cfrac{1}{z-a-\zeta_1(b-a)} \ - \ \cfrac{\zeta_1\zeta_2(b-a)^2}{z-a-(\zeta_2+\zeta_3)(b-a)} \\ &\quad - \ \cfrac{\zeta_3\zeta_4(b-a)^2}{z-a-(\zeta_4+\zeta_5)(b-a)} \ - \ \dots \end{aligned}$$

We recall that $\zeta_n := p_n(1-p_{n-1})$.

More precisely, if μ is supported by N points in $[a, b]$, then necessarily $\zeta_{2N-1}\zeta_{2N} = 0$. Indeed, whenever μ has finite support its moment sequence belongs to the border of M_k for a certain k, which uniquely characterizes the measure (Dette and Studden, 1997, Theorem 1.2.5). Because the canonical moments sequence is finite, the continuous fraction expansion simplifies:

$$\begin{aligned} S(z) &= \cfrac{1}{z-a-\zeta_1(b-a)} \ - \ \cfrac{\zeta_1\zeta_2(b-a)^2}{z-a-(\zeta_2+\zeta_3)(b-a)} \\ &\quad - \ \dots \ - \ \cfrac{\zeta_{2N-3}\zeta_{2N-2}(b-a)^2}{z-a-(\zeta_{2N-2}+\zeta_{2N-1})(b-a)} \ . \end{aligned}$$

Theorem 4.5 shows that the Stieltjes transform can be written as a function of the canonical moments. It immediately follows from Theorem 4.5 and Lemma 4.4 that the numerator and the denominator in Eq. (4.20) follow the same recursive formula with different initial conditions. For $k \geq 0$ the denominator can be compute iteratively:

$$P_*^{(k+1)}(x) = (x - a - (b-a)(\zeta_{2k} + \zeta_{2k+1}))P_*^{(k)}(x) - (b-a)^2\zeta_{2k-1}\zeta_{2k}P_*^{(k-1)}(x) \ , \quad (4.22)$$

where $P_*^{(-1)} = 0$, $P_*^{(0)} = 1$ and $\zeta_0 = 0$. And for $k \geq 1$, the numerator writes:

$$Q^{(k)}(x) = (x - a - (b-a)(\zeta_{2k} + \zeta_{2k+1}))Q^{(k-1)}(x) - (b-a)^2\zeta_{2k-1}\zeta_{2k}Q^{(k-2)}(x) \ , \quad (4.23)$$

with $Q^{(-1)}(x) = 0, Q^{(0)}(x) = 1$. Those formulas are very important for computational purpose. Indeed, the support points of μ thus consist of the roots of $P_*^{(N+1)}$. The sequence of polynomials $(P_*^{(k)})_{1\leq k\leq N+1}$ are actually the monic orthogonal polynomials with respect to the measure μ (Dette and Studden, 1997, p. 84). Moreover, from equations Eq. (4.21), Eq. (4.22) and Eq. (4.23), it follows that the weights are uniquely determined from the moment sequence and the support points $(x_i)_{1\leq i\leq N+1}$ of the measure μ. Consequently, it is enough to consider only the computation of the support points from the canonical moments, the weights being easily recovered afterwards. It implies that a discrete measure $\mu \in \Delta_c^*$ is fully characterized by just its finite support, as the constraints are fixed and the weight associated to a position is determined from the constraints and the position itself. Theorem hereunder follows from the discussion.

Theorem 4.6 ((Dette and Studden, 1997, Theorem 3.6.1))**.** *Let μ denote a measure on the interval $[a, b]$ supported on $N+1$ points with canonical moments $p_1, p_2, \ldots$. Then, the support of μ is the set of $\{x : P_*^{(N+1)}(x) = 0\}$ defined by Equation (4.22).*

Theorem 4.6 gives a first link between the support points of a discrete measure and its moment sequence. In our optimization problem, we would like to determine all measures satisfying a given finite moment sequence. Moreover, the cardinal of the support of the measure is also known. It has been previously shown in Chapter 3 that the extreme points of the moment class are convex combination of at most $N+1$ Dirac masses, when N moment constraints define the space.

Consequently, in the following we consider a fixed moment sequence $\mathbf{c}_N = (c_1, \ldots, c_N) \in M_N$, and a measure μ in Δ_c^* [Eq. (4.2)] supported on at most $N+1$ points, such that its moments up to order n coincide with $\mathbf{c}_N$. Of course, this implies that the canonical moments of μ coincide with $\mathbf{p}_N = (p_1, \ldots, p_N)$, the corresponding sequence of canonical moments related to $\mathbf{c}_N$, as described in Section 4.2.2. We study the polynomial $P_*^{(N+1)}$ as its roots correspond to the support points of a discrete measure supported on exactly $N+1$ points. Notice that $P_*^{(N+1)}$ is computed from a sequence of $2N+1$ canonical moments $(p_1, \ldots, p_{2N+1})$. Naturally, the first N canonical moments $(p_1, \ldots, p_N)$ are set from the moment constraints $\mathbf{p}_N$, as we are looking to discrete measures fitting the corresponding moments. However, the $N+1$ canonical moments left $(p_{N+1}, \ldots, p_{2N+1})$ can be considered as free parameters in this problem of discrete measure reconstruction. Therefore, we define a map that connects a sequence of $(N+1)$ "free" canonical moments to the polynomials $P_*^{(N+1)}$.

Define the set $\mathcal{K}_{N+1} = \left\{\mathbf{p} \in [0,1]^{N+1} \mid p_i \in \{0,1\} \Rightarrow x_k = 0,\ k \geq i\right\}$ which corresponds to a canonical moment sequence (if one equals 0 or 1, then the others above are not defined) and the function:

$$\begin{aligned} K_{\mathbf{p}_N} : \quad \mathcal{K}_{N+1} &\to \mathbb{R}[X] \\ (p_{N+1}, \ldots, p_{2N+1}) &\mapsto P_*^{(N+1)} . \end{aligned} \tag{4.24}$$

The function $K_{\mathbf{p}_N}$ computes, from a sequence of canonical moments $(p_1, \ldots, p_{2N+1}) \in \{\mathbf{p}_N\} + \mathcal{K}_{N+1}$, a polynomial $P_*^{(N+1)}$ with respect to Eq. (4.22). Therefore, the roots of $P_*^{(N+1)}$ correspond to the support points of a measure with canonical moments $\mathbf{p}_N$, paired to the moments sequence $\mathbf{c}_N$. Hence, it warrants that the measures satisfy the moment constraints. Corollary 4.7 below applies Theorem 4.6 but for measures supported on *at most* $N+1$ points, rather than *exactly* $N+1$ points. Given a measure supported on strictly less than $N+1$ points, say k, the question is therefore to know whether it makes sense to evaluate the $N+1$ roots of $P_*^{(N+1)}$, instead of the kth roots of

$P_*^{(k)}$ for $k < N+1$. The benefit of Corollary 4.7 comes from the numerical performance of the related algorithm. We show that all the measures in Δ_c^* can be recovered from the roots of polynomials $P_*^{(N+1)}$ in

$$K_{\mathbf{p}_N}(\mathcal{K}_{N+1}) := \left\{ P_*^{(N+1)} = K_{\mathbf{p}_N}(p_{N+1}, \ldots, p_{2N+1}), \text{for } (p_{N+1}, \ldots, p_{2N+1}) \in \mathcal{K}_{N+1} \right\}. \quad (4.25)$$

We detail in Section 4.3.3 an efficient numerical method to compute the roots of $P_*^{(N+1)}$ that benefits from the three terms recurrence in Eq. (4.22).

Corollary 4.7. *Consider a sequence of moments $\mathbf{c}_N = (c_1, \ldots, c_N) \in M_N$, the extreme set of the moment class $\mathcal{A}_c^*$:*

$$\Delta_c^* = \left\{ \mu = \sum_{i=1}^{N+1} \omega_i \delta_{x_i} \in \mathcal{P}([a,b]) \mid \mathbb{E}_\mu(x^j) = c_j, \ j = 1, \ldots, N \right\},$$

and the set

$$\mathcal{K}_{N+1} = \left\{ (p_{N+1}, \ldots, p_{2N+1}) \in [0,1]^{N+1} \mid p_i \in \{0,1\} \Rightarrow p_k = 0, \ k > i \right\}.$$

Then, there exists a bijection depicted in Figure 4.3 between Δ_c^* and $\mathcal{K}_{N+1}$.

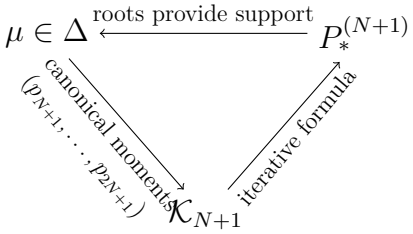


Figure 4.3: Relation between the set of admissible measures and the canonical moments.

Proof. A limit argument is used for the proof. Without loss of generality we can always assume $a = 0$ and $b = 1$ as the problem is invariant using affine transformation. We first consider the case where $\text{card}(\text{supp}(\mu))$ is exactly $N+1$. From Theorem 4.6, the polynomial $P_*^{(N+1)}$ is well defined with $N+1$ distinct roots corresponding to the support of μ. Note that this implies that $(p_1, \ldots, p_{2N-1})$ belongs to $]0,1[^{2N-1}$ and that p_{2N}, p_{2N+1} or p_{2N+2} belong to $\{0,1\}$.

Now, the functions $g(x,z) = 1/(z-x)$ are equicontinuous for z in any compact region which has a positive distance from $[0,1]$. The Stieljes transform is a finite sum of equicontinuous functions and therefore also equicontinuous. Thus, if a measure μ converges weakly to μ^*, the convergence must be uniform in any compact set with positive distance from $[0,1]$ (see Royden (1968)). It is then always possible to restrict ourselves to measures of cardinal $m < N+1$, by letting p_k converge to 0 or 1 for $2m-2 \leq k \leq 2m$. Note that by doing so the polynomials $P_*^{(m)}$ and $P_*^{(N+1)}$ will have the same roots. But, $P_*^{(N+1)}$ and $Q^{(N-1)}$ will have some others roots of multiplicity strictly equal (see Equations (4.20) and (4.22)). The corresponding weights of these roots are vanishing, so that the measures extracted from $P_*^{(m)}$ and $P_*^{(N+1)}$ are the same. □

From a computational point of view, as the proof relies on a limit argument, we can always generate $p_k \in]0,1[$, for $n+1 \leq k \leq 2N+1$. This avoids to deal with

the situation $p_k \in \{0,1\} \Rightarrow p_j = 0$ for $j > k$, that defined the set $\mathcal{K}_{N+1}$. The N first canonical moments of this sequence are set from the moment constraints $\mathbf{c}_N$, while the canonical moments $(p_k)_{N+1 \leq k \leq 2N+1} \in \mathcal{K}_{N+1}$ constitute $N+1$ free parameters, in equal number to the support cardinal of $\mu \in \Delta_c^*$. This is indeed the evidence that this problem is conditioned by $N+1$ parameters. So are measures in Δ_c^* as this demonstrates why the weights of the discrete measures are not free parameters, and can be recovered from the moment constraints and the support points. The computation of $\mathcal{K}_{N+1}$ is easy, it can be done by random generation of $N+1$ numbers in $]0,1[$, yet it allows to generate the support of all the measures in Δ_c^*. This provides a very nice parameterization of the problem that naturally accounts for moment constraints. The optimization set Δ_c^* and all discrete measures it contains can be generated from $N+1$-uplets of $]0,1[^{N+1}$.

The optimization of a QoI over Δ_c^* parameterized with canonical moments are performed with a similar DE solver as for the previous parameterization on weights and support points (we recall that the DE algorithm could be replace by any other means). However, now the parameterization relies on a sequence of canonical moments. Hence, an individual of the evolutionary algorithm is an uplet of $]0,1[^{N+1}$. The main advantage, compared to the parameterization of the QoI with positions and weights, is that when the population evolves and mutates, any new individual remains in the pavement $]0,1[^{N+1}$. Then, it leads to a discrete distribution that satisfies the constraints. Thus, no population members are rejected in the evolutionary solver under this canonical moments parameterization. Because the optimization space is a pavement it is more easily explored. So that, it drastically increases the efficiency of the global optimizer. This outcome also has a elegant geometrical interpretation developed hereunder.

4.2.4 Simple Geometrical Parameterization

From the previous section, it holds that computing roots of polynomials $P_*^{(N+1)}$ in $K_{\mathbf{p}_N}(\mathcal{K}_{N+1})$ [Eq. (4.25)] provides support points of a discrete measure on at most $N+1$ points that satisfies the moment constraints $\mathbf{c}_N$ associated to the canonical moment sequence $\mathbf{p}_N$. However, the discrete measure, and thus the extreme set Δ_c^* in Eq. (4.2), are specified from the support points but also the weights of the discrete measures. It appears from Eq. (4.21) that the weights are uniquely determined from the moment sequence and the support positions of the underlying measure μ. As a consequence, the whole optimization of the QoI over the extreme set Δ_c^* is parameterized with the support points of the measure. This fact can be observed in another way arising directly from the moment constraints. Let μ be a measure in Δ_c^*, such that $\mu = \sum_{i=1}^{N+1} \omega_i \delta_{x_i}$ satisfies the following system of constraints:

$$\begin{cases} \omega_1 & + & \dots & + & \omega_{N+1} & = 1 \\ \omega_1 x_1 & + & \dots & + & \omega_{N+1} x_{N+1} & = c_1 \\ \vdots & & & & \vdots & \vdots \\ \omega_1 {x_1}^N & + & \dots & + & \omega_{N+1} {x_{N+1}}^N & = c_N \end{cases} \tag{4.26}$$

where the N last equations derive from the moment constraints and the first one implies that the measure mass must be one. Because the support points $(x_j)_j$ are distinct, Eq. (4.26) is a Vandermonde system whose determinant is non-zero, and the weights are uniquely determined from the positions and the constraints. As stated, a consequence is that it is always possible to recover the weights from the positions of a discrete measure. In Eq. (4.26) the specific structure of the Vandermonde system can be exploited to construct a progressive algorithm significantly faster ($O(n^2)$) than

regular methods (Björck and Pereyra, 1970). Nevertheless, a Vandermonde system is ill-conditioned (Gautschi, 1990) so that this method is numerically not recommended whenever the number of constraints N increases. Hence, one should favored computing the weights by Eq. (4.21).

Either way, the optimization set Δ_c^* is parameterized by $(x_j)_j$ the support positions of discrete measures. Consequently, the optimization reduces to the following manifold that is explored by the global optimizer.

$$\mathcal{V} = \left\{ \mathbf{x} = (x_1, \dots, x_{N+1}) \in \mathbb{R}^{N+1} : \text{ there exists } (\omega_j) \text{ with } \mu = \sum_{j=1}^{N+1} \omega_j \delta_{x_j} \in \Delta_c^* \right\}.$$

As stated, $\mathcal{V}$ is simply the set of support points of all measures in Δ_c^* respecting the constraints. Now, from Corollary 4.7, we get the geometrical structure of the manifold $\mathcal{V}$. Indeed, it shows that the support points are computed from a polynomial $P_*^{(N+1)}$.

Corollary 4.8 (Stenger et al., 2019b)**.** *The set $\mathcal{V}$ of $(N+1)$ tuples corresponding to the support points of discrete measure with prescribed first N moments $(c_1, \dots, c_n)$ is an algebraic variety of $\mathbb{R}^{N+1}$.*

Proof. The set $\mathcal{V}$ is the zeros locus of the set of polynomials

$$K_{\mathbf{p}_N}(\mathcal{K}_{N+1}) = \left\{ P_*^{(N+1)} = K_{\mathbf{p}_N}(p_{N+1}, \dots, p_{2N+1}), \text{ for } (p_{N+1}, \dots, p_{2N+1}) \in \mathcal{K}_{N+1} \right\},$$

where every polynomial $P_*^{(N+1)}$ is built from a sequence of canonical moments $(p_1, \dots, p_{2N+1})$, the N first being set from the constraints. □

Example. We consider the example previously introduces in Eq. (4.3), that is the moment set

$$\mathcal{A} = \left\{ \mu \in \mathcal{P}([0,1]) \mid \mathbb{E}_\mu[X] = 0.5, \ \mathbb{E}_\mu[X^2] = 0.3 \right\}, \tag{4.27}$$

with extreme points set

$$\Delta = \left\{ \mu = \sum_{i=1}^{3} \omega_i \delta_{x_i} \mid \sum_{i=1}^{3} \omega_i = 1, \ \sum_{i=1}^{3} \omega_i x_i = 0.5, \ \sum_{i=1}^{3} \omega_i x_i^2 = 0.3 \text{ with } \omega_i \geq 0, \ x_i \in [0,1] \right\}.$$

Measures in this set satisfy two moment constraints $\mathbf{c}_2 = (0.5, 0.3)$ corresponding to two canonical moment constraints $\mathbf{p}_2 = (0.5, 0.2)$ from Eq. (4.11). As explained in this section, the optimization set Δ is parameterized by the support points $(x_i)_{1 \leq i \leq 3}$, as the weights can be recovered from the positions and the moment constraints from Eq. (4.26) or Eq. (4.21). Therefore, without considering the canonical moments parameterization, the global solver explores the optimization manifold which writes

$$\mathcal{V} = \left\{ \mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3 : \ \mu = \sum_{i=1}^{3} \omega_i \delta_{x_i} \in \Delta \right\}.$$

This algebraic variety is represented in Fig. 4.4. The coordinates of each point in this algebraic variety correspond to the three roots of a polynomial $P_*^{(3)}$ defined from the sequence of moments up to order five $(0.5, 0.2, p_3, p_4, p_5)$ with (p_3, p_4, p_5) varying over the whole domain $]0,1[^3$. Hence, the whole optimization set Δ is explored without constraints and through a simple parameterization.

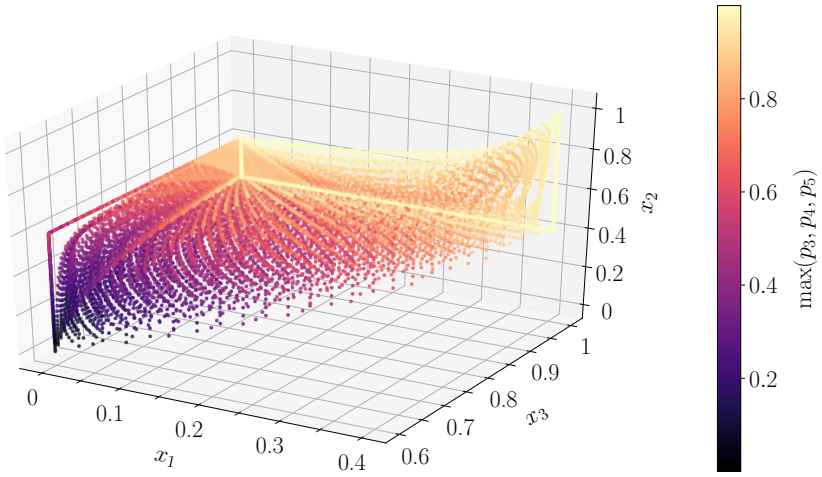


Figure 4.4: The optimization set $\mathcal{V}$ is an algebraic variety, it is the zeros locus of a set of polynomial and can therefore contain singular points. We considered a regular grid for $(p_3, p_4, p_5) \in]0, 1[^3$, that are concatenated to the first two canonical moments $(0.5, 0.2)$ in order to define polynomials $P^*_{(3)}$ whose roots are the coordinates of each points. The regular grid of canonical moments sequence is responsible for the scatter graph, the manifold is not a really a point clouds.

To conclude this section, we established how reparameterizing the extreme points of an unidimensional moment class in Eq. (4.1) with canonical moments. However, real-life computer model often own several inputs, each one being modeled independently either in a moment class or an unimodal moment class. Therefore, the next section is dedicated to generalizing the measure space to fit the OUQ framework introduced in Chapter 3. This includes the unimodal moment class and the product of measure spaces. Most of the works have actually been done in this section, we will show that few adaptations are required to employ canonical moments with unimodal distributions thanks to the very particular form of uniform distribution moments. Moreover, the extension to a product of (unimodal) moment class is simply done by marginal reparameterization.

4.3 Reparameterization of the Optimization Problem

In the general framework introduces in Chapter 3, we study the product of two kinds of measure spaces, the moment class and the unimodal moment class. So far, we studied the parameterization of the moment class in Section 4.1 and 4.2, we now focus on how to adapt the canonical moments parameterization to the unimodal moment class.

4.3.1 Unimodal Moment Class

Consider a unimodal moment class with bounded support $[a, b]$ and mode $\theta \in [a, b]$:

$$\mathcal{A}_c^\dagger = \left\{ \mu \in \mathcal{H}_\theta([a, b]) \mid \mathbb{E}_\mu[X^j] = c_j, \text{ for } 1 \leq j \leq N \right\} , \tag{4.28}$$

with *classical* moment equality constraints. It corresponds to the measure space studied previously in Eq. (4.1) with an additional unimodality constraint. The set of extreme

points is given by Theorem 3.6:

$$\Delta_c^\dagger = \left\{ \mu \in \mathcal{A}^\dagger \mid \sum_{i=1}^{N+1} \omega_j \mathcal{U}(\theta, z_j),\ \omega_j \geq 0,\ z_j \in [a,b] \right\} ,$$

where we recall that $\mathcal{U}(\theta, z_j)$ denotes a uniform distribution on $\mathrm{co}(\{\theta, z_j\})$. Therefore, we are looking for all the tuples $(z_1, \ldots, z_{N+1}) \in \mathbb{R}^{N+1}$ such that the convex combination of uniform distribution satisfies the moment constraints. Now, the nth moment of $\mathcal{U}(\theta, z_j)$ writes (Zwillinger, 2002)

$$m_n^{(j)} = \frac{1}{1+n} \sum_{k=0}^{n} \theta^{n-k} z_j^k$$

so that a measure in $\Delta_c^\dagger$ satisfies the following system of equations

$$\begin{cases} \omega_1 & + \ \ldots \ + & \omega_{N+1} & = 1 \\ \omega_1 \dfrac{(\theta + z_1)}{2} & + \ \ldots \ + & \omega_{N+1} \dfrac{(\theta + z_{N+1})}{2} & = c_1 \\ \vdots & & \vdots & \vdots \\ \omega_1 \dfrac{\sum_{k=0}^{N} \theta^{N-k} z_1^k}{N+1} & + \ \ldots \ + & \omega_{N+1} \dfrac{\sum_{k=0}^{N} \theta^{N-k} z_{N+1}^k}{N+1} & = c_N \ , \end{cases} \tag{4.29}$$

which is equivalent to the system of equations (4.30), obtained by iteratively subtracting the kth equation multiplied by k to the $(k-1)$th equation multiplied by $(k-1)\theta$ for $k = N+1, \ldots, 2$:

$$\begin{cases} \omega_1 & + \ \ldots \ + & \omega_{N+1} & = 1 \\ \omega_1 z_1 & + \ \ldots \ + & \omega_{N+1} z_{N+1} & = 2c_1 - \theta \\ \vdots & & \vdots & \vdots \\ \omega_1 z_1^N & + \ \ldots \ + & \omega_{N+1} z_{N+1}^N & = (N+1)c_N - N\theta c_{N-1} \ . \end{cases} \tag{4.30}$$

One can recognize in this expression the Vandermonde system in Eq. (4.26) associated to the tuple of moment constraints $\mathbf{c}'_N = (2c_1 - \theta,\ 3c_2 - 2\theta c_1,\ \ldots,\ (N+1)c_N - N\theta c_{N-1})$. Therefore, there is a direct equivalence between mixture of Dirac masses with moment sequence $\mathbf{c}'_N$ and mixture of uniform distribution with moments $\mathbf{c}_N$. Hence, using Theorem 4.7 and the canonical moments $\mathbf{p}'_N$ associated to the sequence of moment $\mathbf{c}'_N$, one can compute the support points $(z_1, \ldots, z_{N+1})$ of discrete measures verifying the system Eq. (4.30) which equivalently coincides to the support of uniform mixtures in $\Delta_c^\dagger$ satisfying Eq. (4.29). Once the positions $(z_1, \ldots, z_{N+1})$ are set, one can also compute the weights either by solving the Vandermonde system in Eq. (4.30), or preferably by using Eq. (4.21) with polynomials $P_*^{(N+1)}$ and $Q^{(N+1)}$ computed from the alternative sequence of canonical moments $\mathbf{p}'_N$. We emphasize that this relationship between the moments of a convex combination of Dirac masses and the ones of a convex combination of uniform distributions is remarkable.

The next section focuses on the parameterization of a product measure space. Indeed, we recall that the computer model input measure is multivariate. Because all components are assumed to be independent, the input distribution belongs to a measure space defined as a product of moment classes and/or unimodal moment classes. The canonical moment parameterization of this multivariate space requires further description.

4.3.2 Product Space Parameterization

In the initial engineering problem the computer model takes d input parameters modeled as a d-dimensional random vector evolving in a measure space denoted by $\mathcal{A}_c$. We constructed the measure space in Section 3.3.1 as a product of p moment spaces $\mathcal{A}_{c,i} = \mathcal{A}^*_{c,i}$ for $1 \leq i \leq p$ and $d-p$ unimodal moment spaces $\mathcal{A}_{c,i} = \mathcal{A}^\dagger_{c,i}$ for $p+1 \leq i \leq d$ with modes θ_i, coupled to N_i *classical* moment constraints for every i [see Eq. (4.1) and Eq. (4.28)]. The reparameterization of the optimization space $\mathcal{A}_c = \prod_{i=1}^d \mathcal{A}_{c,i}$ is now obvious, as it consists in marginal parameterization of the extreme points set. The details provided hereunder serve as a reminder for constructing the product space. Those are important before introducing the main algorithm of this thesis for computing a QoI with the canonical moment parameterization. The parameterization for the product space is basically similar to what has been studied previously in this chapter with respect to the tensorization index "$\_i$".

The extreme points set is denoted $\Delta_c = \prod_{i=1}^d \Delta_{c,i}$, where any element in $\Delta_{c,i} \in \mathcal{A}_{c,i}$ is a convex combination of at most $N_i + 1$ Dirac masses for $1 \leq i \leq p$, and convex combination of $N_i + 1$ uniform distribution supported on $\mathrm{co}(\{\theta_i, z_i\})$ for $p+1 \leq i \leq d$. That is,

$$\begin{aligned}
\Delta_{c,i} = \Delta^*_{c,i} &= \left\{ \mu_i \in \mathcal{A}_{c,i} \mid \mu_i = \sum_{k=1}^{N_i+1} \omega_i^{(k)} \delta_{x_i^{(k)}},\ x_i^{(k)} \in [a_i, b_i] \right\} \text{ for } 1 \leq i \leq p\ , \\
\Delta_{c,i} = \Delta^\dagger_{c,i} &= \left\{ \mu_i \in \mathcal{A}_{c,i} \mid \mu_i = \sum_{k=1}^{N_i+1} \omega_i^{(k)} \mathcal{U}(\theta_i, z_i^{(k)}),\ z_i^{(k)} \in [a_i, b_i] \right\} \text{ for } p < i \leq d\ .
\end{aligned} \tag{4.31}$$

On every marginal set $\Delta_{c,i}$ the parameterization depends only the support points, respectively $(x_i^{(k)})_{1 \leq k \leq N_i+1}$ or $(z_i^{(k)})_{1 \leq k \leq N_i+1}$ for any i, and not on the weights that are uniquely determined from the moment constraints and the support points from respectively Eq. (4.26) and Eq. (4.29). Therefore, the optimization takes place in the following manifold

$$\begin{aligned}
\mathcal{V} &= \prod_{i=1}^d \mathcal{V}_i\ , \\
&= \prod_{i=1}^p \left\{ \mathbf{x}_i = \left(x_i^{(1)}, \ldots, x_i^{(N_i+1)}\right) \in [a_i, b_i]^{N_i+1},\ \text{s.t } \mu_i = \sum_{k=1}^{N_i+1} \omega_i^{(k)} \delta_{x_i^{(k)}} \in \Delta_{c,i} \right\} \times \\
&\quad \prod_{i=p+1}^d \left\{ \mathbf{z}_i = \left(z_i^{(1)}, \ldots, z_i^{(N_i+1)}\right) \in [a_i, b_i]^{N_i+1},\ \text{s.t } \mu_i = \sum_{k=1}^{N_i+1} \omega_i^{(k)} \mathcal{U}(\theta_i, z_i^{(k)}) \in \Delta_{c,i} \right\},
\end{aligned} \tag{4.32}$$

which means that the total dimension of the optimization vector is $\sum_{i=1}^d (N_i + 1)$ and that $\mathcal{V}$ is a subset of $\prod_{i=1}^d [a_i, b_i]^{N_i+1}$. Therefore, it grows exponentially with d, the number of input parameters. We emphasize that the canonical moments parameterization consists in exploring each manifold $\mathcal{V}_i$ from sequences of canonical moments in $[0,1]^{N_i+1}$. In the following, we present a pseudo-code algorithm detailing how to optimize efficiently a QoI of a d dimensional computer model using the canonical moments.

4.3.3 Exploratory Algorithms

The general engineering problem has been left behind so far in order to introduce the new parameterization of the optimization space that eases the exploration of the

domain of the optimization. Let ϕ denote a QoI to be optimized over the measure space $\mathcal{A}_c$ presented in Section 4.3.2. The input variables are considered mutually independent and are defined by support bounds, N_i *classical* moment constraints, and possibly some unimodality constraints. The input variables belong in either a moment class Eq. (4.1) or unimodal moment class Eq. (4.28). It is necessary to use a global solver because the QoI does not possess *a priori* any properties such as gradient or smoothness that would ease its optimization. We insist that using a global optimizer is in itself a flaw in the methodology, but the only tool at hand that we have to proceed such optimization.

We introduce in Appendix C.1 the differential evolution (DE) algorithm used for the optimization. The population is composed of a fixed number of vectors in the optimization space $\mathcal{V}$ [Eq. (4.33)] that evolve under given pattern and with a certain amount of randomness. However, the optimization space $\mathcal{V}$ parameterized by the positions of the support of either the convex combination of Dirac masses or uniform distributions is not easily explored, for instance the Mystic toolbox (McKerns et al., 2012) uses this parameterization. Indeed, each individual of the global optimizer is therefore a vector concatenating the support points of discrete measures. Thus, it belongs to the set $\prod_{i=1}^{d}[a_i, b_i]^{N_i+1}$. The constraints are satisfied under the condition that the positions, and the weights generated from them, correspond to an admissible measure in Δ_c. But then, random evolution and mutation of the individuals of the global optimizer leads to new vector of support points in the next generation of population. But, these support point vectors in $\prod_{i=1}^{d}[a_i, b_i]^{N_i+1}$ will not necessarily correspond to measures in Δ_c that satisfy the constraints. Exploring the optimization space this way have been observed to be very ineffective. Under this parameterization and during the optimization process, many support point vectors are not admissible which left few member non rejected in the population, and reduces the performance of the global solver.

It is dear by now that the canonical moments is the right tool to explore such a space. Indeed, the same manifold $\mathcal{V}$ can be parameterized from a canonical moment vector in $\prod_{i=1}^{d}[0,1]^{N_i+1}$. Hence, when evolving and mutating canonical moments vectors, the new population members remain naturally in $\prod_{i=1}^{d}[0,1]^{N_i+1}$, which correspond again to canonical moment sequences of new discrete measures satisfying the constraints. Moreover, the support points of these measures are easily generated from the results of the previous sections of this Chapter. Therefore, no population vectors are rejected under this parameterization which increases drastically the exploration ability and thus the performance of the global optimization of the QoI. The full algorithm to compute the QoI from a canonical moment vector in $[0,1]^{\sum_{i=1}^{d}(N_i+1)}$ is detailed in Algorithm 1. We explain the actual computation of the QoI from the canonical moments, and not the way the solver works. We propose in Appendix C.1 the optimization through a DE solver but any other global optimizer is equally suitable.

We intend to give some explanations on Algorithm 1, the input vector are the N_i+1 *free* canonical moments of each input, so that the total dimension of the input vector is as stated $\sum_{i=1}^{d}(N_i+1)$, every component being in $[0,1]$. Two main loops separate the input variables in moment classes from the ones in unimodal moment classes. In both cases, it is required to compute from the moment constraints $\mathbf{c}_i$ the canonical moment constraints $\mathbf{p}_i$ which are specified for the unimodality case *via* the equivalence between Eq. (4.29) and Eq. (4.30). Then, from the canonical moment sequences one can easily compute the parameters $\zeta_i^{(k)}$ leading to the recursive computation of the polynomials $P_{*i}^{(k)}$ and $Q_i^{(k)}$ in Eq. (4.22) and Eq. (4.23).

Some detailed explanations are required for computing the roots of a polynomial

Algorithm 1: Computation of a QoI from canonical moment parameterization

Input: - lower bounds, $\mathbf{l} = (a_1, \ldots, a_d)$
- upper bounds, $\mathbf{u} = (b_1, \ldots, b_d)$
- moment sequence constraints, $\mathbf{c}_i = (c_i^{(1)}, \ldots, c_i^{(N_i)})$, for $1 \leq i \leq d$

function *QoI*$(p_1^{(N_1+1)}, \ldots, p_1^{(2N_1+1)}, \ldots, p_d^{(N_d+1)}, \ldots, p_d^{(2N_d+1)})$

for $i = 1, \ldots, p$ **do**

using the *Q-D* algorithm, compute the canonical moment constraints associated to the moments $\mathbf{c}_i$: $\mathbf{p}_i = (p_i^{(1)}, \ldots, p_i^{(N_i)})$;

for $k = 1, \ldots, 2N_i + 1$ **do**

$\zeta_i^{(k)} = p_i^{(k)}(1 - p_i^{(k-1)})$;

for $k = 0, \ldots, N_i$ **do**

$P_{i*}^{(k+1)} = (X - a_i - (b_i - a_i)(\zeta_i^{(2k)} + \zeta_i^{(2k+1)}))P_{i*}^{(k)} - (b_i - a_i)^2 \zeta_i^{(2k-1)} \zeta_i^{(2k)} P_{i*}^{(k-1)}$;

for $k = 0, \ldots, N_i - 1$ **do**

$Q_i^{(k+1)} = (X - a_i - (b_i - a_i)(\zeta_i^{(2k+2)} + \zeta_i^{(2k+3)}))Q_i^{(k)} - (b_i - a_i)^2 \zeta_i^{(2k+1)} \zeta_i^{(2k+2)} Q_i^{(k-1)}$;

compute $(x_i^{(1)}, \ldots, x_i^{(N_i+1)})$ roots of $P_{*i}^{(N_i+1)}$;

for $k = 1, \ldots, N_i + 1$ **do**

$\omega_1^{(k)} = \frac{Q_i^{(N_i)}(x_i^{(k)})}{\frac{d}{dx} P_{*i}^{(N_i+1)}(x)|_{x=x_i^{(k)}}}$;

$\mu_i = \sum_{k=1}^{N_i+1} \omega_k \delta_{x_k}$;

for $i = p + 1, \ldots, d$ **do**

$\mathbf{c}'_i = (2c_i^{(1)} - \theta_i, 3c_i^{(2)} - 2\theta_i c_i^{(1)}, \ldots, (N_i + 1)c_i^{(N_i)} - N\theta_i c_i^{(N_i-1)})$;

using the *Q-D* algorithm, compute the canonical moment sequences associated to the constraints $\mathbf{c}'_i$: $\mathbf{p}_i = (p_i^{(1)}, \ldots, p_i^{(N_i)})$;

for $k = 1, \ldots, 2N_i + 1$ **do**

$\zeta_i^{(k)} = p_i^{(k)}(1 - p_i^{(k-1)})$;

for $k = 1, \ldots, N_i$ **do**

$P_{i*}^{(k+1)} = (X - l_i - (u_i - l_i)(\zeta_i^{(2k)} + \zeta_i^{(2k+1)}))P_{i*}^{(k)} - (u_i - l_i)^2 \zeta_i^{(2k-1)} \zeta_i^{(2k)} P_{i*}^{(k-1)}$;

for $k = 0, \ldots, N_i - 1$ **do**

$Q_i^{(k+1)} = (X - a_i - (b_i - a_i)(\zeta_i^{(2k+2)} + \zeta_i^{(2k+3)}))Q_i^{(k)} - (b_i - a_i)^2 \zeta_i^{(2k+1)} \zeta_i^{(2k+2)} Q_i^{(k-1)}$;

compute $(z_i^{(1)}, \ldots, z_i^{(N_i+1)})$ roots of $P_{*i}^{(N_i+1)}$;

for $k = 1, \ldots, N_i + 1$ **do**

$\omega_1^{(k)} = \frac{Q_i^{(N_i)}(z_i^{(k)})}{\frac{d}{dz} P_{*i}^{(N_i+1)}(z)|_{z=z_i^{(k)}}}$;

$\mu_i = \sum_{k=1}^{N_i+1} \omega_k \mathcal{U}(\theta_i, z_k)$;

return $\phi((\mu_1, \ldots, \mu_d))$;

$P_{*i}^{(N_i+1)}$. Let us consider that the sequence of canonical moments $\mathbf{p}_i$ is fixed, we recall that the sequence $(P_{*i}^{(N_i+1)})_{1 \leq k \leq N_i+1}$ is composed of the monic orthogonal polynomials associated to the related measure μ_i (Dette and Studden, 1997). Usually, one of the central problem with orthogonal polynomials is to generate the coefficients in the basic three-terms recurrence relation they are known to satisfy (Marcellán et al., 2006):

$$P_{*i}^{(k+1)} = (X - \alpha_i^{(k)})P_{*i}^{(k)} - \beta_i^{(k)} P_{*i}^{(k-1)} \ .$$

Here, those coefficients are precisely given in Eq. (4.22) and Eq. (4.23) by the canonical moments, for $k \geq 1$

$$\begin{aligned} \alpha_i^{(k)} &= a_i + (b_i - a_i)(\zeta_i^{(2k)} + \zeta_i^{(2k+1)}) \ , \\ \beta_i^{(k)} &= (b_i - a_i)^2 \zeta_i^{(2k-1)} \zeta_i^{(2k)} \ . \end{aligned} \tag{4.33}$$

The coefficient $\beta_i^{(0)}$ multiplies $P_{*i}^{(-1)} = 0$, and hence can be arbitrary. Placing these coefficients $\alpha_i^{(k)}$ on the diagonal and $\sqrt{\beta_i^{(k)}}$ on the two side diagonals of a real symmetric tridiagonal matrix yields the so called Jacobi matrix of the measure μ_i:

$$J(\mu_i) = \begin{bmatrix} \alpha_i^{(0)} & \sqrt{\beta_i^{(1)}} & & \mathbf{0} \\ \sqrt{\beta_i^{(1)}} & \alpha_i^{(1)} & \sqrt{\beta_i^{(2)}} & \\ & \sqrt{\beta_i^{(2)}} & \alpha_i^{(3)} & \ddots \\ & & \ddots & \ddots \\ \mathbf{0} & & & \end{bmatrix} . \tag{4.34}$$

We denote $J_n(\mu_i)$ its principal minor matrix of order n. The zeros of $P_{*i}^{(N_i+1)}$ are precisely the eigenvalues of $J_{N_i+1}(\mu_i)$ (Gautschi, 2006, p. 7). This is one of the reasons for which the recursive coefficients of the three-terms recurrence relation of orthogonal polynomials are of great interest. Fast, well-conditioned, and precise algorithms exist for computing the eigenvalues of a real, symmetric, and tridiagonal matrix (Dhillon and Parlett, 2003) which are much more performing than classical root-finding algorithms that compute the eigenvalues of the companion matrix of $P_{*i}^{(N_i+1)}$ (Horn and Johnson, 2012).

At last in Algorithm 1, the weights are recovered from the positions and the moment sequence using Eq. (4.21) after computing the two polynomials $Q_i^{(N_i)}$ and $P_{*i}^{(N_i+1)}$ from the same recurrence formula. Finally, the complete marginal measure is recovered from $\mu_i = \sum_{i=1}^{N_i+1} \omega_i \delta_{x_i}$ for $1 \leq i \leq p$ or $\mu_i = \sum_{i=1}^{N_i+1} \omega_i \mathcal{U}(\theta_i, z_i)$, for $p < i \leq d$. This allows the calculation of the QoI and its propagation into a global optimizer.

4.3.4 Inequality Constraints

As stated in Section 4.1, it is common to have information, notably from an expert, about the mean or variance of an input. Thus, on *classical* moments of input parameters. That is why all this chapter is focused in enforcing *classical* moment constraints in the moment class $\mathcal{A}_c$ instead of *generalized* moment constraints as in Chapter 3. So far, we restricted the development of the canonical moment parameterization to *equality* constraints. However, uncertainty can affect the moment values. Indeed, consider for instance that engineers possess knowledge on the average value of an input parameter estimated either from data or expert opinion. This value is tainted with epistemic uncertainty, arising from estimation error, or the lack of knowledge of the expert. But one could possesses as well confidence intervals that bring valuable information. The analyst can benefit from accounting for the uncertainty in the moment values. In the following, we propose to replace the restrictive moment equality constraints with inequalities. The new optimization set reads

$$\begin{aligned} \mathcal{A}_{t,i} = \mathcal{A}_{t,i}^{*} &= \left\{ \mu_i \in \mathcal{P}([a_i, b_i]) \mid \alpha_i^{(j)} \leq \mathbb{E}_{\mu_i}[X^j] \leq \beta_i^{(j)} \ , \ 1 \leq j \leq N_i \right\} \ , 1 \leq i \leq p \ , \\ \mathcal{A}_{t,i} = \mathcal{A}_{t,i}^{\dagger} &= \left\{ \mu_i \in \mathcal{H}_{\theta_i}([a_i, b_i]) \mid \alpha_i^{(j)} \leq \mathbb{E}_{\mu_i}[x^j] \leq \beta_i^{(j)} \ , \ 1 \leq j \leq N_i \right\} \ , p < i \leq d \ , \end{aligned} \tag{4.35}$$

and $\mathcal{A} = \prod_{i=1}^{d} \mathcal{A}_{t,i}$. The extreme points $\Delta_{t,i}$ of $\mathcal{A}_{t,i}$ are exactly the same as Eq. (4.31). We now propose a modified version of Algorithm 1 to solve the problem with inequality constraints. For $i = 1, \dots, d$, we denote the moment lower bounds $\boldsymbol{\alpha}_i = (\alpha_i^{(1)}, \dots, \alpha_i^{(N_i)})$ and the moment upper bounds $\boldsymbol{\beta}_i = (\beta_i^{(1)}, \dots, \beta_i^{(N_i)})$. We use Eq. (4.5) to calculate the corresponding moment sequence $\tilde{\boldsymbol{\alpha}}_i = (\tilde{\alpha}_i^{(1)}, \dots, \tilde{\alpha}_i^{(N_i)})$ and $\tilde{\boldsymbol{\beta}}_i = (\tilde{\beta}_i^{(1)}, \dots, \tilde{\beta}_i^{(N_i)})$ after affine transformation to $[0, 1]$. This new algorithm roughly consists in considering the uncertain moments $\mathbb{E}_{\mu_i}[x^j]$ as parameters in the optimization program.

The function computing the QoI in Algorithm 2 has $\sum_{i=1}^{d}(2N_i + 1)$ arguments. The additional parameters in comparison with Algorithm 1 are actually the first $(N_i)_{|i=1,\dots,d}$th moments of the inputs that were previously fixed. They are denoted $\tilde{c}_i^{(k)}$ and vary between $[\tilde{\alpha}_i^{(k)}, \tilde{\beta}_i^{(k)}]$, followed by the $(N_i + 1)$ *free* canonical moments $(p_i^{(N_i+1)}, \dots, p_i^{(2N_i+1)})$ in between $[0, 1]^{N_i+1}$. Therefore, the input vector belongs in the pavement

$$\prod_{k_1=1}^{N_1} [\tilde{\alpha}_1^{(k_1)}, \tilde{\beta}_1^{(k_1)}] \times [0, 1]^{N_1+1} \times \cdots \times \prod_{k_d=1}^{N_d} [\tilde{\alpha}_d^{(k_d)}, \tilde{\beta}_d^{(k_d)}] \times [0, 1]^{N_d+1} \ .$$

The first step in Algorithm 2 consists in calculating the canonical moments up to degree N_i for $i = 1, \dots, d$. This was similar in Algorithm 1 except this time the first N_i moments of the ith input are not fixed and will differ for every individual of the global optimizer population. They are only bounded in between $\tilde{\boldsymbol{\alpha}}_i$ and $\tilde{\boldsymbol{\beta}}_i$. This ensures that the moment constraints are satisfied while the canonical moments from degree $N_i + 1$ up to degree $2N_i + 1$ can vary between $[0, 1]$ in order to generate all the admissible measures in $\Delta_{t,i}$. Once again this new QoI function can be optimized using the DE global solver.

We highlight that one can consider in the same way an uncertainty on the mode θ_i in the unimodal moment space. Simply by considering θ_i as a parameter varying in between its bounds.

Algorithm 2: Computation of a QoI with inequality constraints

Input: - lower bounds, $\mathbf{a} = (a_1, \dots, a_d)$
- upper bounds, $\mathbf{b} = (b_1, \dots, b_d)$
- moments lower bounds, $\tilde{\boldsymbol{\alpha}}_i = (\tilde{\alpha}_i^{(1)}, \dots, \tilde{\alpha}_i^{(N_i)})$ for $i = 1, \dots, d$
- moments upper bounds, $\tilde{\boldsymbol{\beta}}_i = (\tilde{\beta}_i^{(1)}, \dots, \tilde{\beta}_i^{(N_i)})$ for $i = 1, \dots, d$

function *QoI*
$(\tilde{c}_1^{(1)}, \dots, \tilde{c}_1^{(N_1)}, p_1^{(N_1+1)}, \dots, p_1^{(2N_1+1)}, \dots, \tilde{c}_d^{(1)}, \dots, \tilde{c}_d^{(N_d)}, p_d^{(N_d+1)}, \dots, p_d^{(2N_d+1)})$

for $i = 1, \dots, p$ **do**

using the *Q-D* algorithm, compute the canonical moment constraints associated to the moments $\tilde{\mathbf{c}}_i$: $\mathbf{p}_i = (p_i^{(1)}, \dots, p_i^{(N_i)})$;

for $k = 1, \dots, 2N_i + 1$ **do**

$\zeta_i^{(k)} = p_i^{(k)}(1 - p_i^{(k-1)})$;

for $k = 0, \dots, N_i$ **do**

$P_{i*}^{(k+1)} = (X - a_i - (b_i - a_i)(\zeta_i^{(2k)} + \zeta_i^{(2k+1)}))P_{i*}^{(k)} - (b_i - a_i)^2 \zeta_i^{(2k-1)} \zeta_i^{(2k)} P_{i*}^{(k-1)}$;

for $k = 0, \dots, N_i - 1$ **do**

$Q_i^{(k+1)} = (X - a_i - (b_i - a_i)(\zeta_i^{(2k+2)} + \zeta_i^{(2k+3)}))Q_i^{(k)} - (b_i - a_i)^2 \zeta_i^{(2k+1)} \zeta_i^{(2k+2)} Q_i^{(k-1)}$;

compute $(x_i^{(1)}, \dots, x_i^{(N_i+1)})$ roots of $P_{*i}^{(N_i+1)}$;

for $k = 1, \dots, N_i + 1$ **do**

$\omega_1^{(k)} = \frac{Q_i^{(N_i)}(x_i^{(k)})}{\frac{d}{dx} P_{*i}^{(N_i+1)}(x)|_{x = x_i^{(k)}}}$;

$\mu_i = \sum_{k=1}^{N_i+1} \omega_k \delta_{x_k}$;

for $i = p + 1, \dots, d$ **do**

$\tilde{\mathbf{c}}_i' = (2\tilde{c}_i^{(1)} - \theta_i, 3\tilde{c}_i^{(2)} - 2\theta_i \tilde{c}_i^{(1)}, \dots, (N_i + 1)\tilde{c}_i^{(N_i)} - N\theta_i \tilde{c}_i^{(N_i-1)})$;

using the *Q-D* algorithm, compute the canonical moment sequences associated to the constraints $\tilde{\mathbf{c}}_i'$: $\mathbf{p}_i = (p_i^{(1)}, \dots, p_i^{(N_i)})$;

for $k = 1, \dots, 2N_i + 1$ **do**

$\zeta_i^{(k)} = p_i^{(k)}(1 - p_i^{(k-1)})$;

for $k = 1, \dots, N_i$ **do**

$P_{i*}^{(k+1)} = (X - l_i - (u_i - l_i)(\zeta_i^{(2k)} + \zeta_i^{(2k+1)}))P_{i*}^{(k)} - (u_i - l_i)^2 \zeta_i^{(2k-1)} \zeta_i^{(2k)} P_{i*}^{(k-1)}$;

for $k = 0, \dots, N_i - 1$ **do**

$Q_i^{(k+1)} = (X - a_i - (b_i - a_i)(\zeta_i^{(2k+2)} + \zeta_i^{(2k+3)}))Q_i^{(k)} - (b_i - a_i)^2 \zeta_i^{(2k+1)} \zeta_i^{(2k+2)} Q_i^{(k-1)}$;

compute $(z_i^{(1)}, \dots, z_i^{(N_i+1)})$ roots of $P_{*i}^{(N_i+1)}$;

for $k = 1, \dots, N_i + 1$ **do**

$\omega_1^{(k)} = \frac{Q_i^{(N_i)}(z_i^{(k)})}{\frac{d}{dz} P_{*i}^{(N_i+1)}(z)|_{z = z_i^{(k)}}}$;

$\mu_i = \sum_{k=1}^{N_i+1} \omega_k \mathcal{U}(\theta_i, z_k)$;

return $\phi((\mu_1, \dots, \mu_d))$;

Conclusion

In this chapter, we introduce a new parameterization of the extreme points of the measure space that improves the optimization of the QoI. The multidimensional (uni-

modal) moment class presented in Chapter 2 was very general and its measures were defined by *generalized* moment constraints. The first thing was to particularized the measure space in several ways exposed in Section 4.1. Namely, we focused our work on univariate bounded scalar measures in a moment class with finitely many *classical* moment equality constraints. We presented in this case the canonical moment parameterization of the extreme points, which allows efficient generation of discrete measures satisfying the constraints. Fundamentally, we proved that there exists a bijection between the set of all discrete measures supported on at most $N+1$ points satisfying N *classical* moment equality constraints, and the pavement $[0,1]^{N+1}$. In a second phase, we relaxed some of the assumptions concerning the moment class, that were made for clarity reasons. We presented the methodology to adapt the canonical moment parameterization of the extreme points of the moment class to those of the unimodal moment class. It consists in a simple trick involving an alternative moment sequence thanks to the close similarity between these two spaces. Finally, the univariate case was extended to the multivariate case. Actually the theory of canonical moments only applies to scalar measures, and the whole methodology works well here because of the product structure which stands, we recall, when computer model inputs exhibit mutual independence.

Eventually, the only restrictive assumptions left compared to the optimization space in Chapter 2 is that the *generalized* moment constraints were replaced by *classical* moment constraints and the input support was bounded into $[a_i, b_i]$. This serves as the basis for developing the canonical moment parameterization. We acknowledge that these hypotheses are restrictive in an industrial context. Indeed, it is not always usual to have information in *classical* moments of an input parameter. However, from the expert knowledge and available data one can always retrieve a mean or variance, with some confidence intervals that could typically serve as the basis for defining inequality constraints. Moreover, if the input distribution is fully elicited one can also recover its moments. In all generality, enforcing moment constraints eases interpretation, the first two moments correspond equivalently to the mean and the variance of an input distribution. These are familiar quantity for engineers to work with, making the framework quite attractive for practical use (Lemaître et al., 2015; Delage et al., 2018; Gauchy et al., 2019).

We emphasize that the canonical moments parameterization of the (unimodal) moment class does not improve in any way the computational efficiency of the QoI. Algorithm 1 and a direct computation of the QoI from the position and weights of discrete measure require approximately the same computational time. It is actually sightly more expensive for Algorithm 1 because there is an additional step of transforming a canonical moment sequence into a discrete measure. The real advantage of the canonical moments relies in the exploration performance of the global optimizer. We use in this thesis a DE algorithm presented in Appendix C, but any global solver would do. Indeed, the bijection between the set of all discrete measures supported on at most $N+1$ points satisfying N *classical* moment equality constraints and the pavement $[0,1]^{N+1}$ is explicitly determined thanks to canonical moments. Which means one can explore a set of discrete distributions constrained by their moment from a pavement with no constraint. Hence, the exploration of the optimization space is drastically improved, so are the performance of the overall optimization. In particular, because each generation of the global solver evolves without risk of individuals being rejected, so that there is no loss of population member in the DE algorithm as it can happen from a direct parameterization with support points and weights. However, the overall

optimization relies on a global optimizer which has no proof of convergence. Special attention should be paid for this issue, in particular there probably exists better suited solver for this problem that we did not investigate in this thesis.

Under this parameterization, the OUQ framework remains very general, and any quasi-convex lower semicontinuous function can be optimized using the methodology introduced. In the next Chapter we focus on particular quantities of interest that are of some practical interest in engineering studies.

5

Remarkable Quantities of Interest

« *On est puceau de l'horreur comme on l'est de la volupté.* »

Louis-Ferdinand Céline

Contents

Introduction

In the previous chapters, we replaced the abstract topological vector space with practical measure spaces, namely the moment and unimodal moment spaces. Indeed, this thesis is related to important applicative challenges. In this section, we now leave behind the abstract quasi-convex lower semicontinuous function firstly introduced to study more practical and specific QoIs. We present a non-exhaustive list of interesting QoIs, well known by statistician that are also well known for OUQ-type optimization. This chapter builds on the framework introduced in Chapters 3 and 4. We have our disposal a computer code G seen as a black box function (see Eq. (1.1)), with d marginal inputs modeled as independent random variables. We are looking for bounds of a QoI for the computer model output distribution. To compute these bounds, one wishes to exploit the performance of the canonical moment parameterization, so that, every input distribution μ_i is bounded and assumed to belong to either a moment class or a unimodal moment class, with only *classical* moment constraints.

For sake of clarity, we will detail how to obtain bounds on these QoI on various examples. Moreover, in these examples we simplify the notations and often consider that each input distribution μ_i belongs to a moment class. So that, in our examples there are no unimodal measures, which is equivalent to $p = d$ in Eq. (3.7). We also assume that no additional moment constraint is enforced on the joint distribution μ, so that $N = 0$ in Eq. (3.10). The optimization set thus reads

$$\mathcal{A}^* = \left\{ \mu \in \prod_{i=1}^{d} \mathcal{P}(\mathcal{X}_i) \mid \mathbb{E}_{\mu_i}[X^j] \leq 0 \text{ for } 1 \leq j \leq N_i \text{ and } 1 \leq i \leq d \right\} \tag{5.1}$$

Applying the reduction theorem 3.8, upper bounds for any quasi-convex and lower semicontinuous functions are reached on the extreme points of $\mathcal{A}^*$, which are the product of discrete measures supported on $N_i + 1$ points:

$$\Delta^* = \left\{ \mu \in \mathcal{A} \mid \mu_i = \sum_{k=1}^{N_i+1} \omega_i^{(k)} \delta_{x_i^{(k)}} \right\}. \tag{5.2}$$

We insist that this restricted scenario only concerns the examples, and it will be recalled any time necessary, it is set for clarity purposes. Indeed, in the special case where no input is in a unimodal moment class, the optimization of the QoI is straightforward because each evaluation of the QoI is exact (up to the computing error). As every input is discrete, no approximation in the computation of the QoI is needed. In contrast, if at least one input distribution belongs to a unimodal moment class, its extreme points are the convex combinations of uniform distributions, and are therefore continuous measures. Hence, the computation of the QoI requires stochastic simulations such as MC methods, which adds noise to the objective function.

We intend to illustrate the different QoIs applied to the flood model presented in Chapter 2. The flood model takes four input parameters Q, K_s, Z_v, and Z_m, whose initial distributions are introduced in Table 2.1. However, different sources of uncertainty affect these distributions, such as: the choice of the parametric family, errors in the estimation of their parameters, the experimental dataset accuracy, which constitute a non-exhaustive list. Because they are imprecise, the distributions are not considered fixed but belong instead to moment classes. That means they are only characterized by some of their moments, here up to order three. This is less intrusive than setting the whole distribution in terms of the supplied information. These moments up to

order three are given in Table 5.1. The moments in Table 5.1 correspond to those of the initial distributions in 2.1. Consequently, it is possible to compare the optimal bounds on the QoI to the QoI evaluated from the joint initial distribution (noted μ_{init}), as the initial distributions belong themselves to the moment classes. However, taking as fixed constraints the moments of the initial distributions, that we specifically have uncertainty on, can sound a bit flawed. Nevertheless, we recall that this example serves as an illustration, and because the initial distribution parameters are also estimated from the experimental data, the moments of the distribution concur with the physical measurement. Hence, there is no scientific aberration.

Table 5.1: Corresponding moment constraints of the 4 inputs of the flood model.

n°	Variable	Bounds	Mean	Second order moment	Third order moment
1	Q	$[160, 3580]$	1319.42	2.1632×10^6	4.18×10^9
2	K_s	$[12.55, 47.45]$	30	949.137	31422.3
3	Z_v	$[49, 51]$	50	2500.17	125025
4	Z_m	$[54, 55]$	54.5	2970.29	161885.44

In every illustration, the QoI is computed with Algorithm 1 (p.66) using the canonical moment parameterization. This Chapter is organized as a list and every Section exposes a different QoI. However, the sections are not entirely independent; Section 5.3 focuses on the quantile function, and refers to Section 5.1 which deals with measure affine functions and more specifically probabilities of failure. The superquantile presented in Section 5.4 refers to the results of Section 5.3. Finally, Sections 5.5 and 5.6 are mainly independent as they present respectively sensitivity indices and robust Bayesian inference, closely related to the OUQ framework.

5.1 Example of Measure Affine Function

It was shown in Section 3.3.3 that measure affine functions were particularly interesting because they allow relaxation of the lower semicontinuity assumption. Moreover, an affine function is both quasi-convex and quasi-concave. Hence, it is possible to minimize or maximize the quantity of interest (Theorem 3.9). We present first some specific measure affine functions corresponding to a large range of applications.

Proposition 5.1. *Given a polish space $\mathcal{X}$, let $\mathcal{A}$ be a convex subset of $\mathcal{P}(\mathcal{X})$ with generator Δ, and let q be measurable on $\mathcal{X}$ with respect to any measure in $\mathcal{A}$. Then the functional $\mu \mapsto \mathbb{E}_\mu[q] = \int_{\mathcal{X}} q \, \mathrm{d}\mu$ defined for all $\mu \in \mathcal{A}$ is measure affine.*

Proof. Let μ be a measure on $\mathcal{A}$, then there exists a measure $\boldsymbol{\nu}$ supported on Δ with barycenter μ. Therefore, from the barycentric definition Eq. (3.3)

$$\mathbb{E}_\mu[q] = \int_\Delta \mathbb{E}_s[q] \, \mathrm{d}\boldsymbol{\nu}(s) \ ,$$

as $\mu \mapsto \mathbb{E}_\mu[q]$ belongs in the topological dual of $\mathcal{A}$. Therefore $\mu \mapsto \mathbb{E}_\mu[q]$ is measure affine. □

We know that by restraining the analysis to measure affine functions, the lower semicontinuity assumption can be dropped from the reduction theorem, as stated in

Theorem 3.9. However, it remains interesting to have a characterization of the lower semicontinuity of a measure affine function. The lower semicontinuity of the affine function in Proposition 5.1 depends of the property of q. More precisely, it stands that the lower semicontinuity of q (respectively upper semicontinuous) implies the lower semicontinuity (respectively upper semicontinuity) of the mapping $\mu \mapsto \int q\, d\mu$ (Aliprantis and Border, 2007, Theorem 15.5).

Measure affine functionals defined as $\mu \mapsto \mathbb{E}_\mu[q]$ cover a large range of interesting quantities for the statistical analysis of computer codes. We give some examples hereunder.

5.1.1 Moments of G

Consider $q(\mathbf{x}) = G(\mathbf{x})$, then the function $\mu \mapsto \mathbb{E}_\mu[q]$ yields as a QoI the expectation of the computer code G's output. More generally, moments of arbitrary order can be studied, by considering $q(\mathbf{x}) = G(\mathbf{x})^n$, for $n \in \mathbb{N}$. For instance, given $\mu \in \mathcal{A}$, the mean of the computer model can be computed as follows

$$\begin{aligned}\mathbb{E}_\mu[G(\mathbf{X})] &= \int_{\mathcal{X}} G(\mathbf{x})\, \mathrm{d}\mu(\mathbf{x}) \ , \\ &= \int_{\mathcal{X}_1} \cdots \int_{\mathcal{X}_d} G(x_1, \ldots, x_d)\, \mathrm{d}\mu_1(x_1) \ldots \mathrm{d}\mu_d(x_d) \ .\end{aligned}$$

Example. We now simplify the problem for clarity reasons before applying the reduction Theorem. Consider that $N = 0$ in Eq. (3.10) and that every marginal distribution belongs to a moment class [$p = d$ in Eq. (3.7)]. This means that every measure in the measure space $\mathcal{A}^*$ is the product of marginal distributions with N_i moment constraints [see Eq. (5.1)], whose extreme points in Δ^* are products of discrete measures [see Eq. (5.2)]. Therefore,

$$\begin{aligned}\sup_{\mu \in \mathcal{A}^*} \mathbb{E}_\mu[G(\mathbf{X})] &= \sup_{\mu \in \Delta^*} \mathbb{E}_\mu[G(X_1, \ldots, X_d)] \ , \\ &= \sup_{\mu \in \Delta^*} \sum_{k_1=1}^{N_1+1} \cdots \sum_{k_d=1}^{N_d+1} \omega_1^{(k_1)} \ldots \omega_d^{(k_d)} G(x_{k_1}^{(1)}, \ldots, x_{k_d}^{(d)}) \ ,\end{aligned} \tag{5.3}$$

(5.4)

because every X_i follows a distribution μ_i that is discrete. One can see that this expression is convenient to compute as it requires a finite number of calls to the code G to get the exact expectation of the computer model output. Usually and especially when the distributions are continuous, computing the expectation of a function is done by MC methods based on the central limit theorem, yielding an estimation of the expectation and some confidence interval. However, this expression highlights a numerical difficulty. Even when the computer model G is cheap, when proceeding to the optimization, the main cost will arise from the high number of calls to the code G, that is run exactly $\prod_{i=1}^d (N_i + 1)$ times for one evaluation of the QoI. Unfortunately, it also means the complexity grows exponentially with the input dimension d.

Note that whenever one marginal measure is not discrete (such as when μ_i belongs to a unimodal moment class, whose extreme points are thus convex combinations of uniform distributions) then the computation of the expectation requires stochastic algorithms.

5.1.2 Failure Probability

Another major example of measure affine function of the form given in Proposition 5.1 is a probability. Indeed, the choice $q(\mathbf{x}) = \mathbb{1}_C$, the indicator function of a set C, yields a probability. An important example that will be further intensively studied would be $q(\mathbf{x}) = \mathbb{1}_{\{G(\mathbf{x})\leq h\}}$, which yields the PoF at threshold $h \in \mathbb{R}$. In particular, we will show how many other QoIs such as quantiles or superquantiles, are reduced to the optimization of a PoF. More generally, the choice of a loss function $q(\mathbf{x}) = L(G(\mathbf{x}), a)$ where a denotes some decision, yields the expected loss of the decision a.

Example. Consider as an illustration the same setting $\mathcal{A}^*$ [see Eq. (5.1)] as in the previous example with $N = 0$ in Eq. (3.10) and all the marginal distributions belong to a moment class [$p = d$ in Eq. (3.7)]. Then, because the extreme points in Δ^* [see Eq. (5.2)] are products of discrete distributions,

$$\begin{aligned}\sup_{\mu\in\mathcal{A}^*} \mathbb{E}_\mu[\mathbb{1}_{\{G(\mathbf{X})\leq h\}}] &= \sup_{\mu\in\mathcal{A}^*} \mathbb{P}_\mu\left(G(\mathbf{X})\leq h\right) \ , \\ &= \sup_{\mu\in\Delta^*} \mathbb{P}_\mu\left(G(\mathbf{X})\leq h\right) \ , \\ &= \sup_{\mu\in\Delta^*} \sum_{k_1=1}^{N_1+1}\cdots\sum_{k_d=1}^{N_d+1} \omega_1^{(k_1)}\ldots\omega_d^{(k_d)}\mathbb{1}_{\{G(x_{k_1}^{(1)},\ldots,x_{k_d}^{(d)})\leq h\}} \ , \end{aligned} \tag{5.5}$$

which once again is convenient to compute, as it is the exact probability of failure rather than an asymptotically converging estimate. Once again, this expression highlights that the main cost of the optimization arises from the high number of calls to the computer model G, which evaluated on a d-dimensional grid of size $\prod_{i=1}^d (N_i + 1)$. Consequently we recommend from our personal experience the number of input parameters d to be kept under 10 if implemented on a regular computer.

We now illustrate the impact of accounting for a second level of uncertainty in the input distribution of the PoF of the flood model presented in Chapter 2.

5.1.3 Illustration

Consider the flood model presented in the introduction of this chapter. We assume that the input distribution belongs to the moment class $\mathcal{A}^*$ in Eq. (5.1) with corresponding moments described in Table 5.1. The variable of interest $H(\mathbf{X})$ which is the output of the model represents the maximal annual water height of the river. Consider the dyke's height h, the probability that the river exceed the dyke's height is the following PoF

$$\mathbb{P}(H(\mathbf{X}) \geq h) = 1 - \mathbb{P}(H(\mathbf{X}) \leq h) = 1 - F_\mu(h) \ . \tag{5.6}$$

Therefore, the problem boils down to the computation of $F_\mu(h)$, the CDF of the output of the underlying computer model G, for a given threshold h. In the following, we indifferently refer to $\mathbb{P}(H(\mathbf{X}) \geq h)$ or $\mathbb{P}(H(\mathbf{X}) \leq h)$ as the PoF, depending on the context. The measure μ is uncertain, and we consider several levels of information, corresponding to different moment constraints. More precisely, we define the three

following moment classes:

$$\begin{aligned}
\mathcal{A}^{(1)} &= \prod_{i=1}^{4} \left\{ \mu_i \in \mathcal{P}([a_i, b_i]) : \mathbb{E}_{\mu_i}[X] = c_i^{(1)} \right\} , \\
\mathcal{A}^{(2)} &= \prod_{i=1}^{4} \left\{ \mu_i \in \mathcal{P}([a_i, b_i]) : \mathbb{E}_{\mu_i}[X] = c_i^{(1)}, \mathbb{E}_{\mu_i}[X^2] = c_i^{(2)} \right\} , \\
\mathcal{A}^{(3)} &= \prod_{i=1}^{4} \left\{ \mu_i \in \mathcal{P}([a_i, b_i]) : \mathbb{E}_{\mu_i}[X] = c_i^{(1)}, \mathbb{E}_{\mu_i}[X^2] = c_i^{(2)}, \mathbb{E}_{\mu_i}[X^3] = c_i^{(3)} \right\} ,
\end{aligned} \tag{5.7}$$

where $a_i, b_i, c_i^{(1)}, c_i^{(2)}, c_i^{(3)}$ for $1 \leq i \leq 4$, correspond respectively to the bounds and moment constraints given in Table 5.1. Hence, measures in $\mathcal{A}^{(1)}$ are known only by their means, the ones of $\mathcal{A}^{(2)}$ satisfy mean and variance conditions, and the space $\mathcal{A}^{(3)}$ defines measures with their three first moments. It follows that $\mu_{init} \in \mathcal{A}^{(3)} \subset \mathcal{A}^{(2)} \subset \mathcal{A}^{(1)} \subset \mathcal{P}(\prod_{i=1}^{4}[a_i, b_i])$, where μ_{init} is the initial joint distribution given in Table 2.1. Because we aim at computing the worst-case scenario for the PoF with respect to the uncertainty affecting the input measure, the following inequalities hold for any given $h \in \mathbb{R}$:

$$\inf_{\mu \in \mathcal{A}^{(1)}} F_\mu(h) \leq \inf_{\mu \in \mathcal{A}^{(2)}} F_\mu(h) \leq \inf_{\mu \in \mathcal{A}^{(3)}} F_\mu(h) ,$$

which correspond to the output CDF lower envelop in each measure space. The computation of the PoF makes use of the canonical moment parameterization in order for the global optimizer to efficiently explore the moment space $\mathcal{A}^{(j)}$, for $j = 1, 2, 3$ as presented in Algorithm 1, and we use the DE solver presented in Appendix C.1 implemented in python. The optimization cost increases with the number of constraints, for each threshold it takes few seconds to optimize the PoF over $\mathcal{A}^{(1)}$, around 30 seconds over $\mathcal{A}^{(2)}$ and up to 1 minute over $\mathcal{A}^{(3)}$. The results are depicted in Fig. 5.1; the optimal PoF over the set $\mathcal{A}^{(1)}$ is seen to be significantly lower than the other lower bounds. Indeed, a single constraint enforced on the mean means that the optimal solution is supported on at most two points. As has already been observed in the example of Section 3.2.1, this favors a concentration of the input distribution's mass on its most unfavorable value. An additional constraint on the variance gives more realistic bounds on the PoF, corresponding to its optimization over $\mathcal{A}^{(2)}$. Of course, the more constraints we add, the closer we get to the initial distribution CDF which belongs to each $\mathcal{A}^{(j)}$ as its moments match the constraints. One can see that the three constraints defining $\mathcal{A}^{(3)}$ strongly reduce the size of the measure space. Indeed, the corresponding moment space in Eq. (4.4) has a volume decreasing exponentially fast with respect to the constraint number [see Eq. (4.7)] so that the PoF optimum reaches rapidly the value of the initial distribution.

In the following, we focus on enforcing two constraints on every marginal distribution. Indeed, it is the handiest way of defining constraints in the sense that mean and variance are familiar, commonly used quantities. Moreover, they correspond to moment class $\mathcal{A}^{(2)}$, whose size is large enough to encompass the uncertainty affecting the input measure, but also small enough to yield physically interpretable bounds as seen in Fig. 5.1. Additionally, we wish to enforce a most plausible value for the Strickler coefficient K_s. Therefore, this uncertain quantity is now assumed to belong to the unimodal moment class $\mathcal{A}_2^{\dagger}$ with two moment constraints. The following list of measure spaces summarizes the OUQ optimization problem: the optimization space is

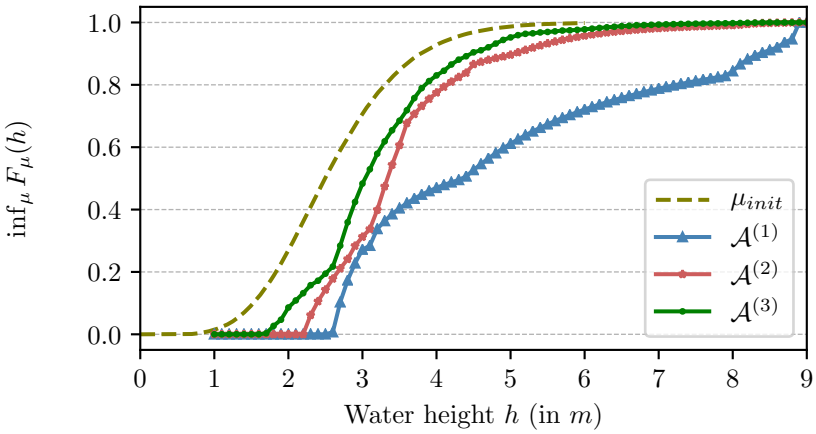


Figure 5.1: The initial CDF of the flood model is compared to the CDF lower envelop over different moment classes presented in Eq. (5.7). This lower envelop is computed as the lowest PoF for a sample of threshold h. This lowest probability of failure encompasses a worst-case scenario with respect to the uncertainty affecting the input distribution.

noted $\mathcal{A}^{uni} = \mathcal{A}_1^* \otimes \mathcal{A}_2^\dagger \otimes \mathcal{A}_3^* \otimes \mathcal{A}_3^*$, with

$$\begin{aligned}
\mathcal{A}_1^* &= \left\{\mu_1 \in \mathcal{P}([160, 3580]) \mid \mathbb{E}_{\mu_1}[X] = 1319.42,\ \mathbb{E}_{\mu_1}[X^2] = 216320\right\} , \\
\mathcal{A}_2^\dagger &= \left\{\mu_2 \in \mathcal{H}_{30}([12.55, 47.45]) \mid \mathbb{E}_{\mu_2}[X] = 30,\ \mathbb{E}_{\mu_2}[X^2] = 949\right\} , \\
\mathcal{A}_3^* &= \left\{\mu_3 \in \mathcal{P}([49, 51]) \mid \mathbb{E}_{\mu_3}[X] = 50, \mathbb{E}_{\mu_3}[X^2] = 2500.17\right\} , \\
\mathcal{A}_4^* &= \left\{\mu_4 \in \mathcal{P}([54, 55]) \mid \mathbb{E}_{\mu_4}[X] = 54.5, \mathbb{E}_{\mu_4}[X^2] = 2970.29\right\} .
\end{aligned} \tag{5.8}$$

where $\mathcal{H}$ is defined in Section 3.2.2. The CDF lower envelop over this space corresponding to the optimal PoF for varying threshold h is illustrated in Fig. 5.2. The initial distribution and the optimized PoF over $\mathcal{A}^{(2)}$ are also depicted in the same figure, so that one can compare the influence of the unimodality constraints on the size of the optimization space. Notice that the inclusion $\mu_{init} \in \mathcal{A}^{uni} \subset \mathcal{A}^{(2)}$, yields $\inf_{\mu \in \mathcal{A}^{uni}} F_\mu(h) \geq \inf_{\mu \in \mathcal{A}^{(2)}} F_\mu(h)$.

The CDF lower envelop over $\mathcal{A}^{uni}$ is closer to the initial distribution. This illustrates how the unimodality constraint shrinks the optimization space size. Moreover, because of this unimodality constraint, the extreme points of $\mathcal{A}_2^\dagger$ are mixtures of uniform distributions $\sum_{k=1}^3 \omega_2^{(k)} \mathcal{U}(30, z_2^{(k)})$, hence continuous measures. The computation of the PoF consequently writes:

$$\sup_{\mu \in \mathcal{A}^{uni}} F_\mu(h) = \sup_{\mu \in \Delta} \sum_{k_1=1}^{3} \cdots \sum_{k_4=1}^{3} \omega_1^{(k_1)} \ldots \omega_4^{(k_4)} \int_{30}^{z_2^{(k_2)}} \mathbb{1}_{\{G(x_1^{(k_1)}, z, x_3^{(k_3)}, x_4^{(k_4)}) \leq h\}} \, \mathrm{d}z \ ,$$

which requires MC methods for its numerical evaluation. Here, we chose a sample size of $n = 10000$. The computation of the PoF is therefore tainted with numerical noise and needs an adapted version of the DE solver presented in Appendix C.2. This increases the computational time as the population member and its trial candidates are evaluated several times whenever the noise variance prevents from distinguishing the best solution. For each threshold the computation time is around 10 minutes, which is a huge increase compared to the previous optimization. Additionally, this noise reduces the performance of the optimizer, therefore we expect the lower bound obtained in this

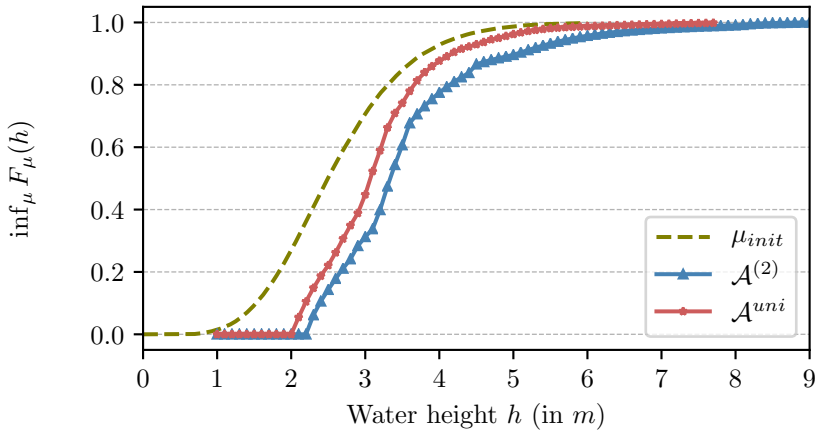


Figure 5.2: The initial CDF of the flood model is compared to the CDF lower envelop over the moment class $\mathcal{A}^{(2)}$ in Eq. (5.7) and the unimodal moment class $\mathcal{A}^{uni}$. This lower envelop is computed as the lowest PoF for a sample of threshold h. As expected $\mathcal{A}^{uni} \subset \mathcal{A}^{(2)}$ implies the minimal unimodal curve to be above the moment curve.

way to be less accurate than in the previous case. It is not possible at this state to know whether the optimum has been reached, although the smoothness of the CDF lower envelop is "reassuring" in the sense that it seems unlikely to happen purely by chance.

The modeling at this state is not completely satisfactory. Indeed, setting a moment constraint in terms of an estimated value seems too restrictive as there is necessarily epistemic uncertainty affecting the estimation of this moment. Therefore, in the following we relax the equality constraints onto an interval, so that the moments can vary in a range of values given in Table 5.2. The variable K_s is still modeled in a unimodal moment class but its mode varies as well. The corresponding optimization space is denoted $\mathcal{A}^{ineq}$. The computation of the PoF is then performed using Algorithm 2 (p.69) and optimized through the noisy DE solver. Notice that once again, the initial distri-

Table 5.2: Corresponding inequality and unimodality moment constraints for the 4 inputs of the flood model.

n°	Variable	Bounds	Mean	Second order moment	Mode
1	Q	$[160, 3580]$	$[1300, 1340]$	$[2.1598\,\text{e}06, 2.1698\,\text{e}06]$	–
2	K_s	$[12.55, 47.45]$	$[29, 31]$	$[900, 1000]$	$\theta_2 \in [29, 31]$
3	Z_v	$[49, 51]$	$[45.5, 50]$	2500.17	–
4	Z_m	$[54, 55]$	$[54.4, 54.6]$	2970.29	–

bution μ_{init} satisfies the constraints and belongs in the space $\mathcal{A}^{ineq}$ which also contains $\mathcal{A}^{uni}$, but not $\mathcal{A}^{(2)}$ due to the unimodality constraints on the Stickler coefficient K_s.

The CDF lower envelop is depicted in Fig. 5.3. As expected from the previous remark, the CDF lower envelop over the space $\mathcal{A}^{ineq}$ is below the one over $\mathcal{A}^{uni}$ which gives lower PoF. And it crosses the CDF lower envelop over $\mathcal{A}^{(2)}$. One can see the impact of relaxing the equality constraints. The bounds on each moment constraint have been set without any expert opinion and act here as an illustration. Increasing the range of variation for each moment would lower the minimal PoF. Nevertheless,

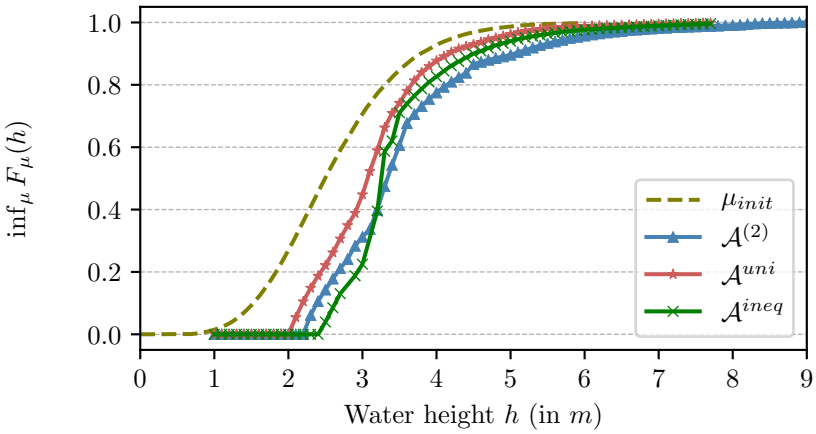


Figure 5.3: We compare the CDF lower envelope over the moment class $\mathcal{A}^{(2)}$, the unimodal moment class $\mathcal{A}^{uni}$ and the unimodal moment class with inequality constraints $\mathcal{A}^{ineq}$. This lower envelop is computed as the lowest PoF for a sample of threshold h. The corresponding quantiles of each curve are depicted in Table 5.3.

the optimum found seems relevant coherent with the inclusion between $\mathcal{A}^{uni}, \mathcal{A}^{(2)}, \mathcal{A}^{ineq}$ and μ_{init}. This good behavior of the algorithm is satisfactory, indeed we recall that the optimization with inequality constraints is more complex as there are more parameters in Algorithm 2 than for equality constraints in Algorithm 1.

This illustration on the flood model allows to visually examine how each optimization space behaves. In particular, the lower CDF envelop is characterized by the size of the moment class. Hence, enforcing two moment constraints in each input appears as a good comprise in terms of information supplied and optimization space size. The unimodality clearly reduces the size of the space and is recommended as soon as this information is relevant. Finally, the inequality constraints are the best way to really account for the uncertainty on the distribution. However, the analyst must first find an appropriate range of variation for each moment as it directly impact the optimal QoI. These bounds on the constraints can derive from confidence intervals in the estimation of the moment from the data. They can also be computed from a robust Bayesian inference, which permits to compute range of variation on moments by integrating the estimation uncertainty together with the prior ignorance.

The PoF being a measure affine function, its computation is quite easy and straightforward. In the next section we investigate more general QoIs.

5.2 Non-Linear Quantities

We briefly extend the affine function presented in Section 5.1 to investigate more general quantities of the form (Berger, 1990)

$$\mu \mapsto F(\mu) = \int q(\mathbf{x}, \varphi(\mu))\, \mu(d\mathbf{x}) \ ,$$

where $\varphi(\mu)$ is measurable. For instance, consider a computer model G, then a very common statistical QoI would be

$$q(\mathbf{x}, \varphi(\mu)) = (G(\mathbf{x}) - \mathbb{E}_\mu[G(\mathbf{x})])^2 \ ,$$

that yields the variance of the output G with input distribution μ. In order to optimize this quantity by applying the reduction theorem, it is needed to linearize the functional. The idea is to substitute the optimization set $\mathcal{A}$ by a subset $\{\mu \in \mathcal{A} \mid \varphi(\mu) = \varphi_0\}$ in order to rewrite the optimization problem as follows

$$\sup_{\mu\in\mathcal{A}} \int q(\mathbf{x}, \varphi(\mu))\, \mu(d\mathbf{x}) = \sup_{\varphi_0} \sup_{\substack{\mu\in\mathcal{A} \\ \varphi(\mu)=\varphi_0}} \int q(\mathbf{x}, \varphi_0)\, \mu(d\mathbf{x}) \ .$$

One can see that whenever $\varphi(\mu)$ is the expectation of some measurable function, the reduction Theorem 3.8 applies to the measure affine function $\mu \mapsto \int q(\mathbf{x}, \varphi_0)\, \mu(d\mathbf{x})$ on the set $\{\mu \in \mathcal{A} \mid \varphi(\mu) = \varphi_0\}$, which is the set $\mathcal{A}$ with additional generalized moment constraints.

Example. Assume as in the previous example that every marginal distribution μ_i belongs to a moment class with N_i constraints, and that there is no additional constraint on the product measure [$N = 0$ in Eq. (3.10)]. Then, the variance of the computer model G writes:

$$\sigma^2(\mu) = \int_{\mathcal{X}} \left(G(\mathbf{x}) - \mathbb{E}_\mu[G(\mathbf{X})]\right)^2 \mathrm{d}\mu(\mathbf{x}) \ .$$

In order to optimize this quantity which is not linear, we temporarily fix the mean of the computer model $\mathbb{E}_\mu[G(\mathbf{X})]$ so that the problem becomes linear. Therefore,

$$\begin{aligned} \sup_{\mu\in\mathcal{A}} \sigma^2(\mu) &= \sup_{\mu\in\mathcal{A}} \int_{\mathcal{X}} \left(G(\mathbf{x}) - \mathbb{E}_\mu[G(\mathbf{X})]\right)^2 \mathrm{d}\mu(\mathbf{x}) \ , \\ &= \sup_{\overline{G}\in\mathbb{R}} \sup_{\substack{\mu\in\mathcal{A} \\ \mathbb{E}_\mu[G(\mathbf{X})]=\overline{G}}} \int_{\mathcal{X}} \left(G(\mathbf{x}) - \overline{G}\right)^2 \mathrm{d}\mu(\mathbf{x}) \ , \end{aligned}$$

and the reduction Theorem 3.8 applies to the linear function

$$\mu \mapsto \int_{\mathcal{X}} \left(G(\mathbf{x}) - \overline{G}\right)^2 \mathrm{d}\mu(\mathbf{x}) \ .$$

But because μ_i satisfies N_i constraints and one more constraint is enforced on the mean of the computer code, the extreme points of the set $\mathcal{A}$ are the product of discrete distribution supported on at most $N_i + 2$ Dirac masses, so that

$$\Delta = \left\{ \mu \in \mathcal{A} \mid \mu_i = \sum_{k=1}^{N_i+2} \omega_i^{(k)} \delta_{x_i^{(k)}} \right\} \ .$$

$$\begin{aligned} \sup_{\mu\in\mathcal{A}} \sigma^2(\mu) &= \sup_{\overline{G}\in\mathbb{R}} \sup_{\substack{\mu\in\Delta \\ \mathbb{E}_\mu[G(\mathbf{X})]=\overline{G}}} \int_{\mathcal{X}} \left(G(\mathbf{x}) - \overline{G}\right)^2 \mathrm{d}\mu(\mathbf{x}) \ , \\ &= \sup_{\mu\in\Delta} \int_{\mathcal{X}} \left(G(\mathbf{x}) - \mathbb{E}_\mu[G(\mathbf{X})]\right)^2 \mathrm{d}\mu(\mathbf{x}) \ , \\ &= \sup_{\mu\in\Delta} \sigma^2(\mu) \ , \\ &= \sup_{\mu\in\Delta} \mathbb{E}_\mu[G(\mathbf{X})^2] - \mathbb{E}_\mu[G(\mathbf{X})]^2 \ , \\ &= \sup_{\mu\in\Delta} \sum_{k_1=1}^{N_1+2} \cdots \sum_{k_d=1}^{N_d+2} \omega_1^{(k_1)} \ldots \omega_d^{(k_d)} \left(G(x_{k_1}^{(1)}, \ldots, x_{k_d}^{(d)}\right)^2 \\ &\quad - \left(\sum_{k_1=1}^{N_1+2} \cdots \sum_{k_d=1}^{N_d+2} \omega_1^{(k_1)} \ldots \omega_d^{(k_d)} G(x_{k_1}^{(1)}, \ldots, x_{k_d}^{(d)}) \right)^2 \ . \end{aligned}$$

5.3 Quantile Function

Another classical QoI, widely used in industrial applications as in safety analysis of nuclear reactors (Wallis, 2007), is the quantile of the output of a computer model G. It is a critical criteria for evaluating safety margins (Iooss and Marrel, 2019). We recall that F_μ denotes the cumulative distribution function of the output of the code, i.e. $F_\mu(h) = \mathbb{P}_\mu(G(\mathbf{X}) \leq h)$. Because the quantile function is not measure affine, this QoI must satisfy the two assumptions of the reduction Theorem 3.8 to be optimized. In the next section we investigate the lower semicontinuity and the quasi-convexity of this function of the input distribution.

5.3.1 Lower Quantile Function

The usual definition of the quantile function corresponds to the one given in the following theorem that proves that the quantile function is a QoI perfectly adapted to the OUQ methodology.

Theorem 5.2. *We suppose that the code G is continuous. Let μ be an element of a convex measure space $\mathcal{A}$, then the lower quantile function*

$$\mu \mapsto Q_p^L(\mu) = \inf\{h : F_\mu(h) \geq p\} , \tag{5.9}$$
$$= \inf\{h : \mathbb{P}_\mu(G(\mathbf{X}) \leq h) \geq p\} , \tag{5.10}$$

is quasi-convex and lower semicontinuous on $\mathcal{A}$ as illustrated in Fig. 5.4.

Proof. A function is quasi-convex if any lower level set is a convex set. Further, it is lower semicontinuous if any lower level set is closed. Hence, we consider for $\alpha \in \mathbb{R}$ the lower level set for $\alpha \in \mathbb{R}$:

$$L_\alpha = \left\{\mu \in \mathcal{A} \mid Q_p^L(\mu) \leq \alpha\right\} ,$$
$$= \{\mu \in \mathcal{A} \mid F_\mu(\alpha) \geq p\} .$$

Indeed, the quantile is the unique function satisfying the Galois inequalities. Therefore,

$$L_\alpha = \left\{\mu \in \mathcal{A} \mid \mu\left(G^{-1}(]-\infty, \alpha])\right) \geq p\right\} .$$

L_α is obviously convex and applying Corollary 15.6 in Aliprantis and Border (2007), L_α is also closed (for the weak topology), as $G^{-1}(]-\infty, \alpha])$ is closed . □

Remark 3. Notice that in this work the quantile is a function of the measure μ. However, the quantile seen as a function of random variable X_μ (with distribution μ) is not quasi-convex, this subtle point is explained in Drapeau and Kupper (2012). We recall that convex combination of random variables involves the convolution of their probability distribution.

Remark 4. The continuity assumption of the computer model G is always verified whenever G is the posterior expectation of a centered Gaussian process with a common covariance kernel. This is convenient as we often replace expensive black box model with Gaussian process regressor (Iooss and Marrel, 2019; De Lozzo, 2015).

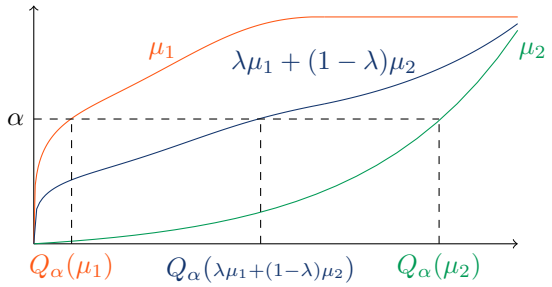


Figure 5.4: Consider two distributions μ_1 and μ_2, their α-quantile $Q_\alpha(\mu_1)$ and $Q_\alpha(\mu_2)$, as well as their mixture distribution $\lambda\mu_1 + (1-\lambda)\mu_2$ for $\lambda \in [0,1]$. Then, the α-quantile of the mixture distribution $Q_\alpha(\lambda\mu_1 + (1-\lambda)\mu_2)$ is necessarily lower than $\max\{Q_\alpha(\mu_1); Q_\alpha(\mu_2)\}$ as illustrated to visualize the quasi-convexity of the quantile function.

Example. Assume that every input μ_i is a moment class with N_i constraints and that there is no additional constraint on the product measure ($N = 0$ in Theorem 3.8). Then as in the previous example, the reduction Theorem applies and the extreme points of the measure space are product of discrete measures supported on at most $N_i + 1$ points. Therefore, the quantile computes as follows

$$\begin{aligned}
\sup_{\mu\in\mathcal{A}} Q_p^L(\mu) &= \sup_{\mu\in\Delta} Q_p^L(\mu) \ , \\
&= \sup_{\mu\in\Delta} \left[\inf_{h\in\mathbb{R}} \{h : F_\mu(h) \geq p\} \right] \ , \qquad (5.11) \\
&= \sup_{\mu\in\Delta} \left[\inf_{h\in\mathbb{R}} \left\{ h : \sum_{k_1=1}^{N_1+1} \cdots \sum_{k_d=1}^{N_d+1} \omega_1^{(k_1)} \ldots \omega_d^{(k_d)} \mathbb{1}_{\{G(x_{k_1}^{(1)},\ldots,x_{k_d}^{(d)})\leq h\}} \geq p \right\} \right] \ . \qquad (5.12)
\end{aligned}$$

We emphasize that in practice we seldom compute directly the quantile of a computer model. Indeed, the computation of the quantile derives from the one of the CDF and it is usual to compute the probability of failure for different level h then to use some binary search algorithm to get the quantile. The next proposition highlights how the maximal quantile over a measure set is connected to the lower envelope CDF of this set. It is interesting as it allows to exploit the result of Section 5.1.2.

Proposition 5.3. *The following duality result holds*

$$\sup_{\mu\in\mathcal{A}} Q_p^L(\mu) = \inf \left\{ h \in \mathbb{R} \mid \inf_{\mu\in\mathcal{A}} F_\mu(h) \geq p \right\} \ .$$

Proof. We illustrate this result in Figure 5.5. we denote by

$$a = \sup_{\mu\in\mathcal{A}} \left[\inf \{h \in \mathbb{R} \ ; \ F_\mu(h) \geq p\} \right] \ ,$$

and

$$b = \inf \left\{ h \in \mathbb{R} \mid \inf_{\mu\in\mathcal{A}} F_\mu(h) \geq p \right\} \ .$$

In order to prove $a = b$, we proceed in two steps. First, we have

$$\begin{aligned}
& \text{for all } h \geq b \ ; \ \inf_{\mu\in\mathcal{A}} F_\mu(h) \geq p \ , \\
\Leftrightarrow \quad & \text{for all } h \geq b \ \text{ and for all } \mu \in \mathcal{A} \ ; \ F_\mu(h) \geq p \ , \\
\Leftrightarrow \quad & \text{for all } \mu \in \mathcal{A} \ \text{ and for all } h \geq b \ ; \ F_\mu(h) \geq p \ , \\
\Rightarrow \quad & \text{for all } \mu \in \mathcal{A} \ ; \ \inf\{h \in \mathbb{R} \mid F_\mu(h) \geq p\} \leq b \ ,
\end{aligned}$$

so that $b \geq a$. Then, because a is the sup of the quantiles,

$$\begin{aligned} & \text{for all } h \geq a \ ; \ \text{ for all } \mu \in \mathcal{A} \ ; \ F_\mu(h) \geq p \ , \\ \Rightarrow \quad & \text{for all } h \geq a \ ; \ \inf_{\mu \in \mathcal{A}} F_\mu(h) \geq p \ , \end{aligned}$$

so that

$$\inf \left[h \in \mathbb{R} \mid \inf_{\mu \in \mathcal{A}} F_\mu(h) \geq p \right] \leq a \ ,$$

and $b \leq a$. □

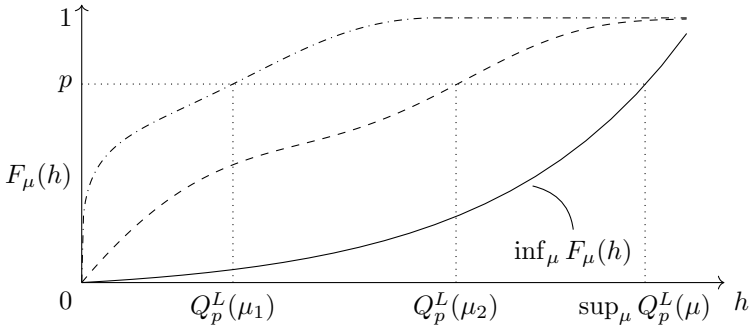


Figure 5.5: Illustration of the Proposition 5.3. Different distributions μ and the associated quantiles $Q_p^L(\mu)$ are plotted in dash-dotted and dashed line. The lower curve represents the CDF lower envelope; one can observe that the maximum quantile $\sup_{\mu \in \mathcal{A}} Q_p^L(\mu)$ corresponds to the p-quantile of the CDF lower envelope: $h \mapsto \inf_{\mu \in \mathcal{A}} F_\mu(h)$.

Proposition 5.3 shows that the maximum quantile over a measure set $\mathcal{A}$ is equal to the quantile associated to CDF lower envelope. In practice we will therefore study the lowest probability of failure $\inf_{\mu \in \mathcal{A}} F_\mu(h)$ for any given threshold h. In the next paragraph we consider the minimization of the quantile that cannot be pursued at the current stage. Indeed, the quantile function introduced in this section is quasi-convex but not quasi-concave. Hence, one cannot apply the reduction theorem in order to minimize this function. However, by considering an alternative version of the quantile its property can suit the requirement for reducing the minimization to the extreme points of a measure space.

5.3.2 Upper Quantile Function

The lower quantile function introduced in Section 5.3.1 can be maximized thanks to Theorem 3.8. Therefore, it is possible to obtain the worst possible quantile over a convex measure space $\mathcal{A}$. In this section, we investigate the minimal value of the quantile of the computer model G. In that way, we have at our disposal bounds around the quantile, which quantifies the range of variation of this QoI over the measure space. However, in order to minimize the quantile, both the upper semicontinuity and quasi-concavity of the optimization function are required. A modified QoI called the upper quantile function is proposed hereunder that is very close to the usual lower quantile function of Theorem 5.2. It is defined by the following theorem.

Theorem 5.4. *Assume that code G is continuous. Then, the upper quantile function $\mu \mapsto Q_p^R(\mu) = \inf\{h : F_\mu(h) > p\}$ is quasi-concave and upper semicontinuous on $\mathcal{A}$.*

Proof. A function is quasi-concave if all its upper level sets are convex. It is upper semicontinuous if all its upper level sets are closed. For $\alpha \in \mathbb{R}$, the upper level set is

$$\begin{aligned} U_\alpha &= \left\{\mu \in \mathcal{A} \mid Q_p^R(\mu) \geq \alpha\right\} \ , \\ &= \{\mu \in \mathcal{A} \mid \forall \varepsilon > 0 : F_\mu(\alpha - \varepsilon) \leq p\} \ , \\ &= \bigcap_{\varepsilon > 0} \{\mu \in \mathcal{A} \mid F_\mu(\alpha - \varepsilon) \leq p\} \ , \\ &= \bigcap_{\varepsilon > 0} \left\{\mu \in \mathcal{A} \mid \mu(G^{-1}(]-\infty, \alpha - \varepsilon])) \leq p\right\} \ , \\ &= \bigcap_{\varepsilon > 0} \left\{\mu \in \mathcal{A} \mid \mu(G^{-1}(]-\infty, \alpha - \varepsilon[)) \leq p\right\} \ . \end{aligned}$$

The last equality deserves some explanation. We prove that the equality holds in two times. For $\varepsilon > 0$, we denote

$$F_c(\varepsilon) = \left\{\mu \in \mathcal{A} \mid \mu(G^{-1}(]-\infty, \alpha - \varepsilon])) \leq p\right\} \ ,$$

$$F_o(\varepsilon) = \left\{\mu \in \mathcal{A} \mid \mu(G^{-1}(]-\infty, \alpha - \varepsilon[)) \leq p\right\} \ .$$

Clearly, $F_c(\varepsilon) \subset F_o(\varepsilon)$ for all $\varepsilon > 0$, so that

$$\cap_{\varepsilon>0} F_c(\varepsilon) \subset \cap_{\varepsilon>0} F_o(\varepsilon) \ .$$

For the reverse inclusion, let μ be an element of $\cap_{\varepsilon>0} F_o(\varepsilon)$. Suppose that μ is not in $\cap_{\varepsilon>0} F_c(\varepsilon)$. Then, there exists a $\varepsilon_0 > 0$ such that $\mu(G^{-1}(]-\infty, \alpha - \varepsilon_0])) > p$. But $\mu(G^{-1}(]-\infty, \alpha - \varepsilon_0])) \leq \mu(G^{-1}(]-\infty, \alpha - \frac{\varepsilon_0}{2}[)) \leq p$, because μ is in $\cap_{\varepsilon>0} F_o(\varepsilon)$ by construction, leading to a contradiction.

To conclude, Aliprantis and Border (2007, Corollary 15.6) proves that $F_o(\varepsilon)$ is closed because $G^{-1}(]-\infty, \alpha - \varepsilon[)$ is open as G is continuous. Hence, U_α is closed as an intersection of closed sets. The convexity of U_α is clear. □

5.3.3 Illustration

In this section, we investigate the quantile of the flood model H. This quantity serves as reference for establishing safety margins concerning the protection dyke's height. For sufficient relability, it is usual to consider the 95% quantile of the model. However, the input distribution is not uniquely defined and is known only by some of its moments. Therefore, we compute the worst-case scenario corresponding to the maximal safety margin that is robust with respect to the uncertainty affecting the input. Applying duality Theorem 5.3, the maximal quantile over $\mathcal{A}$ is to be found on the CDF lower envelop over this space that has already been computed. Therefore, we refer to Section 5.1.3 for a review on this optimization problem. The different quantile values correspond to different moment classes presented in Section 5.1.3. If one wishes to compute a single quantile order, it is not necessary to explore all thresholds on a fine grid. A binary search algorithm can be used on to compute PoF and locate the threshold corresponding to the quantile.

Concerning the quantile values in Table 5.3, it appears that one constraint on the mean gives non-reasonable bounds. In fact, the CDF associated to $\mathcal{A}^{(1)}$ in Fig. 5.1 is discontinuous, which explains why the 95% and 99% quantile are so close. Discontinuity in the lower envelop CDF often occurs when only one constraint is enforced, as illustrated in the following simple example.

Table 5.3: Safety margins for the dyke's height taken as the maximal 95% and 99% quantile of the flood model over different moment class presented in Section 5.1.3.

Initial	$Q_{0.95}(\mu_{init})$	$Q_{0.99}(\mu_{init})$
	4.23	5.11
Moment class	$\sup_\mu Q_{0.95}(\mu)$	$\sup_\mu Q_{0.99}(\mu)$
$\mathcal{A}^{(1)}$	8.81	8.83
$\mathcal{A}^{(2)}$	5.82	7.88
$\mathcal{A}^{(3)}$	4.98	6.64
$\mathcal{A}^{uni}$	4.79	6.32
$\mathcal{A}^{ineq}$	5.18	7.04

Example. Consider two random variables X_1 and X_2 on $[-1, 1]$ with mean 0, and the model $G(x_1, x_2) = x_1 x_2^2$. Then, the code G cannot exceed 1 and whatever the input distribution it holds that $\mathbb{P}(G(\mathbf{X}) \leq 1) = 1$. But the maximal probability $\mathbb{P}(G(\mathbf{X}) = 1)$ over all distribution in $[-1, 1]$ with mean 0 is $\mathbb{P}(G(\mathbf{X}) = 1) = 0.5$. Indeed, it is obtained for X_1 and X_2 equal to the distribution $(\delta_{-1} + \delta_1)/2$. Consequently, the CDF lower envelop has a discontinuity at threshold 1.

The measure space $\mathcal{A}^{(2)}$ gives already much more interesting results. Two constraints are in general sufficient to account for the uncertainty in the input distribution, and are easy to interpret as they correspond to the mean and variance of any input parameter. In some situation a third moment can be helpful; engineers should keep in mind that the moment space size is quite reduced with three constraints, as seen with the quantile of $\mathcal{A}^{(3)}$. Hence, it reduces significantly the uncertainty on the input distribution. The unimodal spaces $\mathcal{A}^{uni}$ defined in Eq. (5.8) and $\mathcal{A}^{ineq}$ for the inequality constraints given in Table 5.2 both enforce two moment constraints on every input.

5.4 Superquantile

In this section we investigate a QoI called superquantile, also known as conditional value at risk or expected shortfall. Historically, this quantity has been widely used in mathematical finance as it satisfies all properties a convex risk measure is required to possess contrarily to the classical quantile (Rockafellar and Royset, 2014). The superquantile is closely related to the quantile, as shown in its definition

Definition 4. The superquantile is defined as

$$\overline{Q}_p(\mu) = \frac{1}{1-p} \int_p^1 Q^L_{p'}(\mu)\, \mathrm{d}p' \; = \mathbb{E}\left[G(X) \mid G(X) \geq Q^L_{p'}(\mu)\right].$$

where $Q^L_{p'}(\mu)$ [Eq. (5.9)] designates the p'-quantile of the underlying computer model G.

The superquantile of order p of a computer model G reads as the average of the quantile function above the level p. Rockafellar and Uryasev (2002) brings the superquantile and the quantile together in a way that supports practical methods of

computation:

$$\begin{aligned} \overline{Q}_p(\mu) &= \min_{\alpha\in\mathbb{R}} \left\{ \alpha + \frac{1}{1-p} \int \max\{0; G(\mathbf{x}) - \alpha\} \, \mathrm{d}\mu(\mathbf{x}) \right\} , \\ Q_p^L(\mu) &= \arg\min_{\alpha\in\mathbb{R}} \left\{ \alpha + \frac{1}{1-p} \int \max\{0; G(\mathbf{x}) - \alpha\} \, \mathrm{d}\mu(\mathbf{x}) \right\} . \end{aligned} \tag{5.13}$$

The superquantile gives some valuable additional information with respect to the quantile. For instance, the quantile is often used in an engineering context of risk assessment to determine safety margins, as shown in the illustration of Section 5.3.3. However, even for high thresholds of order 0.95 or 0.99, the knowledge of the quantile gives no information about the remaining 5% or 1% upper tail distribution. In some badly designed cases, the engineer should be aware of extreme scenario that could happened in the extreme end of the distribution. Rockafellar and Royset (2014) proves that the superquantile is equivalent to the expectation in the upper p-tail distribution (which is the historical definition). Therefore, the superquantile is valuable to know how the distribution spreads above the p-level quantile. In this context, we study how to optimize the superquantile of a computer model G output over a measure space. The following theorem shows that the reduction Theorem applies to the superquantile.

Theorem 5.5. *Let $\mathcal{A}$ be a convex measure space with generator Δ, then*

$$\sup_{\mu\in\mathcal{A}} \overline{Q}_p(\mu) = \sup_{\mu\in\Delta} \overline{Q}_p(\mu) , \tag{5.14}$$

$$\inf_{\mu\in\mathcal{A}} \overline{Q}_p(\mu) = \inf_{\mu\in\Delta} \overline{Q}_p(\mu) . \tag{5.15}$$

Proof. We use the equivalent definition of the superquantile given in Eq. (5.13).

$$\sup_{\mu\in\mathcal{A}} \overline{Q}_p(\mu) = \sup_{\mu\in\mathcal{A}} \left[\min_{\alpha\in\mathbb{R}} \left\{ \alpha + \frac{1}{1-p} \int \max\{0; G(\mathbf{x}) - \alpha\} \, \mathrm{d}\mu(\mathbf{x}) \right\} \right] . \tag{5.16}$$

The proof for the minimum is simpler, that is why we present here only the proof for the supremum of the superquantile. Indeed, permuting the sup and the min in this minmax optimization problem Eq. (5.16) is not always possible. However, Lemma 5.6 gives a sufficient condition to do it

Lemma 5.6 (Minmax permutation (Fan, 1953))**.** *Let $\mathcal{X}$ be a compact Hausdorff space and $\mathcal{Y}$ an arbitrary set. Let f be a real-valued function on $\mathcal{X} \times \mathcal{Y}$ such that, for every $y \in \mathcal{Y}$, $f(x, y)$ is lower semicontinuous on X. If f is convex on X and concave on Y, then*

$$\min_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} f(x, y) = \sup_{y\in\mathcal{Y}} \min_{x\in\mathcal{X}} f(x, y) .$$

Now define H_p the function

$$H_p(\alpha, \mu) := \alpha + \frac{1}{1-p} \int_{\mathcal{X}} \max\{0; G(\mathbf{x}) - \alpha\} \, \mathrm{d}\mu(\mathbf{x}) .$$

H is an concave (affine) function of the measure μ. Moreover (Rockafellar and Uryasev, 2002, Theorem 10) guaranties it is convex in α. Hence, Lemma 5.6 applies and

$$\begin{aligned} \sup_{\mu\in\mathcal{A}} \overline{Q}_p(\mu) &= \min_{\alpha\in\mathbb{R}} \left[\sup_{\mu\in\mathcal{A}} \left\{ \alpha + \frac{1}{1-p} \int \max\{0; G(\mathbf{x}) - \alpha\} \, \mathrm{d}\mu(\mathbf{x}) \right\} \right] , \\ &= \min_{\alpha\in\mathbb{R}} \left\{ \sup_{\mu\in\mathcal{A}} H_p(\alpha, \mu) \right\} . \end{aligned} \tag{5.17}$$

The reduction Theorem 3.9 applies on the measure affine function $\mu \mapsto H_p(\alpha, \mu)$ for all $p, \alpha \in \mathbb{R}$, so that

$$\begin{aligned}\sup_{\mu\in\mathcal{A}} \overline{Q}_p(\mu) &= \min_{\alpha\in\mathbb{R}} \left\{ \sup_{\mu\in\Delta} H_p(\alpha,\mu) \right\} \ , \\ &= \sup_{\mu\in\Delta} \left\{ \min_{\alpha\in\mathbb{R}} H_p(\alpha,\mu) \right\} \ , \\ &= \sup_{\mu\in\Delta} \overline{Q}_p(\mu) \ .\end{aligned}$$

□

From a numerical point of view, the computation of the superquantile is also really close to that of the quantile. We recall that in Section 5.3.3, we computed the quantile using the duality Theorem 5.3 that asserts it is equivalent to computing the CDF lower envelop, hence a PoF. The quantile was then recovered using a binary search algorithm on a grid of threshold values $\alpha \in \mathbb{R}$ and the corresponding PoFs: $\int \mathbb{1}_{\{G(\mathbf{x})\geq\alpha)\}} \, \mathrm{d}\mu(\mathbf{x})$. Similarly in order to compute the optimal superquantile, Eq. (5.17) shows that it is enough to examine the quantity

$$\sup_{\mu\in\mathcal{A}} \int \max\{0; G(\mathbf{x}) - \alpha\} \, \mathrm{d}\mu(\mathbf{x}) = \sup_{\mu\in\mathcal{A}} \int (G(\mathbf{x}) - \alpha) \mathbb{1}_{\{G(\mathbf{x})\geq\alpha)\}} \, \mathrm{d}\mu(\mathbf{x}) \ ,$$

for a fixed superquantile level p and a sample of thresholds $\alpha \in \mathbb{R}$. The optimal superquantile can then be recovered from the formula in Eq. (5.17) on the sample of couple

$$\left(\alpha; \ \sup_{\mu\in\mathcal{A}} \left\{ \alpha + \frac{1}{1-p} \int \max\{0; G(\mathbf{x}) - \alpha\} \, \mathrm{d}\mu(\mathbf{x}) \right\} \right) \ .$$

The only difficulty in this method is to identify appropriate range of variation for the threshold α. However, it is usual to know from expert knowledge and past experiences the behavior of the computer model that can be exploited to target reasonable values for α.

Example. We go back to the example of the previous sections, and we consider $\mathcal{A}^*$ as defined in Eq. (5.1), that is a product of moment classes with extreme points in Δ^*, which is a product of discrete measures $\mu = \prod_{i=1}^d \mu_i$ where $\mu_i = \sum_{k=1}^{N_i+1} \omega_i^{(k)} \delta_{x_i^{(k)}}$. Then, the optimal superquantile $\sup_{\mu\in\mathcal{A}} \overline{Q}_p(\mu)$ reads as follows:

$$\min_{\alpha\in\mathbb{R}} \left\{ \alpha + \frac{1}{1-p} \sup_{\mu\in\Delta} \sum_{k_1=1}^{N_1+1} \cdots \sum_{k_d=1}^{N_d+1} \omega_1^{(k_1)} \ldots \omega_d^{(k_d)} \left(G(x_1^{(k_1)}, \ldots, x_d^{(k_d)}) - \alpha \right) \mathbb{1}_{\{G(x_1^{(k_1)},\ldots,x_d^{(k_d)})\geq\alpha\}} \right\},$$

which is closely related to the expression of the robust quantile in Eq. (5.11) over the same moment class.

5.4.1 Illustration

We consider the flood model H presented in Section 2.1, whose CDF and quantiles have already been studied in an OUQ framework in Section 5.1.3 and 5.3.3. The QoI of interest is the function

$$\mu \mapsto \int (H(\mathbf{x}) - \alpha) \mathbb{1}_{\{H(\mathbf{x})\geq\alpha)\}} \, \mathrm{d}\mu(\mathbf{x}) \ ,$$

that is computed for a sample of $\alpha \in \mathbb{R}$. This QoI is of course optimized using the DE solver in Algorithm 3 and canonical moment parameterization of the QoI in Algorithm 1 and over the space $\mathcal{A}^{(2)}$ introduced in Eq. (5.7). The results are depicted in Fig. 5.6. One can observed in Fig 5.6 that the property $\overline{Q}_p(\mu) \geq Q_p^L(\mu)$ implies

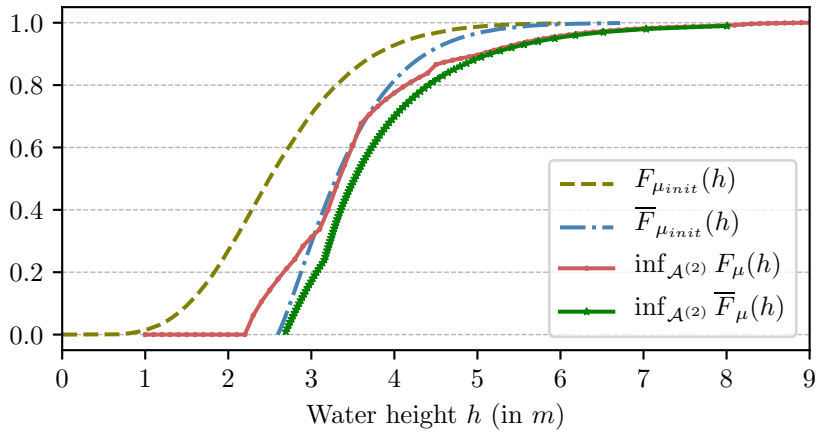


Figure 5.6: This figure depicts the distribution $F_{\mu_{init}}$ and the superdistribution $\overline{F}_{\mu_{init}}$ of the initial input distribution. The superdistribution (Rockafellar and Royset, 2014) is the inverse of the superquantile function. It is computed by integrating the upper p-tail distribution. We also compute the CDF lower envelop and the superdistribution lower envelop over the moment class $\mathcal{A}^{(2)}$.

$\sup_{\mu\in\mathcal{A}} \overline{Q}_p(\mu) \geq \sup_{\mu\in\mathcal{A}} Q_p^L(\mu)$. Moreover, at level 95% and 99% the quantile level and the superquantile are quite close. We recall that these levels serve as reference when adopting safety margins corresponding to the dyke's height for protecting a structure during a flood event. Hence, it means that at level 95% the upper tail of the CDF lower envelop does not affect much in average the value of the robust quantile. This tends to guarantee the security provided by choosing the dyke's height as a 95%-quantile.

Table 5.4: Comparison between the quantile and the superquantile of the flood model with its initial distribution, and in an OUQ analysis over the moment class $\mathcal{A}^{(2)}$.

	$Q_{0.95}(\mu_{init})$	$\overline{Q}_{0.95}(\mu_{init})$
Initial distribution	4.23	4.78
	$\sup_\mu Q_{0.95}(\mu)$	$\sup_\mu \overline{Q}_{0.95}(\mu)$
Moment class $\mathcal{A}^{(2)}$	5.82	5.93

5.5 Sensitivity Index

Global sensitivity analysis aims at determining which uncertain parameters of a computer code mainly drive the output's uncertainty. In that matter, Sobol' indices are widely used as they quantify the contribution of each input to the variance of the

model output (Iooss and Lemaître, 2015; Sobol', 1993). However, because the probability distributions modeling the uncertain parameters are themselves uncertain, we propose to evaluate bounds on Sobol' indices over a class of probability measures. It is not very usual to consider the Sobol index as a function of the input measure, but from this optimization problem rises some nice interpretations. Some work on robust analysis of Sobol' indices might be found in Hart and Gremaud (2019) where the authors study how sensitive the Sobol' indices are to a change in the input distribution, their approach differs from ours and is based on the Frechet derivative on probability distributions. Robustness analysis for global sensitivity based on HSIC has also been studied in Meynaoui et al. (2019).

5.5.1 First Order Sensitivity index

We focus in this paragraph on the well-known first order sensitivity index:

$$S_i(\mu) = \frac{\mathrm{Var}_{\mu_i}(\mathbb{E}_{-i}[Y|X_i])}{\mathrm{Var}(Y)} \ ,$$

where $Y = G(X_1, \dots, X_d)$ denotes the output of the computer model. The first Sobol index measures the effect on the variance of varying X_i alone, but averaged over variations in other input parameters. We are looking for bounds of this index over a class of probability measures that contains all the uncertainty on the parameter distribution. Therefore, whatever the distribution of the ith parameter inside the class, we can guaranty that the parameter's contribution to the variance of the output remains within a certain range. If this range is large, this means that we should temper the interpretation of the Sobol indices as a change of the input distribution can deeply modify the importance of this parameter on the output variation. On the other hand, if the range is small this means that whatever the choice of the input distribution inside the class, the contribution of the ith input to the output variance remains of the same importance (that could be small or high).

This kind of information can help an analyst to better understand the computer model behavior. One should investigate further details regarding the optimization of this function. We emphasize that in the next Theorem, only the distribution of the ith input varies in a measure space. All other input distributions, namely μ_{-i} are set typically to their initial choice. This is justified by the properties of the first order sensitivity index that are required to reduce the optimization to the generator of the measure space.

Theorem 5.7. *Let $\mathcal{X}_i, \mathcal{A}_i$ and Δ_i be defined as in 3.3.1. Then*

$$S_i^+ = \sup_{\mu_i \in \mathcal{A}_i} S_i(\mu) = \sup_{\mu_i \in \Delta_i(N_i+1)} S_i(\mu) \ ,$$

$$S_i^- = \inf_{\mu_i \in \mathcal{A}_i} S_i(\mu) = \inf_{\mu_i \in \Delta_i(N_i+1)} S_i(\mu) \ .$$

Proof. The proof is made for the supremum but is similar for the infimum

$$\sup_{\mu_i \in \mathcal{A}_i} S_i(\mu) = \sup_{\mu_i \in \mathcal{A}_i} \frac{\mathbb{E}_{\mu_i}\left[\left(\mathbb{E}_{-i}[Y|X_i]\right)^2\right] - \left(\mathbb{E}_{\mu_i}\left[\mathbb{E}_{-i}[Y|X_i]\right]\right)^2}{\mathbb{E}_{\mu_i}\left[\mathbb{E}_{-i}[Y^2]\right] - \left(\mathbb{E}[Y]\right)^2} ,$$

$$= \sup_{\mu_i \in \mathcal{A}_i} \frac{\mathbb{E}_{\mu_i}\left[\left(\mathbb{E}_{-i}[Y|X_i]\right)^2\right] - \left(\mathbb{E}[Y]\right)^2}{\mathbb{E}_{\mu_i}\left[\mathbb{E}_{-i}[Y^2]\right] - \left(\mathbb{E}[Y]\right)^2} ,$$

$$= \sup_{\overline{y}_0} \sup_{\substack{\mu_i \in \mathcal{A}_i \\ \mathbb{E}[Y]=\overline{y}_0}} \frac{\mathbb{E}_{\mu_i}\left[\left(\mathbb{E}_{-i}[Y|X_i]\right)^2\right] - \overline{y}_0^2}{\mathbb{E}_{\mu_i}\left[\mathbb{E}_{-i}[Y^2]\right] - \overline{y}_0^2} ,$$

where $\overline{y}_0$ is a real number. This is the same linearization trick employed in Section 5.2. One can also observe due to the squared exponent $\mathbb{E}_{\mu_i}\left[\left(\mathbb{E}_{-i}[Y|X_i]\right)^2\right]$ that the first order sensitivity index is not measure affine with respect to the other inputs μ_{-i}. For that reason, only the ith input is modeled with a second level uncertainty. Now, the function

$$\mu_i \longmapsto \frac{\mathbb{E}_{\mu_i}\left[\left(\mathbb{E}_{-i}[Y|X_i]\right)^2\right] - \overline{y}_0^2}{\mathbb{E}_{\mu_i}\left[\mathbb{E}_{-i}[Y^2]\right] - \overline{y}_0^2} ,$$

is a ratio of two measure affine functions. Proposition 3.11 states that the reduction Theorem 3.8 applies so that

$$\sup_{\substack{\mu_i \in \mathcal{A}_i \\ \mathbb{E}[Y]=\overline{y}_0}} \frac{\mathbb{E}_{\mu_i}\left[\left(\mathbb{E}_{-i}[Y|X_i]\right)^2\right] - \overline{y}_0^2}{\mathbb{E}_{\mu_i}\left[\mathbb{E}_{-i}[Y^2]\right] - \overline{y}_0^2} = \sup_{\substack{\mu_i \in \Delta_i(N_i+1) \\ \mathbb{E}[Y]=\overline{y}_0}} \frac{\mathbb{E}_{\mu_i}\left[\left(\mathbb{E}_{-i}[Y|X_i]\right)^2\right] - \overline{y}_0^2}{\mathbb{E}_{\mu_i}\left[\mathbb{E}_{-i}[Y^2]\right] - \overline{y}_0^2} ,$$

and the result follows. □

Equivalent results can be obtained for the total order index and are developed in the next section, before illustrating the methodology on the flood model.

5.5.2 Total Order Sensitivity Index

The total effect (or total order) index S_{Ti} (Iooss and Lemaître, 2015; Homma and Saltelli, 1996) measures the contribution to the output variance of the i*th* input taking into account all the variance caused by its interaction with any other input. It is defined as follows:

$$S_{Ti}(\mu) = \frac{\mathbb{E}_{-i}\left[\mathrm{Var}_{\mu_i}\left(Y|X_{-i}\right)\right]}{\mathrm{Var}(Y)} = 1 - \frac{\mathrm{Var}_{-i}\left(\mathbb{E}_{\mu_i}\left[Y|X_{-i}\right]\right)}{\mathrm{Var}(Y)} .$$

The total effect indices never sum to one, except when the model is purely additive. The ith total order index also accounts for interactions with the other input X_{-i} and thus depends on their distribution. Obviously, the uncertainty tainting the modeling of μ_{-i} affects the total effect index. Consequently, bounds on the ith total order index S_{Ti} are computed with respect to the variability of all the input distributions except the ith. In other words, the bounds obtained are interpreted as the minimal and maximal total order index the ith input can have, considering the lack of knowledge in all but the ith input distribution. Hence, the lower bound on the ith total effect index corresponds to the minimal variance caused by its interaction between the ith input

and the remaining parameters, and the upper bound on S_{Ti} gives the maximal variance caused by its interaction between the ith input and the other parameters. Note that S_{Ti} also accounts for X_i's influence alone on the output's variance, which is why it is always greater than S_i. Let us define $\mathcal{X}_{-i}, \mathcal{A}_{-i}$, and Δ_{-i} as

$$\begin{aligned} \mathcal{X}_{-i} &= \mathcal{X}_1 \times \dots \mathcal{X}_{i-1} \times \mathcal{X}_{i+1} \times \dots \mathcal{X}_d \ , \\ \mathcal{A}_{-i} &= \mathcal{A}_1 \times \dots \mathcal{A}_{i-1} \times \mathcal{A}_{i+1} \times \dots \mathcal{A}_d \ , \\ \Delta_{-i} &= \Delta_1 \times \dots \Delta_{i-1} \times \Delta_{i+1} \times \dots \Delta_d \ , \end{aligned} \tag{5.18}$$

where $\mathcal{X}_k, \mathcal{A}_k$, and Δ_k are the measure space in Section 3.3.1. Then, we have the following reduction Theorem

Theorem 5.8. *Given $\mathcal{X}_{-i}, \mathcal{A}_{-i}$ and Δ_{-i} in Eq. (5.18). Then:*

$$S_{Ti}^{+} = \sup_{\mu_{-i} \in \mathcal{A}_{-i}} S_{Ti}(\mu) = \sup_{\mu_{-i} \in \Delta_{-i}(N_i+1)} S_{Ti}(\mu) \ ,$$

$$S_{Ti}^{-} = \inf_{\mu_{-i} \in \mathcal{A}_{-i}} S_{Ti}(\mu) = \inf_{\mu_{-i} \in \Delta_{-i}(N_i+1)} S_{Ti}(\mu) \ .$$

Proof. The proof is made for the supremum but is similar for the infimum

$$\begin{aligned} \sup_{\mu_{-i} \in \mathcal{A}_{-i}} S_{Ti}(\mu) &= \sup_{\mu_{-i} \in \mathcal{A}_{-i}} 1 - \frac{\mathbb{E}_{-i}\left[\left(\mathbb{E}_{\mu_i}[Y|X_{-i}]\right)^2\right] - \left(\mathbb{E}_{-i}\left[\mathbb{E}_{\mu_i}[Y|X_{-i}]\right]\right)^2}{\mathbb{E}_{-i}\left[\mathbb{E}_{\mu_i}[Y^2]\right] - \left(\mathbb{E}[Y]\right)^2} \ , \\ &= \sup_{\mu_{-i} \in \mathcal{A}_{-i}} 1 - \frac{\mathbb{E}_{-i}\left[\left(\mathbb{E}_{\mu_i}[Y|X_{-i}]\right)^2\right] - \left(\mathbb{E}[Y]\right)^2}{\mathbb{E}_{-i}\left[\mathbb{E}_{\mu_i}[Y^2]\right] - \left(\mathbb{E}[Y]\right)^2} \ , \\ &= \sup_{\overline{y}_0} \sup_{\substack{\mu_{-i} \in \mathcal{A}_{-i} \\ \mathbb{E}[Y]=\overline{y}_0}} \frac{\mathbb{E}_{\mu_{-i}}\left[\left(\mathbb{E}_{\mu_i}[Y|X_{-i}]\right)^2\right] - \overline{y}_0^2}{\mathbb{E}_{\mu_{-i}}\left[\mathbb{E}_{\mu_i}[Y^2]\right] - \overline{y}_0^2} \ , \end{aligned}$$

where $\overline{y}_0$ is a real number. This is the same linearization trick employed in the Section 5.2. Now, the function

$$\mu_{-i} \longmapsto \frac{\mathbb{E}_{\mu_{-i}}\left[\left(\mathbb{E}_{\mu_i}[Y|X_{-i}]\right)^2\right] - \overline{y}_0^2}{\mathbb{E}_{\mu_{-i}}\left[\mathbb{E}_{\mu_i}[Y^2]\right] - \overline{y}_0^2} \ ,$$

is a ratio of two measure affine functionals. Proposition 3.11 states that the reduction Theorem 3.8 applies so that

$$\sup_{\substack{\mu_{-i} \in \mathcal{A}_{-i} \\ \mathbb{E}[Y]=\overline{y}_0}} \frac{\mathbb{E}_{\mu_{-i}}\left[\left(\mathbb{E}_{\mu_i}[Y|X_{-i}]\right)^2\right] - \overline{y}_0^2}{\mathbb{E}_{\mu_{-i}}\left[\mathbb{E}_{\mu_i}[Y^2]\right] - \overline{y}_0^2} = \sup_{\substack{\mu_{-i} \in \Delta_{-i}(N_i+1) \\ \mathbb{E}[Y]=\overline{y}_0}} \frac{\mathbb{E}_{\mu_{-i}}\left[\left(\mathbb{E}_{\mu_i}[Y|X_{-i}]\right)^2\right] - \overline{y}_0^2}{\mathbb{E}_{\mu_{-i}}\left[\mathbb{E}_{\mu_i}[Y^2]\right] - \overline{y}_0^2} \ ,$$

and the result follows. □

In the following section we illustrate on the flood model the robust computation of first and total order sensitivity indices.

5.5.3 Illustration

The flood model introduced in Chapter 2 has two influential variables Q and K_s and two non influential ones Z_v and Z_m. The computation of the Sobol' indices in Table 2.2 relies in the initial distribution μ_{init}. We intend to take into account the uncertainty in the input measure by considering the moment class $\mathcal{A}^{(2)}$ introduced in Section 5.1.3. In this context, bounds and two moment constraints are enforced on each four input parameters.

The optimal ith first order index S_i is computed with respect to the varying ith input distribution, while the optimal ith total effect index S_{Ti} is computed over varying all but ith input measure. The QoI are computed over the extreme points with the canonical moment parameterization described in Algorithm 1 and optimized using DE algorithm. The sample size is taken sufficiently large $n = 10^6$ to overlook the confidence interval in the estimation of the Sobol' indices. The estimation error is smaller than 10^{-3} using the asymptotic pick-freeze estimator (Prieur and Tarantola, 2015). Results are depicted in Fig. 5.7.

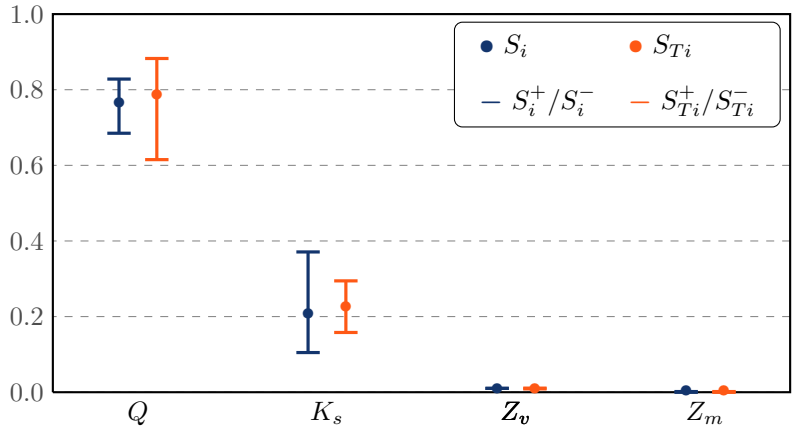


Figure 5.7: We depict the optimal first order Sobol' indices S_i^+ and S_i^- for each inputs varying independently in their moment class. And the optimal total order Sobol' indices S_{Ti}^+ and S_{Ti}^- when the inputs μ_{-i} vary in their moment class.

One can observe in Fig. 5.7 that the bounds on the Sobol indices of Z_v and Z_m remains very low. For the first Sobol' indices S_3 and S_4, it means that for imprecise distribution μ_3 (respectively μ_4) varying in its moment class, the contribution of this input to the variance averaged over the other input remains marginal. For the total order index S_{T3} and S_{T4}, it means that whatever the distribution of the $\mu_1 \times \mu_2 \times \mu_4$ (respectively $\mu_1 \times \mu_2 \times \mu_3$) varying in their respective moment class, the contribution to the variance of Z_v and Z_m considering all the interaction with others input is also negligible.

On the other hand, the result in Fig. 5.7 for Q and K_s the two PII are interesting, it shows that for an imprecise distribution on Q, the initial Sobol' index $S_1 = 0.764$ can decrease as low as $S_1^- = 0.6807$ while the maximal first order index of K_s initially at $S_2 = 0.206$ goes up to $S_2^+ = 0.3752$. Moreover, the total order index S_{T1}^- reaches an even lower bound at 0.611; we recall that the robust total effect index accounts for the uncertainty in all but the ith input distribution. Therefore, considering both indices S_2^+ and S_{T1}^-, one can deduce that the uncertainty on the distribution of K_s has the greatest impact on Sobol' index values. Nevertheless, it is difficult to interpret these results as each sensitivity index varies in a different moment class.

Notice that in Fig. 5.7, even when considering the most penalizing distributions the order ranking of the Sobol' indices remains similar. However, in different situation, in particular by changing the parameters of the Gumbel distribution of Q, it has been observed that the values of the Sobol' indices of the two PII can overlap see (Stenger et al., 2019a).

5.6 Robust Bayesian Analysis

5.6.1 Integration into the Framework

In this section, we explore how the historical robust Bayesian analysis is related to the OUQ framework set in this thesis. Generally speaking, Robust Bayesian analysis (Insua and Ruggeri, 2000) studies the sensitivity of Bayesian analysis to the choice of an uncertain prior distribution. The answer is robust if the inference does not depend significantly on the choice of the input prior distributions. Therefore, a Bayesian analysis is applied to all possible prior distributions from a given class of measures.

The posterior probability distribution can be calculated with Bayes' Theorem by multiplying the prior probability distribution π by the likelihood function $\theta \mapsto l(x \,|\, \theta)$, and then dividing by the normalizing constant, as follows:

$$l(\theta \,|\, x) = \frac{l(x \,|\, \theta)\pi(\theta)}{\int l(x \,|\, \theta)\, \pi(d\theta)}$$

Thus, it is natural to define Ψ the function mapping the prior probability measure to the posterior probability measure. In what follows, $\mathcal{X}$ denotes a Polish space

$$\begin{aligned} \Psi : \quad \mathcal{P}(\mathcal{X}) &\longrightarrow \mathcal{P}(\mathcal{X}) \\ \pi &\longmapsto \Psi(\pi) : \quad C_b(\mathcal{X}) \longrightarrow \mathbb{R} \\ & \qquad\qquad q \longmapsto \Psi(\pi)(q) = \frac{\int_{\mathcal{X}} q(\theta) l(x|\theta)\pi(d\theta)}{\int_{\mathcal{X}} l(x|\theta)\pi(d\theta)} \end{aligned}$$

The functional Ψ has very useful properties:

Lemma 5.9. *If the likelihood function $l(x|\cdot) : \theta \mapsto l(x|\theta)$ is continuous, then Ψ is continuous for the weak* topology in $\mathcal{P}(\mathcal{X})$.*

Proof. let (π_n) be a sequence of probability measure in $\mathcal{P}(\mathcal{X})$ converging in weak* topology towards some probability measure π. The convergence in weak* topology means that for every $q \in C_b(\mathcal{X})$, $\langle \pi_n | q \rangle \to \langle \pi | q \rangle$. But because $l(x|\cdot)$ is continuous the function $q \times l(x|\cdot)$ is also an element of $C_b(\mathcal{X})$, therefore

$$\int_{\mathcal{X}} q(\theta) l(x|\theta)\, \pi_n(d\theta) = \langle q \times l(x|\cdot) | \pi_n \rangle \longrightarrow \langle q \times l(x|\cdot) | \pi \rangle = \int_{\mathcal{X}} q(\theta) l(x|\theta)\, \pi(d\theta) \ ,$$

meaning that $\Psi(\pi_n)$ converges to $\Psi(\pi)$ in the weak* topology. This gives the sequential continuity of Ψ, thus its continuity. Indeed, because $\mathcal{X}$ is polish it is separable and metrizable. So that, $\mathcal{P}(\mathcal{X})$ is also metrizable (Aliprantis and Border, 2007, Theorem 15.12). Hence, it is first-countable (Croom, 2016, Theorem 4.7) which implies it is also sequential. This means that the sequential continuity is equivalent to the continuity. □

The function Ψ can be decomposed as a ratio $\Psi = \Psi_1/\Psi_2$, with

$$\begin{aligned}\Psi_1 : \quad \mathcal{P}(\mathcal{X}) &\longrightarrow \mathcal{P}(\mathcal{X}) \\ \pi &\longmapsto \Psi_1(\pi) : \quad C_b(\mathcal{X}) \longrightarrow \mathbb{R} \\ & \qquad\qquad q \longmapsto \Psi_1(\pi)(q) = \int_{\mathcal{X}} q(\theta) l(x|\theta) \pi(d\theta)\end{aligned}$$

and,

$$\begin{aligned}\Psi_2 : \quad \mathcal{P}(\mathcal{X}) &\longrightarrow \mathbb{R}_+^* \\ \pi &\longmapsto \int_{\mathcal{X}} l(x|\theta)\pi(d\theta)\end{aligned}$$

The main property of both Ψ_1 and Ψ_2 is that they are both linear maps. The posterior distribution, seen as a function of the prior distribution, is therefore the ratio of two linear functions. This is particularly interesting due to the following Proposition, which states that the composition of a quasi-convex function with the ratio of two linear mapping is also quasi-convex.

Proposition 5.10. *Let $\mathcal{A}$ be a convex subset of a topological vector space, and ϕ a quasi-convex lower semicontinuous functional on $\mathcal{A}$. Assume that $\Psi_1 : \mathcal{A} \mapsto \mathcal{A}$ is a linear mapping and $\Psi_2 : \mathcal{A} \mapsto \mathbb{R}_+^*$ is a linear functional. Then, $\phi \circ (\Psi_1/\Psi_2) : \mathcal{A} \mapsto \mathbb{R}$ is also a quasi-convex lower semicontinuous functional.*

Proof. Let π_1, π_2 be in $\mathcal{A}$. Given $\lambda \in [0,1]$, notice that

$$\begin{aligned}\phi\left(\frac{\Psi_1(\lambda\pi_1 + (1-\lambda)\pi_2)}{\Psi_2(\lambda\pi_1 + (1-\lambda)\pi_2)}\right) &= \phi\left(\frac{\lambda\Psi_1(\pi_1) + (1-\lambda)\Psi_1(\pi_2))}{\lambda\Psi_2(\pi_1) + (1-\lambda)\Psi_2(\pi_2)}\right) , \\ &= \phi\left(\beta\frac{\Psi_1(\pi_1)}{\Psi_2(\pi_1)} + (1-\beta)\frac{\Psi_1(\pi_2)}{\Psi_2(\pi_2)}\right) ,\end{aligned}$$

with $\beta = \dfrac{\lambda\Psi_2(\pi_1)}{\lambda\Psi_2(\pi_1) + (1-\lambda)\Psi_2(\pi_2)}$ in $[0,1]$. Hence,

$$\phi\left(\frac{\Psi_1(\lambda\pi_1 + (1-\lambda)\pi_2)}{\Psi_2(\lambda\pi_1 + (1-\lambda)\pi_2)}\right) \leq \max\left\{\phi\left(\frac{\Psi_1(\pi_1)}{\Psi_2(\pi_1)}\right) \ ; \ \phi\left(\frac{\Psi_1(\pi_2)}{\Psi_2(\pi_2)}\right)\right\} .$$

This proves the quasi-convexity of $f \circ (\Psi_1/\Psi_2)$. The lower semicontinuity stands because for $\alpha \in \mathbb{R}$, the lower level set

$$\Gamma_\alpha = \left\{\pi \in \mathcal{A} \mid \phi\left(\frac{\Psi_1(\pi)}{\Psi_2(\pi)}\right) \leq \alpha\right\} = \left\{\pi \mid \frac{\Psi_1(\pi)}{\Psi_2(\pi)} \in \phi^{-1}\left(]-\infty, \alpha]\right)\right\} ,$$

is the inverse image of the lower level set α under the continuous map $\pi \mapsto \Psi_1(\pi)/\Psi_2(\pi)$ according to Lemma 5.9. Therefore, Γ_α is closed. □

Proposition 5.10 proves that all the lower semicontinuous quasi-convex functions presented in the previous Chapters are well suited for robust Bayesian analysis. Hence, any QoI presented above can be used in a Bayesian setting. Therefore, the input distribution of a computer model μ can derive from a Bayesian inference such that $\mu = \Psi_1(\pi)/\Psi_2(\pi)$, where π is an imprecise distribution modeled in a reasonable class of prior. For instance, the optimization of the quantile of posterior distributions inferred from priors in a moment class can be reduced to the extreme points of this class.

Moreover, one can easily see that if the functional ϕ is measure affine then $\phi \circ (\Psi_1/\Psi_2)$ is also the ratio of two measure affine functionals. From Proposition 3.11, it then holds that lower semicontinuity is not necessary to apply the reduction Theorem. Consequently, one can also optimize moments or probabilities of the posterior distribution over a class of prior distributions (Sivaganesan and Berger, 1989).

5.6.2 Illustration

In the following, we consider that the input Q of the flood model H (Section 2.1) is modeled as before using a Gumbel distribution (see Table 2.1). Indeed, extreme value theory (Coles, 2001) justifies the choice of a Gumbel distribution for the maximal annual flow rate. However, in a Bayesian setting, the uncertainty on the location parameter ρ and the scale parameter β of the Gumbel distribution is quantified through a prior distribution $\pi(\rho, \beta)$. The prior distributions are defined through $\rho \sim \mathcal{G}a(500, 2)$ and $1/\beta \sim \mathcal{G}a(280, 2)$, where $\mathcal{G}a(\alpha, \tau)$ is the Gamma distribution with shape parameter α and scale parameter τ.

This choice of prior is questionable, hence, it would be more likely not to consider one specific prior but instead a set of admissible prior distributions. Therefore, we model the prior distribution to belong in a moment class, we specify its mean taken as the mean of the initial Gamma distributions. The bounds are taken equal to the 10% and 90% quantiles of the Gamma distributions for the prior distribution lower and upper bounds. So that it is representative of our *a priori* knowledge on the Gumbel parameters. Doing so the initial prior distribution belongs in the moment class, therefore, one can compare the result of a full Bayesian inference to the robust one.

Table 5.5: Corresponding moment constraints of the parameters ρ, β of the Gumbel distribution of Q.

Variable	Bounds	Mean
ρ	$[944, 1057]$	1000
$1/\beta$	$[517, 603]$	560

This corresponds to two moment classes, ρ belongs to $\widetilde{\mathcal{A}}_1^* = \{\mu \in \mathcal{P}([550, 700]) : \mathbb{E}_\mu[X] = 626.14\}$ and β to $\widehat{\mathcal{A}}_1^* = \{\mu \in \mathcal{P}([150, 250]) \mid \mathbb{E}_\mu[X] = 190\}$. The other parameter's distributions K_s, Z_v, Z_m are set to their previous classes in Eq. (5.8), that is respectively $\mathcal{A}_2^\dagger, \mathcal{A}_3^*$ and $\mathcal{A}_4^*$. Finally, the distribution $\Theta \sim (\rho, \beta, K_s, Z_v, Z_m)$ belongs to the product space $\mathcal{A}^{Bayes} = \widetilde{\mathcal{A}}_1^* \times \widehat{\mathcal{A}}_1^* \times \mathcal{A}_2^\dagger \times \mathcal{A}_3^* \times \mathcal{A}_4^*$.

The Gumbel model and the analytic formulation of the code in 2.1 yield the following close form for the probability of failure conditional on $(\rho, \beta, K_s, Z_v, Z_m)$:

$$\mathbb{P}(H \leq h \mid \Theta) = \exp\left(-\exp\left\{\beta\left(\rho - 300K_s\sqrt{\frac{Z_m - Z_v}{5000}}(h - Z_v)^{5/3}\right)\right\}\right) .$$

Therefore, the Bayesian PoF corresponds to the integrated cost

$$F_\Theta(h) = \mathbb{P}(H \leq h) = \int \mathbb{P}(H \leq h \mid \Theta)\, \pi(\Theta|D)\, d\Theta , \tag{5.19}$$

where $\pi(\Theta|D) \propto l(D|\Theta)\pi(\Theta)$ is the posterior distribution of Θ. The quantity is optimized over the extreme points of the moment class $\mathcal{A}^{Bayes}$, this reads

$$\sup_{\Theta\in\mathcal{A}^{Bayes}} F_\Theta(h) = \sup_{\Theta\in\mathcal{A}^{Bayes}} \sum_{\widetilde{k}_1=1}^{2}\sum_{\widehat{k}_1=1}^{2} \omega_1^{(\widetilde{k}_1)}\omega_1^{(\widehat{k}_1)} \times \frac{l(D|\rho^{(\widetilde{k}_1)},\beta^{(\widehat{k}_1)})}{\sum_{\widetilde{k}_1=1}^{2}\sum_{\widehat{k}_1=1}^{2}\omega_1^{(\widetilde{k}_1)}\omega_1^{(\widehat{k}_1)} l(D|\rho^{(\widetilde{k}_1)},\beta^{(\widehat{k}_1)})} \times \sum_{k_2=1}^{3}\sum_{k_3=1}^{3}\sum_{k_4=1}^{3} \omega_2^{(k_2)}\omega_3^{(k_3)}\omega_4^{(k_4)} \int_{30}^{z_2^{(k_2)}} \mathbb{1}_{\{H(\rho^{(\widetilde{k}_1)},\beta^{(\widehat{k}_1)},z,x_3^{(k_3)},x_4^{(k_4)})\leq h\}}\,\mathrm{d}z \ ,$$

The quantity in Equation Eq. (5.19) is minimized over the product space $\mathcal{A}^{Bayes}$, it is computed using algorithm 1 and optimized with the noisy DE solver in Algorithm 4. The results are depicted in Figure 5.8. The quantile of order 0.95 is equal to $4.58m$ which is only slightly higher that the initial quantile value equal to $4.29m$. However, we emphasize that this result is really sensitive to the bound values enforced on the hyperparameters ρ and β. Indeed, whenever the prior distribution is discrete, the posterior distribution is itself discrete with similar positions. Only the associated weights are adjusted with respect to the likelihood. Moreover, because there is only one mean constraint in the moment class, the positions of the discrete measure tend to reach the truncation bounds, that explains their strong influence.

In all generality, the sensitivity to the data of the most penalizing posterior distribution is weaken (Betro et al., 1994). It is known from the Cromwell's rules that the prior support should cover the whole space of parameters, so that any event has a nonzero mass in order to traduce its possibility of happening. Nevertheless, the worst prior, here discrete, should not be considered as an isolated pathology, it shows directions of instability where spreading mass critically affects the inference. Historically, Freedman (1965) proved the Bernstein–von Mises theorem (that states that the posterior distribution is asymptotically independent to the prior as the data sample grows large) does not hold almost surely when allowing for a very broad range of possible priors.

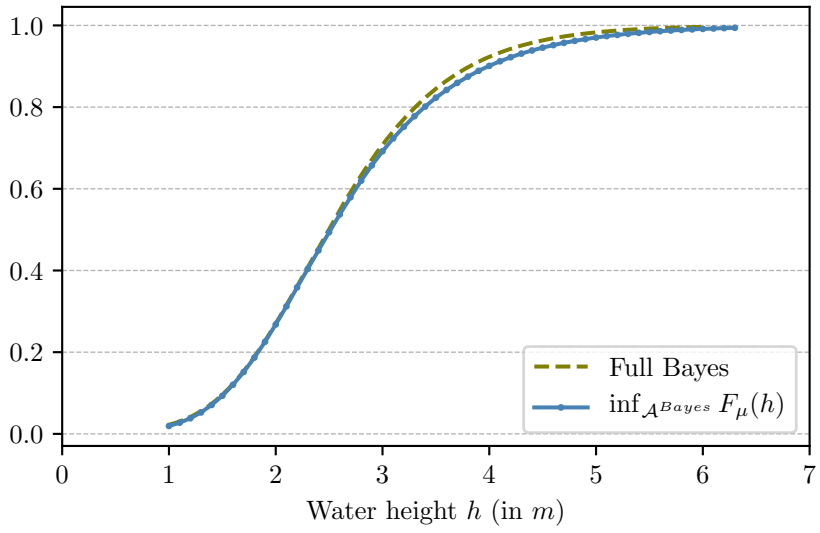


Figure 5.8: The solid line represents the CDF lower envelop over the measure set $\mathcal{A}^{Bayes}$ from Eq. (5.19). The dashed line represents the classical Full Bayes inference of the same quantity when Q is a modeled Gumbel distribution with prior density on its parameters.

Conclusion

In this chapter, several quantities of interest have been studied that suit the OUQ framework developed in the previous chapters. The QoI introduced are valuable for both their mathematical properties and practical relevance. Indeed, some of them such as the quantile serves as the reference risk measure for assessing safety margins in a risk analysis context. Mathematically speaking, the quantile is also a quasi-convex lower semicontinuous function of the input measure. We also study others QoI, for instance the PoF, that are measure affine functions. Such properties allow the application of one of our two main reductions: Theorem 3.8 and Theorem 3.10. Nice computational simplifications are derived by optimizing the QoI only over the extreme points of the considered measure space. In particular, we provide examples where exact computation of the QoI is possible, that is, when the measure space is a product of moment classes and the extreme points are all discrete measures [see Eq. (5.1) and Eq. (5.2)]. When the QoI can be exactly computed for every measure in the extreme points set, the optimization can be done using the DE algorithm. Otherwise, the noisy DE algorithm can be used whenever the evaluation of the QoI is noisy, typically when its computation requires MC methods. This happens when one input measure is continuous, for instance when it is the extreme point of an unimodal moment class.

Illustrations are carefully detailed in this Chapter for the flood model. First to help the understanding of the methodology, and second in order to depict how the optimal QoI can be translated in terms of robustness with respect to the analyst's objectives, such as establishing robust safety margins. The advantage of computing an optimal QoI is hence proven on this toy computer model. It appears that the uncertainty affecting the input distribution can have a dramatic consequence on the QoI. In particular when the size of the measure space is too large, that is when only little information is known on the input distribution. Therefore, an overlarge optimal QoI gives a clear signal that the epistemic uncertainty tainting the input parameter distributions cannot be omitted, and that further analysis or measurements are necessary. The optimization space should encompass precisely the actual uncertainty on the input, so that we do not overestimate or underestimate the QoI bounds. In that matter, the asset of the (unimodal) moment class is that its size is easily controlled by adding additional moments. Moreover, whenever an uncertainty subsists on the moment values, it is recommended to set varying bounds through inequality constraints. Another solution would be to explicitly represent uncertainty on the moment values through Bayesian inference, as shown on the last illustration.

It has been proven that enforcing the mean and variance of every input reduces yet quite a lot the moment space size. This information can be collected from data and/or expert knowledge, providing a real-world setting for this methodology. In our illustration, the moment constraints are derived from the fully elicited input distributions of the model. This method is inherently flawed as the moment constraints are retrieved explicitly from a distribution, the validity of which the analyst doubts. Enforcing inequality moment constraints allows to relax this issue. In the next chapter, we illustrate the optimization of a QoI on a real engineering case study. In particular, we focus on estimating a quantile, which serves as reference for establishing safety margin.

6

Application to an Industrial Computer Experiment

« Everything except language knows the meaning of existence. Trees, planets, rivers, time know nothing else. They express it moment by moment as the universe. Even this fool of a body lives it in a part, and would have full dignity within it but for the ignorant freedom of my talking mind. »

Les Murray

Contents

Introduction

This section aims at presenting the methodology on a real-world case. The computer experiment under study which has been introduced in Chapter 2 is the code CATHARE. We recall that this code is expensive, as many real life engineering computer experiment. Hence, we replaced the computer model by a surrogate model, here a Gaussian regressor with the methodology exposed in Section 2.2. So far, the code is supposed deterministic so that we use the predictor (the mean) of the Gaussian process in order to proceed the optimization of the QoI. Here, we are investigating a quantile. Indeed, the objective for the analyst is to determine a safety margin that guaranties the security of the nuclear power plant structure. As previously, Theorem 5.3 states it is equivalent the output CDF lower envelop over a measure space, it also gives a nice visualization of the optimization result. The CDF optimization of CATHARE over a moment space are depicted in Section 6.1

However, the whole methodology of robust analysis with a second level uncertainty quantification holds on to the computer model. Therefore, if the computer model is not accurate the OUQ analysis has no point, as the uncertainty affecting the model propagates through Algorithm 1 to the QoI bounds. The optimal QoI value becomes hardly interpretable as long as the error tainting the computer code is not handled. Here, even when considering the code CATHARE as an accurate representation of the reality, we have replaced the model with a metamodel whose prediction is imperfect. Accounting for the uncertainty in the metamodel is the purpose of Section 6.2. The methodology basically adds robustness in our approach.

6.1 Results on Cathare

As stated, this section presents the result of an OUQ analysis on the computer code CATHARE, noted G. In Chapter 2, we introduced a simplified model by selecting the nine most influential inputs amongst the initial 27 through the power of a test which rejected the independence between an input X_i and the output $Y = G(\mathbf{X})$ of the computer model. The test is based on a HSIC index. The 18 non-detected inputs are set to their nominal values. Then, a Gaussian process regressor $\mathscr{G}(\mathbf{x}, \theta)$ was built over the 9 most influential inputs conditional on the data set of 1000 simulations. The predictivity coefficient of the surrogate model is equal to $Q^2 = 0.92$. We refer to Section 2.2 for more details. Finally, we have at hand a cheap and deterministic model replacing the original expensive code G, by the predictor of the Gaussian process

$$\begin{aligned} \overline{\mathscr{G}} : \quad & \mathbb{R}^9 \longrightarrow \mathbb{R} \\ & \mathbf{x} \longmapsto \overline{\mathscr{G}}(\mathbf{x}) = \mathbb{E}_\theta[\mathscr{G}(\mathbf{x}, \theta)]. \end{aligned} \tag{6.1}$$

The nine inputs of the computer model are uncertain and modeled as random variables with distributions described in Table 2.3. Of course, these distributions are but one possible representation of reality. They derive from the CIRCE method (Damblin and Gaillard, 2020) and correspond to Bayesian posterior distributions with respect to the experimental database. Many different model could have been chosen. The second level uncertainty affecting the choice of the input distribution is taken into account by considering moment classes instead of one particular but imprecise distribution. Consequently, we present in Table 6.1 the moment constraints that are enforced on every input.

Table 6.1: Corresponding moment constraints of the 9 most influential inputs of the CATHARE model (Iooss and Marrel, 2019).

Variable	Truncation	First moment	Second moment	Physical input
$n°10$	$[0.1, 10]$	1.33	3.02	Interfacial friction in the hot legs
$n°22$	$[0, 12.8]$	6.4	45.39	Wall-liquid friction undersaturated
$n°25$	$[11.1, 16.57]$	13.83	192.22	Wall-liquid friction saturated
$n°2$	$[-44.9, 63.5]$	9.3	1065	Heat transfer in the core
$n°12$	$[0.1, 10]$	1.33	3.02	Interfacial friction in the steam generators inlet plena
$n°9$	$[0.1, 10]$	1.33	3.02	Interfacial friction in the steam generators outlet plena
$n°14$	$[0.235, 3.45]$	0.99	1.19	Interfacial friction in the core
$n°15$	$[0.1, 3]$	0.64	0.55	Interfacial friction in the upper plenum
$n°13$	$[0.1, 10]$	1.33	3.02	Interfacial friction in the downcomer

Each input measure satisfies only two moment constraints. Indeed, we have seen in Fig. 5.1 that it is usually a good compromise to sufficiently reduce the size of the measure space while minimizing the amount of information provided. It is, in a sense, less informative to enforce two moments than eliciting a whole distribution. With these constraints, the corresponding measure space encompasses the uncertainty affecting each input measure. A single constraint defines too large a moment class, leading to an overestimation of the measure of risk with a corresponding most penalized quantile that is not physically likely. On the other hand, with more than two constraints the number of support points of each discrete measure on the extreme points increases, thus lengthening the computational cost. Indeed, Eq. (5.5) shows that when each one of the d-inputs is a discrete measure supported on at most N points, the computation of the PoF involves running the code N^d times on a grid. Thus, adding one more constraint in the inputs of CATHARE implies running 4^9 times the metamodel $\overline{\mathscr{G}}$ instead of 3^9 times, for each evaluation of the PoF.

The constraints correspond, as for the flood model, to the moments of the initial distributions in Table 2.3. This makes sense as the initial distribution choice is data driven from the CIRCE method, so that the resulting moments are themselves inherently fitted to the data. We recall that, engineers are interested by the high-level quantile of the model output, typically of order 95%. Therefore, our risk measurement is the maximal quantile corresponding to the most penalizing input satisfying the information gathered in the constraints. By applying the duality theorem 5.3, it is sufficient to compute the minimal PoF to recover the maximal quantile. Of course, we apply the reduction Theorem to compute the maximal quantile, hence the optimum is found on the extreme points of the moment class. Because every input μ_i is in a moment

class denoted $\mathcal{A}_i$ with two moment constraints detailed in Table 6.1, the input measure $\mu = (\mu_1, \ldots, \mu_9)$ of the computer model $\overline{\mathscr{G}}$ belongs to the product space $\mathcal{A} = \prod_{i=1}^{9} \mathcal{A}_i$ that enforces the mutual independence of the inputs. The extreme distributions are discrete measures supported on at most three points $\mu_i = \sum_{k=1}^{3} \omega_i^{(k)} \delta_{x_i^{(k)}}$, they belongs in $\Delta_i \subset \mathcal{A}_i$. Finally, we denote $\Delta = \prod_{i=1}^{9} \Delta_i$ the extreme points set of $\mathcal{A}$s. The reduction theorem applied to the PoF reads:

$$\begin{aligned} \inf_{\mu \in \mathcal{A}} F_\mu(T) &= \inf_{\mu \in \mathcal{A}} \mathbb{P}_\mu(\overline{\mathscr{G}}(\mathbf{X}) \leq T) \ , \\ &= \inf_{\mu \in \Delta} \sum_{k_1=1}^{3} \cdots \sum_{k_9=1}^{3} \omega_1^{(k_1)} \ldots \omega_9^{(k_9)} \mathbb{1}_{\left\{\overline{\mathscr{G}}\left(x_1^{(k_1)}, \ldots, x_9^{(k_9)}\right) \leq T\right\}} \ . \end{aligned} \tag{6.2}$$

In Fig. 6.1, the computation of the minimal PoF is one day long for each threshold. That is why, the CDF is restricted to a reduced area of interest (quantile in between 0.5 and 0.99), we had already a vague idea where were located the corresponding thresholds, and we parallelized the task so that the computation did not exceed one week on a regular computer. One can compare in Fig. 6.1, the results of the computations performed with both the Mystic framework (McKerns et al., 2012) and our Algorithm 1. We recall that Mystic is an optimization toolbox for Python, that includes the OUQ framework introduced in Owhadi et al. (2013). Both computations were performed with strictly identical solvers, so that computation times were almost identical. For each threshold the computation takes few minutes on a computer equipped with 32Go and an Intel® Core™ i7-6700HQ CPU 2.60GHz x 8, but no parallelization process were implemented so that only one core was really used. Nevertheless, we emphasize that the performance of the DE solver is directly related to the space exploration quality. Therefore, the convergence speed is substantially better with the canonical moment parameterization. In Fig. 6.1, the gap between our method and Mystic is significant, the maximal quantile obtained with the Mystic framework is lower that the result of our optimization. This shows that it faces difficulties to explore efficiently the space of admissible measures and struggles to converge. Notice that there is still no guaranty that the true optimum has been reached by any method, but the canonical moment parameterization is systematically closer for a given computational budget, as the minimal PoF is lower.

The reduced performance of Mystic is entirely due to the parameterization of the optimization space. Indeed, in this toolbox the PoF is expressed directly by the positions and weights of the discrete measures as in Eq. (5.5). Once arbitrary support points are generated, an intermediary transformation of the measure is needed in order to respect the constraints. Typically, in order to respect one mean constrain, the positions are translated until the constraint is satisfied. In addition, whenever a mean and a variance are enforced, the support points are first translated to respect the mean, then scaled to respect the variance, and finally translated again to ensure the mean constraint remains satisfied. This method allows great adaptability to the type of constraint, but during those transformations (translations, scaling, etc.) the support points can be sent out of bounds so that the measure is no more admissible. Hence, in the DE solver, many individuals are rejected for violating the constraints. Because many population vectors are thrown away, the overall performance of the algorithm is reduced. Thus, the difficulty to converge apparent in Figure 6.1 reflects the poor space exploration of Mystic framework.

Meanwhile, our algorithm allows to explore the set of admissible measures without any vector rejection thanks to the canonical moment parameterization of the moment

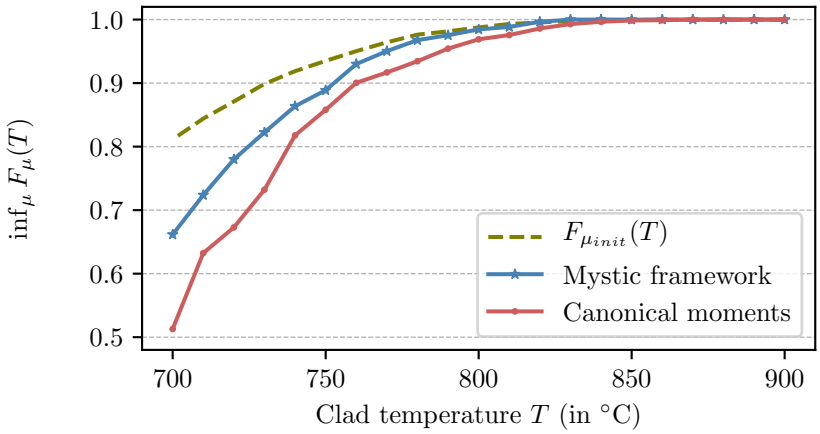


Figure 6.1: The dashed line represents the output CDF with input distributions in Table 2.3. The two solid lines represent the CDF lower envelop over a moment class with constraints in Table 6.1 and optimized respectively with Mystic framework and our methodology.

class. Indeed, our QoI has been parameterized by canonical moments as detailed in Algorithm 1, which allows a natural integration of the moments constraints. Our objective function is ultimately optimized globally using a DE solver, as for Mystic. However, our parameterization provides a free of constraint optimization. Indeed, the DE solver population evolves within the whole $[0,1]^n$ pavement corresponding to canonical moment tuples, where n is related to the number of moment constraints.

The difference between the blue and red curve in Fig. 6.1 is all the more pronounced as the exact same DE solver was used for the optimization. More precisely, the 40 individuals population was initially randomly generated, then evolved up to 800 generations using a best/1/bin strategy for choosing the offsprings (see Appendix C). The convergence criteria was set to 10^{-3} relative change in the objective function for the last 50 iterations. The crossover probability and the scaling factor were set to 0.9.

We highlight that any global optimizer can replace the DE solver. Moreover, notice that there is not guaranty that the true optimum has been reached. Results found are clearly better than Mystic, but this remains one of the main drawbacks of relying on a global optimizer with no proof of convergence. However, several runs of the optimization lead to the same result which is a positive sign.

We compare the different estimations obtained for the 95%-quantile peak cladding temperature for the IBLOCA application in Table 6.2. A 95%-confidence interval for the empirical quantile estimator was constructed with the bootstrap method. The $95\% - 95\%$ Wilks quantile (Wilks, 1941; Robert and Casella, 2005) denoted $Q^{Wilks}_{0.95,0.95}$ is computed from the 1000 simulation sample such that the 95%-quantile $Q^L_{0.95}$ verify

$$\mathbb{P}(Q^L_{0.95} \leq Q^{Wilks}_{0.95,0.95}) \geq 0.95 \ .$$

The *plug-in* approach and *full-Gp* approaches presented in Section2.2 correspond to the classical estimation of the quantile using respectively the predictor of the metamodel and the full Gaussian process (Iooss and Marrel, 2019). The OUQ method corresponds to the maximal quantile, when the input distributions are only defined by their first two moments and bounds (see Table 6.1). It is optimal in the sense that it minimizes the uncertainty in the input distribution considering the amount of information provided. Exploiting this kind of bounds, decision-makers are able to quantify the worst possible impact due to input uncertainty on the risk measure, so that they can adapt their

choices of safety margins. Here, the 95%-quantile goes from 741.46°C with the *full-Gp* approach to 787.85°C when accounting for the most penalizing distribution in the moment class.

Table 6.2: Results for the 95%-quantile estimates by the proposed robust approach (far right) versus parametric approaches (left to right): empirical (based on raw simulation results), Wilks 95% − 95% (based on raw simulation results), Gp plug-in and full Gp.

	Empirical	Wilks 95% − 95%	Plug-in	Full-Gp	OUQ
Value	743.21	748.92	735.83	741.46	787.85
90%-CI	[736.7, 747.41]			[738.76, 744.17]	

We emphasize that the optimization of the QoI over the moment class results in an overestimation of the risk measurement, as it corresponds to a worst-case scenario. This is not necessarily an issue in a context of reliability assessment. It gives a methodological approach to account for all available observations and expert knowledge without the need to calibrate the associated input distributions. The data are exploited by enforcing moment constraints onto the uncertain inputs, which is less intrusive than setting their whole distribution.

The joint distribution associated with the optimal 95%-quantile in Table 6.2 is a discrete measure. One can criticize that it hardly corresponds to a physical, real world, interpretation. In order to address this issue, one can enforce unimodality on every problematic inputs, since the optimal measure corresponds then to a convex combination of uniform distributions. Moreover, the most penalizing discrete distribution should not be considered as an "isolated pathology", it shows penalizing directions where the probability mass should spread out to get the worst-case QoI. This could also be approximated by a continuous measure that would give closely related worst-case scenario. In any case, continuous or discrete, the optimal distribution describes an epistemic uncertainty, i.e., a lack of knowledge, rather than the actual variations of a physical variable.

In this section, the expensive computer model has been replaced by the predictor of the Gp metamodel conditioned by the data $\mathscr{G}(\mathbf{x}, \theta)$. However, this regressor is an imperfect substitute to the model. Therefore, it affects the whole methodology and the optimal quantile value obtained in Table 6.2. In the next Section, we propose to take into account the covariance structure of the Gp to quantify the error induced by the metamodel, in a similar manner as the *full-Gp* approach.

6.2 Accounting for Metamodel Uncertainty

In Section 6.1, the quantile of the CATHARE code was optimized with respect to the uncertainty tainting the input distribution. The results we presented were based on a replacement of the original computer model G by a Gp metamodel predictor $\overline{\mathcal{G}}(\mathbf{x})$. By replacing the original code by a surrogate model, one substantially reduces computing time but loses in model quality. Indeed, the prediction of the metamodel does not perfectly fit the original computer model. This model error inevitably propagates to the QoI. So if one put a lot of efforts for accounting the uncertainty affecting the input distribution, it is necessary to also quantify the impact on the risk measurement

of the metamodel-induced uncertainty. We propose in this section a methodology by exploiting the full structure of the Gp metamodel instead of only the predictor.

This *full-Gp* approach consists basically in computing a sample of PoF in Eq. (6.2) from several draws of the Gp instead of taking only its expectation. We obtain a sample of PoF corresponding to different realization of the Gp, instead of only one PoF derived from the metamodel predictor. Because we have at our disposal a sample of PoF, the choice of a quantity to optimize is unclear. Therefore, we aggregate this sample into one representative value, we choose an appealing statistical indicator of this sample, for instance its mean, a quantile or its minimum or maximum. This is then optimized through the same Algorithm 1. This is known in the literature as robust optimization (Ben-Tal et al., 2009; Beyer and Sendhoff, 2007). For instance, if one chooses the minimum of the sample PoF, then when minimizing this minimum over the measure space the result remains conservative. Indeed, it corresponds to the minimization over all distributions in a measure space of the minimum PoF over several realizations of the Gp conditioned on the data. By doing so, we gain robustness with respect to the metamodel-induced uncertainty. In this situation, the bound obtained corresponds to the most penalizing Gp realization and the most penalizing input distribution which preserves the robustness consistency.

Let us consider as in the previous section the moment constraints of Table 6.1 corresponding to the nine inputs of the Gp metamodel $\mathscr{G}(\mathbf{x}, \theta)$. We recall that no unimodality assumption is made, so that the extreme points of the moment class $\mathcal{A}_i$ are discrete measures. Let denote $\mathbf{x}^{(\mathbf{k})} = (x_1^{(k_1)}, \ldots, x_9^{(k_9)})$ in the computation of the PoF in Eq. (6.2), so that it rewrites

$$\inf_{\mu \in \mathcal{A}} F_\mu(T) = \inf_{\mu \in \Delta} \sum_{k_1=1}^{3} \cdots \sum_{k_9=1}^{3} \omega_1^{(k_1)} \ldots \omega_9^{(k_9)} \mathbb{1}_{\left\{\mathbb{E}_\theta[\mathscr{G}\left(\mathbf{x}^{(\mathbf{k})}, \theta\right)] \leq T\right\}} \,. \tag{6.3}$$

The positions and the weights of the discrete measure are the parameters evolving in the global optimizer. Hence, even when parameterized with canonical moments, each individual of a population in the evolutionary algorithm has the weights and positions set. Therefore, in Eq. (6.3) the code $\mathscr{G}$ is computed over a 9-dimensional grid with 3^9 nodes denoted $\{\mathbf{x}^{(\mathbf{k})}\}_{\mathbf{k} \in \{1;2;3\}^9}$. Nevertheless, the important point is that those nodes are set for each individual in the solver population, so that it is possible to generate different realizations of the Gp metamodel on those nodes. We highlight that this method is only possible in the presence of discrete measures. Indeed, if one input is modeled in an unimodal moment class, one cannot draw several realizations of the Gaussian process using this process as the PoF is computed through MC methods. Here, for n realizations of the Gp we get the sample of PoF

$$\mathscr{Y}_{T,\mu}^n = \left\{ \sum_{k_1=1}^{3} \cdots \sum_{k_9=1}^{3} \omega_1^{(k_1)} \ldots \omega_9^{(k_9)} \mathbb{1}_{\left\{\mathscr{G}\left(\mathbf{x}^{(\mathbf{k})}, \theta_j\right) \leq T\right\}} \right\}_{j=1,\ldots,n} = \left\{ y_{T,\mu}^{(j)} \right\} \,. \tag{6.4}$$

Whenever one is interested in the mean over all realizations of the Gp process then the average PoF writes

$$\overline{\mathscr{Y}}_T(\mu) = \mathbb{E}_\theta \left[\sum_{k_1=1}^{3} \cdots \sum_{k_9=1}^{3} \omega_1^{(k_1)} \ldots \omega_9^{(k_9)} \mathbb{1}_{\left\{\mathscr{G}\left(\mathbf{x}^{(\mathbf{k})}, \theta_j\right) \leq T\right\}} \right] \,, \tag{6.5}$$

which expression has to be compared to the one in Eq. (6.3). The mean or any other statistical indicator, such as a quantile, can then be classically estimated from the

sample of PoF in Eq. (6.4). For instance, the estimator of the mean writes

$$\widehat{\mathscr{Y}}^n_{T,\mu} = \frac{1}{n}\sum_{j=1}^{n} y^{(j)}_{T,\mu} \ , \tag{6.6}$$

and the estimator of the PoF distribution is given as

$$\widehat{\mathscr{F}}^n_{T,\mu}(h) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{y^{(j)}_{T,\mu}\leq h\}} \ , \tag{6.7}$$

with associated estimated p-quantile

$$\widehat{\mathscr{Q}}^{n,p}_{T,\mu} = \inf\left\{h\in\mathbb{R} : \widehat{\mathscr{F}}^n_{T,\mu}(h)\geq p\right\} \ . \tag{6.8}$$

Then, one can proceed the optimization of any of these quantities over the measure space for μ varying in $\mathcal{A}$.

The sample of PoF in Eq. (6.4) is distributed according to a random variable that we should further investigate. The Gp is evaluated on a grid $\{\mathbf{x}^{(\mathbf{k})}\}_{\mathbf{k}\in\{1;2;3\}^9}$, on each of this point the Gp is normally distributed

$$\mathscr{G}(\mathbf{x}^{(\mathbf{k})},\theta) \sim \mathcal{N}(\overline{\mathscr{G}}(\mathbf{x}^{(\mathbf{k})}),\sigma) \ . \tag{6.9}$$

Therefore,

$$\mathbb{1}_{\{\mathscr{G}(\mathbf{x}^{(\mathbf{k})},\theta)\leq T\}} \sim \mathcal{B}(p),$$

where $\mathcal{B}(p)$ designates a Bernoulli distribution with parameter p. The parameter here corresponds to the probability that a random variable with normal distribution in Eq. (6.9) exceeds the threshold T

$$p = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{T} \exp\left(-\frac{1}{2}\frac{\left(x-\overline{\mathscr{G}}(\mathbf{x}^{(\mathbf{k})})\right)^2}{\sigma^2}\right) \ .$$

Therefore, the random variable

$$\sum_{k_1=1}^{3}\cdots\sum_{k_9=1}^{3} \omega_1^{(k_1)}\ldots\omega_9^{(k_9)}\mathbb{1}_{\left\{\mathscr{G}\left(\mathbf{x}^{(\mathbf{k})},\theta_j\right)\leq T\right\}} \ ,$$

is a convex combination of 3^9 dependent Bernoulli distributions. Indeed, the dependence between the Bernoulli originates from the parameters, which derive from the same Gp conditioned on the data.

In the following, we consider $n = 100$ random realizations of the Gp. However, we restrict the study to the six most influential PII. We fix three components (n°15, n°14, n°13) of the Gp to a nominal value, here to their mean value in Table 6.1. Such a restriction was made necessary to keep the computation time manageable on an ordinary computer. Moreover, the random realizations of the Gp must be evaluated on a grid of size N^d where d is the input number, and N the number of support points. Consequently, without restrictions the different Gp realizations must be evaluated on 3^9 points. As this operation involves inverting a covariance matrix of the same size, and must be done for every individual of the DE solver, this is too expensive and requires the need for additional computational power. However, by reducing the problem to 3^6 evaluated points, the computation can be run normally on any computer.

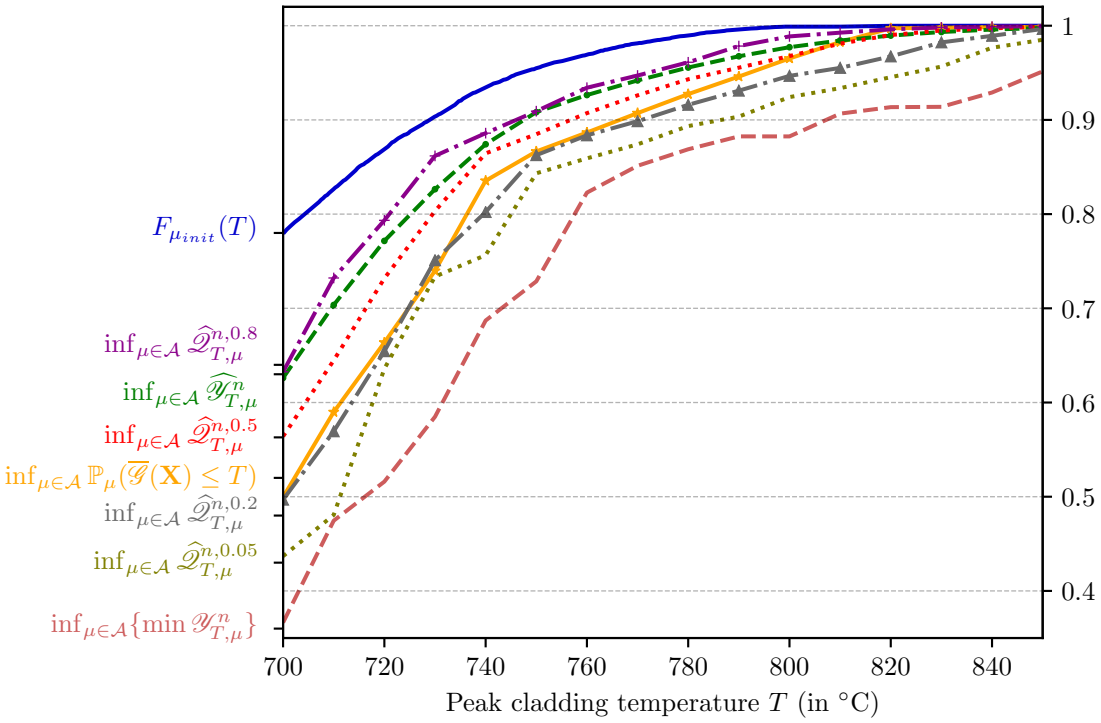


Figure 6.2: One compute a PoF sample $\mathscr{Y}^n_{T,\mu}$ of size $n = 100$. Different quantiles of this sample $\widehat{\mathscr{Q}}^{n,p}_{T,\mu}$ [see Eq. (6.8)] are optimized over the measure space, as well as the sample mean $\widehat{\mathscr{Y}}^{n}_{T,\mu}$ [see Eq. (6.6)]. In solid line, we compute the output initial distribution, and the optimization of the CDF lower envelop from the predictor of the metamodel (reduced to 6 input). Each curve takes approximately two days to compute on a regular computer, without any optimization and parallelization of the algorithm.

It yields a $n = 100$ size sample of PoF corresponding to random realizations of the Gp metamodel [see Eq. (6.4)]. We consider the optimization of different statistical indicator on this PoF sample, for instance the mean as in Eq. (6.5), but also different quantile values, such as median, 20%-quantile, 5%-quantile, and also minimum. One can remark that only low quantile values are considered, indeed, we recall that we aim at minimizing the PoF over the measure space. Each one of these statistical indicators is a QoI that is computed using Algorithm 1 and optimized using the DE solver in Algorithm 3. In order to gain robustness, it is appropriate to consider statistics of the PoF sample with low values. Doing so, we preserve the conservatism of the robust analysis as we account for both the most penalizing input distribution and the most penalizing realization of the Gp metamodel. This overconservatism is interesting for the analyst because the bounds obtained on the PoF represent a worst-case scenario that should never be observed. If this overestimated measure of risk corresponds to safety margins that are "acceptable" for the decision-maker, the analyst can be positively confident about the structure's safety.

The results are depicted in Figure 6.2. The ranking of the curves is coherent with the statistical indicators used to aggregate the PoF sample. We also depict the 95% quantile of each curves in Table 6.3. It appears that the curve for the 5%-quantile and the minimal value of the PoF sample $\mathscr{Y}^n_{T,\mu}$ are not smooth. It suggests that the number of Gp realizations set to $n = 100$ is too small for estimating extreme quantile, so that the reliability of the estimates is associated with large confidence

intervals. It is well known that empirical high quantiles are noisy. However, because they are approximately Gaussian with known variance, one can compute the number of trajectory required to make the curve smoother.

One can also notice that the CDF lower envelop computed from the predictor (the solid yellow line) gives almost similar results as the quantile level of order 20% of the PoF sample. In other words, the predictor returns a PoF lower than almost 80% of the Gp realizations. This result is not a general characteristic, and is specific to each computer model. Here, the optimal quantile deriving from the predictor appears to be rather conservative, but in other situations it could underestimate the true value. For this reason, whenever possible one should always account for the metamodel error.

Table 6.3: Results for the 95%-quantile estimates taking into account the metamodel uncertainty. Those are the quantiles of the CDF lower envelop depicted in Fig 6.2

	95%-quantile (°C)
Initial distribution	747.1
$\inf_{\mu\in\mathcal{A}} \widehat{\mathscr{Q}}^{n,0.8}_{T,\mu}$	772.0
$\inf_{\mu\in\mathcal{A}} \widehat{\mathscr{Y}}^{n}_{T,\mu}$	775.9
$\inf_{\mu\in\mathcal{A}} \widehat{\mathscr{Q}}^{n,0.5}_{T,\mu}$	785.6
$\inf_{\mu\in\mathcal{A}} \mathbb{P}_\mu(\overline{\mathscr{G}}(\mathbf{X}) \leq T)$	792.1
$\inf_{\mu\in\mathcal{A}} \widehat{\mathscr{Q}}^{n,0.2}_{T,\mu}$	803.6
$\sup_{\mu\in\mathcal{A}} \widehat{\mathscr{Q}}^{n,0.05}_{T}(\mu)$	824.1
$\inf_{\mu\in\mathcal{A}}\{\min \mathscr{Y}^{n}_{T,\mu}\}$	849.3

Conclusion

This chapter was dedicated to applying the OUQ methodology to a real-word engineering case. The expensive computer model has been replaced and reduced to a nine dimensional Gaussian process. Each input distribution belongs in a moment class with two constraints enforced. We first computed the maximal quantile over this measure space by using the predictor of the Gp as a deterministic surrogate model. The results obtained give nice interpretable bounds but we do not consider the imperfect prediction of the metamodel.

Hence, we propose a methodology to account for the whole covariance structure of the Gp. Several trajectories of the Gp are drawn yielding a sample of QoI. Then, we optimized a well suited statistical indicator of this sample, this technique is known as robust optimization. The final results is robust regarding both the uncertainty affecting the surrogate model, and the uncertainty affecting the input distribution.

We emphasize that this whole methodology is possible only in the presence of discrete measures. It is required that every input distribution belongs in a moment class as in Eq. (6.3). Indeed, if just one input is in a unimodal moment class, then the computation of the PoF would involve a continuous distribution as the extreme points of the unimodal moment class are mixture of uniform distributions. In that case, the computation of several realizations of the Gp metamodel is much more complicated,

the issue is that there is no fixed point over which evaluating the Gp. A solution would be to compute the PoF using MC methods based on a fixed design of experiment for every individual of the global optimizer.

Finally, this methodology tends to be computationally expensive. In order to account for the uncertainty tainting the metamodel, we had to reduce the problem to the six most influential inputs. However, evolutionary algorithms are easily parallelized, an advantage we did not fully take benefit in this work . Using a computer cluster, one can expect to significantly quicken the computation time of the optimization.

7

Input Dependence and High Dimension

« C'est une triste chose de songer que la nature parle et que le genre humain n'écoute pas. »

Victor Hugo

Contents

Introduction

So far, the application introduced in Chapters 5 and 6 relied upon the assumption that the input parameters of a computer model were mutually independent. To be more specific, in Chapter 3 the theoretical developments assumed that the distribution of each scalar parameter belongs to the space of all probability measures on a Polish space $\mathcal{P}(\mathcal{X})$. In all generality, this Polish space could as well be multidimensional, for instance on a pavement $\prod_i [a_i, b_i]$, so that the joint distribution in $\mathcal{P}(\prod_i [a_i, b_i])$ would account for the dependence between inputs. However, the canonical moment parameterization introduced in Chapter 4 can only be applied to scalar measures. Hence it is convenient that the input distribution μ has independent components, since it is a product of one-dimensional marginal distributions $\mu = (\mu_1, \ldots, \mu_d) \in \prod_{i=1}^{d} \mathcal{P}(\mathcal{X}_i)$. For this reason, every illustration presented in Chapters 5 and 6 assumed independent inputs in order to perform the optimization using the canonical moments parameterization.

This independence assumption, while justified in the applications presented in Chapter 2, can be too restrictive. Indeed, sometimes no guaranty of independence can be ensured. In an engineering context, where physical phenomena inside a black-box computer model can be complex to comprehend, independence between the input variables is indeed a strong assumption.

This Chapter is dedicated to the presentation of several leads that have been explored to account for dependent variables. Needless to say that it is a challenging problem. The three different sections are self-contained, and can be read independently. However, we recommend keeping the chronological order. Indeed, the first section explores the limits of the canonical moments parameterization when dealing with a multivariate random vector in a moment space. For that, we investigate how our OUQ framework can be applied to a neural network with a significantly higher number of parameters compared to the reduced nine-dimensional version of the CATHARE computer model studied in the previous chapter. Section 7.2 focuses on deterministic methods to solve OUQ problems, that warrants conservative results. To that end, we investigate semi-definite programming, which has been historically used to solve moment problems. The conservative bounds it yields are very interesting as it complements the non-conservative approach of the previous sections. The last section is based on theory of copula, a handy tool to model dependent variables. The set of all copula is a compact convex set which possesses its own extreme points. We show how the reduction Theorem can be applied to this set, thanks to the general framework developed in Chapter 3. However, in this setting the marginal distributions of the joint measure μ must be fixed. Hence, the methodology differs in this respect from the original purpose of this thesis.

7.1 Numerical Limitations with Canonical Moments

In this section, we discuss some limitations, principally from a computational point of view, of the canonical moments parameterization for a random vector in moment class. To this matter, we study a neural network (NN), more specifically focusing on a *multilayer perceptron* (Rumelhart et al., 1986). We adopt the probabilistic approach for such tools known as Bayesian neural network, introduced by (Neal, 1996). It is common for NNs to depend on a large number of parameters. This allows us to test in this section how the framework introduced in this thesis can cope with high-dimensional

uncertain inputs. In particular, in Chapter 6, it was particularly expensive to compute OUQ bounds on the exceedance probability of the output of a computer model with only nine inputs. Indeed, whenever the computation of the QoI is reduced to the extreme points written as product of discrete measures, it is equivalent to computing the QoI over a grid of points whose size grows exponentially with the dimension of the input vector [see for instance Eq. (5.3) or Eq. (5.5)].

Here, we deal with a problem involving hundreds of parameters so that the previous methodology cannot be adapted. We investigate a solution where parameters are concatenated by blocks. Doing so, parameters of a same block are no longer mutually independent. We will show that this solution is not fully satisfactory and raises new numerical issues, in particular due to the canonical moment parameterization. First, we detail the construction of a Bayesian NN in Section 7.1.1. Then, in Section 7.1.2, we investigate in an illustration why canonical moments raise problems in high dimension.

7.1.1 Bayesian Neural Network

A multilayer perceptron network takes in a set of inputs x_i, and, computes the corresponding outputs of interest $G_k(\mathbf{x})$, perhaps using some hidden units. In a typical network with one hidden layer, such as the one illustrated in Fig. 7.1, the outputs can be computed as follows:

$$\begin{aligned} G_k(x) &= z_k^{(2)} + \sum_j u_{jk}^{(2)} h_j(\mathbf{x}) \ , \\ h_j(\mathbf{x}) &= \tanh(z_j^{(1)} + \sum_i u_{ij}^{(1)} x_i) \ , \end{aligned} \tag{7.1}$$

where $u_{ij}^{(1)}$ is the weight of the connection between input unit i and hidden unit j, and $u_{jk}^{(2)}$ is the weight of the connection betwen hidden unit j and output unit k. The parameters $z_j^{(1)}$ and $z_k^{(2)}$ are the biases of the hidden and output units. Each output value is thus obtained as the weighted sum of all hidden unit values, plus a bias. Every hidden unit is given by a weighted sum of input unit values, which then passes through a nonlinear activation function, such the hyperbolic tangent used in this example. A nonlinear activation function is useful to represent "hidden features" of the inputs, that are useful to compute the corresponding outputs. If linear activation functions were used, hidden layer can be omitted and a direct connection between the inputs and outputs could be equivalently schemed.

Several works have shown that multilayer perceptrons can be used to approximate any function defined on a compact domain, whenever sufficient hidden units are used (potentially many) (Cybenko, 1989; Funahashi et al., 1989; Hornik et al., 1989). Nevertheless, more complex architectures are commonly used, in particular so-called deep neural networks including a high number of hidden layers, or recurrent networks, in contrast to the feedforward networks used here, whose connections never form cycles, and whose outputs can be computed in a single forward pass. Using complex architectures can improve predictivity while lowering the number of units.

From now on, we use the multilayer perceptron network to define a probabilistic model for regression, but it can also be employed for classification tasks (Neal, 1996). In any case, it is necessary to define the distribution of one or several outputs $y_k = G_k(\mathbf{x})$ given the input $\mathbf{x}$. As is usual in the regression context, no distribution is considered for $\mathbf{x}$, assumed fixed by the user.

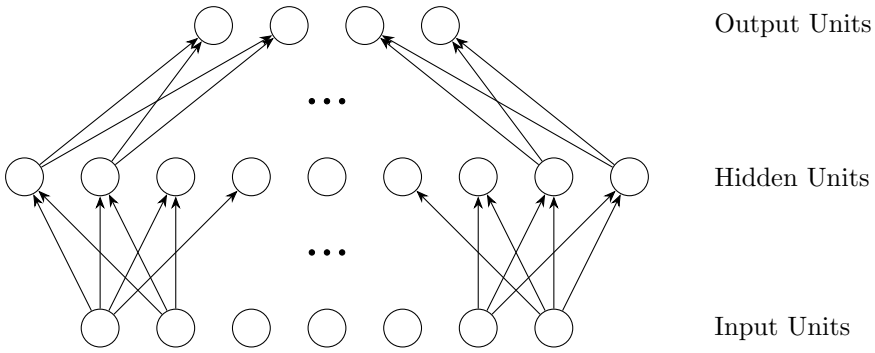


Figure 7.1: A multilayer perceptron with one layer of hidden units. The input units at the bottom are fixed to their values for a particular case. The values of the hidden units are then computed, followed by the values of the output units. The value of a unit is a nonlinear function of the weighted sum of the values received from other units connected to it via arrows.

Likewise, the distribution of the weights needs to be further modeled. For a regression model, the conditional distribution of output unit y_k, given input $\mathbf{x}$, is modeled as a Gaussian, with mean $G_k(\mathbf{x})$ and standard deviation σ_k. The different outputs are usually taken to be mutually independent given the input, so that the conditional probability density of the whole output writes

$$\mathbb{P}(\mathbf{y}|\mathbf{x}) = \prod_k \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-(f_k(\mathbf{x}) - y_k)^2/2\sigma_k^2\right) . \tag{7.2}$$

The standard deviation σ_k might be seen as a noise level, and can be fixed or set as a hyperparameter in the model.

The weights and biases are learnt based on a training set $(\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), \ldots, (\mathbf{x}^{(n)}, \mathbf{y}^{(n)})$ corresponding to some inputs $x^{(i)}$ and their corresponding output $\mathbf{y}^{(i)}$. Standard training procedures adjust the weights and biases in the network in order to minimize an error measure, usually the sum of the squared difference between the network predictions and the training set outputs. We emphasize that minimizing this error is equivalent to maximum likelihood estimation for the Gaussian noise model Eq. (7.2), since minus the log of the model likelihood is proportional to the sum of the squared errors.

Classical issues during the learning, such as overfitting, can be solved for instance by adding a penalty term to the sum of squares, proportional to the sum of squares of the weights and biases for instance. This tends to favor smaller weights and is known as weights decay. This problem is difficult as penalizing too much the learning procedure leads to underfitting the network, ignoring the data.

In a Bayesian approach of NN learning, the objective is to find the predictive distribution for the target values in a new test case, not belonging in the training set. Since the input distribution is not modeled as it is given by the user in this new case, the predictive output distribution writes

$$\begin{aligned} &\mathbb{P}(\mathbf{y}^{(n+1)} \mid \mathbf{x}^{(n+1)}, (\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), \ldots, (\mathbf{x}^{(n)}, \mathbf{y}^{(n)})) \\ &\qquad = \int \mathbb{P}(\mathbf{y}^{(n+1)} \mid \mathbf{x}^{(n+1)}, \boldsymbol{\theta})\ \mathbb{P}(\boldsymbol{\theta} \mid (\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), \ldots, (\mathbf{x}^{(n)}, \mathbf{y}^{(n)}))\, \mathrm{d}\boldsymbol{\theta} , \end{aligned} \tag{7.3}$$

where $\boldsymbol{\theta}$ represents the network parameters (weights and biases). The posterior density of the parameters $\boldsymbol{\theta}$ is proportional to the prior density times the likelihood function.

Of course, in an OUQ framework we will consider that the prior is imprecisely defined. More precisely, we will define moment constraints for the parameters' prior distributions. Because the input distribution is not modeled, the likelihood writes here:

$$l(\boldsymbol{\theta} \; ; \; (\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), \ldots, (\mathbf{x}^{(n)}, \mathbf{y}^{(n)})) = \prod_{i=1}^{n} \mathbb{P}(\mathbf{y}^{(i)} \mid \mathbf{x}^{(i)}, \boldsymbol{\theta}) \; . \tag{7.4}$$

The distribution for the output $y_k^{(i)}$, given the corresponding input and the network parameters $\boldsymbol{\theta}$, is defined for a regression task in Eq. (7.2).

Then, because the user wishes to predict a new case $y_k^{(n+1)}$ and its squared loss error, the best strategy consists in estimating its predictive distribution mean. For a regression model, it writes

$$\widehat{y}_k^{(n+1)} = \int G_k(\mathbf{x}^{(n+1)}, \boldsymbol{\theta}) \; \mathbb{P}(\boldsymbol{\theta} \mid (\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), \ldots, (\mathbf{x}^{(n)}, \mathbf{y}^{(n)})) \, \mathrm{d}\boldsymbol{\theta} \; , \tag{7.5}$$

where the network output function G_k is written with its dependency to the input and the network parameters $\boldsymbol{\theta}$ explicitly. This expression shows how to determine the squared loss error of the output estimation. Bayesian NN are indeed appealing for evaluating confidence in the predictive values of the computer model. Without this framework a NN acts as a deterministic function, which is one of the main drawbacks for artificial intelligence applications: the regressor accuracy is not currently quantified. Other methods, such as the dropout strategy, are commonly used to add confidence intervals to the regressor prediction (Srivastava et al., 2014; Gal and Ghahramani, 2016).

At first sight, Bayesian networks may not appear suitable for use with NN. Indeed, Bayesian inference starts with assigning a joint prior to the model parameters, embodying the prior beliefs about the problem. However, in a multilayer perceptron, the parameters are weights and biases, whose interpretations are obscure. The Bayesian engine is weakened by a choice of poor prior. Reducing the prior information to a single probability distribution necessarily implies some arbitrary choices, such as that of a parametric family of probability measures. This issue calls for a robust analysis of the problem.

Hence, in the following we consider a set of priors for the parameters. Few information are available on the parameters weights. Nevertheless, we have at our disposal estimates that are built during the learning procedure, denoted by θ^*, which we propose to use as the parameter prior means. This corresponds to a moment constraint enforced into the parameter prior distributions, and will be actually the only one enforced. The parameter bounds denoted by $[a_i, b_i]$ can be considered as hyperparameters but will be set in the following for simplicity to 10% of the mean value. As previously stated, we intend to optimize the predictive distribution mean in Eq. (7.5) over this moment class. Let us detail further the procedure. To do so, we focus in the next section on a particular example. We recall that our purpose here is to exhibit limitations of canonical moments use for parameterizing high dimensional problems, therefore, we won't be too extensive in the setting of the illustration.

7.1.2 Illustration of the Problem

Let consider a set of learning points generated inside the $[-20, 20]$ interval, following a one dimensional stationary centered Gaussian process associated to a Gaussian kernel with length scale set to 5, so that the 500 generated points follows a smooth trajectory.

70% of the generated data are used for training, and the remaining 30% for prediction tests. The training and test sets are depicted in Fig. 7.2, as one can see the training and test points are not chosen randomly. The multilayer perceptron learnt on this training set has the following architecture: there is only one input and one output as the data are one dimensional, two hidden layers of 50 units associated to hyperbolic tangent activation function are constructed. This simple architecture is depicted in Fig. 7.3.

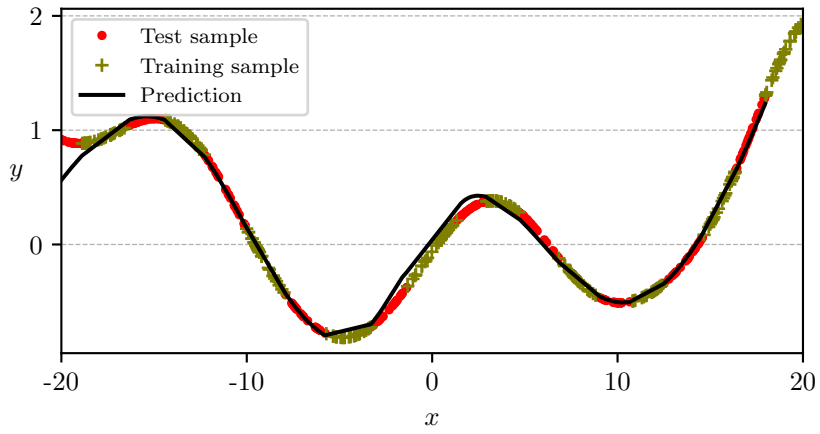


Figure 7.2: The training and test sets are generated from a one dimensional Gaussian process associated to a Gaussian kernel with length scale set to 5. The multilayer perceptron prediction, once trained, is pictured in solid line.

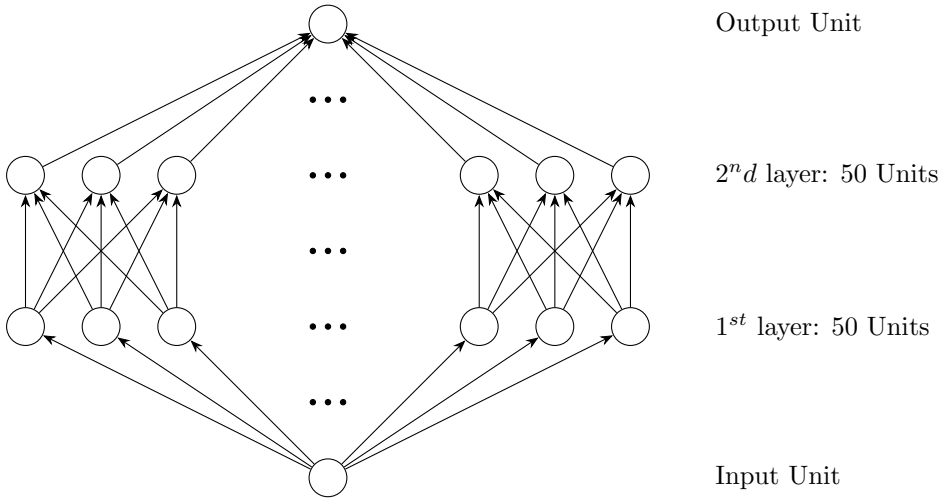


Figure 7.3: A multilayer perceptron with two hidden layers. The input unit at the bottom and the output represents one dimensional real function, more specifically it predicts a Gaussian process with Gaussian kernel and length scale set to 5, conditioned on 350 simulations points.

The total number of parameters is 2 701, and comprise:

- 100 parameters for the function biases and weights $z_j^{(1)}$ and $u_j^{(1)}$ to go from the input unit to the first hidden layer though

$$h_j^{(1)}(x) = \tanh(z_j^{(1)} + u_j^{(1)} x) \ , \ \text{for } 1 \leq j \leq 50 \ ;$$

- 50 biases $z_j^{(2)}$ and 2500 weights $u_j^{(2)}$ from the first hidden layer to the second

$$h_j^{(2)}(x) = \tanh(z_j^{(2)} + \sum_{k=1}^{50} u_{kj}^{(2)} h_k^{(1)}(x)) \ , \text{ for } 1 \leq j \leq 50 \ ;$$

- 1 bias b_1 and 50 weights v_j for defining the output function G_1 from the second hidden layer

$$G_1(x) = z_1^{(3)} + \sum_{j=1}^{50} u_j^{(3)} h_j^{(2)}(x) \ .$$

Even with this simple architecture, one can see that the number of parameters increases quickly. The OUQ framework introduced previously does not seem adapted to deal with such a high number of constraints. Indeed, assuming that all inputs are independent, each parameter belongs to a moment class with a single constraint. Hence, in order to optimize the quantity in Eq. (7.5) over the moment space, the computation of the QoI requires to evaluate the NN precisely 2^{2701} times, as every extreme point is supported on a most two Dirac masses. This is, obviously, computationally intractable. This exponentially growing cost of the grid size required to compute the expectation has already been discussed in Section 5.1.1.

7.1.3 Block-Independence modelization

To overcome this issue, we propose to consider block-independence instead. Hence, we constitute 3 blocks of parameters (one per layer)

$$\begin{aligned}
(u_1^{(1)}, \dots, u_{50}^{(1)}, z_1^{(1)}, \dots, z_{50}^{(1)}) &\sim \boldsymbol{\mu}_1 \in \mathcal{P}(\prod_{i=1}^{100} [a_i, b_i]) \ , \\
(u_{1,1}^{(2)}, \dots u_{1,50}^{(2)}, u_{2,1}^{(2)}, \dots, u_{50,50}^{(2)}, z_1^{(2)}, \dots, z_{50}^{(2)}) &\sim \boldsymbol{\mu}_2 \in \mathcal{P}(\prod_{i=101}^{2650} [a_i, b_i]) \ , \\
(u_1^{(3)}, \dots u_{50}^{(3)}, z_1^{(3)}) &\sim \boldsymbol{\mu}_3 \in \mathcal{P}(\prod_{i=2651}^{2701} [a_i, b_i]) \ .
\end{aligned} \tag{7.6}$$

with $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$, and $\boldsymbol{\mu}_3$ of dimensions $100, 2550$, and 51, respectively. As stated, we enforce one mean constraint on every distribution $\boldsymbol{\mu}_i$ such that $\mathbb{E}_{\boldsymbol{\mu}_i}[\mathbf{X}] = \boldsymbol{\theta}_i^*$, where $\boldsymbol{\theta}_i^*$ are the learnt parameters on the training set of the layer i, hence, it corresponds to a 'vectorized' version of the constraints, yielding the following moment class:

$$\begin{aligned}
\mathcal{A}_1 &= \left\{ \boldsymbol{\mu}_1 \in \mathcal{P}(\prod_{i=1}^{100} [a_i, b_i]) \mid \mathbb{E}_{\boldsymbol{\mu}_1}[\mathbf{X}] = \boldsymbol{\theta}_1^* \right\} \ , \\
\mathcal{A}_2 &= \left\{ \boldsymbol{\mu}_2 \in \mathcal{P}(\prod_{i=101}^{2650} [a_i, b_i]) \mid \mathbb{E}_{\boldsymbol{\mu}_2}[\mathbf{X}] = \boldsymbol{\theta}_2^* \right\} \ , \ , \\
\mathcal{A}_3 &= \left\{ \boldsymbol{\mu}_3 \in \mathcal{P}(\prod_{i=2651}^{2701} [a_i, b_i]) \mid \mathbb{E}_{\boldsymbol{\mu}_3}[\mathbf{X}] = \boldsymbol{\theta}_3^* \right\} \ , \ .
\end{aligned}$$

Notice, that even thought it seems that only one constraint is enforced, the OUQ framework allows only for scalar *generalized* moment constraints, by setting the expectation of some measurable function $\varphi : \mathbb{R}^d \to \mathbb{R}$. So that the constraint $\mathbb{E}_{\boldsymbol{\mu}_i}[\mathbf{X}] = \boldsymbol{\theta}_i^*$ equivalently writes $\mathbb{E}_{\boldsymbol{\mu}_i}[X_j] = \theta_{i,j}^*$ for $1 \leq j \leq \dim(\boldsymbol{\mu}_i)$, where X_j and $\theta_{i,j}^*$ are the

jth components of $\mathbf{X}$ and $\boldsymbol{\theta}_i^*$. Therefore, one joint constraint accounts for respectively $100, 2550$, and 51 scalar constraints on the measures $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$, and $\boldsymbol{\mu}_3$.

Applying the reduction Theorem 3.5 it holds that the extreme points of the moment class $\mathcal{A}_i$ are convex combinations of at most $101, 2551$, and 52 Dirac masses. In order to compute the predictive distribution mean in Eq. (7.5), the multilayer perceptron has to be run $101 \times 2551 \times 52 \approx 13.3 * 10^6$ times. This is much less than considering the parameters independent and this can be run on a regular computer.

However, this block modeling faces new numerical issues. Indeed, as the extreme measures are supported on $101, 2551$, and 52 Dirac masses. It implies computing the roots of polynomials of degree $101, 2551$, and 52, and recovering the weights from the roots using Eq. (4.21). Clearly, it is computationally intractable. A solution could consist in considering that the measure $\boldsymbol{\mu}_i$ is supported on exactly n_i points, with n_i small (for instance 2 or 3), and to run the optimization with those values. Indeed, we know that the optimal solution is a discrete measure supported on *at most* $101, 2551$, and 52 Dirac masses. So that, as far as possible one should seek the optimal solution with small support cardinality. The process is repeated by incrementally increasing n_i by 1 up to respectively $101, 2551$, and 52, for $1 \leq i \leq 3$. As n_i represents the cardinal of the support of measure $\boldsymbol{\mu}_i$. It is expected that the optimum of the QoI can be obtained for measures supported on a small amount of Dirac masses. Indeed, a measure supported on 101 Dirac masses or more spreads the weights "close" to a continuous distribution, which is not necessarily efficient. We recall that OUQ gives bounds in QoI by specifically significant masses on the most penalizing input values.

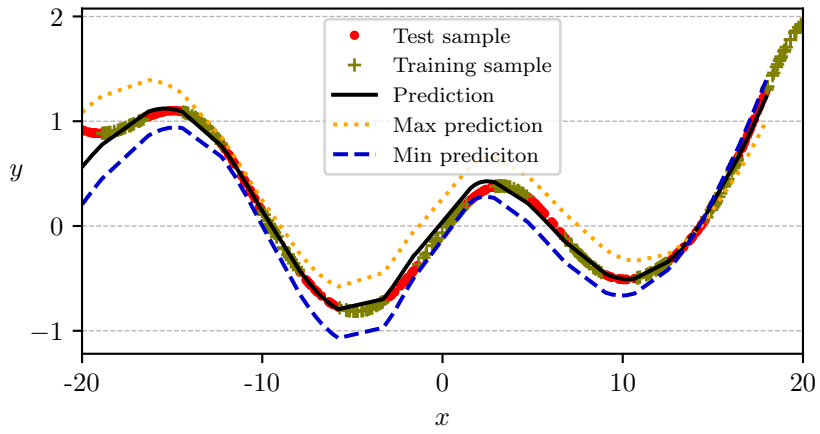


Figure 7.4: We maximize and minimize the difference between the mean posterior prediction and the true value over the moment class. The optimal parameters obtained yield the Max prediction and Min prediction curves

Finally, we illustrate the kind of results one can obtain in Fig. 7.4. Here, each input is assumed supported by at most 3 Dirac masses. In addition, we are looking for bounds on the predictive mean of every new test case [see Eq. (7.5)]. In this example, there are 150 test cases, which means we have to perform the optimization 150 times in order to find the maximal and minimal predictive mean over the measure space. In order to circumvent this issue, we gather all the optimization into a single value, by maximizing and minimizing the signed difference between the mean prediction and the true value. The parameters obtained yield the max and min lines of Fig. 7.4.

7.1.4 Conclusion

This section has highlighted that the OUQ framework is not adapted for high dimension. We recall that whenever each input distribution is discrete, the computation of a PoF is subject to the curse of dimension. Indeed, it requires evaluating the computer model on a grid that is exponentially increasing with the dimension. Therefore, even for simple neural network containing 'only' 2701 parameters (more complex networks can have millions) computation of the PoF is not tractable. We have shown that the grid size reduces when the parameters are considered dependent by blocks. However, one moment constraint on a joint distribution has to be read as several scalar moment constraints in equal number to the dimension of the distribution. Therefore, the extreme points of the multivariate moment class are discrete measures supported by a high number of support points. The difficulty then arises from the canonical moments parameterization as these support points correspond to the roots of polynomials which are complicated to compute. We propose an illustration that gives a rather pessimistic overview of this kind of optimization we have to deal with, proving that our current methodology is not well-suited for high dimensional problem.

In the following sections, we investigate new leads to account for dependency in the input variables. The next section, whose methodology completely differs from the theory exposed previously in this manuscript, offers some attractive features.

7.2 Conservative Optimization Procedure

The optimization of the QoI as proposed in Chapters 3, 4 and 5 relies on the DE algorithm which is a stochastic global optimizer. Whatever the performance of this algorithm, there is no guarantee that the true optimum has been reached. Global optimization is currently based more on heuristic arguments than theoretical asymptotic proofs of convergence. Some algorithm such as simulated annealing do possess theoretical proofs of convergence but exhibit poor performance in practice due to slow converging rates.

In this section, we propose to solve the optimization of a QoI over a measure space using Lasserre hierarchy of moments (Lasserre, 2010). This method differs from all the previous chapters and, as surprising as it sounds, does not depend on the reduction theorem presented in Chapter 3. Similar algorithmic methods based on semi definite programming (SDP) have been explored by Insua and Ruggeri (2000) in the context of robust Bayesian analysis, or in Popescu (2005); Vandenberghe et al. (2007) for solving concentration inequalities. In a more general setting, moment problems were early addressed using the duality approach in Isii (1962). Subsequently, Shapiro (2001) proved strong duality using conic linear programming and finite dimension reduction results. Modern computational approaches Lasserre (2002); Bertsimas et al. (2000) proposed an SDP formulation of the moment problem under the assumption that the moment functions are polynomial representable.

We restrict the QoI to our main application that is the computation of an optimal quantile, or equivalently using Theorem 5.3 to the computation of a minimal PoF over a measure space. Hence, in the following we consider as in the previous chapters a computer model G, and the moment class $\mathcal{A}^{ind}$ with *classical* moment constraint equalities

$$\mathcal{A}^{ind} = \left\{ \mu \in \prod_{i=1}^{d} \mathcal{P}(\mathcal{X}_i) \mid \mathbb{E}_{\mu_i}[X^j] = c_i^{(j)} \text{ for } 1 \leq j \leq N_i \text{ and } 1 \leq i \leq d \right\} . \tag{7.7}$$

We recall that a measure in $\prod_{i=1}^{d} \mathcal{P}(\mathcal{X}_i)$ has mutual independent components due to the product structure of the space. In an OUQ framework, one aims at optimizing the PoF at threshold $h \in \mathbb{R}$ over the measure space $\mathcal{A}^{ind}$: $\inf_{\mu \in \mathcal{A}^{ind}} \mathbb{P}_\mu(G(\mathbf{X}) \leq h) = \inf_{\mu \in \mathcal{A}^{ind}} F_\mu(h)$.

The moment class $\mathcal{A}^{ind}$ serves as reference for the optimization but is not the one under study in the following. As stated, the product of measure spaces $\prod_{i=1}^{d} \mathcal{P}([a_i, b_i])$ implies the independence of the input variables. However, in the next sections we focus on a different measure space. Indeed, we study the larger space $\mathcal{P}(\prod_{i=1}^{d}[a_i, b_i])$ that includes $\prod_{i=1}^{d} \mathcal{P}([a_i, b_i])$. Hence, any joint distribution with dependent input variables also belongs to this space. However, when optimizing the QoI over the space $\mathcal{P}(\prod_{i=1}^{d}[a_i, b_i])$, we will show in Section 7.2.1 that it is too large for practical applications. Indeed, the bounds obtained on the QoI appear to be overconservative. Therefore, in Section 7.2.3, we explore a strategy in order to reduce the size of the optimization space and restrain the input variables dependence.

7.2.1 Generalized Moment Problem

This theory is based on the Lasserre hierarchy of moments (Lasserre, 2010) as we recall now. Let ϕ, and h_j be multivariate polynomials, that are elements of the multivariate polynomial ring at d variables $\mathbb{R}[X_1, \ldots, X_d]$. Let also Γ be a set of indices, $\{\gamma_j \ : \ j \in \Gamma\}$ a set of real numbers, and $\mathbb{K}$ a semi-algebraic set, that is,

$$\mathbb{K} = \left\{ \mathbf{x} \in \mathbb{R}^d \ : \ g_i(\mathbf{x}) \geq 0, \ i = 1, \ldots, N \right\} \ , \tag{7.8}$$

where g_i is a multivariate polynomial for $i = 1, \ldots, N$. We define hereunder a generalized moment problem (GMP) as

$$\begin{aligned} \rho_{mom} = \sup_{\mu \in \mathcal{M}_+(\mathbb{K})} & \int_{\mathbb{K}} \phi \, d\mu \ , \\ \text{s.t.} \quad & \int_{\mathbb{K}} h_j \, d\mu \leq \gamma_j, \ j \in \Gamma \ . \end{aligned} \tag{7.9}$$

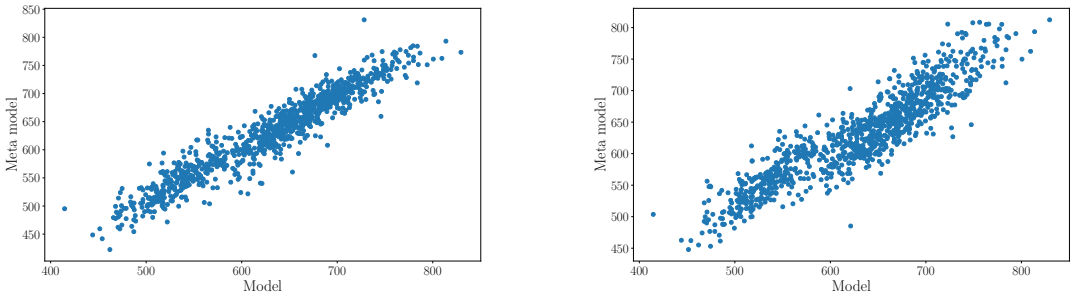
We highlight that almost all functions considered in this problem are polynomials. As restrictive as it sounds, we recall that from the Stone-Weierstrass theorem polynomials are dense under the uniform norm in the space of continuous functions on a compact Hausdorff space, (Feller, 1971, p. 223), so that GMP can be reasonably adapted to a large amount of problems. Lasserre (2010) provides an algorithm to solve a GMP based on semi-definite programming. The purpose of this section is therefore to transform our optimization problem, $\inf_{\mu \in \mathcal{A}} F_\mu(h)$, into a Generalized Moment Problem. We denote G the computer model so that for a fixed threshold h, we have

$$\begin{aligned} \inf_{\mu \in \mathcal{A}^{ind}} F_\mu(h) &= \inf_{\mu \in \mathcal{A}^{ind}} \mathbb{P}(G(\mathbf{X}) < h) \ , \\ &= \inf_{\mu \in \mathcal{A}^{ind}} \mathbb{P}(\mathbf{X} \in \mathcal{S}) \quad \text{where } \mathcal{S}^c = \left\{ \mathbf{x} \in \mathbb{R}^d \mid h - G(\mathbf{x}) \geq 0 \right\} \ , \\ &= 1 - \sup_{\mu \in \mathcal{A}^{ind}} \mathbb{P}(\mathbf{X} \in \mathcal{S}^c) \ . \end{aligned} \tag{7.10}$$

A careful reader should notice the strict inequality in the definition of the CDF, $\mathbb{P}(G(\mathbf{X}) < h)$. This strict inequality is defined so that the set $\mathcal{S}^c$ is closed. This is a necessary assumption to perform the optimization. This does not affect the final result as the corresponding quantile, which is related to the PoF, is identical for both

inequality and strict inequality. Moreover, the problem is rewritten to correspond to a maximum search. Indeed, we later explain why one cannot directly handle a minimum search in a GMP. One can also remark that in order for $\mathcal{S}^c = \{X \in \mathbb{R}^p \mid G(X) \geq h\}$ to be a semi algebraic set [Eq. (7.8)], it is mandatory for the code G to be polynomial.

We address this issue using a polynomial regressor. However, because it is not the main topic of this thesis we only present the methodology employed. In Table 2.3 was introduced the 9 PII of the computer model CATHARE. We also have at our disposal 1000 simulations of the code to create the surrogate model as introduced in Section 2.2. Instead of a Gaussian process metamodel, here, we propose to make use of a polynomial regressor known as polynomial chaos expansion to replace the computer model. We use the work of Blatman (2009) implemented in the OpenTURNS package (Baudin et al., 2017) available for C++/Python. The methodology relies on a least angle regression strategy (LARS) strategy and a corrected leave-one-out analysis in order to select the best degree polynomial with most important monomials. For the computer model CATHARE reduced to the 9 most influential inputs, the optimal polynomial has total degree 3 and writes as the linear combination of 53 monomials. The predictivity coefficient is $Q^2 = 0.84$ (Le Gratiet et al., 2017). The results are presented in Figure 7.5. The predictivity coefficient is rather low, but the code CATHARE is quite challenging to assess. For that reason, we intentionally used the Gaussian process regressor in the previous chapters because it had a better predictive power.



(a) Kriging metamodel, $Q^2 = 0.92$. (b) Polynomial metamodel, $Q^2 = 0.83$.

Figure 7.5: Comparison between the model and the surrogate model prediction over 1000 realizations.

The optimization space $\mathcal{A}^{ind}$ is now rewritten to fit the formalism of the GMP framework. First, we define the input space:

$$\mathbb{K} = \bigotimes_{k=1}^{d} [a_k, b_k] \ .$$

Clearly, the set of all probability measures on $\mathbb{K}$ includes distributions with mutually independent components. This property reads

$$\bigotimes_{k=1}^{d} \mathcal{P}([a_k, b_k]) \subset \mathcal{P}(\mathbb{K}) \ , \tag{7.11}$$

and the measure in $\mathcal{P}(\mathbb{K})$ represent all the input distributions regardless of marginal

dependency. Hence, the assertion $\mu \in \mathcal{A}^{ind}$ implies

$$\mu \in \mathcal{A}^{ind} \Longrightarrow \ \mu \in \mathcal{M}_+(\mathbb{K}) \quad \text{s.t.} \quad \int_{\mathbb{K}} \mathbf{x}^{\alpha_i^{(j)}} \, \mathrm{d}\mu = c_i^{(j)}, \quad \alpha_i^{(j)} = j \times (\delta_{1,i}, \ldots, \delta_{d,i}), \text{for } 1 \leq j \leq N_i \text{ , and } 1 \leq i \leq d \text{ ,} \quad \int_{\mathbb{K}} \mathrm{d}\mu = 1 \text{ ,} \tag{7.12}$$

where $\mathcal{M}_+(\mathbb{K})$ denotes the set of positive measure on $\mathcal{K}$ and $\{\alpha_i^{(j)}\}_{(j,i)}$ are d-dimensional tuples called multi-index. The notation $\mathbf{x}^\alpha$ refers to the monomial of multidegree α, that is $\mathbf{x}^\alpha := x_1^{\alpha_1} ... x_d^{\alpha_d}$. The set $\mathbb{K}$ is a closed semi-algebraic set, indeed, it is more explicit under the following expression which fits the definition in Eq. (7.8)

$$\mathbb{K} = \left\{ \mathbf{x} \in \mathbb{R}^d \mid g_i(\mathbf{x}) \geq 0, \ g_i(\mathbf{x}) = (x_i - a_i)(b_i - x_i), \ i = 1, \ldots, d \right\} \ . \tag{7.13}$$

We may now define the space of optimization in the GMP framework:

$$\mathcal{A}^{GMP} = \left\{ \mu \in \mathcal{P}(\mathbb{K}) \mid \mathbb{E}_\mu[\mathbf{X}^{\alpha_i^{(j)}}] = c_i^{(j)}, \ j = 1, \ldots, N_i \text{ and } i = 1, \ldots, d \right\}. \tag{7.14}$$

The constraints in Eq. (7.14) are identical to the constraint in Eq. (7.7) except that they are written in a d-dimensional framework. However, the main difference lies in the independence of the input variables that is not ensured in the GMP optimization space $\mathcal{A}^{GMP}$. Because $\mathcal{A}^{ind} \subset \mathcal{A}^{GMP}$, optimizing the QoI over $\mathcal{A}^{GMP}$ leads to explore a much bigger space than $\mathcal{A}$, and will likely return overlarge bounds.

The OUQ problem over the space $\mathcal{A}^{GMP}$ consists in searching the quantity $1 - \rho_{mom}$ with

$$\rho_{mom} = \sup_{\mu \in \mathcal{P}(\mathbb{K})} \int_{\mathbb{K}} \mathbb{1}_{\mathcal{S}^c} \, \mathrm{d}\mu \ , \quad \text{s.t.} \quad \int_{\mathbb{K}} \mathbf{x}^{\alpha_i^{(j)}} \, \mathrm{d}\mu = c_i^{(j)} \ , j = 0, \ldots, N_i \text{ , and } i = 1, \ldots, d \text{ ,} \tag{7.15}$$

where we write $c_i^{(0)} = 1$ to enforce any positive measure in $\mathcal{M}_+(\mathbb{K})$ to be a probability measure. Notice that Eq. (7.15) does not fit exactly the expression of a GMP defined in Eq. (7.9), as the indicator function in the integrand is not a polynomial. In order to fit the requirements of a GMP, we need to decompose $\mu \in \mathcal{M}_+(\mathbb{K})$ into a sum $\kappa + \nu$, with $\kappa, \nu \in \mathcal{M}_+(\mathbb{R}^d), \kappa$ supported on $\mathcal{S}^c$ and ν supported on $\mathbb{K}$. This is the strategy proposed in (Lasserre, 2010, Chapter 7). So we consider instead the following GMP:

$$\rho_{mom} = \sup_{\substack{\nu \in \mathcal{M}_+(\mathbb{K}) \\ \kappa \in \mathcal{M}_+(\mathcal{S}^c)}} \int_{\mathcal{S}^c} \mathrm{d}\kappa \ , \quad \text{s.t.} \quad \int_{\mathbb{K}} \mathbf{x}^{\alpha_i^{(j)}} \, \mathrm{d}\nu + \int_{\mathcal{S}^c} \mathbf{x}^{\alpha_i^{(j)}} \, \mathrm{d}\kappa = c_i^{(j)} \ , j = 0, \ldots, N_i \text{ , and } i = 1, \ldots, d \text{ .} \tag{7.16}$$

The multi-measures moment problem in Eq. (7.16) is equivalent to Eq. (7.15). Indeed, consider an arbitrary feasible solution κ, ν such that $\nu(\mathcal{S}^c) > 0$, and let ν_1 and ν_2 be the restriction of ν to $\mathcal{S}^c$ and $\mathbb{K} \backslash \mathcal{S}^c$ respectively. Then, (κ', ν') with $\kappa' := \kappa + \nu_1$ and $\nu' := \nu_2$ are also a feasible solution with value $\kappa'(\mathcal{S}^c) = \kappa(\mathcal{S}^c) + \nu_1(\mathcal{S}^c) \geq \kappa(\mathcal{S})$. Therefore, as we maximize $\mu(\mathcal{S}^c)$ there is no need to impose that ν is supported on $\mathbb{K} \backslash \mathcal{S}^c$. Notice that this justifies why our problem of minimizing a PoF has been reformulated into a maximization problem.

7.2.2 Semi-Definite Programming

Now, as our OUQ optimization problem has been formulated as a GMP, it is time to write the problem in Eq. (7.16) into a semi-definite program. Some definitions and notations are first prerequisite. Let $\mathbf{y} = (y_\alpha)_{\alpha\in\mathbb{N}^d}$ be a sequence of real numbers indexed by the canonical basis of $\mathbb{R}[X_1,\dots,X_d]$ given by

$$v_\infty(X) = \{\mathbf{X}^\alpha \,|\, \alpha \in \mathbb{N}^d\} = \{1, x_1, ..., x_d, x_1^2, x_1x_2, ..., x_1x_d, x_2^2, x_2x_3, ...\} \ .$$

We define the linear functional $L_\mathbf{y} : \mathbb{R}[X_1,\dots,X_d] \to \mathbb{R}$ with :

$$g \mapsto L_y(g) := \sum_{\alpha\in\mathbb{N}^d} g_\alpha y_\alpha$$

where g_α are the coefficients of the polynomial g written in the canonical basis.

Definition 5 (Moment matrix)**.** We define the moment matrix corresponding to a sequence $\mathbf{y} = (y_\alpha)_{\alpha\in\mathbb{N}^d}$, whose columns and lines are indexed by $v_r(X)$ (the truncated canonical basis of $\mathbb{R}[X_1,\dots,X_d]$ representing the polynomials up to degree r) with

$$M_r(y)(\alpha,\beta) := L_y(\mathbf{X}^\alpha\mathbf{X}^\beta) = y_{\alpha+\beta}, \quad \forall\alpha,\beta\in\mathbb{N}^d \text{ with } |\alpha|,|\beta| \le r \ .$$

This matrix size is $s(r) := \binom{d+r}{d}$, for instance let $d = r = 2$, then $M_2(\mathbf{y})$ is written

$$\begin{array}{c} \\ (00) \\ (10) \\ (01) \\ (20) \\ (11) \\ (02) \end{array} \begin{array}{c} \begin{array}{cccccc} (00) & (10) & (01) & (20) & (11) & (02) \end{array} \\ \begin{bmatrix} y_{00} & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{30} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{bmatrix} \end{array}$$

The moment matrix has the following remarkable property:

Proposition 7.1. *Given a polynomial $u \in \mathbb{R}[X_1,\dots,X_d]$ and its coordinate vector $\boldsymbol{u}$ in the canonical basis. We have*

$$\langle \boldsymbol{u}, M_r(y)\boldsymbol{u}\rangle = L_y(u^2) \ .$$

*Moreover if the sequence $\mathbf{y}$ has a corresponding measure μ (*i.e. *whose moments are the sequence $\mathbf{y}$), then $M_r(y) \succeq 0$. Indeed:*

$$\langle \boldsymbol{u}, M_r(y)\boldsymbol{u}\rangle = L_y(u^2) = \int_{\mathbb{R}^d} u(\mathbf{X})^2 \,\mu(\mathrm{d}\mathbf{X}) \ \ge 0.$$

Remark. Usually $M_r(y) \succeq 0$ does not imply that the sequence y is associated to any probability measure. This is some kind of multidimensional Hausdorff moment problem. Hence, some hypothesis on the compactness of the working space are necessary to ensure that the sequence $\mathbf{y}$ is indeed the moments sequence of a measure μ (Lasserre, 2010), as we will see in the following.

Definition 6 (Localizing matrix)**.** Given $g \in \mathbb{R}[X_1,\dots,X_d]$ a polynomial and $\mathbf{y} = (y_\alpha)_{\alpha\in\mathbb{N}^d}$ a sequence of real numbers. The localizing matrix $M_r(gy)$ corresponding to g and y, whose columns and lines are indexed by $v_r(X)$, is defined by:

$$M_r(gy)(\alpha,\beta) \ := \ L_y(g(\mathbf{X})\mathbf{X}^\alpha\mathbf{X}^\beta) = \sum_{\gamma\in\mathbb{N}^d} g_\gamma y_{\gamma+\alpha+\beta} \ ,$$

for all $\alpha,\beta\in\mathbb{N}^d$, with $|\alpha|,|\beta| \le r$.

The localizing matrix has a property equivalent to the moment matrix, which reads:

Proposition 7.2. *Given a polynomial* $u \in \mathbb{R}_r[X_1, \ldots, X_d]$ *and its coordinate vector* $\boldsymbol{u} \in \mathbb{R}^{s(r)}$ *in the canonical basis. We have*

$$\langle \boldsymbol{u}, M_r(gy)\boldsymbol{u} \rangle = L_y(gu^2)$$

If $\mathbf{y}$ *has a corresponding measure* μ *supported on the closed semi algebraic set* $\{\mathbf{x} \in \mathbb{R}^d : g(\mathbf{x}) \geq 0\}$ *then* $M_r(gy) \succeq 0$. *Indeed:*

$$\langle \boldsymbol{u}, M_r(gy)\boldsymbol{u} \rangle = L_y(gu^2) = \int_{\mathbb{R}^d} g(\mathbf{X})u(\mathbf{X})^2 \, \mu(\mathrm{d}\mathbf{X}) \; \geq 0.$$

With all these definitions, we can now recast the optimization problem in Eq. (7.16) as a semi-definite program. Let us denote

$$\begin{aligned} &\mathbf{y}_\alpha = \int_{\mathcal{S}^c} \mathbf{x}^\alpha \, \mathrm{d}\kappa \;\text{ and }\; \mathbf{z}_\alpha = \int_{\mathbb{K}} \mathbf{x}^\alpha \, \mathrm{d}\nu \;, \\ &2v \text{ or } 2v-1 \text{ the degree of polynomial } G - h \;, \\ &2w_j \text{ or } 2w_j - 1 \text{ the degrees of } \{g_j\}_j \text{ in Eq. (7.13)} \;, \end{aligned}$$

so that the generalized moment problem in Eq. (7.16) may be relaxed in the sequence of semi-definite optimization problems

$$\begin{aligned} \rho_n = &\sup_{\mathbf{y},\mathbf{z}} y_0 \\ &s.t.\; y_{\alpha_i^{(j)}} + z_{\alpha_i^{(j)}} = c_i^{(j)}, \text{ for } 0 \leq j \leq N_i, \text{ and } 1 \leq i \leq d \;, \\ &\quad M_n(\mathbf{y})M \succeq 00 \;;\; M_n(\mathbf{z}) \succeq 0 \;, \\ &\quad M_{n-v}((h-G)\mathbf{y}) \succeq 0 \;, \\ &\quad M_{n-w_j}(g_j\mathbf{z}) \succeq 0 \;. \end{aligned} \tag{7.17}$$

The second equation corresponds to the moment constraints, while the three last equations involving the moment and localization matrix try to ensure that the sequences $\mathbf{y}$ and $\mathbf{z}$ are the moment sequences of two measures whose supports are contained in $\mathcal{M}_+(\mathcal{S}^c)$ and $\mathcal{M}_+(\mathbb{K})$, respectively. Moreover, the semi-definite optimization problem Eq. (7.17) corresponds to a relaxation of the GMP in Eq. (7.16), with relaxation degree denoted n. This implies, using Eq. (7.10) and Eq. (7.15), that

$$\rho_n \geq \rho_{mom} \geq \sup_{\mu \in \mathcal{A}^{ind}} \mathbb{P}(\mathbf{X} \in \mathcal{S}^c) \;,$$

and

$$1 - \rho_n \leq 1 - \rho_{mom} \leq \inf_{\mu \in \mathcal{A}^{ind}} F_\mu(h) \;.$$

So that the solution provided by the semi-definite program is actually a deterministic lower bound of the PoF. This method hence completes the stochastic global optimization in Chapters 3 and 4, which yields an upper bound on the true optimum. In this semi-definite problem, the more we increase the degree of relaxation n, the more we add constraints in the semi-definite optimization problem. An immediate consequence is that $(\rho_n)_{n \in \mathbb{N}}$ is a decreasing sequence. Furthermore, Lasserre (2010) gives a criteria to evaluate whether convergence to ρ_{mom} in Eq. (7.16) is reached. The following theorem holds.

Theorem 7.3 ((Lasserre, 2010, p.165)). *Let ρ^* be the optimal value of the semi-definite program (7.17). Then:*

1. *For every $n \geq v, \rho_n \geq \rho_{mom}$ and moreover, $\rho_n \xrightarrow{n} \rho^* \geq \rho_{mom}$*
2. *If ρ_n is attained at an optimal solution $(\mathbf{y}, \mathbf{z})$ which satisfies*

$$\begin{cases} rank\ M_n(\mathbf{y}) &= rank\ M_{n-v}(\mathbf{y}) \\ rank\ M_n(\mathbf{z}) &= rank\ M_{n-1}(\mathbf{z}) \end{cases} \tag{7.18}$$

then $\rho_n = \rho_{mom}$.

Theorem 7.3 states that a rank criterion warrants the convergence of the semi-definite program (7.17), in a finite number of iteration. In practice, convergence is often reached for a low value of n, typically 3 or 4. We used the toolbox *Gloptipoly3* for Matlab developed in Henrion et al. (2009) in order to implement the algorithm.

The semi-definite program (7.17) minimizes the probability $\mathbb{P}_\mu(\mathbf{X} \in \mathcal{S}^c)$ for $\mu \in \mathcal{A}^{GMP}$. We recall that the working space $\mathcal{A}^{GMP}$ differs from the initial space $\mathcal{A}^{ind}$ by the dependence of the input variables. This space is much larger than the space $\mathcal{A}^{ind}$, in practice the minimum is so low that it is no longer interpretable.

In the following, we propose a lead to circumvent this issue. In order to reduce the optimization space size, we add artificial moment constraints to decorrelate the input through their moments.

7.2.3 SDP Limitation

One wishes to reduce the size of the optimization space $\mathcal{A}_{GMP}$ in order to bring it closer in a sense to the space $\mathcal{A}^{ind}$, where the marginal variables are independent. This optimization space is defined with moment constraints, and the constraint number drives the size of $\mathcal{A}_{GMP}$. The idea consists in reducing the size of the set $\mathcal{A}_{GMP}$ by artificially reproducing some characteristics of mutually independent random variables. For this, we add constraints to the marginal cross moments.

Let $\mathbf{y}$ be the sequence of moments associated to a measure $\mu \in \mathcal{M}_+(\mathbb{K})$. We recall that N_i constraints are initially enforced on every input μ_i, such that

$$\mathbb{E}[X_i^j] = \int_{\mathbb{K}} \mathbf{x}^{\alpha_i^{(j)}}\, \mathrm{d}\mu = c_i^{(j)}, \text{ with } \alpha_i^{(j)} = j \times (\delta_{1,i}, \ldots, \delta_{d,i}) \ ,$$

which is equivalent to the notation $y_{\alpha_i^{(j)}} = c_i^{(j)}$. Suppose that all inputs are mutually independent, then it holds that

$$\mathbb{E}[X_1^{j_1} \ldots X_d^{j_d}] = \mathbb{E}[X_1^{j_1}] \ldots \mathbb{E}[X_d^{j_d}] \ , \ j_k = 0, \ldots, N_k \ , \tag{7.19}$$

which is similar to writing that: $y_{(j_1,\ldots,j_d)} = c_{j_1}^{(1)} \ldots c_{j_d}^{(d)}$. Clearly, Eq. (7.19) does not ensure that the variables are independent. In all generality, there is no equivalence between independence and decorrelation of finite order. However, this gives the possibility to add $\prod_{i=1}^d (N_i+1) - 1 - \sum_{i=1}^d N_i$ constraints in the semi-definite program (7.17). By doing so, one can reduce the size of the space $\mathcal{A}_{GMP}$ and ensure the variables are at least decorrelated up to some power. One can also discuss the convergence of such a method. Let the number of constraint N_i, goes to $+\infty$. Then, it has already been stated that as soon as a measure μ_i is supported on a bounded interval, it is uniquely

defined by its moment sequence, so that enforcing all its moments ensures that it is independence from all the other inputs.

We now present and discuss some results. The implementation of the semi-definite program Eq. (7.17) was done in Matlab and greatly simplified thanks to the work of Henrion et al. (2009) using the toolbox *Gloptipoly3*. We explore the strategy that consists in adding new constraints to the moments of the measure $\mu \in \mathcal{M}_+(\mathbb{K})$, in order to simulate the behavior of independent variables and reduce the size of the space $\mathcal{A}^{GMP}$. We evaluate the polynomial metamodel restricted to $d = 4$ variables and we enforce the mean and the variance of each variables. This involves that, to the initially 8 constraints enforced, we can add up to 72 new cross covariance constraints to the space $\mathcal{A}^{GMP}$. In Figure 7.6, we present the results by adding constraints classified by their order; a constraint is said to be of order n if $\sum j_i = n$ in expression $\mathbb{E}[X_1^{j_1}...X_4^{j_4}]$, $0 \leq j_i \leq 2$. It appears that the constraints of maximal order do not have much influence on

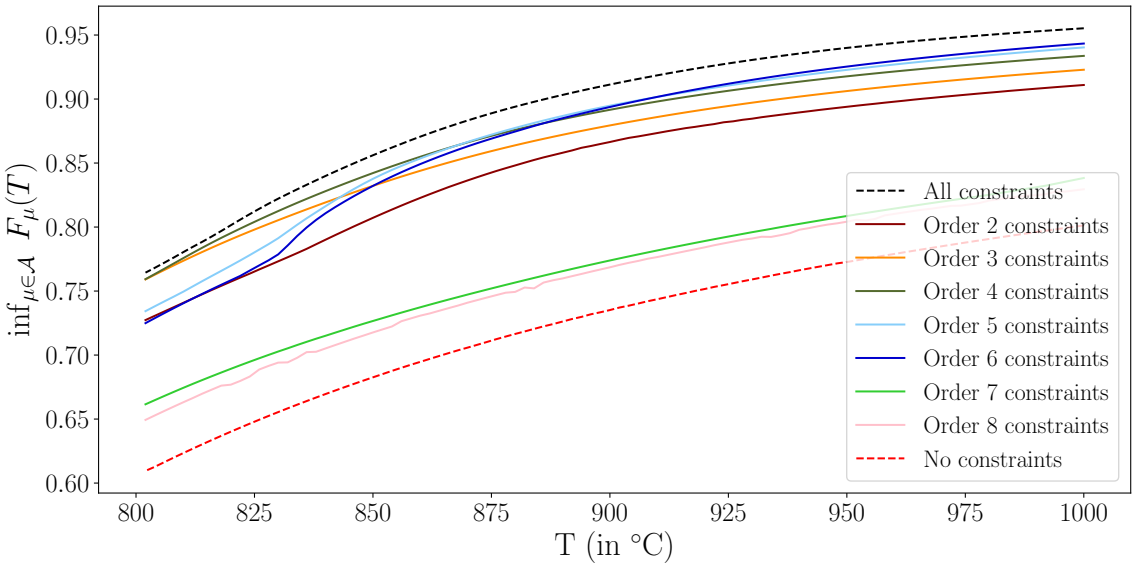


Figure 7.6: Adding new constraints reduces the size of the space of optimization $\mathcal{A}^{GMP}$ thus increasing the objective value. We use a polynomial chaos expansion of CATHARE restricted to the 4th most influential inputs only, with two constraints on the mean and the second moment.

the result, partially because there are less numerous. The addition of all the constraint seems very promising as the CDF lower envelop substantially increases, proving that the space $\mathcal{A}^{GMP}$ have been reduced. For instance the associated robust quantile of order 0.8 worth approximately $1000°C$ without any additional constraint and $820°C$ when adding the 72 cross moment constraints. Constraints of order 7 or 8 seem to have low influence, but we recall that there is only four constraints of order 7 and one constraint of order 8. The number of constraint is different depending of their order, so that it is advised to quantify the quantile taking into account every additional constraints.

Because this strategy seems to perform well, we compare in Fig. 7.7 the result of the SDP optimization with the OUQ framework developed in the previous chapters. To do so, we restricted the uncertain computer model to the six most influential inputs, as in Section 6.2, and we only enforced a single mean constraint on each input distribution, following Table 6.1. The optimization with SDP was run over $\mathcal{A}_{GMP}$ with 57 additional cross moment constraints to decorrelate each marginal. On the other hand, the DE optimization was run over $\mathcal{A}^{ind}$ so that the input variables are independent. One can

notice in Fig. 7.7 that the quantiles are substantially higher than in Chapter 6 with only one constraint. For example, quantile of order 0.9 is equal to 1300°C in Figure 7.7 with one constraint on the mean, compared to 875°C in Figure 7.6 with one additional constraint on the variance. This is natural as the more constraints we add, the smaller the optimization space becomes. The reason for enforcing only one constraint is that otherwise the problem becomes too complex to handle for the *Gloptipoly3* toolbox and does not converge.

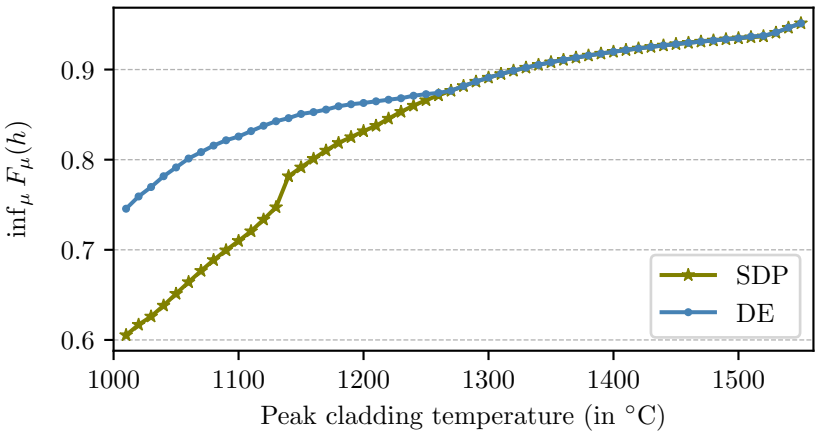


Figure 7.7: Comparison between the semi-definite program and the canonical parameterization solved with two different solvers. The metamodel of the code CATHARE is restricted to 6 variables, and the only constraints enforced are on the mean of each input. No other constraint can be enforced or else the SDP computation fails.

The results of DE solver and SDP are close for high quantiles (≥ 0.88). We highlight that this ensures that the true optimum is reached. Indeed the SDP systematically underestimates the true optimum (Theorem 7.3), while the stochastic optimization with DE algorithm overestimates the optimum. Thus, as long as SDP and stochastic optimization return the same value, the global optimum is known with absolute certainty (up to unavoidable numerical errors). It also means that the addition of constraints in $\mathcal{A}^{GMP}$ is promising, as it narrows the space close to $\mathcal{A}^{ind}$. This solution does not work so well for lower quantiles, the SDP completely underestimates the optimum compared to the DE solver. It appears that the lower the threshold h, the bigger the space $\mathcal{S}^c = \left\{\mathbf{X} \in \mathbb{R}^d \mid G(\mathbf{X}) \leq h\right\}$, so that even with the addition of the cross moment constraints one cannot warrant the mutual independence of the input variables.

7.2.4 Limitations and Conclusions

The objective of this section was to explore an alternative formulation to our optimization problem, seen as a generalized moment problem. This is convenient as GMP can be relaxed then solved using semi-definite programming. While this method is appealing, it also requires the computer model to be well approximated by a polynomial, which can be ill-suited, in particular when the code is not smooth enough.

Lasserre (2010) provides an algorithm to extract the measure solution when the convergence has been reached, but the decomposition of an input measure into the sum of two measures in Eq. (7.16) highly reduces the stability of the algorithm. In practice, it is difficult to extract the solution except on toy examples in low dimension (one

or two). Moreover, the complexity of the calculation grows exponentially with three parameters, the number of inputs, the maximal order of constraints, and the degree of the polynomial. Indeed the size of the moment and localizing matrices depends on those parameters and grows exponentially. While the degree of the polynomial metamodel is fixed (set to 3 for the code CATHARE), the addition of high order constraints constitutes the originality of our method proposed in Section Eq. (7.2.3). This combinatorial explosion limits the effectiveness of the approach, so that in practice one cannot work with dimensions higher than 6, or else the computation fails.

The objective value $\inf_{\mu\in\mathcal{A}} F_\mu(h)$ can be transformed into a semi-definite program, but this leads to the loss of the input mutual independence. The problem being that the space of all joint distribution is too large for the optimization of the QoI whose bounds become overconservative. We have been explored a strategy in order to restore independence. It is based on the addition of constraints to the cross moments of the measure μ. However, we quickly reached the limitations of this method as the dimension increases. This problem could have been anticipated, since the proposed algorithm is deterministic and subject to the curse of dimensionality.

Semi-definite programing is a common tool for solving moment problem (Bertsimas et al., 2000; Lasserre, 2010). One of the main benefits of this methodology is that it gives a conservative estimation of the robust quantile, which is very attractive when assessing safety margins in an engineering framework. However, despite its popularity it does not seem suited to deal with high dimensional problems.

7.3 Copula Class

In this section, we investigate a completely different way of dealing with the dependence of the model input variables. Consider as previously that μ is a d-dimensional joint measure representing the potentially dependent input parameters. It has previously been pointed out that the measure space $\mathcal{P}(\prod_{i=1}^d [a_i, b_i])$ is quite big, meaning the QoI bounds obtained over this space are overlarge. In this section, we investigate a subset where the marginal distributions of the measure are fixed. This differs from the modeling studied so far. Indeed, the purpose from Chapters 1 to 6 was to account for the uncertainty affecting precisely the marginal distribution of μ. As stated, we now consider that the marginal distributions are set. Hence, only the dependence structure between each component of the measure μ is modeled and considered uncertain.

This setting is perfectly adapted to copula theory (Durante and Sempi, 2015; Nelsen, 2006). Basically, a copula is a multivariate cumulative distribution function for which the marginal probability distribution of each variable is uniformly distributed in $[0, 1]$. Although Definition 7 is not exactly similar, Sklar's theorem states that any multivariate joint distribution can be written in terms of univariate marginal distribution functions and a copula which describes the dependence structure between the variables. Copulas are attractive as they allow to easily model and separate the marginals from the dependence structure.

Optimization of a QoI over the set of all copula has already explored in Ghosh and Bhandari (2017) and Hofer and Iacò (2014). The specific case of quantile optimization of a computer model in rather high dimension (≈ 10) was studied in Benoumechiara et al. (2018) using parametric vine copulas (Joe, 1996; Czado, 2010). There are many parametric copula families, but in this work no assumption is made on the nature of the copula. Hence, our work only differs from Benoumechiara et al. (2018) in the departure from the parametric assumption, as we consider here the set of all copulas. However,

we will show that the methodology proposed for optimizing a QoI over the class of all copulas is very limited, and that it hardly extends to input dimensions greater than 3. Nevertheless, this section presents some interesting topological properties of copula sets that suit the OUQ framework, so that it is possible to apply the reduction theorem to the set of all copulas. Let us first define formally a d-dimensional copula:

Definition 7. A d-dimensional copula is a function C from $[0,1]^d$ to $[0,1]$ with the following properties:

- For every $\mathbf{x}$ in $[0,1]^d$, $C(\mathbf{x}) = 0$ if at least one coordinate of $\mathbf{x}$ is 0. If all coordinates of $\mathbf{x}$ are 1 except x_k then $C(\mathbf{x}) = x_k$
- C is d-increasing. This means the C-volume, denoted V_C, of every d-dimensional rectangle is positive $V_C \geq 0$. Where the C-volume of a rectangle $[\mathbf{a}, \mathbf{b}] := \prod_{i=1}^{d} [a_i, b_i]$ is defined as

$$V_C = \sum_{\mathbf{v} \in \mathrm{ver}([\mathbf{a},\mathbf{b}])} \mathrm{sign}(\mathbf{v}) C(\mathbf{v})$$

  where

$$\mathrm{sign}(\mathbf{v}) = \begin{cases} 1, & \text{if } v_j = a_j \text{ for an even number of indices,} \\ -1, & \text{if } v_j = a_j \text{ for an odd number of indices,} \end{cases}$$

  and $\mathrm{ver}([\mathbf{a}, \mathbf{b}]) = \{a_1, b_1\} \times \cdots \times \{a_d, b_d\}$ is the set of vertices of the rectangle $[\mathbf{a}, \mathbf{b}]$.

Sklar's theorem states that every d-copula defines a measure μ on $([0,1]^d, \mathcal{B}([0,1]^d))$ which is d-fold stochastic, i.e, it fulfills for any Borel set $A \subset [0,1]$

$$\mu([0,1] \times [0,1] \times A \times [0,1] \times [0,1]) = \lambda(A).$$

Conversely every d-fold joint distribution function μ with margins $F_1, \ldots, F_d$ defines a copula C such that for all $\mathbf{x}$ in $\mathbb{R}^d$

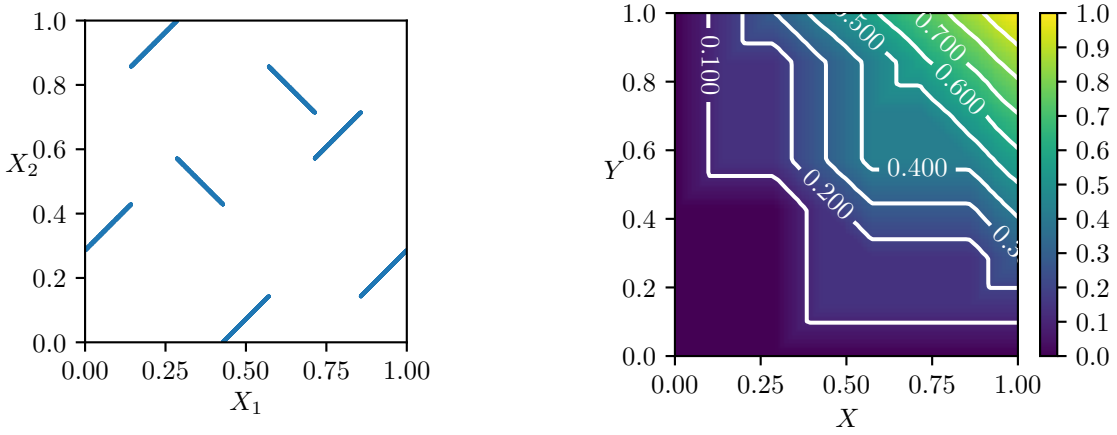
$$\mu(]-\infty, \mathbf{x}]) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{7.20}$$

So that there is a one to one correspondence between the set of d-dimensional probability measures with uniform marginal on $[0,1]$ and the set of all copulas denoted by $\mathcal{C}$. In the following, for the sake of clarity, we simply consider that all marginal distributions of the joint distribution μ are uniform in $[0,1]$. Notice that $\mathcal{C}$ is obviously convex, it is also a compact subset of the space of all continuous real-valued functions with domain in $[0,1]^d$ as demonstrated in (Durante and Sempi, 2015, Theorem 1.7.7), but the first proof of this result goes back to Brown (1965). From the Krein-Milman Theorem, it follows that the copula space $\mathcal{C}$ is the convex hull of its extreme points. However, while the extreme points of the moment class and the unimodal moment class in Chapter 3 are well characterized, the generator of $\mathcal{C}$ is quite challenging to describe. Sufficient conditions for a copula to be an extremal point of $\mathcal{C}$ are given in Ghosh and Bhandari (2017). An important statement of this article is that for any function $f : [0,1] \to [0,1]^d$ with associated graph $\mathcal{G}_f = \{(x, f(x)) : x \in [0,1]\}$, any copula supported on $\mathcal{G}_f$ is an extreme copula. Nevertheless, a full characterization of the extreme copulas of $\mathcal{C}$ is available in Štēpán (1993) and Benes and Štēpán (1991), but the characterization is really abstract did not find any application so far (and up to our knowledge).

We focus instead on a very interesting subclass of copulas denoted $\mathcal{M}$ known under the name shuffle of Min, or permutation copula (Mikusinski et al., 1992; Durante and Sánchez, 2012). In the following, the dimension of the problem is restricted to the case $d = 2$. We will later expose the situation for $d \geq 3$ and although the theory is similar, the numerical implementation does not perform well in dimension higher than two. A shuffle of Min is defined as follows Mikusinski et al. (1992):

Definition 8. A copula C is a shuffle of Min if and only if there is a positive integer n, two partitions $0 = s_0 < \cdots < s_n = 1$ and $0 = t_0 < \cdots < t_n = 1$ of $[0, 1]$ and a permutation σ on $\{1, \ldots, n\}$ such that each $[s_{i-1}, s_i] \times [t_{\sigma(i)-1}, t_{\sigma(i)}]$ is a square in which C deposits a mass of size $s_i - s_{i-1}$ spread uniformly along one of the diagonals. For each $i = 1, \ldots, n$ we let $m(i)$ denote the slope of the diagonal $[s_{i-1}, s_i] \times [t_{\sigma(i)-1}, t_{\sigma(i)}]$ along which the mass in that square is spread. We say that C is the shuffle of Min generated by $(n, \{s_i\}, \{t_i\}, \sigma, m)$

Of course any shuffle of Min is an extreme copula, as stated in Ghosh and Bhandari (2017) as its support is the graph of a real function. However, we use the more restricted version of permutation copula by enforcing the partitions to be regular partition. Hence, $[s_{i-1}, s_i] = [t_{i-1}, t_i] = [\frac{i-1}{n}, \frac{i}{n}]$ and the $[0, 1]$ interval is split according to a grid of step $1/n$. With that assumption, a regular shuffle of Min can be generated by (n, σ, m) as illustrated in Fig. 7.8. The reason for using only regular partitions is that one can always reduce a shuffle of Min to a regular one by taking the lowest common multiple of the interval size as the regular partition grid size. We also refer to Puccetti and Wang (2015) for a better understanding of how shuffle of Min copula are related to optimal transport theory (Villani, 2008). The main property of shuffle of Min states as follows:



(a) Shuffle of Min support, also represents its density

(b) Shuffle of Min distribution, with associated level set.

Figure 7.8: Representation of shuffle of Min copula generated by (n, σ, m) with $n = 7$, $\sigma = (3, 7, 4, 1, 6, 5, 2)$, and $m = (1, 1, -1, 1, -1, 1, 1)$.

Theorem 7.4 ((Nelsen, 2006, Theorem 3.2.2)). *For any $\epsilon > 0$, and for any copula C, there exists a shuffle of Min, which we denote C_ϵ, such that*

$$\sup_{x_1, x_2 \in [0,1]} |C_\epsilon(x_1, x_2) - C(x_1, x_2)| < \epsilon$$

Theorem 7.4 translates that any copula can be uniformly approximated by permutation copulas. More precisely, the proof (in the same reference) requires the partition size n to be $n \geq 4/\epsilon$, so that one must refine the partition to better approximate any copula. Theorem 7.4 also proves that the set of shuffle of Min $\mathcal{M}$ is dense in the set of all copulas $\mathcal{C}$ endowed with the sup norm. It is quite remarkable to highlight a dense subset of $\mathcal{C}$ composed exclusively of extreme points. This shows how interesting are the topological properties of the copula set $\mathcal{C}$. Because the copula set is compact convex and that we have at our disposal a characterization of its extreme points, the following corollary holds:

Corollary 7.5 (Theorem 3.3)**.** *Let $\phi : \mathcal{C} \to \mathbb{R}$ be a quasi-convex lower semicontinuous function, and let $\mathrm{ex}(\mathcal{C})$ denotes the extreme points of the copula set, then*

$$\sup_{C \in \mathcal{C}} \phi(C) = \sup_{C \in \mathrm{ex}(\mathcal{C})} \phi(C) \ .$$

By applying Corollary 7.5, it holds that a PoF on $\mathcal{C}$ finds its optimum on an extreme copula. Moreover, because any copula can be approximated by a Shuffle of Min it is enough to proceed the optimization over the set of shuffle of Min $\mathcal{M}$ which is parametric. Let $\mathcal{C}(F_1, F_2)$ denote the set of all joint distribution $H(x_1, x_2)$ with marginals $F_1(x_1), F_2(x_2)$, then for a continuous function g we have from Eq. (7.20)

$$\int_{\mathbb{R}^2} g(x_1, x_2) \, \mathrm{d}H(x_1, x_2) = \int_{[0,1]^2} g(F_1^{-1}(x_1), F_2^{-1}(x_2)) \ , \mathrm{d}C(x_1, x_2) \ ,$$

where C is a copula. Hence, we can always pretend from now that the marginals F_1, F_2 are uniform. Thus, the optimization problem consists in obtaining bounds of the form:

$$\begin{aligned} \int_{[0,1]^2} g(x_1, x_2) \, \mathrm{d}C_{min}(x_1, x_2) \leq \int_{[0,1]^2} g(x_1, x_2) \, \mathrm{d}C(x_1, x_2) \ , \\ \text{and } \int_{[0,1]^2} g(x_1, x_2) \, \mathrm{d}C(x_1, x_2) \leq \int_{[0,1]^2} g(x_1, x_2) \, \mathrm{d}C_{max}(x_1, x_2) \ . \end{aligned} \tag{7.21}$$

where C_{min} and C_{max} are copulas. Because they can be approximated by shuffle of Min the optimization problem rewrites

$$\begin{aligned} \sup_{C \in \mathcal{C}} \int g(x_1, x_2) \, \mathrm{d}C(x_1, x_2) &= \sup \left\{ \mathbb{E}[g(X)] : X \sim C, \text{ where } C \text{ is a copula } \right\} \ , \\ &= \lim_{k \to \infty} \max_{C_k \in \mathcal{M}_k} \left\{ \mathbb{E}[g(\widetilde{X})] : \widetilde{X} \sim C_k \right\} \ , \end{aligned}$$

where C_k is a shuffle of Min with partition size k, and $\mathcal{M}_k$ denotes the set of all shuffle of Mins with partition size equal to k. In the special case in which we maximize a PoF,*i.e.* $g(\mathbf{x}) = \mathbb{1}_{G(\mathbf{x}) \leq h}$ for $h \in \mathbb{R}$ that is not continuous, we replace g by a continuous approximation g_ϵ defined as

$$g_\epsilon(\mathbf{x}) = \begin{cases} 1 & , \text{ if } G(\mathbf{x}) \leq h \\ 1 - \frac{|G(\mathbf{x}) - h|}{\epsilon} & , \text{ if } |G(\mathbf{x}) - h| \leq \epsilon \\ 0 & , \text{ otherwise} \end{cases} \tag{7.22}$$

Then, it holds that:

$$\begin{aligned} \sup_{C \in \mathcal{C}} \mathbb{P}_C(G(\mathbf{X}) \leq h) &= \lim_{\epsilon \to 0} \lim_{k \to \infty} \max_{C_k \in \mathcal{M}_k} \left\{ \mathbb{E}[g(\widetilde{X})] : \widetilde{X} \sim C_k \right\} \ , \\ &= \lim_{\epsilon \to 0} \lim_{k \to \infty} \max_{C_k \in \mathcal{M}_k} \int_{[0,1]^2} g_\epsilon(x_1, x_2) \, \mathrm{d}C_k(x_1, x_2) \ . \end{aligned} \tag{7.23}$$

There exists a very efficient way of solving the optimization problem Eq. (7.21) in the two dimensional case. We present this algorithm in the following. As stated in Puccetti and Wang (2015) optimization over copula spaces is connected to optimal transport theory (Villani, 2008). Indeed, optimizing the shuffle of Min can be seen as a problem of weight placement in the partition of the pavement $[0,1]^2$. We illustrate the close relationship between the optimization problem Eq. (7.21) and the linear assignment problems of the form

$$\max_{\sigma\in\mathfrak{S}} \sum_{i=1}^{n} a_{i,\sigma(i)} \ , \tag{7.24}$$

where $\mathfrak{S}$ is the set of all permutations of $\{1,\dots,n\}$. The Hungarian algorithm due to Kuhn (1955) is known for solving this problem efficiently. For a detailed description of assignment problems and related solution algorithms we refer to Burkard et al. (2009). The optimization of an expectation over the set of copulas is easily solved whenever the function is piecewise constant as stated in the following Theorem:

Theorem 7.6 ((Hofer and Iacò, 2014, Theorem 2.1)). *Let $n \geq 1, A = \{a_{i,j}\}_{1\leq i,j\leq n}$ be a real valued $n \times n$ matrix and let the function $\widetilde{g}$ be defined as*

$$\widetilde{g}(x_1,x_2) := a_{i,j}(x_1,x_2) \in \left[\frac{i-1}{n},\frac{i}{n}\right[\times \left[\frac{j-1}{n},\frac{j}{n}\right[\ .$$

Then the copula which maximizes

$$\sup_{C\in\mathcal{C}} \int_{[0,1]^2} \widetilde{g}(x_1,x_2)\,\mathrm{d}C(x_1,x_2) \ , \tag{7.25}$$

is given as a shuffle of Min generated by $(n,\sigma^*,(1,\dots,1))$, where σ^* is the permutation which solves the assignment problem

$$\max_{\sigma\in\mathfrak{S}} \sum_{i=1}^{n} a_{i,\sigma(i)} \ .$$

Moreover, the maximal value of Eq. (7.25) is given as

$$\sup_{C\in\mathcal{C}} \int_{[0,1]^2} \widetilde{g}(x_1,x_2)\,\mathrm{d}C(x_1,x_2) = \frac{1}{n}\sum_{i=1}^{n} a_{i,\sigma^*(i)} \ . \tag{7.26}$$

Note that the maximal copula is by no means unique, since for instance the value of the integral in Eq. (7.25) is independent of the choice of the slope vector $m = (1,\dots,1)$ in the generation of the shuffle of Min. It is also possible to derive the lower bound by considering the opposite function $\widetilde{\widetilde{g}} = -\widetilde{g}$. Theorem 7.6 relates to piecewise constant functions but provides a basis for the optimization of a continuous function. Indeed, the following generalization refers to Theorem 7.6.

Theorem 7.7 ((Hofer and Iacò, 2014, Theorem 2.2)). *Let g be a continuous function on $[0,1]^2$, let the sets $I_{i,j}^n$ be given as*

$$I_{i,j}^n = \left[\frac{i-1}{2^n},\frac{i}{2^n}\right[\times \left[\frac{j-1}{2^n},\frac{j}{2^n}\right[\ \textit{for } 1\leq i,j\leq 2^n \ ,$$

for every $n > 1$ and define the functions $\underline{g}_n, \overline{g}_n$ as

$$\begin{aligned} \underline{g}_n(x_1,x_2) &= \min_{(x_1,x_2)\in I_{i,j}^n} g(x_1,x_2), \textit{ for all } (x_1,x_2)\in I_{i,j}^n \ , \\ \overline{g}_n(x_1,x_2) &= \max_{(x_1,x_2)\in I_{i,j}^n} g(x_1,x_2), \textit{ for all } (x_1,x_2)\in I_{i,j}^n \ . \end{aligned} \tag{7.27}$$

Furthermore, let $\overline{C}^n_{max}, \underline{C}^n_{max}$ *be the copulas which maximize respectively*

$$\sup_{C\in\mathcal{C}} \int_{[0,1]^2} \underline{g}_n(x_1,x_2)\ \mathrm{d}C(x_1,x_2) \quad \textit{and} \quad \sup_{C\in\mathcal{C}} \int_{[0,1]^2} \overline{g}_n(x_1,x_2)\ \mathrm{d}C(x_1,x_2)\ ,$$

then

$$\begin{aligned}\int_{[0,1]^2} \underline{g}_n(x_1,x_2)\ \mathrm{d}\underline{C}^n_{max}(x_1,x_2) &\leq \sup_{C\in\mathcal{C}} \int_{[0,1]^2} g_n(x_1,x_2)\ \mathrm{d}C(x_1,x_2) \\ &\leq \int_{[0,1]^2} \overline{g}_n(x_1,x_2)\ \mathrm{d}\overline{C}^n_{max}(x_1,x_2)\ , \qquad (7.28)\end{aligned}$$

for every n*, and*

$$\begin{aligned}\lim_{n\to\infty} \int_{[0,1]^2} \underline{g}_n(x_1,x_2)\ \mathrm{d}\underline{C}^n_{max}(x_1,x_2) &= \lim_{n\to\infty} \int_{[0,1]^2} \overline{g}_n(x_1,x_2)\ \mathrm{d}\overline{C}^n_{max}(x_1,x_2) \\ &= \sup_{C\in\mathcal{C}} \int_{[0,1]^2} g_n(x_1,x_2)\ \mathrm{d}C(x_1,x_2)\ . \qquad (7.29)\end{aligned}$$

By defining the functionial families $\{\underline{g}_n\}_n$ and $\{\overline{g}_n\}_n$ differently, we might get an approximation technique which converges faster to the optimal value, in particular as the minimization and maximization steps in Eq. (7.27) can be costly. One can use for instance the following approximation

$$g_n(x_1,x_2) = g\left(\frac{i}{2^n}, \frac{j}{2^n}\right),\ \text{for all } (x_1,x_2) \in I^n_{i,j}\ ,$$

the drawback being that we loose the upper and lower bounds of the optimal value, that are useful for numerical applications. Indeed, such bounds provide rates of convergence for the method, assuming the Lipschitz continuity of g:

Corollary 7.8 ((Hofer and Iacò, 2014, Corollary 2.1))**.** *Let the assumptions of Theorem 7.7 hold and, in addition assume that* g *is Lipschitz continuous on* $[0,1]^2$ *with parameter* L*. Then*

$$\left| \int_{[0,1]^2} \overline{g}_n(x_1,x_2)\, \mathrm{d}\overline{C}^n_{max}(x_1,x_2) - \int_{[0,1]^2} \underline{g}_n(x_1,x_2)\, \mathrm{d}\underline{C}^n_{max}(x_1,x_2) \right| \leq L\frac{\sqrt{2}}{2^n}$$

The performance of the Hungarian algorithm is quite remarkable; when $n = 10$ for instance, corresponding to a linear assignment problem related to a matrix A of size $2^{10} \times 2^{10}$, it takes few seconds to run on an ordinary computer. The main problem of this methodology is that the linear assignment problem is easily solved in dimension two but becomes NP-hard (Bovet and Crescenzi, 1994) for dimension $d \geq 3$. No equivalent of the Hungarian algorithm exist to date in higher dimensions. This is why we restricted our study to $d = 2$ from the beginning of this section. However, the shuffle of Min can be generalized to any dimension and remains dense and extremal in the set of all d-dimensional copulas. We refer to Ghosh and Bhandari (2017) for a review on that topic.

Nevertheless, being restricted to two dimensional copulas still offers promising prospects for practical applications. To see why, let us consider a d-dimensional computer model G, with joint input distribution μ having fixed marginal, and quantity of interest defined as the 95%-quantile $q_{0.95}$ of the output. One can evaluate the impact of the dependence structure with variables packed in pairs. That is, select two integers i, j, such that $0 \leq i, j \leq d$ and consider that the dependence structure between the ith and

jth variables is imprecise. Then, compute the worst case quantile $\sup_{\mathcal{C}} q_{0.95}(i,j)$ over all two dimensional copulas modeling the dependency between $X_i \sim \mu_i$ and $X_j \sim \mu_j$. It is afterwards possible to classify and detect critical bidimensional dependences which most affect the quantile. It does not mean that the detected pairs of variables are dependent. However, it helps determining which specific couples of parameters engineers should focus their attention on, since they can deeply modify the result of the analysis.

In the following, an illustration is proposed on the flood model presented in Chapter 2, the parameter distributions are fixed to their initial choice in Table 2.1. However, we now consider that the two most influential parameters Q and K_s are possibly dependent. Therefore, we intend to evaluate the maximal 95%-quantile over the set of all copula modeling the bidimensional dependence structure and see how this affects the analysis. To that end, the original computer model $H(Q, K_s, Z_v, Z_m)$ is integrated over the distribution Z_v, and Z_m so that the resulting model $\widetilde{H}$ only depends on the two parameters under study Q and K_s:

$$\widetilde{H}(Q, K_s) = \int_{[49,51]\times[54,55]} H(Q, K_s, x_3, x_4)\, \mathrm{d}\mu_{Z_v}(x_3)\mathrm{d}\mu_{Z_m}(x_4)\ . \tag{7.30}$$

Using our usual trick, we apply Theorem 5.3 so that instead of computing the maximal quantile we evaluate the minimal PoF: $\inf_{\mathcal{C}} F_\mu(h)$, over the copulas set for different threshold h. To that extend, consider the function with support in $[0,1]^2$ defined for $1 \leq i, j \leq 2^n$, and for $h \in \mathbb{R}$ as

$$\widetilde{g}_n(x_1, x_2) = \mathbb{1}_{\widetilde{H}\left(F_Q^{-1}\left(\frac{i}{2^n}\right), F_{K_s}^{-1}\left(\frac{j}{2^n}\right)\right) \leq h} \text{ for all } (x_1, x_2) \in \left[\frac{i-1}{2^n}, \frac{i}{2^n}\right[\times \left[\frac{j-1}{2^n}, \frac{j}{2^n}\right[\ , \tag{7.31}$$

where F_Q^{-1} and $F_{K_s}^{-1}$ are the inverse distribution functions of the variable Q and K_s, respectively. The Hungarian algorithm is directly applied to the function $\widetilde{g}_n$, yieldling, as n goes to ∞, the optimal PoF according to Theorem 7.6 and 7.7. We chose $n = 10$ which yields a matrix A of size 1024×1024 in the corresponding linear assignment problem.

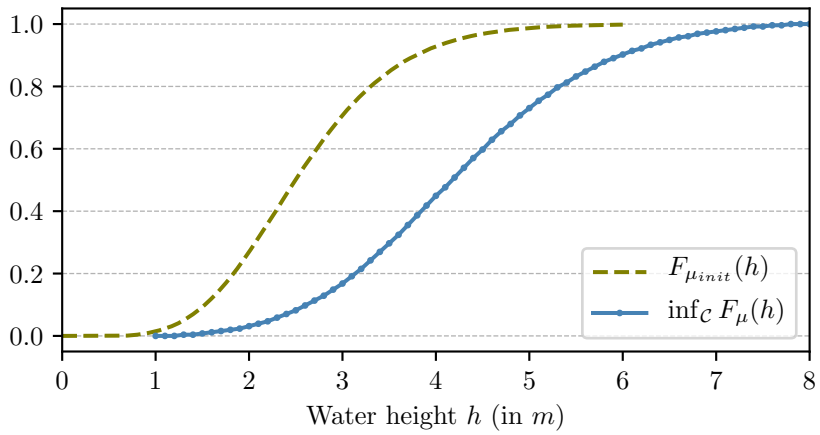
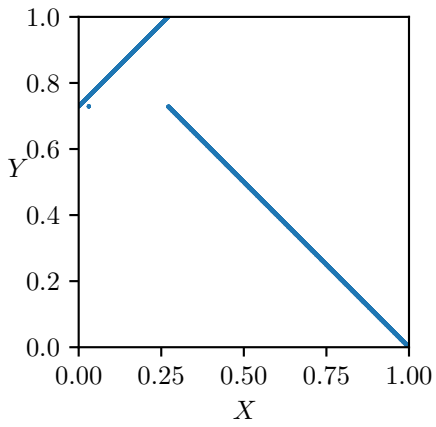
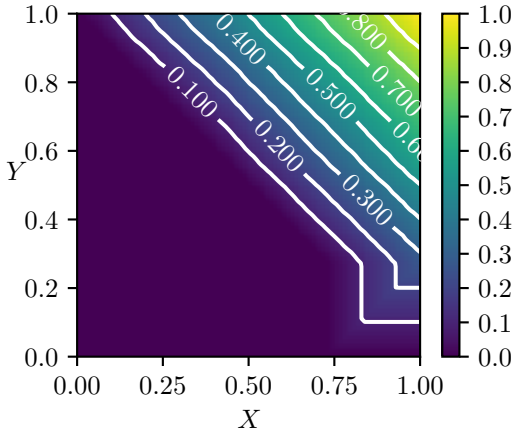


Figure 7.9: This figure represents the lowest CDF over all bidimensional copulas modeling the dependence structure between Q and K_s in the flood model. The marginals are exactly the ones of the initial distribution. However, accounting for the most penalizing dependence affects a lot the output distribution.



(a) Optimal shuffle of Min support, also representing its density



(b) Optimal shuffle of Min distribution, with associated level set.

Figure 7.10: We selected the threshold $h = 5$, the result of the minization of the PoF $\inf_{\mathcal{C}} F_\mu(5)$ using the Hungarian algorithm yields the optimal shuffle of Min here. It is surprising even though the partition size is $2^{10} = 1024$, the optimal shuffle of Min consists in only two parts.

Figure 7.9 shows that the dependence structure between Q and K_s can, in a worst case scenario, drastically changes the 95%-quantile value. Indeed, for the initial independent inputs, the 95%-quantile equals $4.23m$. However, considering the most penalizing dependence structure between the two PII without changing the marginal distributions yield a 95%-quantile equals to $6.51m$. We emphasize that the Hungarian algorithm is applied for every threshold h. For instance, $h = 5$ seeking the optimal PoF $\inf_{C\in\mathcal{C}} F_\mu(5)$ yields an optimal shuffle of Min that is depicted in Fig. 7.10. It is quite surprising for a shuffle of Min with regular partition of size $n = 2^{10} = 1024$ to actually look like a shuffle of Min only separated in two pieces. Our guess is that it is due to the quasi-linearity of the computer model.

Optimization over the copula set is interesting to account for the most penalizing dependence structures between two inputs on the QoI. The topological property of the copulas set makes it perfectly suited for the OUQ framework developed in the previous chapters. However, here we do not have a complete characterization of the extreme points of the copula set. Instead, we have identified a subset of parametric copulas that is both extremal and dense in the set of all copulas with respect to the uniform norm. These remarkable properties, associated to the special structure of shuffle of Min copulas, allow an efficient optimization of the QoI, using techniques derived from combinatorial optimization. Nevertheless, the overall method is strictly limited to two dimensions, as the combinatory algorithm is NP-hard beyond.

Conclusion

In this chapter, we investigated different leads to account for dependence between inputs. We started in Section 7.1 by noticing that the OUQ canonical moments parameterization presented in the previous chapter was not suited for high dimensional problems, with a large number of independent inputs. In that context, we intended to study a second level uncertainty in the input prior distributions of neural network

parameters. From the observation that the numerical cost for optimizing the posterior mean prediction of the NN was exploding, we proposed to concatenate input variables in blocks of dependent inputs. Doing so, new numerical challenges appear such as computing roots of high degree polynomials, corresponding to the support points of multivariate discrete measures supported on a high number of points. It is not really surprising that canonical moments are not adapted for high dimensions as the canonical moment theory developed in Chapter 4 is unidimensional.

Other alternatives to canonical moments have been explored. In Section 7.2 we investigate an historical method for dealing with moment problems. In particular, we make use of the Lasserre hierarchy of moments to transform our problem into a generalized moment problem that can be relaxed and converted into a semi-definite program. During this transformation of the optimization problem, we consider no more a product of measure, but the much larger space of all joint distributions, with any kind of dependence between inputs. Several attempts were made to reduce the size of this optimization space. The degree of relaxation and a rank criterion for convergence allow to compute an optimal QoI that is conservative in opposition to the global optimization presented in the rest of this thesis. However, because the deterministic solver we use is subject to the curse of dimension, the rank criteria for convergence of the algorithm fails for dimensions typically higher than six.

Because the set of all joint distribution is very large, in Section 7.3 we focused on the subclass of joint probability measures with fixed marginals. Therefore, only the uncertain dependence structure between marginal remained to be studied. This dependence structure takes the form of a copula. Hence, the problem reduces to the optimization of QoI over the set of all copulas. The topological structure of this is very interesting, in particular we exposed a dense and parametric subset composed uniquely of extremal copulas known as shuffle of Min copulas. The optimization over this subset can be realized in dimension two through a very efficient combinatory algorithm. However, the problem becomes NP-hard beyond dimension two. For that reason, we recommend in an engineering approach to evaluate the most penalizing dependence structures on the QoI by pairs of inputs.

As one can see, all leads investigated for accounting for dependence are limited by the dimension of the input vector. This remains to this day a challenging problem.

8

Conclusion and Perspectives

« It is no measure of health to be well adjusted to a profoundly sick society. »

Jiddu Krishnamurti

Conclusion

The purpose of this thesis was to provide tools to deal with second level uncertainty in the quantification of a risk measurement from a computer model. After introducing the interest of a second level uncertainty in Chapter 1, we presented a state of the art of methods for dealing with the uncertainty affecting the modeling of the input distribution. In particular, there is a huge literature of works dealing with the input distribution modeling. However, in this thesis we chose to focus on the optimal uncertainty quantification framework. Instead of considering one precise probability distribution for modeling the input uncertainty, we investigate a set of probability measures. This set is assumed to embed the second level uncertainty affecting the input distribution choice. This choice is usually driven by the available information, which should hence also be integrated to the way we construct the measure space.

Nevertheless, considering an uncertain input distribution that belongs to some measure space only addresses part of the problem. Indeed, the final goal remains to compute a QoI while accounting for the input uncertainty. One could try to compute the QoI distribution over the measure space, but this space is in all generality infinite dimensional and not even parametric. The solution adopted to overcome this issue consists in computing bound on the QoI. By doing so, we account only for the worst case scenario leading to a maximal or minimal QoI. In other words, the purpose is to find the distribution that most penalizes the QoI value over the measure space. For this reason it is essential to ensure that the measure space is representative of the uncertainty affecting the input measure. The two terminologies *optimal* and *robust* uncertainty quantification can be employed in this context. *Optimal* has to be taken in the sense that, given information about the input, there exist optimal bounds on uncertainties, and *robust* in the sense that the optimal bounds are independent of the input modeling choice of the engineer.

After exposing several application cases in Chapter 2, we presented in Chapter 3 a generalization of the OUQ framework introduced in Owhadi et al. (2013). Our framework is appealing for two main reasons. The first one is that it has been built upon powerful mathematical result, and the second reason is that it is well-suited for practical applications. The general idea consists in optimizing a quasi-convex lower semicontinuous function over a convex subset of a locally convex topological vector space, that is generated by its extreme points. We then proved that the supremum of the function over the whole optimization subspace can be found on its extreme points. For this matter, we studied two particular measure spaces, namely the moment class and the unimodal moment class. Both sets of measures satisfy *generalized* moment constraints.

Those spaces are by far the most important objects of this thesis. They are attractive in an engineering context as *generalized* moment constraints allow for great adaptability with respect to the available information. For instance, it is possible to enforce quantiles, moments, probabilities etc... The number of constraint is not limited, so that the analyst can shape the measure space to best fit the available information and the uncertainty on the input distribution.

Moreover, the moment class and the unimodal moment class have attractive topological properties. In particular, we have at hand a full characterization of their extreme points over which the optimization of the QoI is performed. More precisely, the extreme points of the moment class (respectively unimodal moment class) are convex combinations of Dirac masses (respectively convex combination of uniform distributions with

support bounded on one side by the mode). Consequently, the extreme points are parametric measures, which is convenient for the optimization of the QoI. This result is what makes these two measure spaces and the OUQ framework so appealing.

Beyond the measure space, the other theoretical advantage of our framework is that it allows a wide choice of optimization functions. Indeed, quasi-convexity embeds a very large class of functions. The lower semicontinuity assumption can also be replaced by upper semicontinuity associated to a compactness assumption of the measure space. If so, the reduction theorem is known as the Bauer maximum principle. One can also dispense with any regularity assumption when studying measure affine functions or ratio of measure affine functions. In addition to the theoretical aspect of the optimization function, we studied in Chapter 5 some useful practical QoIs. Most of them are classical quantities, but we consider it here as functions of the input measure of the computer model. In this context, we have shown that a failure probability is a measure affine function, and a quantile is a quasi-convex lower semicontinuous function. We also studied the optimization of superquantiles and Sobol' indices. Moreover, we embed the robust Bayesian analysis into our theoretical framework, showing that it is a particular case of an OUQ problem. Nevertheless, we emphasize that the literature on robust Bayesian analysis also explores prior measure spaces that are different from the moment class and the unimodal moment class.

A major drawback is that the QoI parameterized over the extreme points is in general deprived of any property that would ease its optimization, such as continuity or convexity. Therefore, the computation of the QoI bounds requires the use of a global optimizer. This implies that the performance of the optimization is directly related to the ability to probe the optimization space. But neither the moment class nor the unimodal moment class are easily explored. This in particular due to the fact that they are parameterized by weights and support points of mixtures of Dirac masses and uniform distributions, respectively, with numerous underlying constraints.

For that reason, we propose an alternative parameterization of these measure spaces in Chapter 4. However, this specific parameterization requires more restrictive assumption so that the *generalized* moment constraints defining the measure space must be specialized into *classical* moment constraints. In practice, this limits the interest of moment classes. Indeed, there usage implies that moments of the uncertain input distribution are known by the analyst, or at least partially known, as it is always possible to enforce inequality constraints. In practice, moments can be estimated from data and/or expert knowledge, retrieved from the original modeling of the input, or even bounded from a robust Bayesian inference. Therefore, the methodology restricted to *classical* moments is not meaningless. Once set, we demonstrated how canonical moments provide a well-suited parameterization for moment classes. In particular, canonical moments define a bijection between the set of all discrete measures supported on $n+1$ points satisfying n moment constraints and the pavement $[0,1]^{n+1}$. Therefore, the global optimization that relied on the exploration of the moment class, now relies on the exploration of a simple pavement. Probing the latter is straightforward.

The main drawback of canonical moments is numerical. Indeed, the bijection between sequences of canonical moments and discrete distributions satisfying the constraints involves the computation of polynomial roots , in equal number to the support' cardinal. The roots of these polynomials, which are orthogonal for the underlying measure, correspond to the support points of the measure. We have at our disposal a well-conditioned algorithm for this task that profits of the three terms recurrence defining the sequence of orthogonal polynomials. However, recovering the weights of

the discrete distribution is a not well-conditioned problem. This prevents from using canonical moments with a large amount of constraints, or to consider blocks of dependent inputs as detailed in Chapter 7.

In short, the canonical moments parameterization provides a powerful way for exploring the generator of the moment class. Of course, many flaws persist, but they are not all directly imputable to the canonical moment parameterization. Indeed, some weaknesses are due to the overall methodology independently of the optimization space parameterization:

- The optimization still requires the use of a global solver. In this thesis, we make use of a DE algorithm, whose performance appears quite satisfying. Nevertheless, global optimizers are based on heuristics and cannot necessarily be shown to converge to the true optimum. We successfully proved that the algorithm had reached the optimum in a simple 6-dimensional case with only one constraint on the mean of each input. Indeed, in Chapter 7, we studied a conservative algorithm based on a semi definite programming of a relaxed equivalent optimization problem. We observed in the simple 6-dimensional case that the conservative SDP and the global optimization over the extreme points returned similar solutions, which are therefore close to the optimum. However, this result is marginal and cannot be adapted in higher dimensions or with more moment constraints, since in the latter the SDP computation fails. In general, there is no proof that the true optimum has been reached;

- The computation of the QoI is subject to the curse of dimensionality. It is particularly obvious when each one of the d-input distribution is discrete. Indeed, in this case any evaluation of the QoI involves runs of the computer model on a d-dimensional grid whose size grows exponentially with the dimension. For this reason, we advise to limit the computation to models with at most 10-dimensional inputs on a regular computer, as we did in Chapter 6. In this chapter, we depicted how to apply the methodology on a computer model used on a real-life industrial case-study. In particular, we have dealt with dimension reduction, surrogate modeling, and consequently accounting for the surrogate model error. Beyond 10 dimensions, we recommend to parallelize computations on a cluster.

This thesis was motivated by important applicative challenges. So that, we endeavored to connect every theoretical aspect with a practical application. As stated, one of the main reasons why the OUQ framework is attractive is that it gathers both strong theoretical results and high potential for practical applications. Nevertheless, the flaws addressed hereabove lead to consider several perspectives and leads for improvement.

Perspectives

Numerical Improvement

There are clear improvement to be done with respect to the numerical optimization aspects. In particular, we only worked with two different global solvers, namely the differential evolution and simulated annealing algorithms. The former was seen to be much more efficient in our cases, while simulated annealing algorithm seemed too complex to tune. For this reason, this thesis only focused on the DE algorithm. However, other global solvers could potentially perform better.

Nevertheless, direct and easy improvement can be done within the DE solver used here. First, this algorithm is suited for parallelization, as any individual in a population can be computed separately. Moreover, in this thesis we regularly computed the CDF lower envelop corresponding to the minimization of a PoF at different thresholds over a measure space. For any threshold, the optimization was completely reset. However, one can make use of the previous best solution at threshold h to generate the initial population locally around for the optimization at threshold $h + \Delta h$. This method works well in practice and accelerates the convergence. We purposely did not employ this strategy in the illustrations of the thesis. Indeed, we have proved that the space exploration was very efficient using the canonical parameterization associated with the DE algorithm without deploying this method.

else we could not have proved that the canonical parameterization associated with the DE algorithm was how we obtained an efficient space exploration.

Computation of the QoI Derivatives

One of the main problem is that the QoI computed over discrete measures is not necessarily continuous. For instance, if one wishes to compute a PoF, then it writes as a sum of indicator functions over a grid of points. That is one of the reasons why a raw global solver is mandatory to perform the optimization. However, when one input is in an unimodal moment class, then the optimization is run over mixtures of uniform distributions, which are continuous. Then, it could be possible to compute the partial derivatives of the PoF with respect to the parameters of the uniform distribution mixture. These derivatives could then be used to boost the performance of the optimization. However, in practice the PoF usually does not depend directly on the parameters of the uniform distribution mixture, but rather on a sequence of canonical moments. This complicates the analytical computation of the PoF derivatives.

Canonical Moments with Generalized Moment Constraints

As stated in the conclusion, the canonical moment parameterization of the moment class is only possible in the presence of *classical* moment constraints. This limits the applicability of moment classes. An interesting perspective would be to generalize the canonical parameterization in presence of *generalized* moment constraints defining the moment class. For instance, by assuming some kind of isomorphism between the *generalized* moments constraints and *classical* moment constraints, or by representing the *general* moment constraint as a series. Although it is not clear how to proceed about this challenging problem, it would stimulate interest in moment classes and the potential of the canonical moments parameterization.

In industrial applications, such as safety analysis of nuclear reactors, it is usual to specify the input distributions through their quantiles instead of their *classical* moments. Quantile constraints can be seen as a specific case of *generalized* moment constraints. However, in this situation the problem greatly simplifies (see Moreno and Cano (1991)). Indeed, consider two real number a and b and a partition $\{[c_i, c_{i+1}]\}_{0 \leq i \leq n}$ of $[a, b]$, then define the measure space specified by known quantiles over the partition: $\mathcal{A} = \{\mu \in \mathcal{P}([a, b]) \mid \mu([c_i, c_{i+1}]) = p_i,\ 0 \leq i \leq n\}$. Notice that there is only n true constraints, the last one being related to the others. There is no need for canonical moment parameterization in order to generate a discrete measure $\mu = \sum_{i=0}^{n} \omega_i \delta_{x_i}$ satisfying the constraints. By construction, the position x_i belongs in $[c_i, c_{i+1}[$, for $0 \leq i \leq n$. The associated weights are obviously set to fit the quantile level such that $\omega_i = p_i$ for

$0 \leq i \leq n$. Therefore, because the weights are fixed, the optimization of the QoI is equivalent to finding the most penalizing positions in every element of the partition. This can be done using any global optimizer where the QoI is parameterized directly by the positions of the discrete measure.

Gaussian Process Error

When accounting for the metamodel-induced uncertainty in Chapter 6, the proposed approach consisted in simulating several realizations of the Gaussian process over a finite grid of points. However, as the dimension of the input vector increases, the size of the grid grows exponentially. We rapidly reached numerical issues, as computing new realizations of a Gaussian process involves inverting the covariance matrix between the process at the evaluation points. An efficient generation of conditional trajectories of a Gaussian process over large number of evaluation points would improve the optimization (Le Gratiet, 2013).

Moreover, we have at our disposal a grid of evaluation points only in presence of discrete input distributions. If one input is unimodal, the method cannot be applied as it requires generating continuous realizations of the Gaussian process. Another perspective would be to set a finite design of experiment over which we compute the computer model. This way, we get fixed points over which computing several trajectories of the Gaussian process.

An other approach would consist in an augmented MC in order to incorporate the Gp metamodel uncertainty. Indeed, a MC sampling is necessary to compute the QoI as soon as at least one distribution is continuous. Another MC sampling is required for computing several trajectories of the computer metamodel. One could consider merging the two MC loops into a single algorithm.

Stochastic Process

Rather than considering probability distribution, one would like to study Stochastic process. Consider for instance a Gaussian process that could represent a physical quantity like a wind speed vector field evolving in time. In that context, one would like to predict the energy production of a wind turbine. The model would take for input the Gaussian process and return the energy. The wind field being uncertain, we can model the set of all Gaussian process with only a finite number of covariances fixed. Then, one can optimize the energy of the wind turbine over the set of all Gaussian process satisfying the constrained covariances using the spectral measure representation.

More precisely, let us consider a complex stationary process $X := (X_n; n \in \mathbb{Z})$, and assume that it is centered, so that

$$\begin{cases} \forall t \in \mathbb{Z}, & \mathbb{E}(X_t) &= 0 \ ; \\ \forall (t,s) \in \mathbb{Z}^2, & \mathbb{E}(X_t \overline{X_s}) &= \gamma^X_{t-s} \ ; \end{cases}$$

where γ^X is the autocorrelation function. Let $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$, so that $\partial\mathbb{D}$ denotes the unit circle. By Bochner's theorem (Priestley, 1981; Brockwell et al., 1991), there exists a unique probability measure known as the *spectral* measure $\mu_X \in \mathcal{P}(\partial\mathbb{D})$ such that:

$$\forall k \in \mathbb{Z}, \ \gamma^X_k = \int_{\partial\mathbb{D}} e^{ik\theta} \mu_X(\mathrm{d}\theta) \ . \tag{8.1}$$

Because covariance uniquely characterizes the class of Gaussian processes (Rasmussen and Williams, 2005), there exists a unique Gaussian process X^G with the same autocovariance as X. It is called the Gaussian representative. Now suppose we are dealing

with the set of all Gaussian process for which partial information is available in the form of finitely many covariances $\mathbf{c} = (c_1, \ldots, c_n)$, such that we consider the set of all spectral measures:

$$\mathcal{D} = \left\{ \mu_X \in \mathcal{P}(\partial\mathbb{D}) \mid \int_{\partial\mathbb{D}} e^{ik\theta} \mu_X(\mathrm{d}\theta) = c_k,\ 1 \leq k \leq n \right\} .$$

Then, the measure space $\mathcal{D}$ is the space of all probability measure on $\partial\mathbb{D}$ defined by trigonometric moment constraints. Therefore, its extreme points are discrete measures. In addition, there exists a theory of trigonometric canonical moments that is suited for exploring efficiently this space.

New Applications

Beyond the two computer models presented in this thesis, one toy case and one real-life engineering model, it would be valuable to test the methodology on a large scope of different models. Here, the applications are mainly related to risk management analysis, but other fields are also interested by OUQ such as mechanical engineering (Bonnet et al., 2020) or small particle hypervelocity impact (Kamga et al., 2014). It is not only interesting but also important to see how this work can be combined to existing search in OUQ such as in Han et al. (2015) or Sullivan et al. (2013).

Part II

Appendix

A

Robustness in Perturbed Law Indices

The content of this appendix is related to the following paper which results from the 6th months internship of Clément Gauchy that I supervised during my PhD at EDF R&D:

Gauchy, C., Stenger, J., Sueur, R., and Iooss, B. 2019. An Information Geometry Approach for Robustness Analysis in Uncertainty Quantification of Computer Codes. *Preprint. hal-02425477.*

This appendix deals with the problem of robustness analysis which is an emerging field in the domain of uncertainty quantification. It consists of analysing the response of a computer model with uncertain inputs to the perturbation of one or several of its input distributions. Thus, a practical robustness analysis methodology should rely on a coherent definition of a distribution perturbation. This paper addresses this issue by exposing a rigorous way of perturbing densities. The proposed methodology is based on the Fisher distance on manifolds of probability distributions. A numerical method to calculate perturbed densities in practice is presented. This method comes from Lagrangian mechanics and consists of solving an ordinary differential equations system. This perturbation definition is then used to compute quantile-oriented robustness indices. The resulting Perturbed-Law based Indices (PLI) are illustrated on several numerical models. This methodology is also applied to an industrial study (simulation of a loss of coolant accident in a nuclear reactor), where several tens of the model physical parameters are uncertain with limited knowledge concerning their distributions.

A.1 Introduction

During the last decades, two major trends in industrial and research practices have led to a rise in importance of uncertainty quantification (UQ) methodologies (De Roc-

quigny et al., 2008; Smith, 2014; Ghanem et al., 2017). The first is the replacement of full-scale physical experiments, considered costly and difficult to implement, by numerical models. This choice raises the issue of a potential mismatch between computer codes and the physical reality they aim to simulate. The second trend consists in accounting for the risks in an increasing number of industrial activities, this implies that those risks should be evaluated from a quantitative point of view. In both situations, the quantification of uncertainties can be conducted by considering as a vector of random variables, named $\mathbf{X} = (X_1, ..., X_d)$, the uncertain inputs of the computer code represented by a function $G(\cdot)$. The most widespread approach consists of running $G(\cdot)$ with different combinations of inputs in accordance with their range of variation, in order to study the related uncertainty on the output $Y = G(X_1, ..., X_d)$ or to estimate a specific quantity of interest (QoI). A QoI is is a statistical quantity derived from Y, e.g. a performance as the mean of Y or a risk criterion as a high-level quantile of Y.

As an example, the nuclear industry faces major issues as facilities age and regulatory authorities' requirements strengthen (Bucalossi et al., 2010; Mousseau and Williams, 2017). For example, the operators have to study the "Loss of Coolant Accident" (LOCA) resulting in a break on the primary loop of pressurized water nuclear reactors. This scenario can be simulated using system thermal-hydraulic computer codes, which include tens of physical parameters such as condensation or heat transfer coefficients (Mazgaj et al., 2016; Sanchez-Saez et al., 2018). Yet, the values of these parameters are known with a limited precision (Larget, 2019) as they are calculated by the way of other quantities measured via small-scale physical experiments.

Some other variables are only observed during periodic inspections, such as the characteristics of pumps in hydraulic systems.

Various methods coming from the UQ domain are useful in considering these uncertainties in the system safety analysis. First of all, some methods aim at improving the exploration of the input domain $\mathcal{X}$ by using specific designs of experiments, such as the space filling designs (Fang et al., 2005). Such a design allows to cover an input domain as evenly as possible with a fixed number of code runs as well to limit unexplored areas as much as possible.

For the estimation of some specific QoI, such as a probability of threshold exceedance by the output or an α-order quantile of the output, Monte Carlo type methods are often preferred. In particular, accelerated Monte Carlo methods (e.g. importance sampling or subset simulation) target the most informative areas of $\mathcal{X}$ in the sampling algorithm in order to estimate the QoI while controlling its estimation error (Morio and Balesdent, 2016). As a preliminary or concomitant stage, global sensitivity analysis is also essential in order to eliminate non-influential parameters and to rank influential parameters according to their impact on the QoI (Iooss and Lemaître, 2015; Iooss and Marrel, 2019).

All these approaches are useful to deal with the existence of uncertainties in applied problems. However, industrial (e.g. nuclear facilities) operators have to face the difficulty of justifying their risk assessment methodologies not merely by providing simulation results. Such a justification has to demonstrate that the computed values overestimate the actual risks which most of the time cannot be calculated. This principle of conservatism, which can be easily implemented when dealing with very simple monotonic physical models, can be hard to be adapted to computer codes simulating complex and non monotonic physical phenomena. It is also not always straightforward to apply this principle when implementing UQ methods based on a set of computer experiments providing a whole range of values for the output quantity Y.

To address this issue, the new UQ branch of robustness analysis has emerged during the recent years in the field of sensitivity analysis. It consists of evaluating the impact of the choice of the inputs' distributions and, more precisely, by analyzing the QoI variations with respect to this choice. A first solution would consider a whole set of input laws and analysing the related output distributions. For global sensitivity analysis, Hart and Gremaud (2019) uses "optimal perturbations" of the probability density functions to analyze the robustness of the variance-based sensitivity indices (called Sobol' indices (Sobol', 1993)). Meynaoui et al. (2019) and Chabridon (2018) propose approaches to deal with the so-called second-level uncertainty, i.e. uncertainty on the parameters of the input distributions. Another approach, called optimal uncertainty quantification, avoids specifying the input probability distributions, turning the problem to the definition of constraints on moments (Owhadi et al., 2013; Stenger et al., 2020). This solution is out of scope of the present work which considers that the initial input probability, that has been defined by the user, is of practical importance.

In practical engineering uncertainty quantification studies, input distributions are truncated as it corresponds to physical parameters with known domain of validity. It is therefore natural to assume no uncertainty on the support of the input random variables. In this paper, we also assume their mutual independence. Keeping in mind that our goal is to directly deal with the input distributions (without considering second-level uncertainty), one particularly interesting solution has been proposed in the context of reliability-oriented sensitivity analysis by Lemaître (2014) (see also Lemaître et al. (2015); Sueur et al. (2016)) with the so called Perturbed-Law based Indices (PLI). A density perturbation consists of replacing the density f_i of one input X_i by a perturbed one $f_{i\delta}$, where $\delta \in \mathbb{R}$ represents a shift of a moment (e.g. the mean or the variance). Amongst all densities with shifted mean or variance of a δ value, $f_{i\delta}$ is defined as the one minimizing the Kullback-Leibler divergence from f_i. This method has been applied on the computation of a probability of failure (Iooss and Le Gratiet, 2019; Perrin and Defaux, 2019), a quantile (Sueur et al., 2017; Larget, 2019) and a superquantile (Iooss et al., 2020; Larget and Gautier, 2020) as the QoI.

However, this method is not fully satisfactory. Indeed, the minimal Kullback-Leibler divergence can significantly vary between different inputs' distribution of even two different parameters of the same density, so that some densities are more perturbed than others. Moreover, some distributions do not have defined moments. As in Perrin and Defaux (2019), an iso-probabilistic operator can be applied to transform all the input random variables into centered normalized Gaussian ones. It allows to make perturbations comparable when applied in this standard space, but it remains difficult to translate this interpretation in the initial physical space which is the one of interest for the practitioners. Note that another type of robustness analysis has been proposed in quantitative finance by Cont et al. (2010). These authors investigate whether the estimated QoI is sensitive to a small perturbation of the empirical distribution function. For this purpose, they define the robustness of a QoI as its continuity with respect to the Prokhorov distance on the set of integrable random variables.

The goal of this paper is to propose a novel approach for perturbing probability distribution. It relies on density perturbation based on the Fisher distance (Costa et al., 2012) as a measure of dissimilarity between the initial density f_i and the perturbed one $f_{i\delta}$. This distance defines a geometry on spaces of probability measures called information geometry (Nielsen, 2013). The statistical interpretation of the Fisher distance provides an equivalence between perturbation of non-homogeneous quantities and consequently a coherent framework for robustness analysis. To present this approach, we

first review the existing density perturbation methods in Section A.2. Section A.3 is then dedicated to the description of our method and the discussion of our numerical tools. Section A.4 illustrates our methodology of density perturbation on the practical robustness index PLI. An analytical application and an industrial case study are presented in Section A.5. The last section gives conclusions and some research perspectives.

A.2 Previous approaches of density perturbation for UQ robustness analysis

The method of Lemaître et al. (2015) has been later called PLI by (Sueur et al., 2016), as it is based on the idea of perturbing the inputs' densities. It aims at providing a practical counterpart to the general idea of analyzing the output QoI of a model in a UQ framework when one or several parameters of the input probabilistic model (considered as the reference one) is changed. This can be seen as a way to take into account an "error term" one could add to an imperfectly known input distribution.

A.2.1 Kullback-Leibler divergence minimization

To build a perturbed distribution $f_{i\delta}$ from a distribution f_i, the approach of Lemaître et al. (2015) is non-parametric. It is mainly thought to analyze perturbations on the most common characteristics of input laws which are the mean and variance. To illustrate it in the case of a mean perturbation, we assume the random variable $X_i \sim f_i$ has mean $\mathbb{E}[X_i] = \mu$. By definition, the perturbed density will have a $\mu + \delta$ mean. But this is obviously not sufficient to fully determine the perturbed law and especially to explicitly access the value of $f_{i\delta}$ on the whole domain of X_i. Amongst all densities with a mean equal to $\mu + \delta$, $f_{i\delta}$ is defined as the solution of the minimization problem

$$f_{i\delta} = \underset{\pi \in \mathcal{P},\, s.t\, \mathbb{E}_\pi[\mathbf{X_i}] = \mathbb{E}_{f_i}[X_i] + \delta}{\arg\min} KL(\pi || f) \ , \tag{A.1}$$

where $\mathcal{P}$ is the set of all probability measures absolutely continuous with respect to f_i. This approach basically consists of perturbing the chosen parameter while changing the initial model as little as possible. With this definition, "changing" the model is understood as an increase of entropy, the Kullback-Leibler divergence between two densities f and π being

$$KL(\pi || f) = \int \log\left(\frac{\pi(x)}{f(x)}\right) f(x)\, dx \ . \tag{A.2}$$

TThis method can be applied on higher order moments (for instance moments of order 2, to define variance perturbation) and, more generally, to constraints that can be expressed as a function of the perturbed density, as quantiles (Lemaître, 2014). Notice that, in the case of an initial Gaussian distribution, the perturbed distribution remains Gaussian with a mean shift of δ.

In the general case, this method has several drawbacks: First of all, the likelihood ratio between $f_{i\delta}$ and f_i might not have an analytic form, which leads to numerical difficulties. Moreover, this method requires defined moments for the initial density. Finally, the main difficulty concerns the interpretation of the results obtained from this PLI method. Indeed, each uncertain input of the UQ model is perturbed with a

range of δ values. To interpret the QoI shift resulting of these perturbations in the standard space, a clear understanding of the physical meaning of each perturbation is necessary. Low interpretability of the perturbed density can appears for some physical parameters, e.g. for uncertainties on the state of the system coming from a variability of the quantity throughout the operating process. In this case, the probability distribution of the uncertain quantity can be regarded in terms of relative frequency of occurrence. But it can be more difficult when it comes to constant physical parameters known with a limited accuracy.

We recall that all input random variables are assumed mutually independent. Nonetheless, the effect of perturbations can be considered only for each variable individually and in absolute terms (as a same δ shift might have completely different impacts for different input densities). This methodology thus yields difficulty to compare the relative impact of perturbations between different inputs.

A.2.2 Standard space transformation

To interpret the δ shift on the input distribution and especially to allow a comparison between inputs according to the impact on the QoI of a same perturbation, an equivalence criterion between inputs is required. An idea developed by Perrin and Defaux (2019) consists of applying perturbations in the so-called *standard space* (instead of the initial physical space) in which all input laws are identical, making all perturbations equivalent. Finally, the perturbed densities are obtained by applying the reverse transformation as the one used to transform inputs in the standard space.

In the case of independent inputs, the required distribution transformation is a simple inverse probability transform. Given a random vector $\mathbf{X}$ with cumulative distribution function F, the transform is the random vector $\mathbf{S} = \Phi^{-1}(F(\mathbf{X}))$, where Φ is the cumulative distribution function of the standard Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. Consequently, $\mathbf{S}$ follows a standard Gaussian distribution whatever the initial distribution F. In the Kullback-Leibler divergence minimization framework (see Section A.2.1), a perturbation of the mean simply consists of a mean shift without changing the standard deviation. Hence this leads to an analytical expression for the perturbed density $f_{i\delta}$ thanks to the variable change formula (Stirzaker, 2003, p.318):

$$f_\delta(x) = e^{\frac{-\delta^2 + 2\delta\Phi^{-1}(F(x))}{2}} f(x) \ . \tag{A.3}$$

This simple formula makes the perturbed density and the likelihood ratio easy to compute.

However, similar perturbations in the standard space implies very different ones in the physical space according to the initial distribution. As an example, Figure A.1 depicts two Kullback-Leibler divergences (approximated with Simpson's rule (Abramowitz and Stegun, 1964)) between a particular distribution (the Triangular $\mathcal{T}(-1, 0, 1)$ [1] and the Uniform one $\mathcal{U}[-1, 1]$) and its associated distribution in the standard space. The results show that the Kullback-Leibler divergence behaves very differently in the physical space, depending of the original distribution, even though the same perturbation is applied in the standard space. For example, there is no general rule to estimate the mean of the physical perturbed input for a given mean perturbation in the standard space. Such difficulties are even more significant when considering perturbations on other parameters than the mean. For instance, there is no general equivalence in the

[1] the triangular distribution $\mathcal{T}(-1, 0, 1)$ is parametrized by its minimum a, mode b and maximum c

physical space between perturbations applied on the mean and on the standard deviation of the same input in the standard one. Hence, it seems generally impossible to convert in a simple way the results given by this method into a relationship between input and output physical quantities, making these results difficult to interpret.

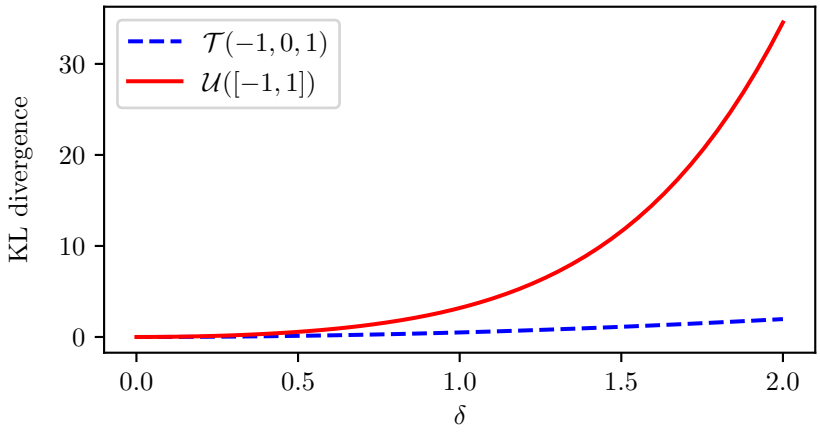


Figure A.1: Kullback-Leibler divergence between the initial distribution and the perturbed one for perturbation levels $\delta \in [0, 2]$. Further description on the perturbed distribution can be found in Lemaître (2014) and Lemaître et al. (2015).

A.3 A perturbation method based on information geometry

The Kullback-Leibler divergence can be interpreted as the power of a hypothesis test with null hypothesis "X_i follows the distribution f_i" and an alternative hypothesis "X_i follows distribution $f_{i\delta}$" (Eguchi and Copas, 2006). For this reason, it seems to be an appropriate tool to measure how far a perturbed density is from its initial reference and thus to provide a formal counterpart to the dim idea of "uncertainty on the distribution". It is especially well suited to compare Gaussian distributions, which requires, in a robustness analysis context, an additional transformation to embed inputs in a standard space as these are physical quantities with potentially non Gaussian distributions. This additional operation, which also provides an equivalence between non-homogeneous input variables, makes it impossible to interpret in terms of physical inputs the perturbations of the related standard ones.

A.3.1 Fisher distance

To allow intuitive understandings of the consequence of these perturbations on the output distribution, it is necessary to base our perturbation method on a metric which allows at the same time to compare perturbations on different parameters of the same distribution and on different inputs of the UQ model. In particular it should not depend on the representation of the input distribution, which means being independent of the parametrization. The Fisher distance has all these advantages. It is based on the local scalar product induced by the Fisher information matrix in a given parametric space and defines a Riemannian geometry on the corresponding set of probability measures

as on any Riemannian manifold with its associated metric. Consider the parametric density family $\mathcal{S} = \{f_\theta, \theta \in \Theta \subset \mathbb{R}^r\}$. We recall that every input variables represent physical parameters with known domain of validity, therefore for all θ in Θ, the support of f_θ is assumed fixed. The metric associated to the coordinate function θ, called the Fisher (or Fisher - Rao) metric, is defined as:

$$I(\theta) = \mathbb{E}\left[\nabla_\theta \log f_\theta(X)(\nabla_\theta \log f_\theta(X))^T\right] \ ,$$

where $I(\theta)$ is the Fisher information matrix evaluated in θ for this statistical model. The Fisher information, well known for instance in optimal design, Bayesian statistics and machine learning, is a way of measuring the amount of information that an observable random variable X carries about an unknown parameter θ of the distribution of X. The Fisher information matrix defines the following local inner product in $\mathcal{S}$ for $u \in \mathbb{R}^r$ and $v \in \mathbb{R}^r$:

$$\langle u, v\rangle_\theta = u^T I(\theta) v \ . \tag{A.4}$$

Given two distributions f_{θ_1} and f_{θ_2} in the manifold $\mathcal{S}$, a path from f_{θ_1} to f_{θ_2} is a piecewise smooth map $q : [0, 1] \to \Theta$ satisfying $q(0) = \theta_1$ and $q(1) = \theta_2$. Its length $l(q)$ satisfies the following equation:

$$l(q) = \int_0^1 \sqrt{\langle \dot{q}(t), \dot{q}(t)\rangle_{q(t)}} dt \ , \tag{A.5}$$

where $\dot{q}$ is the derivative of q. Alike, the energy $E(q)$ of a path is defined by the equation:

$$E(q) = \int_0^1 \frac{1}{2}\langle \dot{q}(t), \dot{q}(t)\rangle_{q(t)} dt \ . \tag{A.6}$$

The distance between f_{θ_1} and f_{θ_2}, called the Fisher distance, is defined as the minimal length over the set of paths from f_{θ_1} to f_{θ_2}, denoted by $\mathcal{P}(f_{\theta_1}, f_{\theta_2})$:

$$d_F(f_{\theta_1}, f_{\theta_2}) = \inf_{q \in \mathcal{P}(f_{\theta_1}, f_{\theta_2})} l(q) \ . \tag{A.7}$$

The path γ minimizing this length - or equivalently minimizing the energy - is called a geodesic (Costa et al., 2012). The specific choice of the Fisher information matrix for a Riemannian metric matrix leads to a very interesting statistical interpretation, as shown in Amari (2016, p.27). It is directly related to the Cramer-Rao lower bound (Rao, 1945) which states that, for any unbiased estimator $\widehat{\theta}$ of θ, the covariance matrix $\mathrm{Var}(\widehat{\theta})$ is bounded by $I(\theta)^{-1}$. This means that the Fisher information is the maximum amount of information about the value of a parameter one can extract from a given sample. More formally, under some regularity conditions [given by (Newey and McFadden, 1994, Theorem 3.3)], if $x_1, .., x_n$ are n independent observations distributed according to a density f_θ, the maximum likelihood estimator $\widehat{\theta}_n$ of θ converges weakly to a normal law with mean θ and covariance $\dfrac{I(\theta)^{-1}}{n}$. The density of $\widehat{\theta}_n$ denoted by $p(\widehat{\theta}_n, \theta)$ writes

$$p(\widehat{\theta}_n, \theta) = \frac{1}{\sqrt{(2\pi)^n \det(I(\theta))}} \exp\left(-\frac{n(\widehat{\theta}_n - \theta)^T I(\theta)(\widehat{\theta}_n - \theta)}{2}\right) . \tag{A.8}$$

When n is large, this probability density is proportional to $(\widehat{\theta}_n - \theta)^T I(\theta)(\widehat{\theta}_n - \theta)$ which is the local inner product defined in equation Eq. (A.4). Therefore, the Fisher distance

between two distributions with parameters θ and θ' can be constructed as a measure of the risk of confusion between them. In other words, the Fisher distance between two distributions f_θ and $f_{\theta'}$ represents the separability of the two distributions by a finite sample of independent observations sampled from the f_θ distribution.

We illustrate the Fisher distance on a simple example. Consider the statistical manifold of univariate normal distributions $\mathcal{S} = \{\mathcal{N}(\mu, \sigma^2), (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*\}$. The Fisher information matrix has the analytical form (Costa et al., 2012):

$$I(\mu, \sigma) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix}. \tag{A.9}$$

We can apply the change of coordinate $\phi(\mu, \sigma) \rightarrow (\frac{\mu}{\sqrt{2}}, \sigma)$, so that the related geometry is the hyperbolic geometry in the Poincaré half-plane (Stillwell, 1997), in which the geodesic and distance between two normal distributions are known analytically. Geometrically, the geodesics are the vertical lines and the half circle centered on the line $\sigma = 0$.

Further details on the interpretation of information geometry can be found in Costa et al. (2012). Figure A.2 shows the position of four Gaussian distributions in the $(\frac{\mu}{\sqrt{2}}, \sigma)$ half-plane. It is clear that the Gaussian distributions C and D are more difficult to be distinguished than the distributions A and B although in both cases the KL divergence is the same. The hyperbolic geometry induced by the Fisher information provides a representation in accordance with this intuition. Indeed, the two dashed curves are the geodesics respectively between points A and B, and points C and D. We observe that the Fisher distance between A and B is greater that the distance between C and D. This illustrates how information geometry provides a proper framework to measure statistical dissimilarities in a space of probability measures.

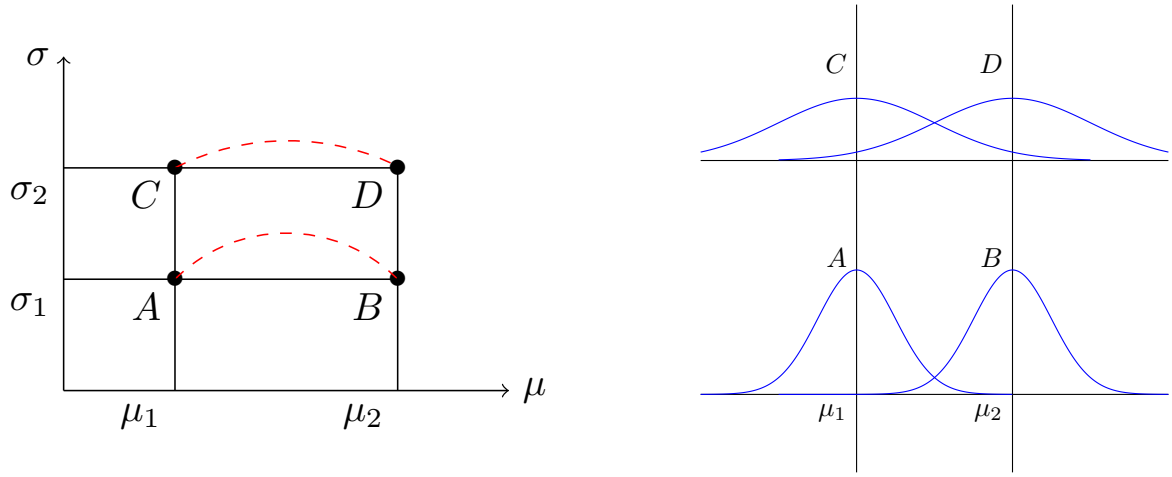


Figure A.2: Representation of four Gaussian distributions in the parameter space on the left, and their respective distributions on the right. Although $KL(A||B) = KL(C||D)$, it is easier to distinguish A from B than C from D. The dashed curved lines are two geodesics in $(\frac{\mu}{\sqrt{2}}, \sigma)$ plane with different lengths.

The Fisher distance provides a satisfactory grounding to our notion of density perturbation. We define a perturbation of a density f to be of magnitude δ if the Fisher distance between f and the perturbed density f_δ is equal to δ. The set of

all perturbations of f at level δ is then the Fisher sphere of radius δ centered in f, whenever this perturbation is applied to one or another of the parameters. This implies that, in this framework, we do not consider one specific perturbed distribution but a non finite set of probability densities. The next section is dedicated to the development of a numerical method to compute the Fisher spheres of radius δ centered in f.

A.3.2 Computing Fisher spheres

As detailed in Section A.3.1, geodesics are defined as the solution of a minimization problem. More specifically a geodesic is a path with minimal length or energy (denoted E). Given a smooth map $q : [0,1] \to \mathcal{S}$, we have

$$E(q) = \int_0^1 \frac{1}{2} \langle \dot{q}(t), \dot{q}(t) \rangle_{q(t)} dt \ . \tag{A.10}$$

In the following we denote $L(t, q, \dot{q}) = \frac{1}{2} \langle \dot{q}(t), \dot{q}(t) \rangle_{q(t)}$ and L is called the Lagrangian of the system. The energy of a path can be rewritten as

$$E(q) = \int_0^1 L(t, q, \dot{q}) dt \ . \tag{A.11}$$

A necessary condition for the path q to minimize the energy E is to satisfy the Euler-Lagrange equation (see Gelfand and Fomin (2012) for details):

$$\frac{\partial L}{\partial q} = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \right) \ . \tag{A.12}$$

We denote $p = \dfrac{\partial L}{\partial \dot{q}}$ and obtain by derivation of the quadratic form $L(t, q, \dot{q}) = \frac{1}{2} \dot{q}^T I(q) \dot{q}$ that $p = I(q)\dot{q}$, and $\dot{q} = I^{-1}(q)p$. Then, inspired by Lagrangian mechanics theory (Arnold, 1997, p.65), the Hamiltonian $H(p, q)$ defined by

$$\begin{array}{rcl} H(p,q) & = & p^T \dot{q} - L(t, q, \dot{q}) = p^T I^{-1}(q) p - \frac{1}{2} \dot{q}^T I(q) \dot{q} \\ & = & \frac{1}{2} p^T I^{-1}(q) p \end{array} \tag{A.13}$$

is constant whenever q is a geodesic. Eq. Eq. (A.13) is derived from the Euler Lagrange equation and implies that (p, q) follows a system of Ordinary Differential Equation (ODE) called Hamilton's equations:

$$\left\{ \begin{array}{rcccl} \dot{q} & = & \dfrac{\partial H}{\partial p} & = & I^{-1}(q) p \, , \\ \dot{p} & = & -\dfrac{\partial H}{\partial q} & = & \dfrac{\partial L(t, q, I^{-1}(q)p)}{\partial q} \, . \end{array} \right. \tag{A.14}$$

The objective is to determine any geodesics q satisfying $q(0) = \theta_0$ and $d_F(f, q(1)) = \delta$, it corresponds to computing the Fisher sphere centered in f_{θ_0} with radius δ. The only degree of freedom left to fully solve the ODE system Eq. (A.14) is the initial velocity $p(0)$. Notice that the Hamiltonian is equal to the kinetic energy as $p = I(q)\dot{q}$. As the Hamiltonian is constant on a geodesic, we have for all t:

$$\frac{1}{2} \langle \dot{q}(t), \dot{q}(t) \rangle_{q(t)} = k \ , \tag{A.15}$$

where k is non-negative. The length of q is therefore equal to

$$\int_0^1 \sqrt{\langle \dot{q}(t), \dot{q}(t) \rangle_{q(t)}} dt = \sqrt{2k} \ , \tag{A.16}$$

so that $\delta = \sqrt{2k}$. Therefore, Eq. Eq. (A.13) rewrites:

$$\delta = \sqrt{2k} \iff p^T I^{-1}(q) p = \delta^2. \tag{A.17}$$

Taking equation Eq. (A.17) at initial state $t = 0$, we can determine all the initial velocity such that $d_F(q(0), q(1)) = \delta$. Those velocities are needed to solve the ODE system Eq. (A.14) and compute the geodesics.

Generally, computing the geodesic between two given distributions is a challenging problem. Methods relying on shooting algorithms have been developed in that matters. Our framework overcomes this problem as we compute the entire Fisher sphere. In the next section, we focus on numerical methods for computing geodesics by solving the systems of ODE Eq. (A.14). These methods are illustrated by computing Fisher spheres in the Gaussian manifold $\mathcal{S} = \{\mathcal{N}(\mu, \sigma^2), (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*\}$.

A.3.3 Numerical results

The Hamilton equations Eq. (A.14) are solved with numerical approximation methods. Figure A.3 illustrates our numerical resolution method in the Gaussian case, that is when $\mathcal{S} = \{\mathcal{N}(\mu, \sigma^2), (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*\}$. In order to solve Eq. (A.14), we compare two different numerical methods: namely, the explicit Euler algorithm and the Adams-Moulton algorithm. We recall that in the Gaussian case we have at our disposal an exact analytical expression of the Fisher sphere detailed in Costa et al. (2012). The Fisher sphere is centered in $\mathcal{N}(0, 1)$ with radius $\delta = 1$. Notice that there is no observable difference between the two methods in Figure A.3. Hence, a better way to estimate the numerical error is required. We recall that the Hamiltonian value is conserved along the geodesics. Therefore, it is possible to quantify the performance of the numerical approximation by computing the value $\Delta(t) = \dfrac{H(p(t), q(t)) - H(p(0), q(0))}{H(p(0), q(0))}$ for $t \in [0, 1]$. Δ represents the relative variation of the Hamiltonian along the path q computed with our numerical methods.

Figure A.4 displays the value of $\Delta(t)$ for $t \in [0, 1]$ for one arbitrary geodesic shown in Figure A.3. The relative error for the Adams Moulton method is negligible while the maximum relative error for the explicit Euler scheme is around 0.3%. Hence, in the Gaussian case the Adams Moulton scheme is preferred. Nevertheless, some instabilities have been observed in practice mainly due to the truncation of the distribution support which impair the Hamiltonian consistency. Symplectic method (Amari and Nagaoka, 2000; Leimkuhler and Reich, 2005) and more particularly symplectic Euler algorithm could help to assess this problem by forcing the Hamiltonian constant. This will be the subject of a future work. Moreover, the truncation can lead to other numerical errors when the radius δ is too large. Indeed, the normalization factor of some truncated distribution can become smaller than the computer machine precision.

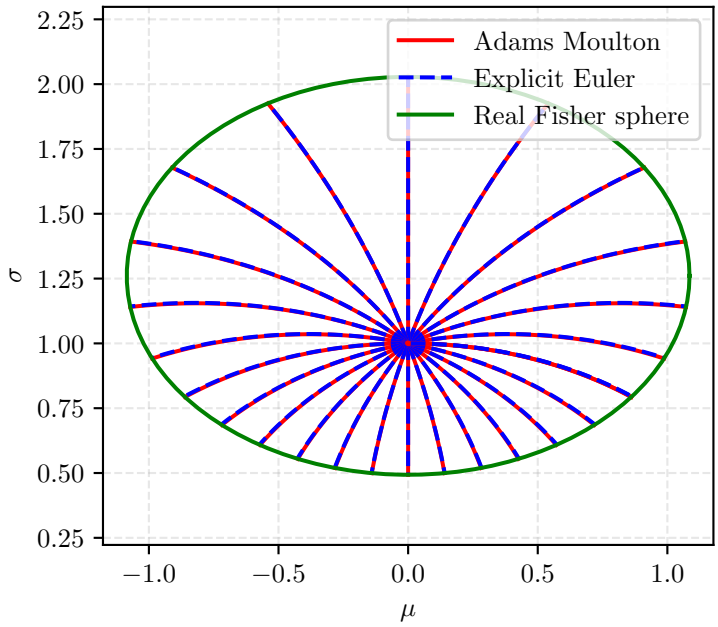


Figure A.3: Geodesics in the Gaussian information geometry computed with Euler explicit and Adams Moulton methods. The radius δ is equal to 1.

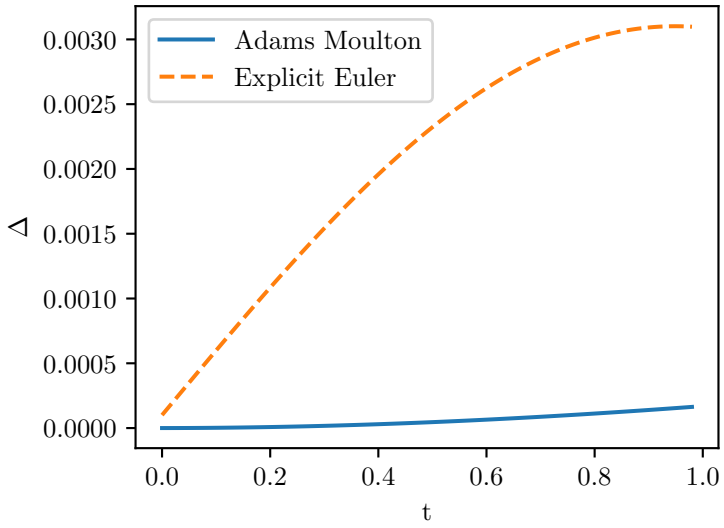


Figure A.4: Relative variation of the Hamiltonian Δ along a geodesic for two different numerical scheme.

A.4 Application to Perturbed-Law based Indices

The UQ robustness analysis explained in Section A.1 and Section A.2 aims at quantifying the impact of a lack of knowledge on an input distribution in UQ of model outputs. In Section A.3, a coherent formal definition of density perturbation has been proposed. We now illustrate the interest of this solution for the definition of a practical robustness analysis methodology. Analyzing the effect of perturbing an input density first requires defining an index which summarizes this effect on the QoI.

A.4.1 Definition of the Perturbed-Law based Index

A PLI aims to measure the impact of the modification of an input density on some events affecting the QoI such as a quantile or a threshold exceedance probability of the model output (Lemaître et al., 2015; Sueur et al., 2016). In the following, we focus on a quantile of order α, which is often used in practical applications as a risk measure (Mousseau and Williams, 2017; Delage et al., 2018; Larget, 2019).

Given the random vector $\mathbf{X} = (X_1, ..., X_d) \in \mathcal{X}$ of our d independent uncertain input variables, $G(\cdot)$ our numerical model and $Y = G(\mathbf{X}) \in \mathbb{R}$ the model output, the quantile of order α of Y is:

$$q^\alpha = \inf\{t \in \mathbb{R}, F_Y(t) \geq \alpha\} \ , \tag{A.18}$$

where F_Y is the cumulative distribution function of the random variable Y. In order to compute the i-th PLI, we change the density f_i of X_i into a density $f_{i\delta}$, where $\delta \in \mathbb{R}^+$ represents the level of the perturbation. The perturbed quantile then writes:

$$q^\alpha_{i\delta} = \inf\{t \in \mathbb{R}, F_{Y,i\delta}(t) \geq \alpha\} \ , \tag{A.19}$$

where $F_{Y,i\delta}$ is the cumulative distribution function corresponding to the input variable X_i sampled from $f_{i\delta}$. The PLI is then simply defined as the relative change in the output quantile generated by the perturbation :

$$S_{i\delta} = \frac{q^\alpha_{i\delta} - q^\alpha}{q^\alpha} \ . \tag{A.20}$$

This definition slightly differs from the one proposed in previous studies (Lemaître et al., 2015; Sueur et al., 2017). Indeed, after several applications of the PLI, it has been found more convenient to compute directly the relative variation of the quantile when submitted to a density perturbation. The interpretation is straightforward.

In a lot of applications, for instance in nuclear safety exercises, the computer models are costly in terms of CPU time and memory. Only a limited number of N code runs is then available for the estimation of all the PLIs. We then have a sample $\mathcal{Y}_N = \{y^{(n)}\}_{1\leq n\leq N}$ of N outputs of the model from a sample $\mathcal{X}_N = \{\mathbf{X}^{(n)} = (x_1^{(n)}, ..., x_d^{(n)})\}_{1\leq n\leq N}$ of N independent realizations of $\mathbf{X}$. The estimation of the quantile is based on the empirical quantile estimator denoted $\widehat{q}^\alpha_N = \inf\{t \in \mathbb{R}, \widehat{F}^N_Y(t) \leq \alpha\}$ where $\widehat{F}^N_Y(t) = \dfrac{1}{N}\displaystyle\sum_{n=1}^{N} \mathbb{1}_{(y^{(n)}\leq t)}$ is the empirical estimator of the cumulative density function of Y. In order to estimate the perturbed quantile $\widehat{q}^\alpha_{N,i\delta}$ from the same sample $\mathcal{X}_N$, we use the so-called reverse importance sampling mechanism from (Hesterberg, 1996) to compute $\widehat{F}^N_{Y,i\delta}$ (Delage et al., 2018):

$$\widehat{F}^N_{Y,i\delta}(t) = \frac{\displaystyle\sum_{n=1}^{N} L_i^{(n)} \mathbb{1}_{(y^{(n)}\leq t)}}{\displaystyle\sum_{n=1}^{N} L_i^{(n)}} \ , \tag{A.21}$$

with $L_i^{(n)}$ the likelihood ratio $\dfrac{f_{i\delta}(x_i^{(n)})}{f_i(x_i^{(n)})}$. The estimator of the PLI is then

$$\widehat{S}_{N,i\delta} = \frac{\widehat{q}^\alpha_{N,i\delta} - \widehat{q}^\alpha_N}{\widehat{q}^\alpha_N} \ . \tag{A.22}$$

As presented in Section A.3, the Fisher sphere of radius δ and centered in the initial input distribution f_i, denoted by $\partial\mathcal{B}_F(f_i,\delta) = \{g,\ d_F(f_i,g) = \delta\}$, provides a good definition for perturbing distributions. This means that we do not consider one specific perturbation at level δ, but a whole set of perturbed distributions $\partial\mathcal{B}_F(f_i,\delta)$. Over this set, we compute the maximum $S_{i\delta}^+$ and the minimum $S_{i\delta}^-$ of the PLI for any distributions in $\partial\mathcal{B}_F(f_i,\delta)$:

$$S_{i\delta}^+ = \max_{g\in\partial\mathcal{B}_F(f_i,\delta)} S_i(g)\ , \tag{A.23}$$

$$S_{i\delta}^- = \min_{g\in\partial\mathcal{B}_F(f_i,\delta)} S_i(g)\ , \tag{A.24}$$

where $S_i(g)$ is the PLI with g as the perturbed density for the variable X_i.

Among all perturbed distributions at level δ, we investigate the one that deviates the quantile the most from its original value. Thus, these two quantities $S_{i\delta}^+$ and $S_{i\delta}^-$ measure the robustness of the numerical code taking into account uncertainties tainting the input distribution.

A.4.2 Theoretical properties of the estimator

In this section, we investigate some theoretical aspects of the PLI estimator $\widehat{S}_{N,i\delta}$. As it is based on the quantile estimators, we first focus on the asymptotic properties of the estimator $\left(\widehat{q}_N^\alpha, \widehat{q}_{N,i\delta}^\alpha\right)$. Detailed proof of the following results are reported in the Section A.7.

Theorem A.1. *Suppose that F_Y is differentiable at $q^\alpha = F_Y^{-1}(\alpha)$ with $F_Y'(q^\alpha) > 0$ and that $F_{Y,i\delta}$ is differentiable at $q_{i\delta}^\alpha = F_{Y,i\delta}^{-1}(\alpha)$ with $F_{Y,i\delta}'(q_{i\delta}^\alpha) > 0$. We denote $\Sigma = \begin{pmatrix} \sigma^2 & \tilde{\theta}_i \\ \tilde{\theta}_i & \tilde{\sigma}_{i\delta}^2 \end{pmatrix}$ with*

$$\sigma_i^2 = \frac{\alpha(1-\alpha)}{F_Y'(q^\alpha)^2}\ , \tag{A.25}$$

$$\tilde{\sigma}_{i\delta}^2 = \frac{\mathbb{E}\left[\left(\frac{f_{i\delta}(X_i)}{f_i(X_i)}\right)^2 (\mathbb{1}_{(G(\boldsymbol{X})\le q_{i\delta}^\alpha)} - \alpha)^2\right]}{F_{Y,i\delta}'(q_{i\delta}^\alpha)^2}\ ,$$

$$\tilde{\theta}_i = \frac{\mathbb{E}\left[\frac{f_{i\delta}(X_i)}{f_i(X_i)}\mathbb{1}_{(G(\boldsymbol{X})\le q^\alpha)}\mathbb{1}_{(G(\boldsymbol{X})\le q_{i\delta}^\alpha)}\right] - \alpha\mathbb{E}[\mathbb{1}_{(G(\boldsymbol{X})\le q_{i\delta}^\alpha)}]}{F_Y'(q^\alpha)F_{Y,i\delta}'(q_{i\delta}^\alpha)}\ .$$

Suppose that the matrix Σ is invertible and $\mathbb{E}\left[\left(\frac{f_{i\delta}(X_i)}{f_i(X_i)}\right)^3\right] < +\infty$. Then

$$\sqrt{N}\left(\begin{pmatrix}\widehat{q}_N^\alpha \\ \widehat{q}_{N,i\delta}^\alpha\end{pmatrix} - \begin{pmatrix} q^\alpha \\ q_{i\delta}^\alpha\end{pmatrix}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma)\ .$$

The PLI $S_{i\delta}$ is a straightforward transformation of the joint distribution $(q^\alpha, q_{i\delta}^\alpha)^T$. To obtain the almost sure convergence of $\widehat{S}_{N,i\delta}$ to $S_{i\delta}$, it suffices to apply the continuous-mapping theorem to the function $s(x,y) = \dfrac{y-x}{x}$.

Theorem A.2. *Given the assumptions of theorem A.1, we have*

$$\sqrt{N}(\widehat{S}_{N,i\delta} - S_{i\delta}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, d_s^T\Sigma d_s) \text{ with } d_s = \begin{pmatrix} -q^\alpha/{q_{i\delta}^\alpha}^2 \\ 1/q^\alpha \end{pmatrix}. \tag{A.26}$$

Notice that the asymptotic variance relies on the α initial quantile and perturbed quantile, which are precisely what we want to estimate. Hence, Theorem A.2 cannot be used for building asymptotic confidence intervals. However, the convergence properties are important for the method credibility and acceptance. In practice, the estimation error can be measured using bootstrapping (Efron, 1979).

A.4.3 Practical implementation of the methodology

As already discussed, in practical applications, the computer model is often costly and cannot be reevaluated. The main limitation of the previously exposed estimator $\widehat{S}_{N,i\delta}$ arises from the available sample size which is finite. Therefore, at a certain level of perturbation, there might not be enough sample points to correctly compute the perturbed quantile (and its confidence interval). One of the key issue of the methodology is to determine how far the input distribution should be perturbed. We propose to adapt the empirical criterion from Iooss et al. (2020) in order to establish a maximal perturbed level δ_{max}. The number of points in the output sample $\mathcal{Y}_N$, smaller or larger than the δ-perturbed quantile has to be sufficient. A value of $N_{\mathcal{Y}} = 10$ has been chosen (from several numerical tests) as the smallest size for computing the PLI-quantile. As soon as a distribution on the Fisher sphere exceeds this criteria, the corresponding radius is taken as δ_{max}.

The estimation of the quantity of interest $S^+_{i\delta}$ and $S^-_{i\delta}$ is summarized as follows:

- Choose a level of perturbation δ, an input number $i \in [\![1; d]\!]$ and a sample of K points on the Fisher sphere of radius δ centered in f_i using the numerical method of Section A.4.2.

- For each $\{f^{(k)}_{i\delta}\}_{1\leq k\leq K}$ sampled on the Fisher sphere, estimate $q^{\alpha,(k)}_{i\delta}$ using the reverse importance sampling technique based on the sample $\mathcal{X}_N$. Verify that the number of point in the output sample below or above the perturbed quantile $q^{\alpha,(k)}_{i\delta}$ satisfies the stopping criteria $N_{\mathcal{Y}}$. Then, compute the PLI estimator $\widehat{S}^{(k)}_{N,i\delta}$.

- The estimators $\widehat{S}^+_{N,i\delta}$ and $\widehat{S}^-_{N,i\delta}$ of the quantity of interest $S^+_{i\delta}$ and $S^-_{i\delta}$ are taken as the maximal and minimal value of the PLI sampled on the Fisher sphere $\{\widehat{S}^{(k)}_{N,i\delta}\}$.

We emphasize that this approach only restricts to expensive computer models. Indeed, the bootstrap variance of the estimated quantile with reverse importance sampling tends to become very large as illustrated in Iooss et al. (2020). This is due to the likelihood ratio that punctually explodes. Thus, when dealing with a cheap code, one can directly resample over the perturbed distribution in order to estimate the output quantile. In this situation, there is no limiting level of perturbation δ_{max}.

The code for the new version of the PLI, called OF-PLI (for Optimal Fisher-based PLI) in the following, is available at https://github.com/JeromeStenger/PLI-Technometrics. In future works, the OF-PLI confidence intervals (computed *via* bootstrap) will provide valuable additional information such as confidence intervals. They are not pictured in the following application as it requires at this stage further investigations. The code for computing the old version of the PLI, called E-PLI (for Entropy-based PLI) in the following, is available in the sensitivity package of the R software.

A.5 Perturbed-Law based Indices in engineering studies

The PLI, as defined in the previous sections, allow to assess to what extent the output quantile can be impacted by an error of magnitude δ in the characterization of an input distribution. In the next subsection, we compare in a toy example the newly introduced methodology (OF-PLI) to the previous one (E-PLI). Moreover, as the PLI are based on a change in the input distribution, it differs from global sensitivity measures (Iooss and Lemaître, 2015) which evaluate the effect of input variability for a fixed probabilistic model. To study the potential coherence and divergence between the two approaches, we compare Sobol' indices and OF-PLI results in Section A.5.2 on an analytical model. In the third subsection, we illustrate the use of the OF-PLI as a support in nuclear safety analysis of a pressurized water nuclear reactor.

A.5.1 A toy example: the Ishigami function

The Ishigami function Ishigami and Homma (1990) is used as an example for uncertainty and sensitivity analysis methods, in particular because it exhibits strong non-linearity and non-monotonicity. In this section, we apply the methodology introduced in Section A.4.3 to estimate OF-PLI and compare our results to the E-PLI. The Ishigami function, which takes three input random variables (X_1, X_2, X_3) normally distributed $\mathcal{N}(0,1)$, is defined with the following analytical formula:

$$G(x_1, x_2, x_3) = \sin(x_1) + 7\sin(x_2)^2 + 0.1x_3^4 \sin(x_1)\ . \tag{A.27}$$

We intend to evaluate the impact of a perturbed input distribution to the 95%-quantile. In this simple example where the function is cheap to evaluate, we do not use the reverse importance sampling estimator of the quantile as proposed in Section A.4.3. We rather draw new samples of size $N = 2000$ directly from the perturbed input distributions in order to compute the output perturbed quantile. We chose a number of $K = 100$ trajectories over each Fisher sphere for computing the minimum and maximum of the OF-PLI. The OF-PLI are computed for perturbation levels δ varying in $[0, 0.9]$. We emphasize that the choice $\delta_{max} = 0.9$ is arbitrary. Indeed, there is here no actual limit for the maximal perturbation level as the OF-PLI are computed by resampling from the perturbed distribution. We also compute the 95%-confidence intervals calculated from 50 values of $\widehat{S}^+_{N,i\delta}$ and $\widehat{S}^-_{N,i\delta}$.

The OF-PLI results are depicted in Figure A.5. It appears that the third input has most impact in particular for shifting the quantile to the right. On the other hand, the second input has more impact for perturbing the quantile to the left. Our results coincide to the well known behavior of the Ishigami function in terms both of non-linearity of the model and primary influence of the third input.

Because, the maximum and minimum of the OF-PLI are taken over the Fisher sphere, we depict in Figure A.6 the distribution of the OF-PLI over the Fisher sphere with radius $\delta = 0.9$ of the third input. One can see that in this situation the maximum and minimum are found for respectively high variance and low variance with no change of the mean.

These results are be compared to the E-PLI (see Section A.2). We recall that the inputs are all normally distributed so that their is no need to apply the inversion distribution function. Therefore, perturbing the mean (respectively the variance) of

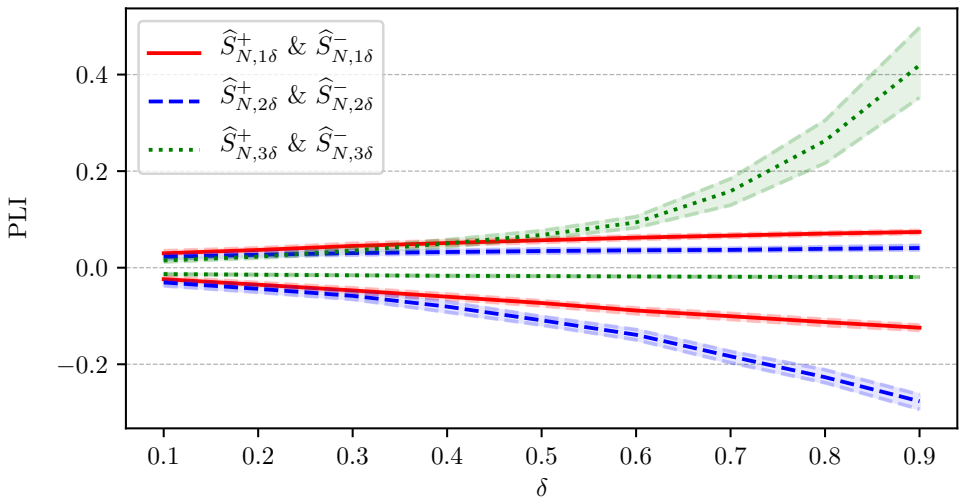


Figure A.5: Minimum and maximum of the OF-PLI over the Fisher sphere over $K = 100$ trajectories for δ varying in $[0, 0.9]$, and their 95%-confidence intervals.

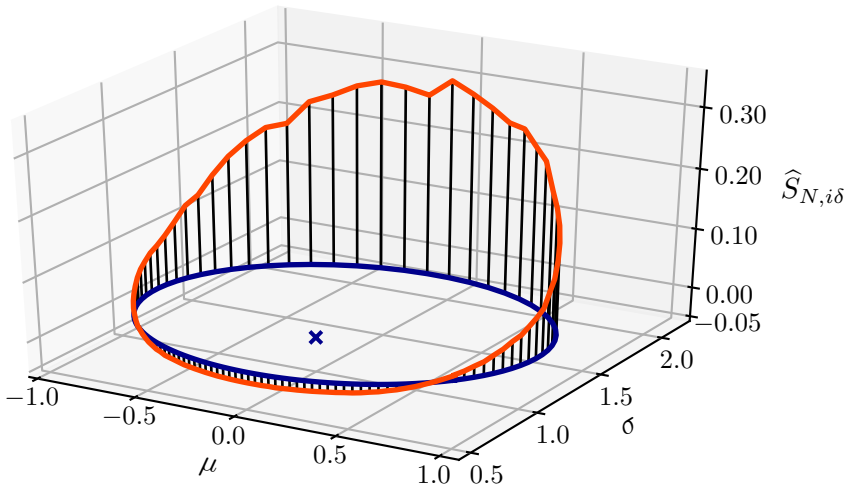


Figure A.6: Value of the OF-PLI $\hat{S}_{N,i\delta}$ (red line) for the third input of the Ishigami model ($N = 100$, $i = 3$) over a Fisher sphere of radius $\delta = 0.9$ (blue line).

the input variable is equivalent to drawing straight horizontal trajectories (respectively vertical) in the parameters space (see Fig. A.3). Results are depicted in Figure A.7, the mean of the Gaussian is perturbed in $[-1, 1]$ and its variance in $[0, 4]$. This corresponds to the range of variation of these parameters for the Fisher sphere radius varying in $[0, 0.9]$. We compare the third input between the two methodology, we detected in Figure A.6 that the maximal OF-PLI was reached for high variance and no mean perturbation which is coherent with the results in Figure A.7. However, one misses the true impact a perturbed density can induce in any situation where the maximal and minimal OF-PLI are not obtained in these two axes, such as, for instance, the first variable. Hence, the E-PLI, restricted to two directions of the Fisher sphere, have limited interpretation.

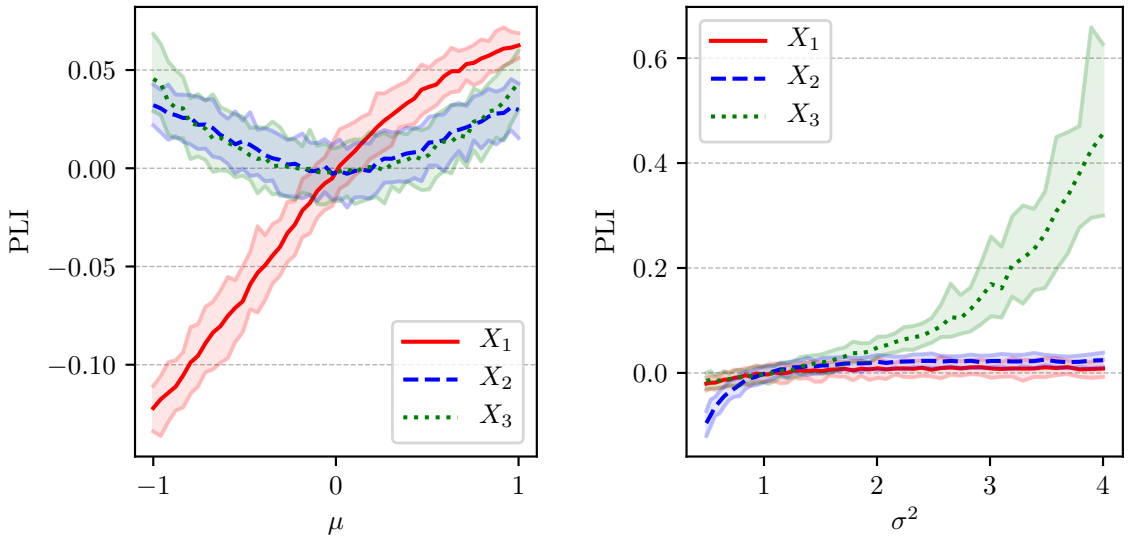


Figure A.7: Computation of the E-PLI. Left: perturbation of the mean of the Gaussian distribution. Right: perturbation of the variance of the Gaussian distribution.

A.5.2 An analytical model: flood risk of an industrial site

The model of interest concerns a flooded river simulation, which is especially useful in assessing the risk of submergence of a dike protecting industrial sites nearby a river. To this purpose, we use a model implementing a simplified version of the 1D hydrodynamical equations of Saint Venant. This model computes H, the maximal annual water level of the river, from four parameters Q, K_s, Z_m and Z_v, which are considered uncertain:

$$H = \left(\frac{Q}{300 K_s \sqrt{2.10^{-4}(Z_m - Z_v)}} \right)^{0.6} . \tag{A.28}$$

The inputs are modeled as random variables with associated truncated distributions given in Table A.1 (Iooss and Lemaître, 2015).

Table A.1: Input variables of the flood model with their associated probability distributions.

Input n°	Name	Description	Probability distribution	Truncation
1	Q	Maximal annual flowrate	Gumbel $\mathcal{G}(1013, 558)$	$[500, 3000]$
2	K_s	Strickler coefficient	Normal $\mathcal{N}(30, 7.5)$	$[15, +\infty]$
3	Z_v	River downstream level	Triangular $\mathcal{T}(50)$	$[49, 51]$
4	Z_m	River upstream level	Triangular $\mathcal{T}(55)$	$[54, 56]$

In global sensitivity analysis, Sobol' indices are the most popular sensitivity measures because they are easy to interpret: each Sobol' index represents a share of the output variance and all indices sum to 1 under assumption of independent inputs (Sobol', 2001; Saltelli and Tarantola, 2002; Prieur and Tarantola, 2015). They will be then compared to the results of our robustness analysis framework in order to illustrate their difference. However, these conventional Sobol' indices focus on the central part of the distribution (variance of the output). We then also compute the target Sobol indices (Marrel and Chabridon, 2020), i.e. Sobol' indices applied to the indicator function of exceeding a given threshold (chosen here as the 95%-quantile of the output).

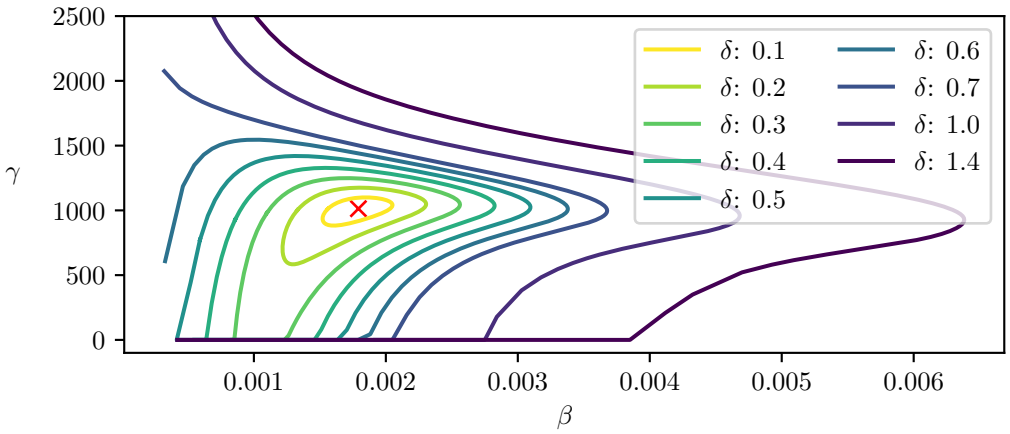
To compute the first order and total Sobol' indices of the inputs of the flood model (Eq. (A.28)), the asymptotically efficient pick-freeze estimator (Prieur and Tarantola, 2015) is used with elementary Monte Carlo matrix of size 10^6. It gives a total cost of $N = 6 \times 10^6$ model runs and a standard deviation of the indices' estimation error smaller than 10^{-3}. As shown in Table A.2, in the central part of the distribution (conventional Sobol' indices), we observe that the variable Q is clearly more influential than the variable K_s whereas Z_v and Z_m appear to have almost no influence on the output. From the target Sobol' indices, we observe that, in the extreme part of the distribution (close to the 95%-quantile), Q and K_s have the same total effect (due to a strong interaction between them in order the output exceeds the threshold).

Table A.2: Sobol' indices estimates of the flood model inputs.

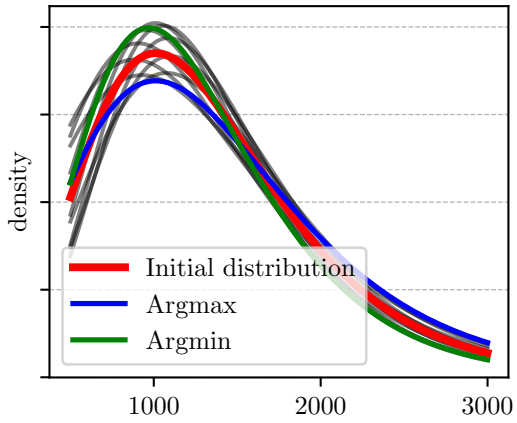
Inputs	Q	K_s	Z_v	Z_m
First-order Sobol' indices	0.713	0.254	0.006	0.006
Total Sobol' indices	0.731	0.271	0.008	0.008
First-order target Sobol' indices	0.242	0.125	0.002	0.002
Total target Sobol' indices	0.867	0.739	0.119	0.121

We compute the OF-PLI (w.r.t. a quantile of order $\alpha = 0.95$) for the flood model inputs with the methodology of Section A.4.3 for increasing Fisher spheres radius $\delta \in [0, 1.4]$ with step 0.1. The spheres are respectively centered on the distributions of Table A.1. On each of these spheres, we compute the OF-PLI for $K = 100$ different perturbed distributions using a sample of $N = 2000$ points distributed according to the initial distribution. The maximal radius $\delta_{max} = 1.4$ derives from the stopping criteria in Section A.4.3. More precisely, the criterion is reached for the first input Q at perturbation level $\delta > 1.4$, meaning there are lower than $N_{\mathcal{Y}} = 10$ sample points above the maximal perturbed quantile. The Figure A.8 depicts how the Fisher sphere centered in the variable Q deforms and how the perturbed densities spread around the initial distribution. Figures A.8b and A.8c indicates that the maximal value of the OF-PLI is obtained by putting weight to the right hand side of the distribution queue (the distributions minimizing and maximizing the OF-PLI are colored green and blue). This behavior was here predictable as the the height river is a growing function of the river flow (see Eq. Eq. (A.28)). However, this analysis can give substantial information in an real world engineering study. At last, one can observe (Fig. A.8a) that the Fisher sphere flatten to the boundary of the parameters' domain. This characteristic is peculiar to each probability distribution, for instance it never not happen for the non-truncated normal distribution.

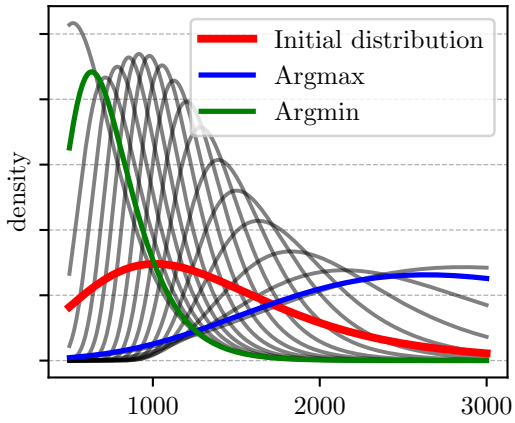
The results of the OF-PLI, displayed in Figure A.9, confirm those of the target Sobol' indices (see Table A.2): the variables 3 and 4, corresponding to Z_v and Z_m, are much less influential on the output quantile of level $\alpha = 0.95$ than the variables 1 and 2, corresponding to Q and K_s. Moreover, perturbations of Q and K_s seem to have comparable effects on the 95%-quantile of H although they have significantly different contributions to the output variance. On the other hand, compared to target Sobol' indices, OF-PLI provide more informative results with their evolution as a function of δ. This clearly shows how a lack of knowledge on an input uncertainty can have a low or high impact on the value of a risk measure. In conclusion, this example confirms the interest of the OF-PLI as it conveys complementary information compared to existing sensitivity indices. Notice that the flat parts visible on some curves are due to approximation errors attributed to the low number of sample points N and high



(a) Deformation of the Fisher sphere for increasing radius δ.



(b) Densities over the Fisher sphere ($\delta = 0.1$).



(c) Densities over the Fisher sphere ($\delta = 1.4$).

Figure A.8: Analysis of the Fisher metric based perturbation of the truncated Gumbel distribution of the variable Q (see Table A.1).

quantile level (0.95).

A.5.3 A nuclear safety case

This industrial application concerns the study of the peak cladding temperature (PCT) of fuel rods in case of loss of coolant accident caused by an intermediate-size break in the primary loop (IB-LOCA) in a nuclear pressurized water reactor. According to operation rules, this temperature must remain below a threshold to prevent any deterioration of the reactor state. The thermal-hydraulic transient caused by this accidental scenario is simulated with the CATHARE2 code (Geffraye et al., 2011), providing a temperature profile throughout time for the surface of the nuclear core assemblies (Mazgaj et al., 2016). The thermal-hydraulic model involves boundary and initial conditions, and many physical parameters (heat transfer coefficient, friction coefficient, etc.) whose exact values are unknown. The probability distributions of these inputs can be obtained from data, from expert knowledge or recovered by solving inverse problems on an experimental database Baccou et al. (2019).

The input uncertainties are propagated inside this model and the UQ objective consists of estimating a high-order quantile of the PCT (model output). This α-quantile

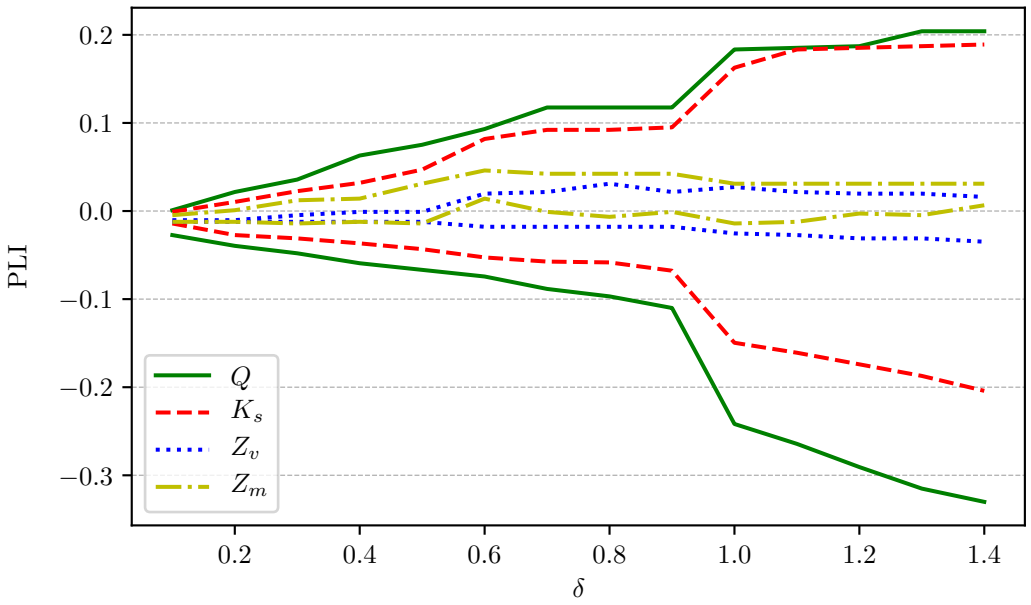


Figure A.9: Maximum and minimum estimated value of the OF-PLI $\widehat{S}^{+}_{N,i\delta}$ and $\widehat{S}^{-}_{N,i\delta}$ for the different variables of the flood model.

is interpreted as a pessimistic estimate of the PCT. Like any scientific approach, this methodology is based on hypotheses, which regulatory authorities ask to evaluate the impact on exhibited results. Indeed, nuclear power operators are required to conduct studies in such a way to ensure that actual risks are overestimated. By this "conservatism principle" they are bound to choose the most pessimistic assumption each time a modeling decision has to be made. In deterministic studies, this simply consists of taking the most penalizing values for each of the input variables. This way, the resulting computation is supposed to simulate a worst case scenario for the examined risk. It is, however, not straightforward to implement such a principle when the numerical code is complex with interactions between inputs and non-monotonic effects of inputs. It is even harder to extend this rationale to a UQ framework aiming to represent all potential scenarios with related occurrence plausibility. Recent works (Larget, 2019) have shown that the E-PLI can be useful to support a discussion on the choice of the input distributions.

In our case, we study a reduced scale mock-up of a pressurized water reactor with 7 uncertain inputs given in Table A.3 (Delage et al., 2018). To compute the OF-PLI, an input-output sample of size $N = 1000$ is available, coming from a space filling design of experiments (Fang et al., 2005) (whose points in $[0,1]^d$ have been transformed to follow the inputs' probability distributions). More precisely, a Latin Hypercube Sample minimizing the L^2-centered discrepancy criterion (Jin et al., 2005) has been used. The OF-PLI (with respect to a quantile of order $\alpha = 0.95$) will then be estimated without any additional code run (see Section A.4.1).

Figure A.10 presents the maximum and minimum values of our two estimators $\widehat{S}^{+}_{N,i\delta}$ and $\widehat{S}^{-}_{N,i\delta}$. We compute Fisher spheres with radius δ sampled uniformly in $[0.1, 0.5]$, all respectively centered on the initial input distributions. On every sphere, $K = 100$ perturbed densities are sampled. The OF-PLIs are finally estimated on a 1000-sized dataset. The stopping criterion of A.4.3 gives a maximal admissible OF-PLI of 4%, this value is determined from the maximal admissible quantile such that there is $N_{\mathcal{Y}} = 10$

Table A.3: Input variables of the CATHARE2 code with their associated probability distributions.

Variable number	Input name	Probability distribution
1	STMFSCO	Uniform $\mathcal{U}([-44.9, 63.5])$
2	STBAEBU	Truncated Log Normal $\mathcal{LN}(0, 0.76)$ on $[0.1, 10]$
3	STOIBC1	Truncated Log Normal $\mathcal{LN}(0, 0.76)$ on $[0.1, 10]$
4	STOIBC3	Truncated Log Normal $\mathcal{LN}(0, 0.76)$ on $[0.1, 10]$
5	STOIDC	Truncated Log Normal $\mathcal{LN}(0, 0.76)$ on $[0.1, 10]$
6	STOICO	Truncated Log Normal $\mathcal{LN}(-0.1, 0.45)$ on $[0.23, 3.45]$
7	CLFBR	Truncated Normal $\mathcal{N}(6.4, 4.27)$ on $[0, 12.8]$

sample points above it. Actually, one can see that $\widehat{S}^{+}_{N,7\delta}$ is close to this maximal admissible value.

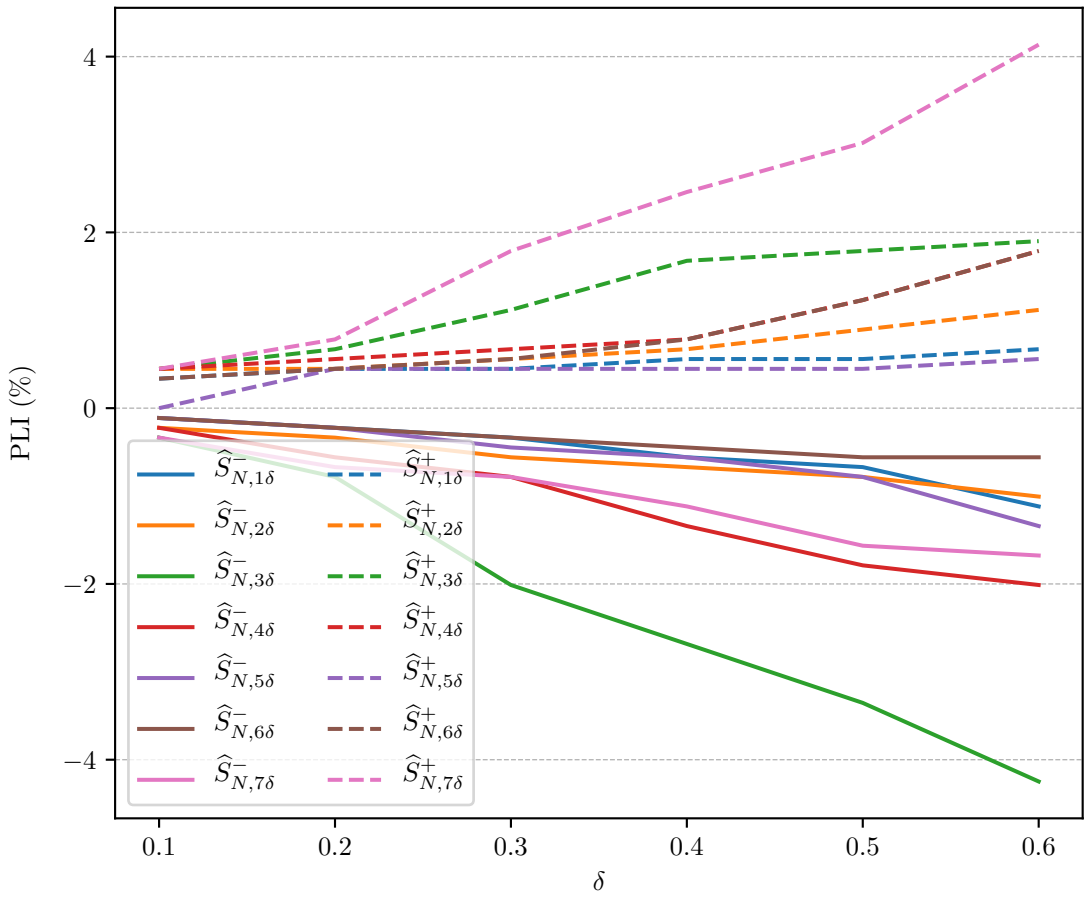


Figure A.10: Bootstrap mean of the maximum and minimum of the OF-PLI $S_{i\delta}$ for the CATHARE2 code. The confidence interval are not shown for the sake of clarity.

Studies previously conducted on the same application (Delage et al., 2018) lead to similar results concerning the most influential inputs on the quantile of the PCT: strong impact of variables 3 and 4 and weak influence of variables 1, 2 and 5. In com-

parison with these studies based on the standard space transformation, our information geometry perturbation methodology leads to a reduced evaluated influence of variable 7. In fact, as it is the only Gaussian distribution, the reverse transformation from the standard space to the physical one operates differently for this input than for the others. Finally, according to the values of $\widehat{S}^{+}_{N,3\delta}$ and $\widehat{S}^{+}_{N,7\delta}$, the variables 3 and 7 appear to be the most influential inputs on the quantile of the PCT. This behavior, which was not observed with the standard space transformation, is probably due to the fact that the standard space approach allows perturbing only one of the probability distribution parameters (for example the expected value). Contrarily, our estimator corresponds to the maximal quantile deviation over a whole set of equivalent perturbations. This shows two main advantages of our newly developed methodology: it prevents the interpretation bias induced by the standard space transformation and it allows for an exhaustive exploration of density perturbations for a given δ.

A.6 Conclusion

Based on the Fisher distance, we have defined an original methodology to perturb input probability distributions in the peculiar case of mutual independent input random variables. The Fisher information is an intrinsic characteristic of probability measure and in particular does not depend on a specific chosen parametric representation. This fundamental property makes it the proper mathematical tool to compare perturbations on different uncertain physical inputs of a computer model, but also on different parameters of the same input distribution. It is even possible to get rid of all references to a parametric sub-domain of the set of probability measures on $\mathcal{X}$, as a non-parametric extension of the Fisher distance is proposed by Holbrook et al. (2017).

Nevertheless, based on the PLI, our method provides useful information on the most influential uncertainties regarding the distributions of input variables, or the so-called "epistemic uncertainties". This is in particular crucial not only in making decisions concerning further research programs aiming at gaining better knowledge about these variables, but also to bring strong backing arguments to operators safety demonstrations. Indeed, we argue that this methodology is adequate for uncertainty studies with poorly reliable input laws identification or when an improved level of robustness is demanded about the choice of input distributions. In the target application (nuclear licensing), our aim is not only to exhibit safety margin values for the simulated accidents but also to prove the methodology as a whole does not induce any risk of underestimating these values. Hence we do not only look for a worst case assessment method, but for a more global understanding of how a potential error on an input's distribution affects the output. In that perspective, a practical option to increase the conservatism of UQ studies is to replace one or several input distributions by penalized deterministic values or by a penalized versions of the distributions themselves. This nevertheless implies to justify the choice of the variables for which this penalization is done (see, e.g., Larget and Gautier (2020)).

Further investigations are still to be completed as this method increases the numerical complexity and the computational time compared to the previous method of Lemaître et al. (2015). Indeed, several Monte Carlo loops are needed to compute the maximal and minimal PLI over Fisher spheres. There is ongoing work about the improvement of the estimation of the maximum and the minimum of the PLI on a Fisher sphere. There is known numerical issue with the reverse important sampling strategy as the likelihood ratio tends to explode as well as the confidence intervals. Moreover

the method consists in sampling trajectories over the Fisher sphere, but one could benefit of a more advanced strategy by optimizing directly the PLI over the sphere, *via* gradient descent along this manifold for instance. The crucial problem of probabilistic dependencies between inputs should also be explored to extend our framework to the non independent-input case, works in robustness analysis dealing with dependent input can be found for instance in Pesenti et al. (2019). Moreover, using a distance in a complex space such as the space of probability density functions instead of a moment perturbation makes our methodology harder to interpret from a physicist's perspective. Thus, it is crucial to clearly define the statistical interpretation of the Fisher distance, i.e. the link with the statistical tests theory. Last but not least, the numerical difficulties illustrated in Section A.3.3 prevents us from having a complete degree of freedom on the δ value.

A.7 Proof of Theorem A.1

We study the consistency and asymptotic normality of specific M and Z-estimators in order to establish the proof of Theorem A.1. We suppose this theory is known so that the details can be kept to the bare minimum. Further readings can be found in Chapters 5.2 and 5.3 of Van der Vaart (2000). Given a sample $(\mathbf{X}^{(n)})_{n\in(1,\dots,N)}$ where $\mathbf{X}$ is a d-dimensional random vector, we define

$$\begin{aligned}
\eta &= \frac{\alpha}{1-\alpha}\ , \\
m_\theta(x) &= -(G(x)-\theta)\mathbb{1}_{(G(x)\leq\theta)} + \eta(G(x)-\theta)\mathbb{1}_{(G(x)>\theta)}\ , \\
M_N(\theta_1,\theta_2) &= \frac{1}{N}\sum_{n=1}^{N} m_{\theta_1}(\mathbf{X}^{(n)}) + \frac{f_{i\delta}(X_i^{(n)})}{f_i(X_i^{(n)})} m_{\theta_2}(\mathbf{X}^{(n)})\ , \\
\hat{\theta}_N &= \arg\max M_N(\theta_1,\theta_2)\ .
\end{aligned} \tag{A.29}$$

$\hat{\theta}_N$ is defined such that its two components correspond respectively to the estimators $\hat{q}_N^\alpha$ and $\hat{q}_{N,i\delta}^\alpha$ of the quantile and the perturbed quantile. The map $\theta \mapsto \nabla_\theta M_N(\theta)$ with $\theta = (\theta_1,\theta_2)^T$ has two non decreasing components (it is a sum of non decreasing maps). Now, by definition of $\hat{\theta}_N$ and concavity of $M_n(\theta)$, it holds that $\nabla_\theta M_N(\hat{\theta}_N) = 0$. Furthermore, we have that $\nabla_\theta M_N(\theta) \xrightarrow{P} ((1+\eta)F_Y(\theta_1)-\eta, ((1+\eta)F_{Y,i\delta}(\theta_2)-\bar{L}_N\eta)^T$ with $\bar{L}_N = \frac{1}{N}\sum_{n=1}^{N}\frac{f_{i\delta}(X_i^{(n)})}{f_i(X_i^{(n)})}$, and this limit is a strictly non decreasing function. Therefore, the assumptions of Lemma 5.10 in (Van der Vaart, 2000, p.47) are satisfied, proving the consistency of the estimator $\hat{\theta}_N \xrightarrow{P} (q^\alpha, q_{i\delta}^\alpha)^T$.

The asymptotic normality is studied via the map $\bar{m}_\theta(x) \mapsto m_{\theta_1}(x) + \frac{f_{i\delta}(x)}{f_i(x)} m_{\theta_2}(x)$ which is Lipschitz for the variable θ with Lipschitz constant $h(x) = \max(1,\eta)\left(1+\frac{f_{i\delta}(x_i)}{f_i(x_i)}\right)$. The function h belongs in L^2 if $\mathbb{E}\left[\left(\frac{f_{i\delta}(X_i)}{f_i(X_i)}\right)^2\right] < +\infty$. The map $\bar{m}_\theta$ is also differentiable in $\theta_0 = \arg\max_{\theta\in\Theta}\mathbb{E}[\bar{m}_\theta(X)]$ with gradient:

$$\nabla_{\theta_0}\bar{m}_{\theta_0}(x) = ((1+\eta)\mathbb{1}_{(G(\mathbf{x})\leq\theta_1)} - \eta,\ \frac{f_{i\delta}(x_i)}{f_i(x_i)}((1+\eta)\mathbb{1}_{(G(\mathbf{x})\leq\theta_2)} - \eta))^T\ . \tag{A.30}$$

Moreover, the map $\theta \to \mathbb{E}[\bar{m}_\theta(\mathbf{X})]$ admits the following Hessian:

$$V_{\theta_0} = \begin{pmatrix} (1+\eta)F'_Y(q^\alpha) & 0 \\ 0 & (1+\eta)F'_{Y,i\delta}(q^\alpha_{i\delta}) \end{pmatrix} , \tag{A.31}$$

which is symmetric definite non negative whenever $F'_Y(q^\alpha) > 0$ and $F'_{Y,i\delta}(q^\alpha_{i\delta}) > 0$. Hence, Theorem 5.23 in (Van der Vaart, 2000, p.53) applies. It proves the asymptotic normality of the estimator $(\hat{q}^\alpha, \hat{q}^\alpha_{i\delta})^T$.

B

Conditional Quantile Sequential Estimation for Stochastic Codes

The content of this appendix is related to the following paper, that I helped correct in the beginning of my PhD:

Labopin-Richard, T., Gamboa, F., Garavier, A. and Stenger, J. 2019. Conditional Quantile Sequential Estimation for Stochastic Codes. In *Journal of Statistical Theory and Practice 13*.

This appendix analyzes an algorithm for the sequential estimation of a conditional quantile in the context of real stochastic codes with vector-valued inputs. The algorithm is based on kk-nearest neighbors smoothing within a Robbins-Monro estimator. We discuss the convergence of the algorithm un- der some conditions on the stochastic code. We provide non-asymptotic rates of convergence of the mean squared error and we discuss the tuning of the algorithm's parameters.

B.1 Introduction

Computer code experiments have encountered, in the last decades, a growing interest among statisticians in several fields (see Santner et al. (2013); Kennedy and O'Hagan (2000); Sacks et al. (1989); Oakley (2004); Jala et al. (2014); Bect et al. (2012) and references therein). In the absence of noise, a numerical black box $g : \mathbb{R}^d \to \mathbb{R}$ maps an *input vector* X to a numerical output $Y = g(X) \in \mathbb{R}$. When the black box does include some randomness, the code is called *stochastic* and the model is as follows: a random vector $\varepsilon \in \mathbb{R}^m$, called *random seed*, models the stochasticity of the function, while X is a random vector. The random seed and the input are assumed to be stochastically independent. The map g (which satisfies some regularity assumption specified below)

is defined on $\mathbb{R}^d \times \mathbb{R}^m$ and outputs

$$Y = g(X, \varepsilon) \,, \tag{B.1}$$

hence yielding possibly different values for the same input X. One observes a sample of pairs (X, Y), without having access to the details of g. In the context of computer experiments, those observations are often expensive (for example when g has a high computational complexity) and one aims at learning rapidly some properties of interest on g.

We focus in this work on the estimation of the conditional quantile of the output Y given the input X. For a given level $\alpha \in [1/2, 1)$ and for every possible input $x \in \mathbb{R}^d$, the target is

$$\theta^*(x) := q_\alpha\Big(g(x, \varepsilon)\Big) \,, \quad x \in \mathbb{R}^d \,,$$

where $q_\alpha(Z) := F_Z^{-1}(\alpha)$ is the quantile of level α of the random variable Z and $F_Z^{-1}(u) := \inf\{x : F_Z(x) \geq u\}$ is the generalized inverse of the cumulative distribution function of Z. Notice that we restrict $\alpha \in [1/2, 1)$ as the case $\alpha \in (0, 1/2]$ can be tackled in the same way considering $-Z$. Our goal is to estimate the conditional quantile for different values of x at the same time.

The algorithm

For a fixed value of x, there are several well-known procedures to estimate the quantile $\theta^*(x)$. Given a sample $(Y_i^x)_{i=1\ldots n}$ of $Y^x := g(x, \varepsilon)$, the empirical quantile is a solution. For a sequential estimation, one may use a Robbins Monro Robbins and Monro (1951) estimator. This method permits to iteratively approximate the zero of a function $h : \mathbb{R} \to \mathbb{R}$ by a sequence of estimators defined by induction: $\theta_0 \in \mathbb{R}^d$ and for all $n \geq 0$,

$$\theta_{n+1} = \theta_n - \gamma_{n+1} H(\theta_n, Z_{n+1}) \,.$$

Here, (γ_n) is the learning rate (a deterministic step-size sequence), (Z_n) is an i.i.d sample of observations, and H is a noisy version of h. Denoting $\mathcal{F}_n := \sigma(Z_1, \ldots Z_n)$ the sigma-field induced by the observations, H is such that

$$\mathbb{E}\Big(H(\theta_n, Z_{n+1}) | \mathcal{F}_n\Big) = h(\theta_n) \,.$$

Classical conditions for the the choice of the step sizes (γ_n) are

$$\sum_n \gamma_n^2 < \infty, \text{ and } \sum_n \gamma_n = \infty \,.$$

These conditions ensure the convergence of the estimates under weak assumptions. For example, convergence in mean squared is studied in Robbins and Monro (1951), almost sure consistency is considered in Blum (1954); Schreck et al. (2016), asymptotic rate of convergence are given in Fabian (1968); Ruppert (1991); Sacks (1958), while large deviations principles are investigated in Woodroofe (1972). There has been a recent interest on non-asymptotic results. Risk bounds under Gaussian concentration assumption (see Frikha and Menozzi (2012)) and finite time bounds on the mean squared error under strong convexity assumptions (see Moulines and Bach (2011); Schreck et al. (2016) and references therein), have been given. Quantile estimation

corresponds to the choice $h : t \mapsto F(t) - \alpha$, where F is the cumulative distribution function of the target distribution. One can show that the estimator

$$\begin{cases} \theta_0 \in \mathbb{R} \\ \theta_{n+1} = \theta_n - \gamma_{n+1} \left(\mathbb{1}_{Z_{n+1} \leq \theta_n} - \alpha \right) \end{cases} \tag{B.2}$$

is consistent and asymptotically Gaussian (see Duflo and Wilson (1997) chapters 1 and 2 for proofs and details). It is important to remind, however, that the lack of strong convexity prevents most non-asymptotic results to be applied directly, except when the density is lower-bounded. We nevertheless mention that Godichon et al. prove in Cardot et al. (2017); Godichon (2016) such non-asymptotic results for the adaptation of algorithm (B.2) to the case where Z is a random variable on an Hilbert space of dimension higher than 2.

Of course, unless x can take a small number of different values, it is not possible to use this algorithm with a sample of Y^x for each possible input value x. Even more, when the code has a high computational complexity, the overall number of observations (all values of x included) must remain small, and we need an algorithm using only one limited sample $(X_i, Y_i)_{i=1\ldots n}$ of (X, Y). Then, the problem is more difficult. For each value of x, we need to estimate quantile of the conditional distribution given x using a *biased* sample. To address this issue, we propose to embed Algorithm Eq. (B.2) into a non-parametric estimation procedure. For a fixed input x, the new algorithm only takes into account the pairs (X_i, Y_i) for which the input X_i is close to x, and thus (presumably) the law of Y_i close to that of Y^x. To set up this idea, we use the k-nearest neighbors method, introducing the sequential estimator:

$$\begin{cases} \theta_0(x) \in \mathbb{R} \\ \theta_{n+1}(x) = \theta_n(x) - \gamma_{n+1} \left(\mathbb{1}_{Y_{n+1} \leq \theta_n(x)} - \alpha \right) \mathbb{1}_{X_{n+1} \in \chi^d_{n+1}(x)} \,, \end{cases} \tag{B.3}$$

where

- $\chi^d_n(x)$ is the subset of $\{X_1, \ldots, X_n\}$ made of the k_n nearest neighbors of x for the euclidean norm on $\mathbb{R}^d$, where $(k_n)_n$ is a fixed sequence. Denoting by $||X - x||_{(i,n)}$ the i-th order statistic of a sample $\left(||X_i - x||\right)_{i=1\ldots n}$ of size n, we have

$$\left\{ X_{n+1} \in \chi^d_{n+1}(x) \right\} = \left\{ ||X_{n+1} - x|| \leq ||X - x||_{(k_{n+1},n)} \right\} .$$

  In this work, we discuss choices of the form $k_n = \lfloor n^\beta \rfloor$ for $0 < \beta < 1$, $n \in \mathbb{N}^*$.

- (γ_n) is the deterministic steps sequence. We focus here on the choice $\gamma_n = n^{-\gamma}$ with $0 < \gamma \leq 1$.

The k-nearest neighbors method of localization first appears in Stone (1976, 1977) for the estimation of conditional expectations. In Bhattacharya and Gangopadhyay (1990), Bhattacharya et al. apply it to the (non-recursive) estimation of the conditional quantile function for real-valued inputs. Regarding the computational cost of the algorithm (B.3), naive implementations of the search for nearest neighbors require $O(n)$ operations at round n, which means that the overall complexity is quadratic. However, the smart use of quad-trees (a hierarchical partition of space) permits to reduce the cost of an iteration to $O(\log(n))$, and in practice the algorithm has almost a linear complexity.

Remark that if the number of neighbors k_n is small, then few observations are used and the estimation is highly noisy; on the other hand, if k_n is large, then values of Y_i may be used that have a distribution significantly different from the target. The challenge is thus to tune k_n so as to reach an optimal balance between bias and variance.

In this work, this tuning is combined with the choice of the learning rate. The main objective of this work is to optimize the choice of the two parameters β and γ of Algorithm Eq. (B.3) that monitor the learning rate γ_n and the number of neighbors k_n. The paper is organized as follows: Section B.2 deals with the stability, and with the almost sure convergence of the algorithm. Furthermore, it contains the main result of our paper: a non-asymptotic inequality on the mean squared error from which an optimal choice of parameters is derived. In Section B.3, we present some numerical simulations to illustrate our results. The technical points of the proofs are deferred to Section B.4, while Section B.6 summarizes the notation and constants used in this paper.

B.2 Main Results

After giving some notation and technical assumptions, we explain in this section how to tune the parameters of the algorithm. We also provide conditions allowing theoretical guarantees of convergence.

B.2.1 Notation

The constants appearing in the sequel are of three different types:

1) (L, U) denote lower- and upper bounds for the support of random variables. They are indexed by the names of those variables;

2) $(N_i)_{i\in\mathbb{N}^*}$ are integers denoting the first ranks after which some properties hold;

3) $(C_i)_{i\in\mathbb{N}^*}$ are positive real numbers used for other purposes.

Without further precision, constants of type 2) and 3) only depend on the model, that is, on g and on the distribution of (ε, X). Further, we denote by $C_i(u)$ or $N_i(u)$, $u \in \mathcal{P}(\{\alpha, x, d\})$ (the power set of a $\{\alpha, x, d\}$), constants depending on the model, on the probability level α, on the point x and on the dimension d. The values of all the constants are summarized in Section B.6.

For any random variable Z, we denote by F_Z its cumulative distribution function. We denote by $\mathcal{B}_x$ the set of the balls of $\mathbb{R}^d$ centred at x. For $B \in \mathcal{B}_x$, we denote by r_B its radius and for $r_B > 0$, we call Y^B a random variable with distribution $\mathcal{L}(Y|X \in B)$.

Remark 5. If the pair (X, Y) has a density $f_{(X,Y)}$ with respect to Lebesgue measure and if the marginal density $f_X(x)$ is positive, then the density of $\mathcal{L}(Y|X = x)$ is

$$f_{Y|X=x} = \frac{f_{(X,Y)}(x, .)}{f_X(x)} ,$$

and when $B = \{x\}$,

$$Y^B \overset{\mathcal{L}}{=} Y^x = g(x, \varepsilon) \sim \mathcal{L}(Y|X = x) .$$

B.2.2 Almost Sure Convergence

In order to prove the convergence of our algorithm, we make two assumptions. The first one, a continuity assumption on the code, can hardly be avoided for our k-nearest neighbors to be valid. The second one is convenient for the simplicity of the analysis.

Assumption A1 For all x in the support of X (that we will denote $\mathrm{Supp}(X)$ in the sequel), there exists a constant $M(x)$ such that the following inequality holds :

$$\forall B \in \mathcal{B}_x,\ \left\| F_{Y^B} - F_{Y^x} \right\|_\infty \leq M(x)\, r_B \ .$$

In words, we assume that the stochastic code is sufficiently smooth. The law of two responses corresponding to two different but close inputs are not completely different. The assumption is clearly required, since we want to approximate the law $\mathcal{L}\left(Y|X=x\right)$ by the law $\mathcal{L}(Y|X \in \chi_n^d(x))$.

Remark 6. If we consider random vector supported by $\mathbb{R}^d \times \mathbb{R}$, we can show that Assumption **A1** holds, for example, as soon as (X,Y) has a regular density with respect to Lebesgue measure. In all cases, it is easier to prove this assumption when the couple (X,Y) has a density: see Subsection 3.1 for an example.

Assumption A2 The law of X has a density with respect to Lebesgue measure, and this density is lower-bounded by a constant $C_{input} > 0$ on $\mathrm{Supp}(X)$.

This hypothesis implies in particular that the law of X has a compact support of volume at most $\frac{1}{C_{input}}$. This kind of assumptions is usual in k-nearest neighbors context (see for example Gadat et al. (2016)). The following theorem studies the almost sure convergence of our algorithm.

Theorem B.1. *Let x and α be fixed. Under Assumptions **A1** and **A2**, Algorithm (B.3) is almost surely convergent whenever $\frac{1}{2} < \gamma \leq \beta < 1$.*

Comments on parameters. In the Theorem B.1, we assume that $0 < \beta < 1$. This means that the number of neighbors goes to $+\infty$ and $||X-x||_{(k_n,n)} \to 0$, as $P(X \in [x-\xi, x+\xi]) > 0,\ \forall \xi > 0$. Obviously, the "localization" condition $k_n/n \to 0$ requires $\beta < 1$: it is quantitatively exploited in Lemma B.9. The condition $\beta \geq \gamma$ can be informally understood in this way. When considering Algorithm (B.2), we deal with the *global learning rate* $\gamma_n = n^{-\gamma}$. In Algorithm (B.3), since for a fixed input x, there is not an update at each step n, one may define an *effective learning rate* γ_{k_n} as follows. At step k, $\theta_k(x)$ has a probability of $\mathbb{P}\Big(X_{n+1} \in \chi_{n+1}^d(x)\Big) \approx k^\beta/k$ to be updated (see Lemma B.6). Up to step n, the estimator is thus updated a number of times approximately equal to

$$N = \sum_{k \leq n} k^{\beta-1} = O\Big(n^\beta\Big) \ .$$

Thus, one has to wait on average up to step $O\Big(n^{\frac{1}{\beta}}\Big)$ in order to reach n updates. Hence, on average, the estimator of the quantile at x evolves with Robbins-Monro iterations roughly equivalent to

$$\theta_{k_n}(x) = \theta_{k_n-1}(x) + \gamma_{k_n}\left(\mathbb{1}_{Y_{k_n} \leq \theta_{k_n}(x)} - \alpha\right) \ ,$$

with the learning rate

$$\gamma_{k_n} = \frac{1}{\left(n^{\frac{1}{\beta}}\right)^{\gamma}} = \frac{1}{n^{\frac{\gamma}{\beta}}} .$$

This is a well-known fact that this algorithm has a *good* behaviour if, and only if, the sum

$$\sum_n \gamma_{k_n} = \sum_n \frac{1}{n^{\frac{\gamma}{\beta}}} ,$$

is divergent. That is if, and only if $\beta \geq \gamma$. At last, the condition $\frac{1}{2} < \gamma \leq 1$ is a classical assumption on the Robbins Monro algorithm to be consistent (see for example in Robbins and Monro (1951)). Here, we restrict the condition to $\gamma < 1$ because we need $1 > \beta \geq \gamma$. The proof of Theorem B.1, in Section B.4, gives rigorous foundations to this heuristic discussion.

B.2.3 Rate of Convergence of the Mean Squared Error

We now study the rate of convergence of the mean squared error $a_n(x) := \mathbb{E}\left((\theta_n(x) - \theta^*(x))^2\right)$. Two rather technical assumptions are required.

Assumption A3 The code function g takes its values in a compact interval $[L_Y, U_Y]$.

Under Assumption **A3**, Lemma B.13 (see Section B.4) explains why if $\beta \geq \gamma$, then $\theta_n(x)$ is almost-surely bounded in an fixed interval $[L_{\theta_n}, U_{\theta_n}]$, and that $|\theta_n(x) - \theta^*(x)|$ is upper-bounded by

$$\sqrt{C_1} := \max\left(U_Y - L_Y + (1-\alpha),\ U_Y + \alpha - L_Y\right) = U_Y - L_Y + \alpha .$$

Assumption A4 For all x, the law of $g(x, \varepsilon)$ has a density with respect to Lebesgue measure which is lower-bounded by a constant $C_g(x) > 0$ on its support.

Lemma B.2. *Denoting* $C_2(x, \alpha) := \min\left(C_g(x), \frac{1-\alpha}{U_Y+\alpha-L_Y},\right)$, *it holds under Assumption* ***A3*** *and* ***A4*** *that for all* n *in* $\mathbb{N}^*$

$$\left[F_{Y^x}(\theta_n(x)) - F_{Y^x}(\theta^*(x))\right]\left[\theta_n(x) - \theta^*(x)\right] \geq C_2(x,\alpha)\left[\theta_n(x) - \theta^*(x)\right]^2. \tag{B.4}$$

Proof. When $\theta_n(x) \in [L_Y, U_Y]$, it is obvious that Inequality (B.4) holds for $C_2 := C_g(x)$. When $\theta_n(x) \in [L_{\theta_n}, L_Y]$, we have

$$L_{\theta_n} \leq \theta_n(x) \leq L_Y \leq \theta^*(x) ,$$

and then $F_{Y^x}(\theta_n(x)) = 0$. Thus,

$$\begin{aligned}
(\theta_n(x) - \theta^*(x))(F_{Y^x}(\theta_n(x)) - F_{Y^x}(\theta^*(x))) &= (\theta_n(x) - \theta^*(x))^2 \frac{(0-\alpha)}{\theta_n(x) - \theta^*(x)} \\
&= (\theta_n(x) - \theta^*(x))^2 \frac{\alpha}{\theta^*(x) - \theta_n(x)} \\
&\geq (\theta_n(x) - \theta^*(x))^2 \frac{\alpha}{U_Y + \alpha - L_Y} \\
&\geq (\theta_n(x) - \theta^*(x))^2 \frac{1-\alpha}{U_Y + \alpha - L_Y} \\
&\geq C_2(x,\alpha)(\theta_n(x) - \theta^*(x))^2 .
\end{aligned}$$

The last case $\theta_n(x) \in [U_Y, U_{\theta_n}]$ can be treated similarly, using that $C_2(x,\alpha) \leq \frac{1-\alpha}{U_Y+\alpha-L_Y}$. □

This lemma is useful to deal with non-asymptotic inequality for the mean squared error. It is the substitute of the strong convexity assumption on the function to minimize, which is often made in the analysis of Robins-Monro stochastic approximation (see for example inMoulines and Bach (2011)) but which does not hold for quantile estimation.

Theorem B.3. *Under hypothesis **A1**, **A2**, **A3** and **A4**, the mean squared error $a_n(x)$ of the algorithm (B.3) satisfies the following inequality : $\forall(\gamma,\beta,\zeta)$ such that $0<\gamma\leq\beta<1$ and $1>\zeta>1-\beta$, $\forall n > N_0 := 2^{\frac{1}{\zeta-(1-\beta)}}$,*

$$a_n(x) \leq \exp\left(-2C_2(x,\alpha)(\kappa_n-\kappa_{N_0})\right) C_1 + \sum_{k=N_0+1}^{n} \exp\left(-2C_2(x,\alpha)\left(\kappa_n-\kappa_k\right)\right) d_k + C_1 \exp\left(-\frac{3n^{1-\zeta}}{8}\right),$$

where for $j\in\mathbb{N}^*$, $\kappa_j = \sum_{i=1}^{j} i^{-\zeta-\gamma}$ and

$$d_n = C_1 \exp\left(-\frac{3n^{1-\zeta}}{8}\right) + 2\sqrt{C_1}M(x)C_3(d)\gamma_n\left(\frac{k_n}{n}\right)^{\frac{1}{d}+1} + \gamma_n^2\frac{k_n}{n}.$$

Here, $C_3(d)>0$ is a constant depending on the dimension d and on the distribution of X (as recalled in Apprendix B.6).

Sketch of proof : Following Moulines and Bach (2011), the idea of the proof is to establish a recursive inequality on $a_n(x)$, that is for $n \geq N_0$,

$$a_{n+1}(x) \leq a_n(x)(1-c_{n+1}) + d_{n+1}$$

where for all $n\in\mathbb{N}^*$, $0<c_n<1$ and $d_n>0$. We use the technical Lemma B.12. In this purpose we begin by expanding the square

$$\begin{aligned}(\theta_{n+1}(x)-\theta^*(x))^2 = (\theta_n(x)-\theta^*(x))^2 &+ \gamma_{n+1}^2\left[(1-2\alpha)\mathbb{1}_{Y_{n+1}\leq\theta_n(x)}+\alpha^2\right]\mathbb{1}_{X_{n+1}\in\chi^d_{n+1}(x)}\\ &- 2\gamma_{n+1}(\theta_n(x)-\theta^*(x))\left(\mathbb{1}_{Y_{n+1}\leq\theta_n(x)}-\alpha\right)\mathbb{1}_{X_{n+1}\in\chi^d_{n+1}(x)}.\end{aligned}$$

Taking the expectation conditionally to $\mathcal{F}_n := \sigma(X_1,\ldots,X_n,Y_1,\ldots,Y_n)$, using $(1-2\alpha)\mathbb{1}_{Y_{n+1}\leq\theta_n(x)}+\alpha^2\leq 1$ and $\alpha = F_{Y^x}(\theta^*(x))$, we obtain thanks to the Bayes formula that

$$\begin{aligned}\mathbb{E}_n\left((\theta_{n+1}(x)-\theta^*(x))^2\right) \leq \mathbb{E}_n\left((\theta_n(x)-\theta^*(x))^2\right) &+ \gamma_{n+1}^2 P_n\\ &- 2\gamma_{n+1}\left(\theta_n(x)-\theta^*(x)\right)\\ &\times P_n\left[F_{Y^{B_n^{k_{n+1}}(x)}}(\theta_n(x)) - F_{Y^x}(\theta^*(x))\right],\end{aligned} \tag{B.5}$$

where $P_n := \mathbb{P}_n\left(X_{n+1}\in\chi^d_{n+1}(x)\right)$ and $B_n^{k_{n+1}}(x)$ is the ball of $\mathbb{R}^d$ centred in x and of radius $||X-x||_{(k_{n+1},n)}$. We rewrite this inequality so as to highlight the presence of two different contributions to the risk:

1) First, the quantity $F_{Y^{B_n^{k_{n+1}}(x)}}(\theta_n(x)) - F_{Y^x}(\theta_n(x))$ represents the *bias error* (due to the use of a biased sample of F_{Y^x}). Using Assumption **A1**, it can be upper-bounded as

$$|F_{Y^{B_n^{k_{n+1}}(x)}}(\theta_n(x)) - F_{Y^x}(\theta_n(x))| \leq M(x)||X - x||_{(k_{n+1},n)} \ .$$

Moreover, by Assumption **A3**, $|\theta_n(x) - \theta^*(x)| \leq \sqrt{C_1}$. Thus,

$$\left|2\gamma_{n+1}(\theta_n(x) - \theta^*(x))P_n\left[F_{Y^{B_n^{k_{n+1}}(x)}}(\theta_n(x)) - F_{Y^x}(\theta_n(x))\right]\right| \leq 2\gamma_{n+1}\sqrt{C_1}M(x)P_n||X - x||_{(k_{n+1},n)} \ .$$

2) The second quantity, $F_{Y^x}(\theta_n(x)) - F_{Y^x}(\theta^*(x))$ represents the *on-line learning error* (due to the use of a stochastic optimization algorithm). Thanks to Assumption **A4** we obtain

$$(\theta_n(x) - \theta^*(x))\left[F_{Y^x}(\theta_n(x)) - F_{Y^x}(\theta^*(x))\right] \geq C_2(x,\alpha)\left[\theta_n(x) - \theta^*(x)\right]^2 \ .$$

Taking the expectation in Inequality Eq. (B.5) yields

$$\begin{aligned} a_{n+1}(x) \leq a_n(x) &- 2\gamma_{n+1}C_2(x,\alpha)\mathbb{E}\left[(\theta_n(x) - \theta^*(x))^2 P_n\right] + \gamma_{n+1}^2\mathbb{E}(P_n) \\ &+ 2\gamma_{n+1}M(x)\sqrt{C_1}\mathbb{E}(||X - x||_{(k_{n+1},n)}P_n) \ . \end{aligned}$$

This inequality reveals a problem : thanks to Lemmas B.6 and B.11 (and thus thanks to assumption **A2**) we can deal with the last two terms, but we are not able to evaluate directly $\mathbb{E}\left[(\theta_n(x) - \theta^*(x))^2 P_n\right]$. In order to solve this problem, we use a truncation parameter ζ_n. Instead of writing a recursive inequality on $a_n(x)$ we write such inequality with the quantity $b_n(x) := \mathbb{E}\left[(\theta_n(x) - \theta^*(x))^2 \mathbb{1}_{P_n > \zeta_n}\right]$. Choosing $\zeta_n = n^{-\zeta}$, we have to tune another parameter but thanks to **A3** and deviation inequalities recalled in Lemma B.9, we obtain a recursive inequality on $a_n(x)$ from the one on $b_n(x)$, for $n \geq N_0$.

Comments on the parameters. We choose $0 < \beta < 1$ for the same reasons as in Theorem B.1. Regarding γ, the inequality is true for all $0 < \gamma \leq \beta$ (which is unusual, as you can see in Godichon (2016) for example). We will nevertheless see in the sequel that this is not because the inequality is true that $a_n(x)$ converges to 0. We will discuss later *good* choices for (γ, β).

Compromise between the two errors. This analysis emphasizes the necessity of a compromise on β to deal with the two previous errors. Indeed,

- the *bias error* gives the term

$$\exp\left(-2C_2(x,\alpha)(x)\sum_{k=N_0+1}^{n}\frac{1}{k^{\zeta+\gamma}}\right) \ ,$$

of the inequality. This term decreases to 0 if and only if $\gamma + \zeta < 1$ which implies $\beta > \gamma$. It suggests that β should not be chosen too small.

- the *on-line learning error* gives the term $(k_n/n)^{1/d+1} = n^{-(1-\beta)(1+1/d)}$ in the remainder. For the remainder to decrease to 0 with the faster rate, we then need β to be as small as possible compared to 1. It suggests that β should not be too large.

The rate of convergence of the mean squared error can be deduced from this theorem. We study the order of the remainder d_n in order to exhibit the dominating terms. It appears that d_n is the sum of three terms. The first one, with a exponential decay, is always negligible as soon as n is large enough, since $1 > \zeta$. The two other are powers of n. Comparing their exponent, we can find the dominating term in function of γ and β. Actually, there exists a rank $N_1(x, d)$ and some constants C_5 and $C_6(x, d)$ such that, for $n \geq N_0 + 1$,

- if $\beta \leq 1 - d\gamma$, then $d_n \leq C_5 n^{-2\gamma+\beta-1}$,
- if $\beta > 1 - d\gamma$, then $d_n \leq C_6(x, d) n^{-\gamma+(1+\frac{1}{d})(\beta-1)}$.

Plugging these inequalities into Theorem B.3 leads to the following result.

Corollary B.4. *Under assumptions of Theorem B.3, there exist ranks $N_4(x, \alpha, d)$ and constants $C_7(x, \alpha, d)$ and $C_8(x, \alpha)$ such that for all $n \geq N_4(x, \alpha, d)$,*

- *when $\beta > 1 - d\gamma$ and $1 - \beta < \zeta < \min\left(1 - \gamma, \left(1 + \frac{1}{d}\right)(1 - \beta)\right)$,*

$$a_n(x) \leq \frac{C_7(d, x, \alpha, \zeta, \gamma)}{n^{-\zeta+\left(1+\frac{1}{d}\right)(1-\beta)}} ;$$

- *when $\beta \leq 1 - d\gamma$, and $\zeta > \max(\beta - \gamma, \gamma - 1)$,*

$$a_n(x) \leq \frac{C_8(x, \alpha)}{n^{\gamma-\beta+1-\zeta}} .$$

Remark 7. For other values of γ and β, the derived inequalities do not imply the convergence to 0 of $a_n(x)$.

From this corollary, the *optimal* choices for (β, γ) can be derived, or more precisely parameters for which our upper-bound on the mean squared error decreases with the fastest rate.

Corollary B.5. *Under the same assumptions as in Theorem B.3, the optimal choice is $\gamma = \frac{1}{1+d}$ with $\zeta > \beta - \frac{1}{1+d} > 0$ as small as possible. With such parameters, there exists a constant $C_9(x, \alpha, d)$ such that $\forall n \geq N_4(x, \alpha, d)$,*

$$a_n(x) \leq \frac{C_9(x, \alpha, d)}{n^{\frac{2}{1+d}+\frac{1-\beta-\zeta}{2}-\beta}} .$$

Comments on the constant $\mathbf{C_9(x, \alpha, d)}$. Like all the other constants of this paper, we know the explicit expression of $C_9(x, \alpha, d)$. For a numerical example, see Subsection B.3.1.

Notice that the constant $C_9(x, \alpha, d)$ depends on x only through the lower bound $C_g(x)$ and the smoothness parameter $M(x)$. Often, $C_g(x)$ and $M(x)$ do not really

Table B.1: Expected precision for the MSE when $N = 1000$

d	1	2	3
η_ζ=0.3	0.088	0.28	0.5
η_ζ=0	0.031	0.1	0.17

depend on x (see for example Subsection 3.1). In these cases (or when we can easily find a bound of $C_g(x)$ and $M(x)$ which do not depend on x), our result is uniform in x. Then, it is easy to deal with the integrated mean squared error and conclude that

$$\int_X a_n(x) f_X(x) dx \leq \frac{C_9(\alpha, d)}{n^{\frac{2}{1+d}+\frac{1-\beta-\zeta}{2}-\beta}} .$$

When α increases to 1, we try to estimate an extremal quantile. Then, $C_2(x, \alpha)$ becomes smaller and then $C_9(x, \alpha, d)$ increases: the bound deteriorates. This is because when α is large, the probability to sample on the right side of the quantile is small and the algorithm is less accurate.

Let us now comment on the dependency on the dimension d. The constant $C_9(x, d, \alpha)$ decreases when the dimension d increases. Nevertheless, this tendency to decrease is too small to balance the behavior of the rate of convergence which is in $n^{\frac{-2}{1+d}}$, an illustration of the well-known curse of dimensionality.

Comment on the rank $\mathbf{N_4}(\mathbf{x}, \alpha, \mathbf{d})$. This rank is the maximum of four ranks. There are two kinds of ranks. The ranks $(N_i)_{i\neq 0}$ depend on constants of the problem but are reasonably small, because the largest of them is the rank after which exponential terms are smaller than power of n terms, or smaller power of n terms are smaller than bigger power of n terms. They often appear to be much smaller than N_0, which tends to be the limiting factor relevant for identifying optimal parameters (and at this stage the reasoning is no longer non-asymptotic).

The rank N_0 is completely different. It was introduced in the first theorem because we could not deal with $a_n(x)$ directly. In fact it is the rank after which the deviation inequality, allowing us to use $b_n(x)$, is guaranteed to hold. It depends on the gap between ζ and $1 - \beta$. The optimal ζ to obtain the rate of convergence of the previous corollary is $\zeta = 1 - \beta + \eta_\zeta$ with η_ζ as small as possible. The constant η_ζ appears on the rank N_0 and also on the rate of convergence (under the assumption that $N_4 = N_0$ which is the case most of time)

$$\forall n \geq N_0 = \exp\left(2\eta_\zeta^{-1}\right),\ a_n(x) = \mathcal{O}\left(n^{\frac{-2}{1+d}+\frac{\eta_\zeta}{2}+\beta}\right) .$$

The smaller η_ζ, the faster the rate of convergence, but also the larger the rank after which the inequalities hold.

Let us give an example. For a budget of $N = 1000$ calls to the code, one may choose $\eta_\zeta = 0.3$ for the inequality to be theoretically true for $n = N$. Table B.1 gives the theoretical precision for different values of d and compares it with the ideal case where $\eta_\zeta = 0$.

We can observe that, when $\eta_\zeta > 0$, the precision increases with the dimension faster than when $\eta_\zeta = 0$. Moreover, as soon as $\frac{1}{1+d} < \eta_\zeta/2$ ($d = 6$ for our previous example), the result does not allow to conclude that a_n decreases to 0 with this choice of η_ζ.

Nonetheless, our simulation study (see next section) seem to indicate that this difficulty could be only an artifact of the proof: the introduction of ζ_n is required by

the difficulty to compute $\mathbb{E}\Big[(\theta_n(x)-\theta^*)P_n\Big]$. In practice, the optimal rate of convergence for optimal parameters is reached early (see Section 3).

B.3 Numerical Simulations

In this part we present some numerical simulations to illustrate our results. The following (simplistic) examples are chosen so as to be able to evaluate clearly the strengths and weaknesses of our algorithm: the constants can be computed and the results can be interpreted easily. To begin with, we deal with dimension 1. We study two stochastic codes, differing by their smoothness.

B.3.1 Dimension 1: Square Function

The first toy example is the very smooth code

$$g(X,\varepsilon) = X^2 + \varepsilon$$

where $X \sim \mathcal{U}([0,1])$ and $\varepsilon \sim \mathcal{U}([-0.5,0.5])$. We try to estimate the quantile of level $\alpha = 0.95$ for $x = 0.5$ and initialize our algorithm to $\theta_1 = 0.3$. We first check that our assumptions are fulfilled in this case. The conditional distribution of the output given $X = x$ is $\mathcal{U}\Big([-\frac{1}{2}+x^2;\frac{1}{2}+x^2]\Big)$, and

$$f_{(X,Y)}(u,v) = \mathbb{1}_{[-\frac{1}{2}+u^2,\frac{1}{2}+u^2]}(v)\,\mathbb{1}_{[0,1]}(u)\ .$$

Moreover, the code function g takes its values in the compact set $[L_Y,U_Y] = [-\frac{1}{2};\frac{3}{2}]$. Let us study assumption **A1**. If $a,b>0$ and if $B=[x-a,x+b]$ is an interval containing x, then

$$\begin{aligned}|F_{Y^B}(t) - F_{Y^x}(t)| &\leq \left|\frac{\int_{-\infty}^t\int_B f_{(X,Y)}(z,y)dydz}{\int_B f_X(z)dz} - \int_{-\infty}^t f_{(X,Y)}(x,y)dy\right| \\ &\leq \frac{\int_{-\frac{1}{2}}^t\int_{x-a}^{x+b}\left|\mathbb{1}_{[-\frac{1}{2}+z^2;\frac{1}{2}+z^2]} - \mathbb{1}_{[-\frac{1}{2}+z^2;\frac{1}{2}+z^2]}\right|(y)dzdy}{\mu(B)}\ .\end{aligned}$$

Now, we have to distinguish the cases in function of the localization of t. There are lots of cases, but computations are nearly the same. That is why we will develop only one case here. When $t\in[-\frac{1}{2};x^2-\frac{1}{2}]$, we have

$$\begin{aligned}|F_{Y^B}(t) - F_{Y^x}(t)| &\leq \frac{\int_{x-a}^{x+b}\int_{-\frac{1}{2}}^t\left|\mathbb{1}_{[-\frac{1}{2}+z^2;\frac{1}{2}+z^2]} - \mathbb{1}_{[-\frac{1}{2}+z^2;\frac{1}{2}+z^2]}\right|(y)}{a+b} \\ &= \frac{\int_{x-a}^{x+b}\left(\mathbb{1}_{z\geq x}(0) + \mathbb{1}_{z\leq x}(t-z^2+\frac{1}{2})\mathbb{1}_{z\geq\sqrt{t+\frac{1}{2}}}\right)dz}{a+b} \\ &= \frac{\int_{x-a}^{x}(t+\frac{1}{2}-z^2)dz}{b+a}\ .\end{aligned}$$

There are again two different cases. Since $t\in[-\frac{1}{2};x^2-\frac{1}{2}]$, we always have $(t+\frac{1}{2})^{\frac{1}{2}}\leq x$. But the position of $(t+1/2)^{1/2}$ relative to $(x-a)$ is not always the same. If $t\in[-\frac{1}{2};-\frac{1}{2}(x-a)^2]$, we get

$$
\begin{aligned}
|F_{Y^B}(t) - F_{Y^x}(t)| &\leq \frac{\int_{x-a}^{x+b}(t - z^2 + \frac{1}{2})dz}{b+a} \\
&\leq \left(t + \frac{1}{2}\right) a - \frac{x^3}{3} + \frac{(x-a)^3}{3} \\
&\leq (x-a)^2 a - x^2 a + a^2 x - \frac{a^3}{3} \\
&\leq -a^2 x + \frac{2a^3}{3} \\
&\leq 0 + r_B \times 1^2 \times \frac{2}{3} ,
\end{aligned}
$$

as $0 < a < 1$. Finally, in this case, **A1** is true with $M(x) = 2/3$. We can compute exactly in the same way for the other cases and we always find an $M(x) \leq 2/3$. The assumption **A2** is also satisfied, taking $C_{input} = 1$. We have already explained that assumption **A3** is true for $[L_Y, U_Y] = [-1/2, 3/2]$. Finally assumption **A4** is also satisfied with $C_g(x) = 1$ and $C_2(x, \alpha) = 0.02$.

B.3.1.1 Almost sure convergence

Let us first deal with the almost sure convergence. We plot in Figure B.1, for $(\beta, \gamma) \in [0, 1]^2$, the relative error of the algorithm. Best parameters are clearly in the area $\beta > \gamma \geq 1/2$. We can even observe that for $\beta \approx 1$, $\beta \leq \gamma$ or $\gamma < 1/2$, the algorithm does not converge almost surely (or very slowly). This is in accordance with our theoretical results. Nevertheless, we can observe a kind of continuity for γ around $1/2$: in practice, the convergence becomes really slow only when γ is significantly far away from $1/2$.

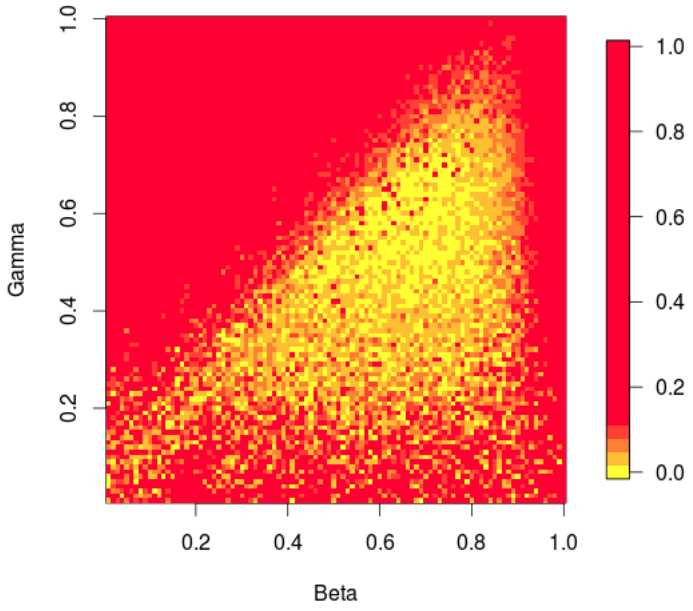


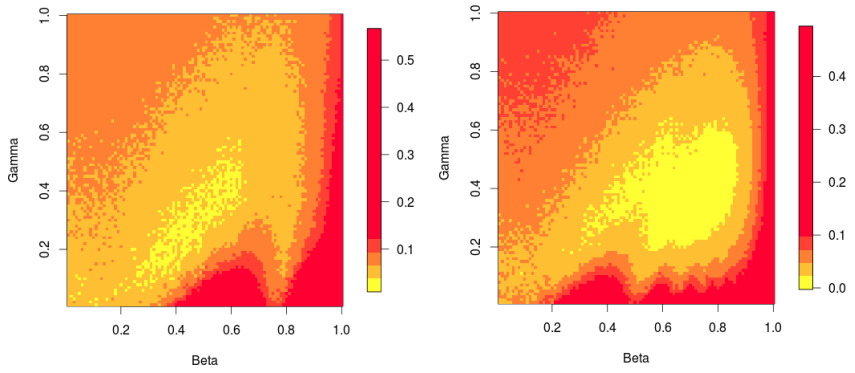
Figure B.1: Relative error for $n = 5000$ dependence on β and γ.

Table B.2: Constant values

Constant	α	$M(x)$	C_{input}	$C_g(x)$	$C_2(x,\alpha)$	$U_Y - L_Y$
Value	0.95	$\frac{2}{3}$	1	1	0.02	2
Constant	$\sqrt{C_1}$	$C_3(d)$	$C_4(d)$	$C_5(x,d)$	$C_6(x,d)$	$C_9(x,d,\alpha)$
Value	2.95	7.39	2	1.95	12	180

B.3.1.2 Mean Square Error (MSE)

Let us study the best choice of β and γ in terms of L^2-convergence. We plot in Figure B.2 the mean squared error in function of γ and β (we estimate the MSE by a Monte Carlo method of 100 iterations).



(a) Mean square error, $n = 50$. (b) Mean square error, $n = 200$.

Figure B.2: Mean squared error in function of β and γ for the square function.

Simulations confirm that the theoretical optimal area $\gamma = 0.5$ and $\beta = \gamma + \eta_\beta$ gives the smallest MSE. Nevertheless, it seems that in practice we can relax the condition that *the gap η_β between β and γ is as small as possible*. Indeed, when η_β is reasonably big, simulations show that we are still in the optimal area.

In this case, we have at hand all the parameters to compute the theoretical bound of our theorems. In particular, in corollary B.5, we get

$$a_n(x) \leq \frac{C_9(x,d,\alpha)}{n^{\frac{1}{1+d}-\eta}} .$$

Table B.2 summarizes the value of the constants needed to compute the theoretical bound in this case.

For $N = 1000$, we obtain the bound $a_N(x) \leq 5.8$ which is over-pessimistic compared to the practical results. We can then think to a way to improve this bound. First of all, the constant $C_2(x,\alpha)$ is in fact not so small. Indeed, we have to take a margin in the proof, for the case where $\theta_n(x)$ goes out of $[L_Y, U_Y]$. This happens only with a very small probability. If we do not take this case into account, we have $C_2(x,\alpha) = 1$. Then $C_9(x,\alpha,d) \approx 3.7$ and then, for $N = 1000$, the bound is 0.11. Practical results are still better (we can observe that for $n = 50$, we already have a MSE inferior to 0.05), but the gap is less important.

B.3.2 Dimension 1 - Absolute Value Function

Let us see what happens when the function g is less smooth with respect to the first variable. We study the code

$$g(X, \varepsilon) = |X| + \varepsilon \ ,$$

where $X \sim \mathcal{U}([-1, 1])$ and $\varepsilon \sim \mathcal{U}([-0.5, 0.5])$. We want to study the conditional quantile in $x = 0$ (the point for which the differentiability fails). Assumptions can be checked as above. Since the almost surely convergence is true and gives really same kind of plots than the previous case, we only study the convergence of the MSE. In that purpose, we plot in Figure B.3 the MSE (estimated by 100 iterations of Monte Carlo simulations) in function of γ and β, for n=300 (the discontinuity constraints us to make more iterations to have a sufficient precision) and $\theta_1 = 0.3$. Conclusions are the same than in the previous example concerning the best parameters. Nevertheless, we can observe that the lack of smoothness implies some remarkable behaviour around $\gamma = 1$.

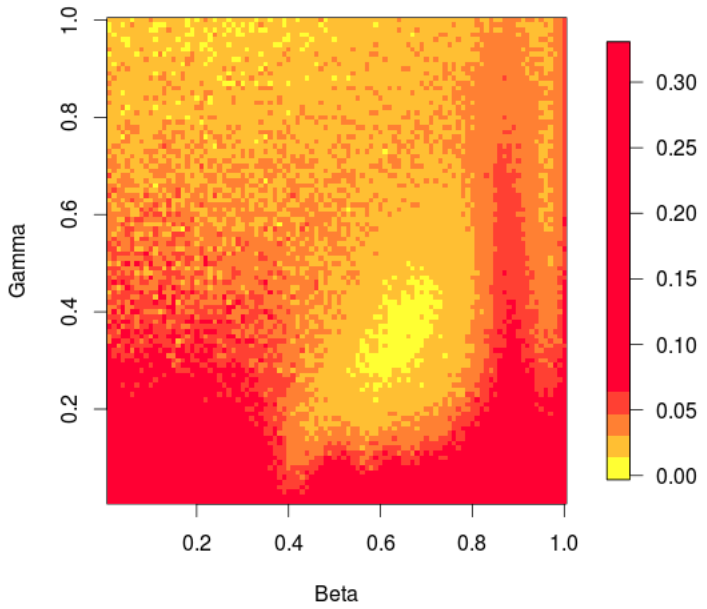


Figure B.3: MSE for $n = 300$ in function of β and γ for absolute value function.

B.3.3 Dimensions 2 and 3

In dimension d, we showed that theoretical optimal parameters are $\gamma = \frac{1}{1+d}$ and $\beta = \gamma + \eta$. To see what happens in practice, we still plot Monte Carlo estimations (200 iterations) of the MSE in function of γ and β.

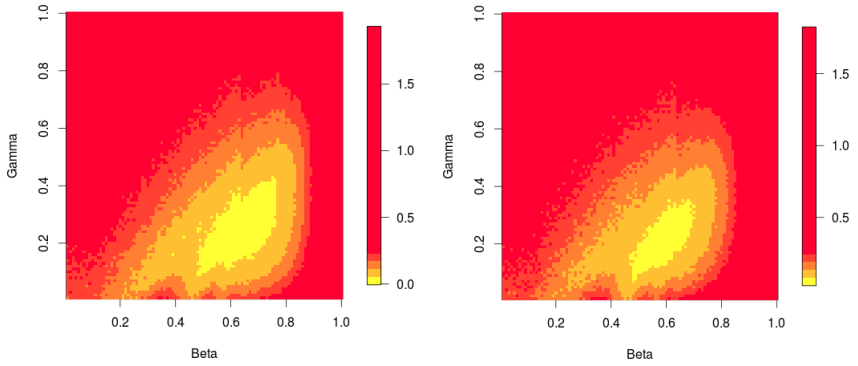
B.3.3.1 Dimension 2

In dimension 2, we study two codes :

$$g_1(X, \varepsilon) = ||X||^2 + \varepsilon \text{ and } g_2(X, \varepsilon) = X_1^2 + X_2 + \varepsilon \ ,$$

where $X = (X_1, X_2) \sim \mathcal{U}([-1, 1]^2)$ and $\varepsilon \sim \mathcal{U}([-0.5, 0.5])$. In each case, we choose $n = 400$ and want to study the quantile in the input point $x = (0, 0)$ and initialize our

algorithm in $\theta_1 = 0.3$. In Figure B.4, we can see that $\beta = 1$ and $\gamma = 1$ are still really bad parameters. As in our theoretical results, $\gamma = \frac{1}{1+d} = \frac{1}{3}$ seems to be the best choice. Nevertheless, even if it is clear that $\beta < \gamma$ is a bad choice, the experiments seems to show that best parameter β is strictly superior to γ, more superior than in theoretical case, where we take β as close as possible of γ. As we said before, in practice, N_0 seems not to be the true limit rank. Indeed, with only $n = 400$ iterations, in this case, the MSE, in the optimal parameters case reaches 0.06.



(a) MSE, $n = 400$, $d = 2$, norm function g_1.

(b) MSE, $n = 400$, $d = 2$, function g_2.

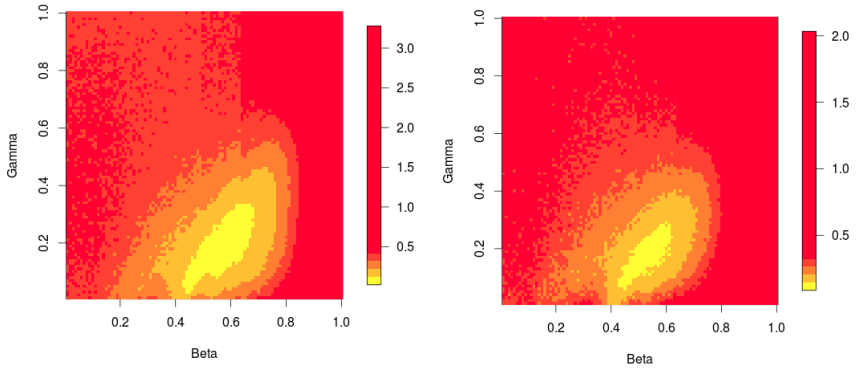
Figure B.4: Mean square error in function of β and γ.

B.3.3.2 Dimension 3

In dimension 3, we study the two codes

$$g_1(X, \varepsilon) = ||X||^2 + \varepsilon \quad \text{and} \quad g_2(X, \varepsilon) = X_1^2 + X_2 + \frac{X_3^3}{2} + \varepsilon \ ,$$

where $X = (X_1, X_2, X_3) \sim \mathcal{U}([-1, 1]^3)$ and $\varepsilon \sim \mathcal{U}([-0.5, 0.5])$. In each case, we choose $n = 500$ and want to study the quantile in the input point $(0, 0, 0)$. The interpretation of Figure B.5 are the same than in dimension 2. The scale is not the same, the convergence is slower again but with $n = 500$ we nevertheless obtain a MSE of 0.10.



(a) MSE, $n = 500$, $d = 3$, norm function g_1.

(b) MSE, $n = 500$, $d = 3$, function g_2.

Figure B.5: Mean squared error in function of β and γ.

B.4 Technical Lemmas and Proofs

B.4.1 Technical Lemmas and Notation

For sake of completeness, we start by recall some well-known facts on order statistics.

Lemma B.6. *When X has a density with respect to Lebesgue measure, denoting $P_n = \mathbb{P}(X \in \chi^d_{n+1}(x)|X_1, \dots X_n)$, we have the following properties*

1) $P_n = F_{||X-x||}\left(||X-x||_{(k_{n+1},n)}\right)$,

2) $P_n \sim \text{Beta}(k_{n+1}, n-k_{n+1}+1)$,

3) $\mathbb{E}(P_n) = k_{n+1}/(n+1)$,

4) $\mathbb{E}(P_n^2) = \left(2k_{n+1}n - k_{n+1}^2 + 3k_{n+1} + k_{n+1}n^2\right)/\left((n+1)^2(n+2)\right)$.

where we denote $F_{||X-x||}$ the cumulative distribution function of the random vector $||X-x||$, $||X-x||_{(k_{n+1},n)}$ the k_{n+1} order statistic of the sample $(||X_1-x||, \dots, ||X_n-x||)$ and $\text{Beta}(a,b)$ *the beta distribution with parameters a and b.*

Proof. Conditionally to $X_1, \dots, X_n$, the event $\{X \in \chi^d_{n+1}(x)\}$ is equivalent to the event $\{||X-x|| \leq ||X-x||_{(k_{n+1},n)}\}$. Then,

$$\begin{aligned} P_n &= \mathbb{P}(X \in \chi^d_{n+1}(x)|X_1 \dots X_n) \\ &= \mathbb{P}_X\left(||X-x|| \leq ||X-x||_{(k_{n+1},n)}|X_1 \dots X_n\right) \\ &= F_{||X-x||}\left(||X-x||_{(k_{n+1},n)}\right) . \end{aligned}$$

Since X has a density, the cumulative distribution function $F_{||X-x||}$ is continuous. Indeed, using the sequential characterization we get for a sequence (t_n) converging to t

$$\begin{aligned} F_{||X-x||}(t_n) &= \mathbb{P}(X \in B_d(x,t_n)) \\ &= \int_{\mathbb{R}^d} f(z)\mathbb{1}_{B_d(x,t_n)}(z)\,\mathrm{d}z . \end{aligned}$$

Since f is integrable, the Lebesgue theorem allows us to conclude that

$$\lim_n \int_{\mathbb{R}^d} f(z)\mathbb{1}_{B_d(x,t_n)}(z)\,\mathrm{d}z = \int_{\mathbb{R}^d} \lim_n f(z)\mathbb{1}_{B_d(x,t_n)}(z)\,\mathrm{d}z = \mathbb{P}(X \in B_d(x,t)) ,$$

so the cumulative distribution function is continuous. Then thanks to classical result on statistics order and quantile transform (see David and Nagaraja (2003)), we get

$$P_n = F_{||X-x||}\left(||X-x||_{(k_{n+1},n)}\right) \sim U_{(k_{n+1},n)} \sim \beta(k_{n+1}, n-k_{n+1}+1) ,$$

where we denoted $U_{(k_{n+1},n)}$ the k_{n+1} statistic order of a independent sample of size n distributed like a uniform law on $[0,1]$. □

Let us now recall some deviation results.

Lemma B.7. *We denote $\mathcal{B}(n,p)$ the binomial distribution of parameters n and p, for $n \geq 1$ and $p \in [0,1]$. Then, if $Z \sim \mathcal{B}(n,p)$, we get*

$$\begin{aligned} \mathbb{P}\left(\frac{Z}{n} < \frac{p}{2}\right) &\leq \exp\left(-\frac{3np}{32}\right) , \\ \mathbb{P}\left(\frac{Z}{n} > 2p\right) &\leq \exp\left(-\frac{3np}{8}\right) . \end{aligned}$$

Proof. Let (Z_i) be an independent sample of Bernoulli of parameter p and let

$$Z = \frac{1}{n}\sum_{k=1}^{n} Z_i \ .$$

We apply the Bernstein's inequality (see for example Theorem 8.2 in Devroye et al. (2013)) to conclude that

$$\mathbb{P}(Z - p < -\zeta p) \leq \exp\left(-\frac{3np\zeta^2}{8}\right) \ ,$$
$$\mathbb{P}(Z - p > \zeta p) \leq \exp\left(-\frac{3np\zeta^2}{8}\right) \ .$$

The results follow by taking $\zeta = \frac{1}{2}$ in the first case and $\zeta = 1$ in the second case. □

We now give some technical lemma useful to prove our main results.

Lemma B.8. *Suppose $\beta \geq \gamma$. Then, for every $C > 0$, we get*

$$\mathbb{P}\left(\sum_n \gamma_n \mathbb{1}_{X_n \in \chi_n^d(x)} \leq C\right) = 0 \ .$$

Proof. Let us denote F the cumulative function distribution of $||X_n - x||$ and $U_n = F(||X_n - x||)$, we get

$$\sum_n \gamma_n \mathbb{1}_{X_n \in \chi_n^d(x)} = \sum_n \gamma_n \mathbb{1}_{U_n \in \chi_n^d(0)} \ .$$

Hence, it is enough to show the desired result for $x = 0$ and $X_j = U_j \sim \mathcal{U}([0,1])$.

Let ζ be a positive real number. Let N be an integer such that

$$\sum_{n \geq N} \exp\left(-\frac{3k_n}{16}\right) \leq \zeta \ . \tag{B.6}$$

We set

$$\Omega := \left\{\forall n \geq N, \ \sum_{j=1}^{n} \mathbb{1}_{U_j \leq \frac{k_n}{2n}} \leq k_n\right\} \ .$$

On this event, for every $n \geq N$, there are at most k_n elements U_i such that U_i is inferior to $\frac{k_n}{2n}$. Thus, if an element satisfies $U_j \leq \frac{k_n}{2n}$, it belongs to the k_n-nearest neighbors of 0. Then, defining $Z_n := \sum_{j=1}^{n} \mathbb{1}_{U_j \leq \frac{k_n}{2n}} \sim \mathcal{B}\left(n, \frac{k_n}{2n}\right)$,

$$\begin{aligned}
\mathbb{P}\left(\overline{\Omega}\right) &\leq \sum_{n \geq N} \mathbb{P}\left(\sum_{j=1}^{n} \mathbb{1}_{U_j \leq \frac{k_n}{2n}} > k_n\right) \\
&\leq \sum_{n \geq N} \mathbb{P}\left(\frac{Z_n}{n} > \frac{k_n}{n}\right) \\
&\leq \sum_{n \geq N} \exp\left(-\frac{3k_n}{16}\right) \leq \zeta \ .
\end{aligned} \tag{B.7}$$

by using the second inequality of Lemma B.7 and Equation (B.6). But, as we noticed above, on the event Ω, we have $\mathbb{1}_{U_n \in \chi_n^d(0)} \geq \mathbb{1}_{U_n \leq \frac{k_n}{2n}}$; and thus

$$\mathbb{P}\left(\Omega \cap \sum_{n \geq N} \gamma_n \mathbb{1}_{U_n \in \chi_n^d(0)} \leq C\right) \leq \mathbb{P}\left(\sum_{n \geq N} \gamma_n \mathbb{1}_{U_n \leq \frac{k_n}{2n}} \leq C\right) \ . \tag{B.8}$$

Let now $(I_k)_k$ be a partition of $[\![N, +\infty[\![$ such that

$$\forall k \geq 1,\ \sum_{n\in I_k} \gamma_n \frac{k_n}{2n} \in [2C, 2C+1]\ .$$

Such a partition exists since, as $\beta \geq \gamma$, the sum $\sum_n \gamma_n \frac{k_n}{n}$ is divergent. Then,

$$\mathbb{E}\left[\sum_{n\in I_k} \gamma_n \mathbb{1}_{U_n \leq \frac{k_n}{2n}}\right] = \sum_{n\in I_k} \gamma_n \mathbb{E}\left(\mathbb{1}_{U_n \leq \frac{k_n}{2n}}\right) = \sum_{n\in I_k} \gamma_n \frac{k_n}{2n}\ ,\ \geq 2C\ .$$

and by independence, and since the variance of a Bernoulli variable is upper-bounded by its expectation,

$$\mathrm{Var}\left[\sum_{n\in I_k} \gamma_n \mathbb{1}_{U_n \leq \frac{k_n}{n}}\right] \leq \mathbb{E}\left[\sum_{n\in I_k} \gamma_n \mathbb{1}_{U_n \leq \frac{k_n}{2n}}\right] = \sum_{n\in I_k} \gamma_n \frac{k_n}{2n} \leq 2C+1\ .$$

Chebyshev's inequality yields:

$$\begin{aligned}\mathbb{P}\left(\sum_{n\in I_k} \gamma_n \mathbb{1}_{U_n \leq \frac{k_n}{2n}} \leq C\right) &\leq \mathbb{P}\left(\mathbb{E}\Big[\sum_{n\in I_k} \gamma_n \mathbb{1}_{U_n \leq \frac{k_n}{2n}}\Big] - \sum_{n\in I_k} \gamma_n \mathbb{1}_{U_n \leq \frac{k_n}{2n}} \geq 2C - C\right) \\ &\leq \frac{2C+1}{C^2} \leq \frac{7}{9}\end{aligned}$$

since $C \geq 3$. Thus,

$$\mathbb{P}\left(\bigcap_k \left\{\sum_{n\in I_k} \gamma_n \mathbb{1}_{U_n \leq \frac{k_n}{2n}} \leq C\right\}\right) = 0\ .$$

and hence

$$\mathbb{P}\left(\sum_{n\geq N} \gamma_n \mathbb{1}_{U_n \leq \frac{k_n}{2n}} \leq C\right) = 0\ . \tag{B.9}$$

Thanks to (B.7), (B.8) and (B.9), we get

$$\mathbb{P}\left(\sum_n \gamma_n \mathbb{1}_{U_n \in \chi_n^d(0)} \leq C\right) \leq \mathbb{P}\left(\sum_{n\geq N} \gamma_n \mathbb{1}_{U_n \in \chi_n^d(0)} \leq C\right) \leq \mathbb{P}(\overline{\Omega}) + 0 \leq \zeta\ ,$$

which holds for all $\zeta > 0$. □

Lemma B.9. *Denoting A_n the event $\{X_1, \ldots, X_n \mid P_n > \zeta_n\}$ where $\zeta_n = \frac{1}{n^\zeta}$ and the parameter ζ satisfies $1 > \zeta > 1 - \beta$, we have for $n \geq 2^{1/(\zeta-(1-\beta))}$,*

$$\mathbb{P}(A_n^C) \leq \exp\left(-\frac{3n^{1-\zeta}}{8}\right)\ .$$

Proof. Thanks to the Lemma B.6, we obtain

$$\begin{aligned}\mathbb{P}(A_n^C) &= \mathbb{P}(\beta(k_{n+1}, n - k_{n+1} + 1) \leq \zeta_n) \\ &= I_{\zeta_n}(k_{n+1}, n - k_{n+1} + 1)\ ,\end{aligned}$$

where we denote I_ζ the incomplete β function. A classical result (see Abramowitz and Stegun (1964)) allows us to write this quantity in terms of the binomial distribution

$$\mathbb{P}(A_n^C) = \mathbb{P}(\mathcal{B}(n, \zeta_n) \geq k_{n+1}) \ .$$

Thanks to Lemma B.7, we know that

$$\mathbb{P}(\mathcal{B}(n, \zeta_n) \geq k_{n+1}) \leq \exp\left(-\frac{3n\zeta_n}{8}\right) \leq \exp\left(-\frac{3n^{1-\zeta}}{8}\right) \ ,$$

as soon as $k_{n+1}/n \geq 2\zeta_n$, which is true as soon as $n \geq 2^{1/(\zeta-(1-\beta))}$ because $\zeta > 1-\beta$. □

Lemma B.10. *Under hypothesis of Theorem B.1, $||X-x||_{(k_{n+1},n)}$ converges almost surely to 0.*

Proof. Let u be a positive number.

$$\begin{aligned} p_u := \mathbb{P}(X \in \mathcal{B}(x,u)) &= \int_{\mathcal{B}(x,u)} f(t)dt \\ &\geq \mu_X\left(\mathcal{B}(x,u)\right) = C_1 \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)} \\ &= C_{input} C_4(d) u^d =: q_u \ . \end{aligned} \tag{B.10}$$

Let Z be a random variable of law $\mathcal{B}(n, p_u)$. Since $||X-x||_{(k_{n+1},n)} > u$ implies that there are at the most k_{n+1} elements of the sample which satisfy $X \in \mathcal{B}(x, q_u)$, we get :

$$\mathbb{P}(||X-x||_{(k_{n+1},n)} > u) = \mathbb{P}(Z < k_{n+1}) \ .$$

Thanks to equation *Eq.* (B.10), and denoting $\tilde{Z}$ a random variable of law $\mathcal{B}(n, q_u)$, we have

$$\mathbb{P}(||X-x||_{(k_{n+1})} > u) \leq \mathbb{P}(\tilde{Z} < k_{n+1}) \ .$$

Lemma B.7 implies that $\mathbb{P}(||X-x||_{(k_{n+1})} > u)$ is the general term of a convergent sum. Indeed, when n is large enough, then $k_{n+1}/n < q_u/2$ because k_{n+1}/n converges to 0 ($\beta < 1$). The Borel-Cantelli Lemma then implies that $||X-x||_{(k_{n+1},n)}$ converges almost surely to 0. □

Lemma B.11. *With the same notation as above,*

$$\mathbb{E}(P_n||X-x||_{(k_{n+1},n)}) \leq C_3(d)\left(\frac{k_{n+1}}{n+1}\right)^{1+\frac{1}{d}} \ .$$

Proof. Let us denote $\tilde{F}$ and $\tilde{f}$ the cumulative and density distribution function of the law of $||X-x||$.

$$\begin{aligned} \mathbb{E}(||X-x||_{(k_{n+1},n)}P_n) &= \mathbb{E}\left(||X-x||_{(k_{n+1},n)}\tilde{F}\left(||X-x||_{(k_{n+1},n)}\right)\right) \\ &= \int y\tilde{F}(y) f_{||X-x||_{(k_{n+1},n)}}(y)dy \ , \end{aligned}$$

with

$$f_{|X-x|_{(k_{n+1},n)}}(y) = \frac{n!}{(k_{n+1}-1)!(n-k_{n+1})!}\tilde{F}(y)^{k_{n+1}-1}\left(1-\tilde{F}(y)\right)^{n-k_{n+1}}\tilde{f}(y)\ .$$

Then we get

$$\begin{aligned}\mathbb{E}(||X-x||_{(k_{n+1},n)}P_n) &= \int y\tilde{F}(y)^{k_{n+1}}(1-\tilde{F}(y))^{n-k_{n+1}}\tilde{f}(y)\frac{n!}{(k_{n+1}-1)!(n-k_{n+1})!}\\ &= \frac{k_{n+1}}{n+1}\mathbb{E}\left(||X-x||_{(k_{n+1}+1,n+1)}\right)\ .\end{aligned}$$

We denote $U_{|.|}$ the upper bound of the support of $||X-x||$, and write

$$\mathbb{E}(||X-x||_{(k_{n+1}+1,n+1)}) = \int_0^{U_{|.|}}\mathbb{P}(||X-x||_{(k_{n+1}+1,n+1)} > u)du\ .$$

Using same arguments that in Lemma B.1, denoting $C_{10}(d) = \sqrt[d]{\frac{2(k_{n+1}+1)}{(n+1)C_{input}C_4(d)}}$,we get

$$\begin{aligned}I := \int_0^{U_{|.|}}\mathbb{P}(||X-x||_{(k_{n+1}+1,n+1)} > u)du &= \int_0^{C_{10}(d)}\mathbb{P}(\mathcal{B}(n+1,q_u) < k_{n+1}+1)du\\ &\quad + \int_{C_{10}(d)}^{U_{|.|}}\mathbb{P}(\mathcal{B}(n+1,q_u) < k_{n+1}+1)du\\ &\leq \int_0^{C_{10}(d)}1du\\ &\quad + \int_{C_{10}(d)}^{U_{|.|}}\exp\left(-\frac{3(n+1)C_{input}C_4(d)u^d}{32}\right)du\ ,\end{aligned}$$

where we use Lemma B.7 in the second integral because $u > C_{10}(d)$ implies $\frac{k_{n+1}+1}{n+1} < \frac{q_u}{2}$. Then, we obtain

$$\begin{aligned}I &\leq C_{10}(d) + \int_{C_{11}(d)}^{+\infty}\exp\left(-\frac{3(n+1)C_{input}C_4(d)u^d}{32}\right)du\\ &\leq C_{10}(d) + \int_0^{+\infty}\frac{u^{d-1}}{C_{10}(d)^{d-1}}\exp\left(-\frac{3(n+1)C_{input}C_4(d)u^d}{32}\right)du\\ &= C_{10}(d) + \frac{C_{11}(d)}{C_{10}(d)^d}\frac{32}{3(n+1)dC_{input}C_4(d)}\left[-\exp\left(-\frac{3(n+1)C_{input}C_4(d)u^d}{32}\right)\right]_0^{+\infty}\\ &= C_{10}(d)\left(1+\frac{3(n+1)dC_{input}C_4(d)}{32C_{10}(d)^d}\right)\\ &= \sqrt[d]{\frac{2(k_{n+1}+1)}{(n+1)C_{input}C_4(d)}}\left(1+\frac{16}{3d(k_{n+1}+1)}\right)\\ &= \sqrt[d]{\frac{k_{n+1}}{n+1}}\left[\sqrt[d]{\frac{2}{C_{input}C_4(d)}}\sqrt[d]{\frac{k_{n+1}+1}{k_{n+1}}}\left(1+\frac{16}{3d(k_{n+1}+1)}\right)\right]\\ &\leq \sqrt[d]{\frac{k_{n+1}}{n}}\sqrt[d]{\frac{4}{C_{input}C_4(d)}}\left(1+\frac{8}{3d}\right)\\ &=: C_3(d)\sqrt[d]{\frac{k_{n+1}}{n+1}}\ ,\end{aligned}$$

because for $n \geq 1$, we get $k_n \geq 1$. □

Lemma B.12. *Let (b_n) be a real sequence. If there exist sequences $(c_n)_{n\geq 1} \in [0,1]^{\mathbb{N}}$ and $(d_n)_{n\geq 1} \in]0,+\infty[^{\mathbb{N}}$ such that*

$$\forall n \geq N_0,\ b_{n+1} \leq b_n(1-c_{n+1}) + d_{n+1}\ ,$$

then for all $n \geq N_0+1$,

$$b_n \leq \exp\left(-\sum_{k=1}^{N_0+1} c_k\right) b_{N_0} + \sum_{k=N_0+1}^{n} \exp\left(-\left(\sum_{j=1}^{n} c_j - \sum_{j=1}^{k} c_j\right)\right) d_k\ .$$

Proof. This inequality appears in Moulines and Bach (2011) and references therein. It can be proved by induction using that $\forall x \in]0,+\infty[,\ \exp(x) \geq 1+x$. □

Let us first prove the following consequence of Assumption **A3**.

Lemma B.13. *Under assumption **A3**, if $\beta \geq \gamma$, then for all x and for all $n \geq 1$,*

$$\theta_n(x) \in [L_Y - (1-\alpha), U_Y + \alpha], a.s.$$

Proof. Suppose that $\theta_n(x)$ leaves the compact set $[L_Y, U_Y]$ by the right at step N_0. By definition, $\theta_{N_0-1} \leq U_Y$ and consequently $\theta_{N_0} \leq U_Y + \alpha\gamma_{N_0}$. At next step, since $\theta_{N_0} > U_Y$, we have $Y_{N_0+1} \leq \theta_{N_0}$ and then

$$\theta_{N_0+1} \leq U_Y + \alpha\gamma_{N_0} - (1-\alpha)\gamma_{N_0+1}\mathbb{1}_{X_{N_0+1}\in\chi^d_{N_0+1}(x)}\ .$$

Then, the algorithm either does not move (if $X_{N_0+1} \notin \chi^d_{N_0+1}(x)$) or comes back in direction of $[L_Y, U_Y]$ with a step of $(1-\alpha)\gamma_{N_0+1}$. Then, if

$$\sum_{n\geq 0} \gamma_n \mathbb{1}_{X_n\in\chi^d_n(x)} = +\infty\ a.s\ ,$$

the algorithm almost surely comes back to the compact set $[L_Y, U_Y]$. Thanks to Lemma B.8, we know that, since $\beta \geq \gamma$, the previous sum diverges almost surely. A similar result holds when the algorithm leaves the compact set by the left and finally we have shown that almost surely as $\gamma_n \leq 1$,

$$\theta_n(x) \in [L_Y - (1-\alpha), U_Y + \alpha] =: [L_{\theta_n}, U_{\theta_n}]\ .$$

□

B.4.2 Proof of Theorem B.1 : Almost Sure Convergence

To prove this theorem, we adapt the classical analysis of the Robbins-Monro algorithm (see Blum (1954)). In the sequel we do not write $\theta_n(x)$ but θ_n to make the notation less cluttered.

B.4.2.1 Martingale decomposition

In this sequel, we still denote $H(\theta_n, X_{n+1}, Y_{n+1}) := \left(\mathbb{1}_{Y_{n+1}\leq\theta_n} - \alpha\right)\mathbb{1}_{X_{n+1}\in\chi^d_{n+1}(x)}$, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$ and $\mathbb{P}_n$ and $\mathbb{E}_n$ the probability and expectation conditionally to $\mathcal{F}_n$. We introduce

$$\begin{aligned} h_n(\theta_n) :&= \mathbb{E}(H(\theta_n, X_{n+1}, Y_{n+1})|\mathcal{F}_n) \\ &= \mathbb{P}_n(Y_{n+1} \leq \theta_n \cap X_{n+1} \in \chi^d_{n+1}(x)) - \alpha\mathbb{P}_n(X_{n+1} \in \chi^d_{n+1}(x)) \\ &= P_n\left[F_{Y^{\chi^d_{n+1}(x)}}(\theta_n) - F_{Y^x}(\theta^*(x))\right] . \end{aligned}$$

Then,

$$T_n = \theta_n + \sum_{j=1}^{n} \gamma_j h_{j-1}(\theta_{j-1}) = \theta_0(x) - \sum_{j=1}^{n} \gamma_j \xi_j ,$$

with $\xi_j = H(\theta_{j-1}, X_j, Y_j) - h_{j-1}(\theta_{j-1})$ is a martingale increment. It is bounded in $\mathbb{L}^2(\mathbb{R})$. Since

$$\sup_n |\xi_n| \leq \alpha + (1+\alpha) = 1 + 2\alpha,$$

the Burkholder inequality gives the existence of a constant C such that

$$\mathbb{E}(|T_n|^2) \leq \mathbb{E}\left(\left(\sum_{j=1}^{n} \gamma_j \xi_j\right)^2\right) \leq C\,\mathbb{E}\left(\sum_{j=1}^{n} (\gamma_j \xi_j)^2\right) \leq C(1+2\alpha)^2 \sum_{j=1}^{n} \gamma_j^2 < \infty .$$

B.4.2.2 The sequence (θ_n) converges almost surely

First, let us prove that

$$\mathbb{P}(\theta_n \to \infty) + \mathbb{P}(\theta_n \to -\infty) = 0. \tag{B.11}$$

Let us suppose that this probability is positive (we name Ω_1 the non-negligible set where $\theta_n(\omega)$ diverges to $+\infty$ and the same arguments would show the result when the limit is $-\infty$). Let ω be in Ω_1. We have $\theta_n(\omega) \leq \theta^*$ for only a finite number of n.

Let us show that on an event $\Omega \subset \Omega_1$ with positive measure, for n large enough, $h(\theta_n(\omega)) > 0$. First, we know that P_n follows a Beta distribution. This is why $\forall n,\ \mathbb{P}(P_n = 0) = 0$. Then, the Borel-Cantelli Lemma gives that

$$\mathbb{P}(\exists N\ \forall n \geq N\ P_n > 0) = 1 .$$

As Ω_1 has a positive measure, we know that there exists $\Omega_2 \subset \Omega_1$ with positive measure such that $\forall \omega \in \Omega_2$, $\theta_n(\omega) \to +\infty$ and for all n large enough, $P_n(\omega) > 0$. Since

$$h(\theta_n(\omega)) = P_n\left(F_{Y^{B_n^{k_{n+1}}(x)}}(\theta_n(\omega)) - \alpha\right) ,$$

we have now to show that on $\Omega \subset \Omega_2$ of positive measure,

$$F_{Y^{B_n^{k_{n+1}}(x)}}(\theta_n(\omega)) - \alpha > 0 .$$

As $\theta_n(\omega)$ diverges to $+\infty$, we can find D such that for n large enough, $\theta_n(\omega) > D > \theta^*$. Then,

$$\begin{aligned} F_{Y^{B_n^{k_{n+1}}(x)}}(\theta_n(\omega)) - \alpha &= F_{Y^{B_n^{k_{n+1}}(x)}}(\theta_n(\omega)) - F_{Y^x}(\theta^*(x)) \\ &= F_{Y^{B_n^{k_{n+1}}(x)}}(\theta_n(\omega)) - F_{Y^{B_n^{k_{n+1}}(x)}}(D) \\ &\quad + F_{Y^{B_n^{k_{n+1}}(x)}}(D) - F_{Y^x}(D) + F_{Y^x}(D) - F_{Y^x}(\theta^*(x)) . \end{aligned}$$

First, $F_{Y^{B_n^{k_{n+1}(x)}}}(\theta_n(\omega)) - F_{Y^{B_n^{k_{n+1}(x)}}}(D) \geq 0$ because a cumulative distribution function is non-decreasing. Then, we set $\eta = F_{Y^x}(D) - F_{Y^x}(\theta^*(x))$ which is a finite value. To deal with the last term, we use our assumption **A1**.

$$F_{Y^{B_n^{k_{n+1}(x)}}}(D) - F_{Y^x}(D) \geq -M(x)||X - x||_{(k_{n+1},n)} \ .$$

We know, thanks to Lemma B.10, that $||X - x||_{(k_{n+1},n)}$ converges almost surely to 0. Then, there exists a set $\Omega_3 \subset \Omega_1$ of probability strictly non-negative such that forall ω in Ω_3, the previous reasoning is true. And for $\zeta < \frac{\eta}{L}$, there exists rank $N(\omega)$ such that if $n \geq N$,

$$F_{Y^{B_n^{k_{n+1}(x)}}}(D) - F_{Y^x}(D) \geq 0 - L\zeta + \eta > 0 \ . \tag{B.12}$$

Finally, for $\omega \in \Omega_3$ (set of strictly non-negative measure), we have shown that after a certain rank, $h(\theta_n(\omega)) > 0$. This implies that on Ω_3 of positive measure,

$$\lim_n \left[\theta_n(\omega) + \sum_{j=1}^{n} \gamma_{j-1} h_{j-1}(\theta_{j-1}(\omega)) \right] = +\infty \ ,$$

which is absurd because in the previous part we proved that T_n is almost surely convergent. Then θ_n does not diverge to $+\infty$ or $-\infty$.

Now, we will show that (θ_n) converges almost surely. In all the sequel of the proof, we reason ω by ω like in the previous part. To make the reading more easy, we do not write ω and Ω any more. Thanks to Equation (B.11) and to the previous subsection, we know that, with probability positive, there exists a sequence (θ_n) such that

$$\begin{cases} (a)\ \theta_n + \sum_{j=1}^{n} \gamma_{j-1} h(\theta_{j-1}) \text{ converges to a finite limit} \\ (b)\ \liminf \theta_n < \limsup \theta_n \ . \end{cases}$$

Let us suppose that $\limsup \theta_n > \theta^*$ (we will find a contradiction, the same argument would allow us to conclude in the other case). Let us choose c and d satisfying $c > \theta^*$ and $\liminf \theta_n < c < d < \limsup \theta_n$. Since the sequence (γ_n) converges to 0, and since (T_n) is a Cauchy sequence, we can find a deterministic rank N and two integers n and m such that $N \leq n < m$ implies

$$\begin{cases} (a)\ \gamma_n \leq \dfrac{(d-c)}{3(1-\alpha)} \\ (b)\ \left| \theta_m - \theta_n - \sum_{j=n}^{m-1} \gamma_j h(\theta_{j-1}) \right| \leq \dfrac{d-c}{3} \ . \end{cases}$$

We choose m and n so that

$$\begin{cases} (a)\ N \leq n < m \\ (b)\ \theta_n < c,\ \theta_m > d \\ (c)\ n < j < m \Rightarrow c \leq \theta_j \leq d \ . \end{cases} \tag{B.13}$$

This is possible since beyond N, the distance between two iterations will be either

$$\alpha\gamma_n \leq \frac{\alpha(d-c)}{3(1-\alpha)} < (d-c) \ ,$$

because $\alpha < \frac{3}{5}$ or

$$(1-\alpha)\gamma_n \leq \frac{1}{3}(d-c) < (d-c) \ .$$

Moreover, since c and d are chosen to have an iteration inferior to c and an iteration superior to b, the algorithm will necessarily go through the segment $[c, d]$. We then take n and m the times of enter and exit of the segment. Now,

$$\begin{aligned}\theta_m - \theta_n &\leq \frac{d-c}{3} + \sum_{j=n}^{m-1} \gamma_{j+1} h_j(\theta_j) \\ &\leq \frac{d-c}{3} + \gamma_{n+1} h(\theta_n) \ ,\end{aligned}$$

because $n < j < m$, we get $\theta^* < c < \theta_j$ and we have already shown that in this case, $h_j(\theta_j) > 0$. We then only have to deal with θ_n. If $\theta_n > \theta^*$, we can apply the same result and then

$$\theta_m - \theta_n \leq \frac{d-c}{3} \ ,$$

which is in contradiction with (b) of equation Eq. (B.13). When $\theta < \theta^*$,

$$\begin{aligned}\theta_m - \theta_n &\leq \frac{d-c}{3} + \gamma_n h_{n-1}(\theta_{n-1}) \\ &\leq \frac{d-c}{3} + \gamma_n(1-\alpha) \\ &\leq \frac{d-c}{3} + \frac{d-c}{3} < (d-c) \ ,\end{aligned}$$

which is still a contradiction with (b) of Eq. (B.13). We have shown that the algorithm converges almost surely.

B.4.2.3 The algorithm converges almost surely to θ^*

Again we reason by contradiction. Let us name θ the limit such that $\mathbb{P}(\theta \neq \theta^*) > 0$. With positive probability, we can find a sequence (θ_n) which converges to θ such that

$$\begin{cases}(a)\ \theta^* < \zeta_1 < \zeta_2 < \infty \\ (b)\ \zeta_1 < \theta < \zeta_2 \ ,\end{cases}$$

(or $-\infty < \zeta_1 < \zeta_2 < \theta^*$ but arguments are the same in this case). Then, for n large enough, we get

$$\zeta_1 < \theta_n < \zeta_2 \ .$$

Finally, on the one hand, (T_n) and (θ_n) are convergent, and we also know that the sum $\sum \gamma_{j+1} h_j(\theta_j)$ converges almost surely. Let us then show that on the other hand, $h(\theta_n) = P_n(F_{Y^{B_n^{k_{n+1}}}(x)}(\theta_n) - \alpha)$ is lower bounded. First we know thanks to Lemma B.9, that for $1 > \zeta > 1 - \beta$ and $\zeta_n = \frac{1}{n^\zeta}$,

$$\mathbb{P}(P_n \leq \zeta_n) \leq \exp\left(-\frac{3n^{1-\zeta}}{8}\right) \ .$$

This is the general term of a convergent sum. Therefore, the Borel-Cantelli Lemma gives

$$\mathbb{P}(\exists N\ \forall n \geq N\ P_n > \zeta_n) = 1 \ .$$

Moreover, as we have already seen in Equation Eq. (B.12), since $\theta_n > \zeta_1 > \theta^*$,

$$F_{Y^{B_n^{k_{n+1}}(x)}}(\theta_n) - \alpha \geq 0 - M(x)||X - x||_{(k_{n+1},n)} + F_{Y^x}(\zeta_1) - F_{Y^x}(\theta^*(x)) \ .$$

Then, when n is large enough so that

$$||X - x||_{(k_{n+1},n)} \leq \frac{F_{Y^x}(\zeta_1) - F_{Y^x}(\theta^*(x))}{2M(x)}$$

holds, we have

$$F_{Y^{B_n^{k_{n+1}}(x)}}(\theta_n) - \alpha \geq \frac{F_{Y^x}(\zeta_1) - F_{Y^x}(\theta^*(x))}{2} \ .$$

Finally there exists a set Ω of positive probability such that, $\forall \omega \in \Omega$

$$\sum_{k=1}^{n} \gamma_{k+1} h_k(\theta_k) \geq \frac{F_{Y^x}(\zeta_1) - F_{Y^x}(\theta^*(x))}{2} \sum_{k=1}^{n} \gamma_{k+1} P_k \geq \sum_{k=1}^{n} \frac{1}{(k+1)^{\gamma+\zeta}} \ ,$$

which is a contradiction (with the one hand point) because the sum is divergent ($\gamma+\zeta < 1$).

B.4.3 Proof of Theorem B.3 : Non-asymptotic Inequality on the Mean Squared Error.

Let x be fixed in $[0,1]$. We want to find an upper-bound for the mean squared error $a_n(x)$ using Lemma B.12. In the sequel, we will need to study $\theta_n(x)$ on the event A_n of the Lemma B.9. Then, we begin to find a link between $a_n(x)$ and the mean squared error on this event.

$$\begin{aligned} a_n(x) &= \mathbb{E}\left[(\theta_n(x) - \theta^*(x))^2 \mathbb{1}_{A_n}\right] + \mathbb{E}\left[(\theta_n(x) - \theta^*(x))^2 \mathbb{1}_{A_n^C}\right] \\ &\leq \mathbb{E}\left[(\theta_n - \theta^*(x))^2 \mathbb{1}_{A_n}\right] + C_1 \mathbb{P}(A_n^C) \\ &\leq \mathbb{E}\left[(\theta_n(x) - \theta^*(x))^2 \mathbb{1}_{A_n}\right] + C_1 \exp\left(-\frac{3n^{1-\zeta}}{8}\right) \ , \end{aligned} \tag{B.14}$$

thanks to Lemma B.9 and for $n \geq N_0$.

Let us now study the sequence $b_n(x) := \mathbb{E}\left[(\theta_n(x) - \theta^*(x))^2 \mathbb{1}_{A_n}\right]$. First, for $n \geq 0$,

$$b_{n+1}(x) \leq \mathbb{E}\left[(\theta_{n+1}(x) - \theta^*(x))^2\right] \ .$$

But,

$$\begin{aligned} (\theta_{n+1}(x) - \theta^*(x))^2 &= (\theta_n(x) - \theta^*(x))^2 \\ &+ \gamma_{n+1}^2 \left[(1-2\alpha)\mathbb{1}_{Y_{n+1} \leq \theta_n(x)} + \alpha^2\right] \mathbb{1}_{X_{n+1} \in \chi_{n+1}^d(x)} \\ &- 2\gamma_{n+1}(\theta_n(x) - \theta^*(x)) \left(\mathbb{1}_{Y_{n+1} \leq \theta_n(x)} - \alpha\right) \mathbb{1}_{X_{n+1} \in \chi_{n+1}^d(x)} \ . \end{aligned}$$

Taking the expectation conditional to $\mathcal{F}_n$, as $[(1-2\alpha)\mathbb{1}_{Y_{n+1} \leq \theta_n(x)} + \alpha^2] \leq 1$, we get

$$\begin{aligned} \mathbb{E}_n\left((\theta_{n+1}(x) - \theta^*(x))^2\right) &\leq \mathbb{E}_n\left((\theta_n(x) - \theta^*(x))^2\right) + \gamma_{n+1}^2 \mathbb{P}_n\left(X_{n+1} \in \chi_{n+1}^d(x)\right) \\ &- 2\gamma_{n+1}\left(\theta_n(x) - \theta^*(x)\right)\left[\mathbb{P}_n\left(Y_{n+1} \leq \theta_n(x) \cap X_{n+1} \in \chi_{n+1}^d(x)\right)\right. \\ &\left. - \alpha \mathbb{P}_n\left(X_{n+1} \in \chi_{n+1}^d(x)\right)\right] \ . \end{aligned}$$

Using the Bayes formula, we get

$$\mathbb{E}_n\left((\theta_{n+1}(x)-\theta^*(x))^2\right) \leq \mathbb{E}_n\left((\theta_n(x)-\theta^*(x))^2\right) + \gamma_{n+1}^2 P_n \\ - 2\gamma_{n+1}\left(\theta_n(x)-\theta^*(x)\right) P_n\left[F_{Y^{B_n^{k_{n+1}}(x)}}(\theta_n(x)) - F_{Y^x}(\theta^*(x))\right] ,$$

Let us split the double product into two terms representing the two errors we made by iterating our algorithm.

$$\mathbb{E}_n\left((\theta_{n+1}(x)-\theta^*(x))^2\right) \leq (\theta_n(x)-\theta^*(x))^2 + \gamma_{n+1}^2 P_{n+1} \\ - 2\gamma_{n+1}\left(\theta_n(x)-\theta^*(x)\right) P_{n+1}\left[F_{Y^{B_n^{k_{n+1}}(x)}}(\theta_n(x)) - F_{Y^x}(\theta_n(x))\right] \\ - 2\gamma_{n+1}\left(\theta_n(x)-\theta^*(x)\right) P_n\left[F_{Y^x}(\theta_n(x)) - F_{Y^x}(\theta^*(x))\right] . \tag{B.15}$$

We now use our hypothesis. By **A1**,

$$|F_{Y^{B_n^{k_{n+1}}(x)}}(\theta_n(x)) - F_{Y^x}(\theta_n(x))| \geq M(x)||X-x||_{(k_{n+1},n)} ,$$

and by **A3**,

$$|\theta_n(x)-\theta^*(x)| \leq \sqrt{C_1} .$$

Thus,

$$-2\gamma_{n+1}(\theta_n(x)-\theta^*(x))P_n\left[F_{Y^{B_n^{k_{n+1}}(x)}}(\theta_n(x)) - F_{Y^x}(\theta_n(x))\right] \\ \leq 2\gamma_{n+1}\sqrt{C_1}M(x)P_n||X-x||_{(k_{n+1},n)} .$$

On the other hand, thanks to **A4** we know that,

$$\left(\theta_n-\theta^*(x)\right)\left[F_{Y^x}(\theta_n(x)) - F_{Y^x}(\theta^*(x))\right] \geq C_2(x,\alpha)\left[\theta_n(x)-\theta^*(x)\right]^2 .$$

Coming back to Equation Eq. (B.15), we get

$$\mathbb{E}_n\left((\theta_{n+1}(x)-\theta^*(x))^2\right) \leq \left(\theta_n(x)-\theta^*(x)\right)^2\left(\mathbb{1}_{An} + \mathbb{1}_{\bar{An}}\right) + \gamma_{n+1}^2 P_n \\ - 2\gamma_{n+1}\left(\theta_n(x)-\theta^*(x)\right)^2 C_2(x,\alpha)P_n \\ + 2\gamma_{n+1}M(x)\sqrt{C_1}||X-x||_{(k_{n+1},n)}P_n .$$

To conclude, we take the expectation

$$b_{n+1}(x) \leq C_1\mathbb{P}(A_n^C) + b_n(x) - 2\gamma_{n+1}C_2(x,\alpha)\mathbb{E}\left[P_n\left(\theta_n(x)-\theta^*(x)\right)^2\right] \\ + \gamma_{n+1}^2\mathbb{E}(P_n) + 2\gamma_{n+1}\sqrt{C_1}M(x)\mathbb{E}\left[P_n||X-x||_{(k_{n+1},n)}\right] .$$

But, by definition of A_n,we get

$$-2\gamma_{n+1}C_2(x,\alpha)\mathbb{E}\left[P_{n+1}\left(\theta_n(x)-\theta^*(x)\right)^2\right] \leq -\gamma_{n+1}\zeta_n C_2(x,\alpha)\mathbb{E}\left[\left(\theta_n(x)-\theta^*(x)\right)^2\mathbb{1}_{A_n}\right] \\ = -2\gamma_{n+1}\zeta_n C_2(x,\alpha)b_n(x);.$$

Finally,

$$b_{n+1}(x) \leq b_n(x)\left(1-2C_2(x,\alpha)\gamma_{n+1}\zeta_n\right) + e_{n+1} ,$$

with

$$e_{n+1} := C_1\mathbb{P}(A_n^C) + \gamma_{n+1}^2\mathbb{E}(P_n) + 2\gamma_{n+1}\sqrt{C_1}M(x)\mathbb{E}\left[P_n||X-x||_{(k_{n+1},n)}\right] .$$

Now using Lemmas B.11, B.9 and B.6 we get for $n \geq N_0$ with

$$e_n \leq d_n := C_1 \exp\left(-\frac{3n^{1-\zeta}}{8}\right) + 2\sqrt{C_1} M(x) C_3(d) \gamma_n \left(\frac{k_n}{n}\right)^{\frac{1}{d}+1} + \gamma_n^2 \frac{k_n}{n} .$$

The conclusion holds thanks to Lemma B.12, for $n \geq N_0 + 1$,

$$b_n(x) \leq \exp\left(-2C_2(x,\alpha)(\kappa_n - \kappa_{N_0})\right) b_{N_0}(x) + \sum_{k=N_0+1}^{n} \exp\left(-2C_2(x,\alpha)\left(\kappa_n - \kappa_k\right)\right) d_k . \tag{B.16}$$

But thanks to Assumption **A3**, we have already shown that $b_{N_0}(x) \leq a_{N_0}(x) \leq C_1$. To conclude, we re-inject Equation (B.16) in Equation (B.14) and obtain for $n \geq N_0 + 1$,

$$\begin{aligned} a_n(x) \leq \exp\left(-2C_2(x,\alpha)(\kappa_n - \kappa_{N_0})\right) C_1 + \sum_{k=N_0+1}^{n} \exp\left(-2C_2(x,\alpha)\left(\kappa_n - \kappa_k\right)\right) d_k \\ + C_1 \exp\left(-\frac{3n^{1-\zeta}}{8}\right) . \end{aligned}$$

B.4.4 Proof of Corollary B.4 : Rate of Convergence

In this part, we will denote

$$T_n^0 := C_1 \exp\left(\frac{-3n^{1-\zeta}}{8}\right) , \; T_n^1 := \exp\left(-2C_2(x,\alpha)(\kappa_n - \kappa_{N_0})\right)$$

and

$$T_n^2 := \sum_{k=N_0+1}^{n} \exp\left(-2C_2(x,\alpha)\left(\kappa_n - \kappa_k\right)\right) d_k .$$

We want to find a simpler expression for those terms to better see their order in n. First, considering T_n^1 we see that $a_n(x)$ can converge to 0 only when the sum

$$\sum_{k\geq 1} \frac{1}{k^{\gamma+\zeta}} = +\infty.$$

This is why we must first consider $\zeta \leq 1 - \gamma$. As $\zeta < 1 - \beta$, we have to take $\beta > \gamma$.

Remark 8. The frontier case $\zeta = 1 - \gamma$ is possible but the analysis shows that it is a less interesting choice than $\zeta < 1 - \gamma$ (there is a dependency in the value of $C_2(x,\alpha)$ but the optimal rate is the same as the one in the case we study). In the sequel, we only consider $\zeta < 1 - \gamma$.

Let us upper-bound T_n^1. As $x \mapsto 1/x^{\zeta+\gamma}$ is decreasing, we get

$$\begin{aligned} T_n^1 &= \exp\left(-2C_2(x,\alpha) \sum_{k=N_0+1}^{n} \frac{1}{k^{\zeta+\gamma}}\right) \\ &\leq \exp\left(-2C_2(x,\alpha) \int_{N_0+1}^{n+1} \frac{1}{t^{\zeta+\gamma}} dt\right) \\ &\leq \exp\left(-2C_2(x,\alpha) \frac{(n+1)^{1-\zeta-\gamma} - (N_0+1)^{1-\zeta-\gamma}}{(1-\zeta-\gamma)}\right) . \end{aligned}$$

Then, T_n^1 (just like T_n^0) is exponentially small when n grows up. To deal with the second term T_n^2 we first study the order in n of d_n. d_n is composed of three terms :

$$d_n \leq C_1 \exp\left(-\frac{3n^{1-\zeta}}{8}\right) + 2\sqrt{C_1}M(x)C_3(d)n^{-\gamma+(\beta-1)(1+\frac{1}{d})} + n^{-2\gamma+\beta-1} .$$

The first one is negligible (exponentially decreasing). Let us compare the two others which are powers of n. Comparing their exponents, we get that there exists constants C_5 and $C_6(d)$ (their explicit form is given in the Section B.6) such that

- if $\beta \leq 1 - d\gamma$, then for $n \geq N_0 + 1$,

$$d_n \leq C_5(x, d)n^{-2\gamma+\beta-1} ,$$

- if $\beta > 1 - d\gamma$, then for $n \geq N_0 + 1$,

$$d_n \leq C_6(x, d)n^{-\gamma+(1+\frac{1}{d})(\beta-1)} .$$

Remark 9. Let us detail how one can find C_5 (it is the same reasoning for C_6). If $\beta \leq 1 - d\gamma$, we know that when n will be big enough, the dominating term of d_n will be the one in $n^{-2\gamma+\beta-1}$. Then, it is logical to search a constant $C_5(x, d)$ such that $\forall n \geq N_0 + 1$,

$$d_n \leq \frac{C_5(x, d)}{n^{2\gamma-\beta+1}} .$$

Such a constant has to satisfy, for all $n \geq N_0 + 1$,

$$C_5(x, d) \geq C_1 \exp\left(-\frac{3}{8}n^{1-\zeta}\right) n^{2\gamma-\beta+1} + \frac{2\sqrt{C_1}M(x)C_3(d)}{n^{-\gamma+(1-\beta)/d}} + 1 .$$

Since $\beta \leq 1 - d\gamma$, the map $x \mapsto \frac{2\sqrt{C_1}M(x)C_3(d)}{n^{-\gamma+(1-\beta)/d}}$ is positive and decreasing. Then its maximum is reached for $n = N_0 + 1$. Moreover, the map $x \mapsto C_1 \exp\left(-\frac{3}{8}n^{1-\zeta}\right) n^{2\gamma-\beta+1}$ is also positive and is decreasing on an $[A, +\infty[$. It also has a maximum. The previous inequality is then true for

$$C_5(x, d) := \max_{n \geq N_0+1} C_1 \exp\left(-\frac{3}{8}n^{1-\zeta}\right) n^{2\gamma-\beta+1} + \frac{2\sqrt{C_1}M(x)C_3(d)}{(N_0+1)^{-\gamma+(1-\beta)/d}} + 1 .$$

Let us study the two previous cases.

Study of T_n^2 when $\beta > 1 - d\gamma$:

To upper-bound these sums, we use arguments from Cardot et al. (2017), which studies the stochastic algorithm to estimate the median on an Hilbert space. The main arguments are comparisons between sums and integrals. Indeed, for $n \geq N_0 + 2$ and $n \geq N_3$ where N_3 is such that

$$\forall n \geq N_3, \ \lfloor \frac{n}{2} \rfloor \geq N_0 + 1 ,$$

$$
\begin{aligned}
T_n^2 &= C_6(x,d)\sum_{k=N_0+1}^{n-1}\exp\left(-2C_2(x,\alpha)\sum_{j=k+1}^{n}\frac{a}{j^{\zeta+\gamma}}\right)\frac{1}{k^{\gamma+\left(1+\frac{1}{d}\right)(1-\beta)}}+\frac{C_6(x,d)}{n^{\gamma+(1+\frac{1}{d})(1-\beta)}}\\
&= C_6(x,d)\sum_{k=N_0+1}^{\lfloor\frac{n}{2}\rfloor}\exp\left(-2C_2(x,\alpha)\sum_{j=k+1}^{n}\frac{a}{j^{\zeta+\gamma}}\right)\frac{1}{k^{\gamma+\left(1+\frac{1}{d}\right)(1-\beta)}}\\
&+ C_6(x,d)\sum_{k=\lfloor\frac{n}{2}\rfloor+1}^{n-1}\exp\left(-2C_2(x,\alpha)\sum_{j=k+1}^{n}\frac{a}{j^{\zeta+\gamma}}\right)\frac{1}{k^{\gamma+\left(1+\frac{1}{d}\right)(1-\beta)}}+\frac{C_6(x,d)}{n^{\gamma+(1+\frac{1}{d})(1-\beta)}}\\
&=: S_1+S_2+S_3\ .
\end{aligned}
$$

First, the function $x\mapsto x^{-\zeta-\gamma}$ is decreasing on $]0,+\infty[$ then

$$
\begin{aligned}
S_2 &\le C_6(x,d)\sum_{k=\lfloor\frac{n}{2}\rfloor+1}^{n-1}\exp\left(-2C_2(x,\alpha)\int_{k+1}^{n+1}\frac{1}{x^{\zeta+\gamma}}dx\right)\frac{1}{k^{\gamma+(1+\frac{1}{d})(1-\beta)}}\\
&= C_6(x,d)\exp\left(-2C_2(x,\alpha)\frac{(n+1)^{1-\gamma-\zeta}}{1-\gamma-\zeta}\right)\\
&\qquad\sum_{k=\lfloor\frac{n}{2}\rfloor+1}^{n-1}\exp\left(-2C_2(x,\alpha)\frac{(k+1)^{1-\gamma-\zeta}}{1-\gamma-\zeta}\right)\frac{1}{k^{\gamma+(1+\frac{1}{d})(1-\beta)}}\ .
\end{aligned}
$$

Then, taking, $1-\beta<\zeta<\min((1-d\gamma),\left(1+\frac{1}{d}\right)(1-\beta))$, we have since $k\ge\lfloor\frac{n}{2}\rfloor+1$

$$
\begin{aligned}
S_2 \le C_6(x,d)\exp&\left(-2C_2(x,\alpha)\frac{(n+1)^{1-\gamma-\zeta}}{1-\gamma-\zeta}\right)\left(\frac{2}{n}\right)^{(1+\frac{1}{d})(1-\beta)-\zeta}\\
&\sum_{k=\lfloor\frac{n}{2}\rfloor+1}^{n-1}\exp\left(-2C_2(x,\alpha)\frac{(k+1)^{1-\gamma-\zeta}}{1-\gamma-\zeta}\right)\frac{1}{k^{\gamma+\zeta}}\ .
\end{aligned}
$$

Now, since for $k\ge1$,

$$
\left(\frac{1}{k}\right)^{\zeta+\gamma}\le\left(\frac{2}{k+1}\right)^{\zeta+\gamma}\ ,
$$

we get

$$
\begin{aligned}
S_2 \le C_6(x,d)\exp&\left(-2C_2(x,\alpha)\frac{(n+1)^{1-\gamma-\zeta}}{1-\gamma-\zeta}\right)\left(\frac{2}{n}\right)^{(1+\frac{1}{d})(1-\beta)-\zeta}2^{\zeta+\gamma}\\
&\sum_{k=\lfloor\frac{n}{2}\rfloor+1}^{n-1}\exp\left(-2C_2(x,\alpha)\frac{(k+1)^{1-\gamma-\zeta}}{1-\gamma-\zeta}\right)\frac{1}{(k+1)^{\gamma+\zeta)}}\ .
\end{aligned}
$$

Since the function $x\mapsto\exp\left(2C_2(x,\alpha)\frac{n^{1-\zeta-\gamma}}{1-\zeta-\gamma}\right)$ is decreasing on $\left[\frac{2C_2(x,\alpha)}{\gamma+\zeta},+\infty\right[$, we also define the integer $N_1(x,\alpha)$ the rank such that

$$
\forall n\ge N_1(x,\alpha),\ \lfloor\frac{n}{2}\rfloor+1\ge\frac{2C_2(x,\alpha)}{\zeta+\gamma}\ .
$$

For $n\ge N_1(x,\alpha)$ we get

$$
\begin{aligned}
S_2 &\leq C_6(x,d)\exp\left(-2C_2(x,\alpha)\frac{(n+1)^{1-\gamma-\zeta}}{1-\gamma-\zeta}\right)\frac{2^{(1+\frac{1}{d})(1-\beta)+\gamma}}{n^{(1+\frac{1}{d})(1-\beta)-\zeta}} \\
&\times \sum_{k=\lfloor\frac{n}{2}\rfloor+1}^{n-1}\int_{\lfloor\frac{n}{2}\rfloor+2}^{n}\exp\left(-2C_2(x,\alpha)\frac{x^{1-\gamma-\zeta}}{1-\gamma-\zeta}\right)\frac{1}{x^{\gamma+\zeta}}dx \\
&\leq \frac{C_6(x,d)}{2C_2(x,\alpha)}\exp\left(-2C_2(x,\alpha)\frac{(n+1)^{1-\gamma-\zeta}}{1-\gamma-\zeta}\right)\frac{2^{(1+\frac{1}{d})(1-\beta)+\gamma}}{n^{(1+\frac{1}{d})(1-\beta)-\zeta}} \\
&\times \left[\exp\left(2C_2(x,\alpha)\frac{n^{1-\zeta-\gamma}}{1-\zeta-\gamma}\right)-\exp\left(2C_2(x,\alpha)\frac{(\lfloor\frac{n}{2}\rfloor+2)^{1-\zeta-\gamma}}{1-\zeta-\gamma}\right)\right] \\
&\leq \frac{C_6(x,d)}{2C_2(x,\alpha)}\frac{2^{(1+\frac{1}{d})(1-\beta)+\gamma}}{n^{(1+\frac{1}{d})(1-\beta)-\zeta}} =: \frac{C_7(x,d,\alpha)}{2}\frac{1}{n^{-\zeta+(1+\frac{1}{d})(1-\beta)}}\,.
\end{aligned}
$$

Let us now deal with the term S_1. As $k \leq \lfloor\frac{n}{2}\rfloor$, we have

$$
\sum_{j=k+1}^{n}\frac{1}{j^{\zeta+\gamma}} \geq \frac{n}{2}\frac{1}{n^{\zeta+\gamma}}\,.
$$

Then,

$$
\begin{aligned}
S_1 &= C_6(x,d)\sum_{k=N_0+1}^{\lfloor\frac{n}{2}\rfloor}\exp\left(-2C_2(x,\alpha)\sum_{j=k+1}^{n}\frac{a}{j^{\zeta+\gamma}}\right)\frac{1}{k^{\gamma+(1-\beta)\left(1+\frac{1}{d}\right)}} \\
&\leq C_6(x,d)\sum_{k=1}^{\lfloor\frac{n}{2}\rfloor}\exp\left(-C_2(x,\alpha)n^{1-\zeta-\gamma}\right)\frac{1}{k^{\gamma+(1-\beta)\left(1+\frac{1}{d}\right)}} \\
&\leq C_6(x,d)\exp\left(-C_2(x,\alpha)n^{1-\zeta-\gamma}\right)\sum_{k=1}^{\lfloor\frac{n}{2}\rfloor}\frac{1}{k^{\gamma+(1-\beta)\left(1+\frac{1}{d}\right)}}\,.
\end{aligned}
$$

Thanks to the exponential term, S_1 is insignificant compared to S_2 whatever is the behaviour of the sum $\sum_k k^{-\gamma-(1-\beta)\left(1+\frac{1}{d}\right)}$, and so is T_1^n. Then, denoting $N_2(d,x)$ the rank after which we have

$$
S_3 + S_1 + T_n^1 + T_n^0 \leq \frac{C_7(x,\alpha,d)}{2n^{(1+\frac{1}{d})(1-\beta)-\zeta}}\,,
$$

we get, in the case where $\beta > 1-\gamma$ and $1-\beta < \zeta < \min((1-\gamma), \left(1+\frac{1}{d}\right)(1-\beta))$, for $n \geq \max\left(N_0, N_1(x,\alpha), N_2(d,x)\right)$

$$
a_n(x) \leq \frac{C_7(x,\alpha,d)}{n^{-\zeta+\left(1+\frac{1}{d}\right)(1-\beta)}}\,.
$$

Study of T_n^2 when $\beta \leq 1-d\gamma$:

Using the same arguments, we conclude that for $1-\beta < \zeta < \min(1-\beta+\gamma, 1-\gamma)$ and $n \geq \max(N_0, N_1(x,\alpha), N_2(d,x))$ (see Section B.6 for precise definitions of these ranks), there exists a constant $C_8(x,\alpha,d)$ such that the mean squared error satisfies

$$
a_n(x) \leq \frac{C_8(x,\alpha,d)}{n^{\gamma-\beta+1-\zeta}}\,.
$$

B.4.5 Proof of Corollary B.5 : Choice of Best Parameters β and γ

Let us now optimize the rate of convergence obtained in previous theorem. When $\beta \geq \gamma$ and $\beta \leq 1 - d\gamma$, the rate of convergence is of order $n^{-\gamma+\beta-1+\zeta}$. To optimize it, we have to choose ζ as small as possible. Then, we take $\zeta = 1 - \beta + \eta_\zeta$. The rate becomes $n^{-\gamma+\eta_\zeta}$. Then, we have also to choose γ as small as possible. In this area, there is only one point in which γ is the smallest, this is the point $(\gamma, \beta) = (\frac{1}{1+d}, \frac{1}{1+d})$. Since we have to take $\beta > \gamma$, the best couple of parameters, in this area, is $(\frac{1}{1+d}, \frac{1}{1+d} + \eta_\beta)$. These parameters follow a rate of convergence of $n^{\frac{-1}{1+d}+\eta}$.

When we are in the second area, the same kind of arguments allows us to conclude to the same optimal point with the same rate of convergence.

In Figure B.6, we use the numerical simulations of Section 3 to illustrate the previous discussion.

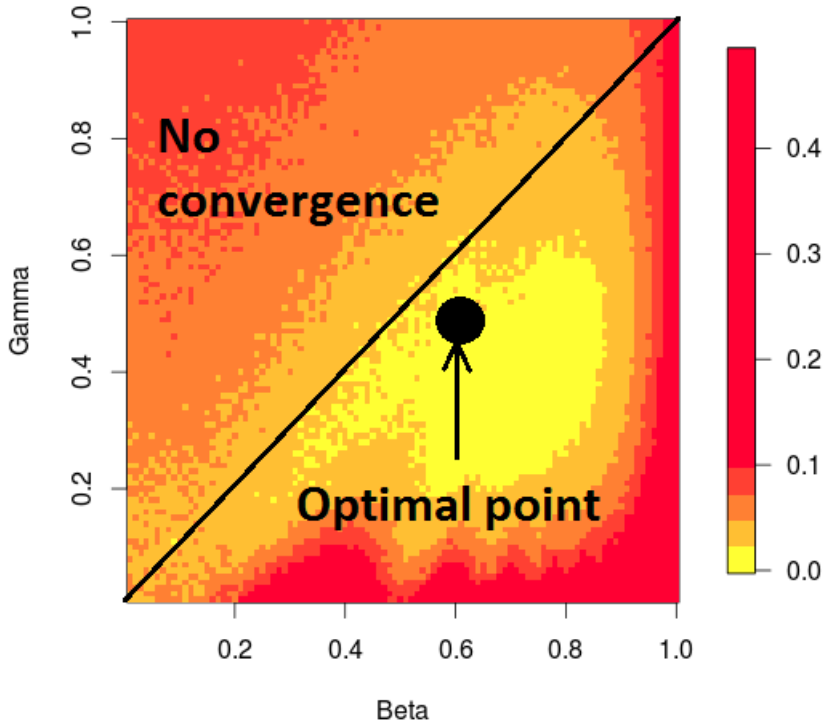


Figure B.6: Theoretical behaviour of the MSE in function of β and γ, $n = 200$.

We have finally shown that

$$a_n(x) \leq \frac{C_9(x, \alpha, d)}{n^{\frac{1}{1+d}-\eta}} ,$$

where the constant is the minimal constant between $C_7(x, \alpha, d)$ and $C_8(x, \alpha, d)$ computed with optimal parameters (γ, β, ζ).

B.5 Conclusion and Perspectives

In this paper, we proposed a sequential method for the estimation of a conditional quantile of the output of a stochastic code where inputs lie in $\mathbb{R}^d$. We introduced a

combination of k-nearest neighbors and Robins-Monro estimator. This algorithm has two parameters: the number of neighbors $k_n = \lfloor n^\beta \rfloor$ and the learning rate $\gamma_n = n^{-\gamma}$. By deriving a bias-variance decomposition of the risk, we showed that our algorithm is convergent for $\frac{1}{2} < \gamma < \beta < 1$ and we studied its mean squared error non-asymptotic rate of convergence. Moreover, we proved that the choice $\gamma = \frac{1}{1+d}$ and $\beta \gtrsim \gamma$ leads to the best rate of convergence. Numerical simulations show that the algorithm tuned with those theoretically optimal parameters is a powerful and accurate estimator of the conditional quantiles, even in dimension $d > 1$.

The theoretical guarantees are shown under strong technical assumptions, but our algorithm is a general methodology to solve the problem. Relaxing the conditions will be the object of a future work. Moreover, the proof that we propose constrained us to use an artefact parameter ζ which implies that the non-asymptotic inequality is theoretically true for large values of n, even if simulations confirm that this problem does not exist in practice. A second perspective is then to find a better way to prove this inequality for smaller n. Finally, it would be of great interest to derive non-asymptotic lower-bounds for the mean squared error of the algorithm.

B.6 Recap of the Constants

Let us sum up all the constants we need in this paper.

B.6.1 Constants of the Model

We denote:

- $M(x)$ the constant of continuity of the model, that is

$$\forall B \in \mathcal{B}_x,\ \forall t \in \mathbb{R},\ |F_{Y^B}(t) - F_{Y^x}t)| \leq M(x) r_B\ .$$

- C_{input} is the positive lower bound of the density of the inputs law f_X.

- $C_g(x)$ is the positive lower bound of the density of the law of $g(x, \varepsilon)$.

B.6.2 Compact Support

We denote:

- $[L_Y, U_Y]$ the compact in which are included the values of g.

- $[L_X, U_X]$ the compact in which is included the support of the distribution of X.

- $[L_{\theta_n}, U_{\theta_n}] := [L_Y - (1-\alpha), U_Y + \alpha]$ the segment in which θ_n can take its values $(\forall x)$.

- $U_{|.|}$ the upper bound of the compact support of the distribution of $||X - x||$ $(\forall x)$.

B.6.3 Real Constants

We denote:

- $\sqrt{C_1} := U_Y + \alpha - L_Y$. C_1 is the uniform in ω and x bound of $(\theta_n(x) - \theta^*(x))^2$.
- $C_2(x,\alpha) := \min\left(C_g(x), \frac{1-\alpha}{U_Y+\alpha-L_Y}\right)$ is the constant such that

$$\left[F_{Y^x}\left(\theta_n(x)\right) - F_{Y^x}\left(\theta^*(x)\right)\right]\left[\theta_n(x) - \theta^*(x)\right] \geq C_2(x,\alpha)\left(\theta_n(x) - \theta^*(x)\right)^2 \ .$$

- $C_3(d) := \sqrt[d]{2}\left(1 + \frac{8}{3d} + \frac{1}{\sqrt[d]{C_{input}}C_4(d)}\right)$.
- $C_4(d) := \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)+1}$.
- $C_5(x,d) := \max\limits_{n\geq N_0+1} C_1 \exp\left(-\frac{3}{8}n^{1-\zeta}\right) n^{2\gamma-\beta+1} + \frac{2\sqrt{C_1}M(x)C_3(d)}{(N_0+1)^{-\gamma+(1-\beta)/d}} + 1.$
- $$C_6(x,d) := \max_{n\geq N_0+1} C_1 \exp\left(-\frac{3}{8}n^{1-\zeta}\right) n^{\gamma+(1+\frac{1}{d})(1-\beta)} + 2\sqrt{C_1}M(x)C_3(d) + \frac{1}{(N_0+1)^{\gamma-\frac{1}{d}(1-\beta)}}.$$
- $$C_5^{optim} := \max_{n\geq N_0+1} C_1 \exp\left(-\frac{3}{8}n^{\left(\frac{1}{1+d}+\eta_\beta\right)-\eta_\zeta}\right)(N_0+1)^{\frac{1}{1+d}-\eta_\beta+1} + 1 + \frac{1}{(N_0+1)^{-\frac{1}{1+d}+\frac{1}{d}(1-\frac{1}{1+d}-\eta_\beta)}}.$$
- $$C_6^{optim}(x,d) := \max_{n\geq N_0+1} C_1 \exp\left(-\frac{3}{8}n^{\left(\frac{1}{1+d}+\eta_\beta\right)-\eta_\zeta}\right) n^{\left(1+\frac{1}{d}\right)-\frac{1}{d(1+d)}-\eta_\beta\left(1+\frac{1}{d}\right)} + 2\sqrt{C_1}M(x)C_3(d) + \frac{1}{(N_0+1)^{-\frac{1}{d}+\frac{1}{d(1+d)}+\frac{1}{1+d}+\frac{\eta_\beta}{d}}}.$$
- $C_7(x,\alpha,d) := \frac{2^{(1+\frac{1}{d})(1-\beta)+\gamma}C_6(x,d)}{C_2(x,\alpha)}$.
- $C_8(x,\alpha) := \frac{2^{2\gamma-\beta+1}C_5(x,d)}{C_2(x,\alpha)}$.
- $C_9(x,\alpha,d) := \min\left(\frac{2^{1+\frac{1}{d}-\frac{1}{d(1+d)}-\eta_\beta(1+\frac{1}{d})}C_5^{optim}(x,d)}{C_2(x,\alpha)}, \frac{2^{\frac{1}{1+d}-\eta_\beta+1}C_6^{optim}(x,d)}{C_2(x,\alpha)}\right)$.
- $C_{10}(d) := \sqrt[d]{\frac{2(k_n+1)}{(n+1)C_{input}C_4(d)}}$.

B.6.4 Integer Constants

We denote :

- $N_0 := 2^{\frac{1}{\zeta-(1-\beta)}}$.
- $N_1(x,\alpha)$ is the rank such that $n \geq N_1(x,\alpha)$ implies

$$\lfloor\frac{n}{2}\rfloor + 1 \geq \frac{2C_2(x,\alpha)}{\zeta+\gamma}.$$

- $N_2(x, \alpha, d)$ is the integer such that $\forall n \geq N_2(x, \alpha, d)$,

  a) If $\beta \leq 1 - d\gamma$,

  $$S_3 + S_1 + T_n^1 + T_n^0 \leq \frac{C_7(x, \alpha, d)}{2n^{\left(1+\frac{1}{d}\right)(1-\beta)-\zeta}} \ ,$$

  where $T_n^1 := \exp\left(-2C_2(x, \alpha) \sum_{k=N_0+1}^{n} k^{-\gamma-\zeta}\right)$, $T_n^0 := C_1 \exp\left(\frac{-3n^{1-\zeta}}{8}\right)$,
  $S_3 := \frac{C_6(x,d)}{n^{\gamma+(1+\frac{1}{d})(1-\beta)}}$ and
  $S_1 := C_6(x, d) \exp(-2C_2(x, \alpha)n^{1-\zeta-\gamma}) \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} k^{-\gamma-(1-\beta)(1+1/d)}$.

  b) If $\beta > 1 - d\gamma$,

  $$S_3 + S_1 + T_n^1 + T_n^0 \leq \frac{C_8(x, \alpha, d)}{2n^{\gamma-\beta+1-\zeta}} \ ,$$

  where $T_n^1 := \exp\left(-2C_2(x, \alpha) \sum_{k=N_0+1}^{n} k^{-\gamma-\zeta}\right)$, $T_n^0 := C_1 \exp\left(\frac{-3n^{1-\zeta}}{8}\right)$,
  $S_3 := \frac{C_5}{n^{2\gamma-\beta+1)}}$ and $S_1 := C_5 \exp(-2C_2(x, \alpha)n^{1-\zeta-\gamma}) \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} k^{-\gamma-(1-\beta)(1+1/d)}$.

- N_3 is the rank such that $\forall n \geq N_3$, $\lfloor \frac{n}{2} \rfloor \geq N_0 + 1$.
- $N_4(x, \alpha, d) := \max(N_0 + 2, N_1(x, \alpha), N_2(x, \alpha, d), N_3)$.

C

Differential Evolution Optimizer

We present hereunder the differential evolutionary (DE) solver used in our application for optimizing the QoI. Two versions are introduced, the first one is the classical algorithm in presence of a deterministic objective function. The second one is our own adaptation of the algorithm when noise affects the function to optimize. All the graphics illustrating the algorithm are taken from Price et al. (2005) that gives a complete overview of DE solvers. Our choice of this algorithm has been influenced by Mystic framework, which mainly relies on this efficient algorithm for solving OUQ problems.

C.1 Differential Evolution Algorithm

Price et al. (2005) presented the first Differential Evolution (DE) algorithm in 1995. Like many evolutionary algorithms, the DE algorithm is a population-based optimizer starts by sampling the objective function at multiple, random points. Points are sampled within a domain defined by its bounds, the objective function taking infinite values outside the domain. For each generated population, the vectors are numbered from 0 to $N_p - 1$. Then, the DE perturbs vectors with the scaled difference of two randomly selected population vectors. To produce a trial vector, $\mathbf{u}_0$, DE adds the scaled random vectors difference to a third, randomly selected, population member. Then, the trial vector u_0 competes with the population vector corresponding to the same index, here number 0. If better, the trial vector is kept as vector number 0 for the next generation, else it is rejected. The procedure is repeated until all N_p population vectors have been tested against a randomly generated trial vector. The pseudo code 3 summarizes the procedure. We propose a graphical presentation of the algorithm for better understanding in Fig. C.1 and C.2.

To complete the DE algorithm, another interesting characteristic is added, called crossover. It consists in replacing some components of the trial vector by components of

Algorithm 3: Differential Evolution Algorithm

while *(convergence criterion not yet met)* **do**
  $\%\mathbf{x_i}$ *defines a vector of the current population*;
  $\%\mathbf{y_i}$ *defines a vector of the new population*;
  for $i = 0, \ldots, N_p$ **do**
    r1=rand(N_p) % *select a random index from* $1, \ldots, N_p$;
    r2=rand(N_p) % *select a random index from* $1, \ldots, N_p$;
    r3=rand(N_p) % *select a random index from* $1, \ldots, N_p$;
    $\mathbf{u}_i = \mathbf{x}_{r3} + F * (\mathbf{x}_{r1} - \mathbf{x}_{r2})$;
    if $\phi(\mathbf{u}_i) \leq \phi(\mathbf{x}_i)$ **then**
      $\mathbf{y}_i = \mathbf{u}_i$;
    else
      $\mathbf{y}_i = \mathbf{x}_i$;

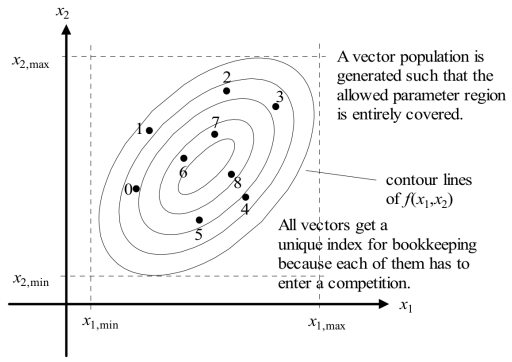
other population vectors. Given a crossover probability C_r the trial vector $\mathbf{u}_i$ mutates through the following pattern

$$\mathbf{u}_i = u_{i,j} = \begin{cases} u_{i,j} \text{ if } U(0,1) \leq C_r \ , \\ x_{i,j} \text{ otherwise} \ . \end{cases} \tag{C.1}$$

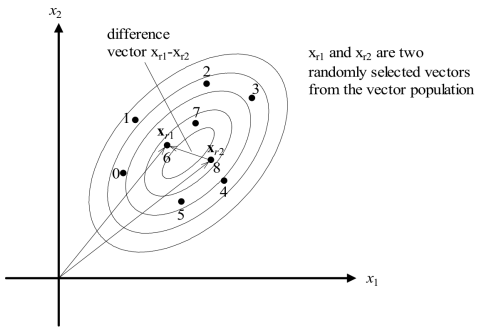
Crossover was originally introduced in order to decrease the likelihood of grouping. Empirical evidence shows that crossover merely speeds the convergence of the algorithm. Nevertheless, it plays an important role in many evolutionary algorithm.

There are several strategy for creating trial candidates (Chakraborty, 2008). In our work, we mostly use the best/1/bin strategy, meaning the difference is used to mutate the best member, so that in fact $\mathbf{x}_{r3} = \mathbf{x}_{best}$ in Algorithm 3 with $\mathbf{x}_{best} = \arg\min_{x_i} \phi(\mathbf{x_i})$.

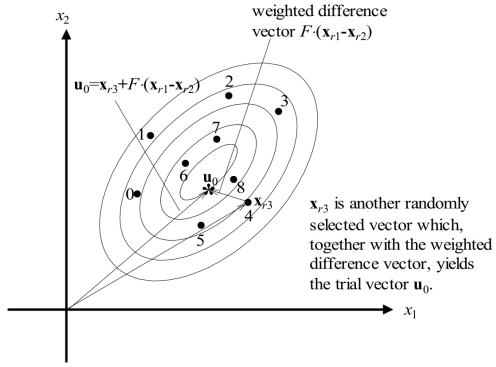
The DE algorithm quickly became a very popular global optimizer due to its empirically good performance, even though no proof of convergence exists to date. In the next section, we explore an adapted version of this algorithm to handle noisy optimization functions.



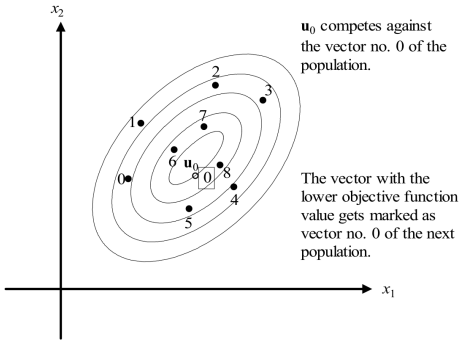
(a) Initializing the DE population.



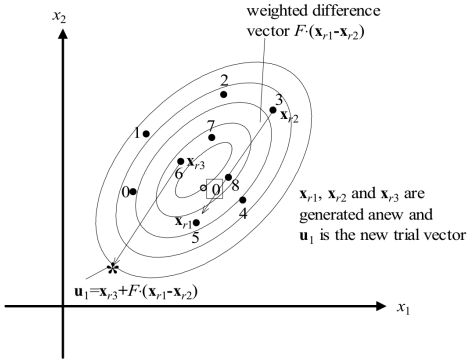
(b) Generating the perturbation: $\mathbf{x}_{r1} - \mathbf{x}_{r2}$.



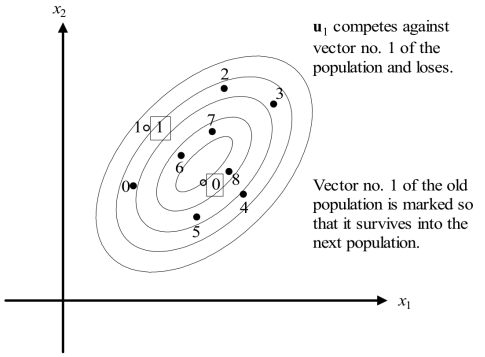
(c) Mutation.



(d) Selection. Because it has a lower function value, $\mathbf{u}_0$ replaces the vector with index 0 in the next generation.

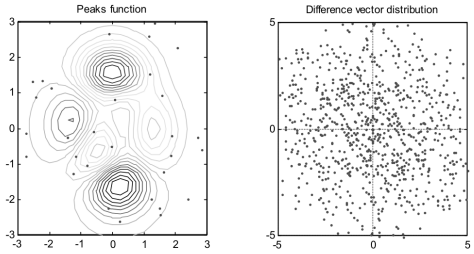


(e) A new population vector is mutated with a randomly generated perturbation.

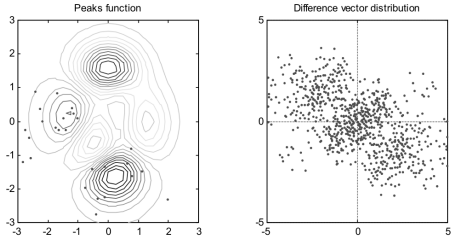


(f) Selection. This time, the trial vector loses.

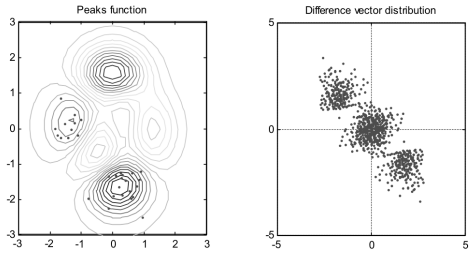
Figure C.1: Differential Evolution step by step (Price et al., 2005)



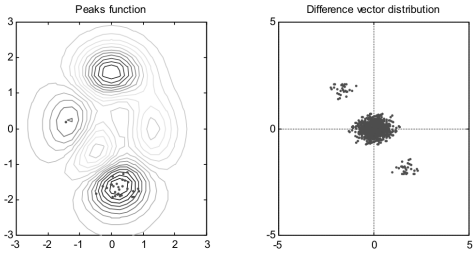
(a) Generation 1: DE's population and difference vector distributions.



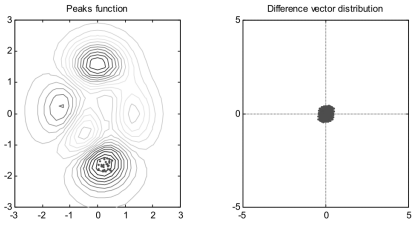
(b) Generation 6: The population coalesces around the two main minima.



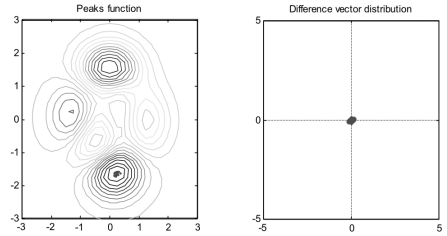
(c) Generation 12: The difference vector distribution contains three main clouds – one for local searches and two for moving between the two main minima.



(d) Generation 16: The population is concentrated on the main minimum.



(e) Generation 20: Convergence is imminent. The difference vectors automatically shorten for a fine-grained, local search.



(f) Generation 26: The population finishes to converge.

Figure C.2: Differential Evolution heuristic of convergence. The distribution of vector differences shows all combinations of vector differences rearranged around the origin (Price et al., 2005)

C.2 Noisy Differential Evolution Algorithm

In this section, we investigate an adaptation of the DE algorithm presented above to account for a noisy function.

One should contextualize what is needed in this thesis. We optimize a QoI on the output of a computer model, parameterized or not by the canonical moments, using a DE solver. We recall that, thanks to the reduction theorems, the optimal QoI is found on the extreme points of a measure space under some conditions described in Chapter 3. In the case where all inputs of the computer model are modeled to belong in a moment class, the extreme points write as convex combination of Dirac masses. Then, computing the QoI, for instance a probability or a quantile, over discrete distributions can be done analytically. So that the computation of the QoI is exact for discrete distribution, thus it is equivalent to the optimization of a deterministic function which refers to the DE algorithm in Section C.1.

However, whenever an input variable is modeled to belong to an unimodal moment class, the extreme points are continuous measures, that write as mixtures of uniform distributions. Therefore, the computation of the QoI, for instance a probability or a quantile requires MC methods. Hence, the QoI is approached by a stochastic function. In order to account for this noise tainting the evaluation of the QoI, we propose a noisy differential evolutionary algorithm that relies on the following generalization of the classical version. This algorithm assumes that the noise affecting the function is a zero-mean Gaussian white noise, whose single hyperparameter is its standard deviation σ. The general idea is that several evaluations of the function with identical inputs reduce the variance of the estimation and improves the signal-to-noise ratio.

As previously, every individual of a population evolves with a scaled difference of two randomly selected population vectors, added to a third randomly selected population member, and possibly mutates [see Eq. (C.1)]. It produces a trial vector that competes with the one of the current generation. But then, when evaluating the function value for these two input vectors, one should account for the noise. In particular, if the two values are distinguishable in a sense that remains to define, then one can easily decide which one to keep for the next generation. However, if their value are too close then one should reduce the variance of their estimation until it becomes clear which one to keep at next generation.

Therefore, one need to define a rule for choosing when and how many new evaluations of the function on the two competing points are necessary. If a candidate solution has been sampled n times, the standard deviation associated to the noise is reduced by $\sqrt{n}$ times. This operation, although beneficial, is clearly computationally expensive, and can significantly slow down the optimisation algorithm. That is why we propose an adaptive strategy of re-sampling, we make use of the scheme proposed in Iacca et al. (2012).

When the offspring $\mathbf{u}_i$ is generated and competes against its ancestor $\mathbf{x}_i$, the value $\delta = |\phi(\mathbf{x}_i) - \phi(\mathbf{u}_i)|$ is computed. If $\delta > \sigma$ the candidate solution with best value is simply chosen for the subsequent generation. This choice can be justified considering that for a given Gaussian distribution, 95.4% of the samples fall into an interval whose amplitude is 4σ and has at its centre the mean value of the distribution. That is why, if the difference between two fitness values is greater that 2σ, it is likely that the point which seems to have better fitness is truly the best candidate solutions.

On the other hand if $\delta < 2\sigma$, noise bands related to the two candidate solutions overlap, and determining the best solution based on only one evaluation is impossible.

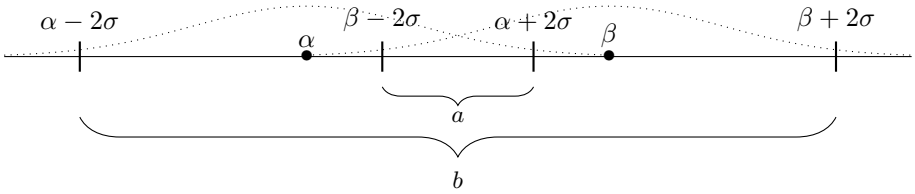


Figure C.3: Graphical representation of the parameter $\nu = a/b$.

Algorithm 4: Noisy Differential Evolution Algorithm

while *(convergence criterion not yet met)* **do**
- %$\mathbf{x_i}$ *defines a vector of the current population*;
- %$\mathbf{y_i}$ *defines a vector of the new population*;
- **for** $i = 0, \ldots, N_p$ **do**
  - r1=rand(N_p) % *select a random index from* $1, \ldots, N_p$;
  - r2=rand(N_p) % *select a random index from* $1, \ldots, N_p$;
  - r3=rand(N_p) % *select a random index from* $1, \ldots, N_p$;
  - $\mathbf{u}_i = \mathbf{x}_{r3} + F * (\mathbf{x}_{r1} - \mathbf{x}_{r2})$;
  - **if** $|\phi(\mathbf{u}_i) - \phi(\mathbf{x}_i)| \leq 2\sigma$ **then**
    - $\mathbf{y}_i = \arg\min\{\phi(\mathbf{x}_i); \phi(\mathbf{u}_i)\}$;
  - **else**
    - $\alpha = \min\{\phi(\mathbf{x}_i); \phi(\mathbf{u}_i)\}$;
    - $\beta = \max\{\phi(\mathbf{x}_i); \phi(\mathbf{u}_i)\}$;
    - $\nu = \frac{\alpha+2\sigma-(\beta-2\sigma)}{\beta+2\sigma-(\alpha-2\sigma)}$;
    - $n_s = \left\lceil \left(\frac{1.96}{2(1-\nu)} \right)^2 \right\rceil$;
    - $n_s = \max\{n_s; 20\}$;
    - **for** $j = 0, \ldots, n_s$ **do**
      - perform resampling of $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{u}_i)$;
    - average the sample as $\overline{\phi}(\mathbf{x}_i)$ and $\overline{\phi}(\mathbf{u}_i)$;
    - $\mathbf{y}_i = \arg\min\{\overline{\phi}(\mathbf{x}_i); \overline{\phi}(\mathbf{u}_i)\}$;

In this case, we define $\alpha = \min\{\phi(\mathbf{x}_i); \phi(\mathbf{u}_i)\}$ and $\beta = \max\{\phi(\mathbf{x}_i); \phi(\mathbf{u}_i)\}$, the following index is calculated:

$$\nu = \frac{\alpha + 2\sigma - (\beta - 2\sigma)}{\beta + 2\sigma - (\alpha - 2\sigma)} \tag{C.2}$$

The index ν represents the intersection of two intervals, characterized by a center in the fitness value and half amplitude 2σ, with respect to their union. In other words, ν is normalized measure of the noise band overlap. This index vary between 0 and 1. The limit condition 0 means that the overlap is limited and thus pairwise ranking given by the single sample estimations is most likely correct. The complementary limit condition, 1 means that the interval overlap is almost total and the two fitness values are too close to be distinguished in the noisy environment. In other words, it can be seen as a reliability measure of a pairwise solution ranking in the presence of noisy values. A graphical representation is depicted in Fig. C.3.

On the basis of the calculated value of ν, a set of additional realizations n_s ar sampled for both the current individual x_i and its competitor u_i. The function is

computed n_s times, this value being determined by

$$n_s = \left\lceil \left(\frac{1.96}{2(1-\nu)} \right)^2 \right\rceil \ , \tag{C.3}$$

where 1.96 is the upper critical value of a normal distribution associated with a confidence level equal to 0.975. Thus, n_s represents the minimum amount of samples which ensure a reliable characterisation of the noise distribution, i.e. the MC sample size which guarantees that the average function value is a sufficiently accurate estimate of its expected value under the sampling distribution.

However, since for $\nu \to 1$, this would result in $n_s \to \infty$, a saturation value for n_s is set in order to avoid infinite loops. Here, we enforce $n_s \leq 20$. The noisy differential evolutionary algorithm has its pseudo code presented in 4. The hyperparemeter σ corresponds to the variance of the noisy function ϕ, it can be estimated by evaluating the function ϕ on a sample test.

D

Résumé Étendu de la Thèse

D.1 Introduction

Dans une étude d'ingénierie, la modélisation numérique des phénomènes physiques est devenue pratique courante. Lorsque l'ingénieur doit étudier un système physique, la réalisation d'expériences répétées est souvent trop coûteuse et complexe à mettre en œuvre. C'est pourquoi les études modernes préfèrent remplacer l'approche expérimentale par des modèles numériques. Ces modèles numériques reproduisent la physique au travers d'équations complexes. Par exemple lors d'une étude de sûreté nucléaire, une expérience simulant un accident en centrale peut être trop coûteuse, dangereuse voir même impossible à mettre en place. Il est alors nécessaire de modéliser numériquement sa physique afin de prédire les évolutions thermiques et mécaniques du système en situation d'accident. Cela s'applique à d'autres domaines d'ingénierie, par exemple un ingénieur en analyse financière qui souhaiterait déterminer la rentabilité d'un projet de parc éolien peut prédire au travers d'un modèle numérique la production du système avant son installation.

La plupart des systèmes physiques ne sont que partiellement connus, d'une part car les équations gouvernant la physique ne sont pas toujours bien maîtrisées mais également car les paramètres du système comme les paramètres physiques d'entrées ou les conditions aux limites ne sont pas parfaitement connus. Dans ce contexte, il est essentiel d'évaluer la précision du modèle numérique. Assurer la fiabilité d'une étude d'ingénierie nécessite donc de prendre en compte toutes les sources d'incertitude pouvant affecter le résultat du code de calcul.

Nous introduisons dans un premier temps la méthode générale de quantification d'incertitude présentement utilisée en ingénierie.

D.2 Modèle numérique

D.2.1 Fonction boîte noire

Dans cette thèse, peu d'hypothèses sont faites sur le code de calcul utilisé. Comme les phénomènes physiques peuvent être complexes à étudier, nous considérons le code comme une fonction boîte noire. Cela signifie qu'étant donné un jeu de paramètres d'entrée, la seule information disponible est la valeur de sortie de la fonction. La formulation mathématique du problème se résume de la manière suivante :

$$\begin{array}{rccc} G: & \mathcal{X} \subset \mathbb{R}^d & \to & \mathcal{Y} \subset \mathbb{R} \\ & \mathbf{x} & \mapsto & y = G(\mathbf{x}) \end{array} \tag{D.1}$$

où G représente le code de calcul et $\mathbf{x} = (x_1, \ldots, x_d)$ sont les valeurs des variables d'entrées dans $\mathbb{R}^d$. Chaque entrée représente un paramètre physique ou environnemental. Étant physiques, les paramètres sont le plus souvent bornés, de telle sorte que l'espace d'entrée $\mathcal{X}$ peut être considéré compact. Bien que le code G soit vu comme une fonction boîte noire c'est-à-dire que l'opérateur n'a pas accès à la physique, il est toutefois possible de déterminer comment une variable d'entrée impacte le résultat de sortie.

Dans cette thèse, le code G est également considéré déterministe. Cela le différencie des codes dit stochastiques qui retournent différentes valeurs de sortie pour un même jeu de paramètres d'entrées. Le code qui reproduit avec plus ou moins de précision les phénomènes physiques, peut être très coûteux. Un appel du code pouvant prendre jusqu'à plusieurs jours de calcul.

L'utilisation d'un modèle de calcul déterministe n'empêche pas l'analyse d'être affectée par différentes sources d'incertitude. La science visant à quantifier l'impact de ces incertitudes sur le résultat d'une expérience numérique est connue sous le nom de quantification d'incertitude (en anglais uncertainty quantification - UQ -) (De Rocquigny et al., 2008; Sullivan, 2015).

D.2.2 Metamodèles

Dans des problèmes d'ingénierie complexes, tel que ceux rencontrés dans des études de sûreté nucléaire, les codes de calcul sont coûteux à évaluer. L'étude des incertitudes devient alors difficile à réaliser. Afin de contourner le problème, la méthode la plus connue consiste à remplacer le code de calcul coûteux par une fonction mathématique peu coûteuse appelée métamodèle (De Lozzo, 2015; Kleijnen and Sargent, 2000). Idéalement, le métamodèle est un substitut fidèle du code. Différentes méthodes de construction existent.

Premièrement, le métamodèle est construit à partir d'un nombre fini et limité d'appels au code. Cet échantillon doit être aussi représentatif que possible du modèle de calcul dans tout le domaine de variation de ses paramètres d'entrées. C'est pourquoi il est habituel d'utiliser des plans d'expérience qui garantissent de bien remplir l'espace des paramètres d'entrée avec un budget donné d'appel au code (Fang et al., 2005). Lorsque la dimension du vecteur d'entrée $\mathbf{x}$ est importante, il est possible d'utiliser des méthodes de *screening* pour distinguer rapidement quelles entrées sont significatives et les trier par influence croissante. Par la suite, le métamodèle peut être construit sur l'échantillon disponible en considérant uniquement les variables les plus influentes (Iooss and Marrel, 2019), les autres paramètres étant alors fixés à une valeur par défaut.

Plusieurs familles de métamodèle sont étudiées dans la littérature (splines, réseaux de neurones, polynôme du chaos, etc...), mais le krigeage, basé sur les processus gaussiens (Sacks et al., 1989; Le Gratiet et al., 2017; Rasmussen and Williams, 2005), est particulièrement intéressant en quantification d'incertitude. Dans ce cadre, le code de calcul est considéré être la réalisation d'un processus gaussien qui est alors conditionné par les observations du code. Cela donne la possibilité de réaliser une inférence statistique puisque qu'un processus gaussien conditionné reste gaussien. De ce fait, des formules analytiques existent pour le prédicteur d'un tel processus, ainsi que l'écart type de la prédiction.

D.3 Sources d'incertitudes

Il est important de pouvoir quantifier la qualité de prédiction d'un modèle numérique. C'est pourquoi toutes les sources possibles d'incertitude qui affectent les résultats du code de calcul doivent être prises en compte. Depuis la réalité physique jusqu'à la construction du modèle, un grand nombre de sources d'incertitude viennent affecter les résultats de la simulation. La classification des différentes sources d'incertitude peut se faire de différentes manières, un état de l'art en la matière se trouve dans Kiureghian and Ditlevsen (2009).

Lorsqu'une source d'incertitude est identifiée, il est intéressant de savoir s'il est possible d'en réduire les conséquences. C'est-à-dire s'il est possible, dans un budget *raisonnable*, d'améliorer notre confiance dans les résultats du modèle. De ce fait, on distingue du point de vue de l'ingénieur deux sources d'incertitude :

- *Incertitude aléatoire.* Elle réfère au caractère aléatoire intrinsèque à tout système physique. Par exemple, le vent qui traverse un champ éolien, ou le résultat du lancer d'une pièce. Il est communément admis qu'un budget réaliste ne permet pas de réduire ce type d'incertitude. Elle affecte le vecteur d'entrée $\mathbf{x}$ du modèle, c'est-à-dire les paramètres physiques et pas directement le modèle de calcul.

- *Incertitude épistémique.* Elle provient du manque d'information de l'analyste. Cette incertitude peut être réduite grâ à un budget raisonnable, par exemple en collectant plus d'information, sous forme de mesures expérimentales, d'avis d'expert, etc. Cette incertitude affecte les paramètres d'entrées comme le modèle numérique.

Certaines écoles ne distinguent pas les incertitudes aléatoires des incertitudes épistémiques. En effet, leur différence repose sur la notion de budget *raisonnable* et donc à la capacité d'acquérir de nouvelles informations (JCGM, 2008). La distinction des deux n'est pas toujours évidente. Par exemple, Fu (2012) étudie le coefficient de Strickler qui mesure la friction du lit d'une rivière. Ce coefficient est teinté à la fois d'incertitude épistémique en raison du faible nombre de mesure expérimentale, et d'incertitude aléatoire (ou variabilité) en raison des transformations naturelles du lit de la rivière au fil des saisons. De la même manière, en disposant d'une quantité phénoménale de capteurs, il serait possible de décrire de manière très fine les vents traversant un champ d'éolienne ou le résultat d'un lancer de pièce. Cette classification fournit toutefois une manière pragmatique d'identifier les incertitudes sur lesquelles il est utile d'allouer un budget supplémentaire pour améliorer la confiance de la prédiction. Les sources d'incertitudes peuvent également se classifier de la manière suivante :

- *Variabilité.* Cela se rapporte à la variabilité naturelle d'un système, par exemple, la taille d'une pièce mécanique fabriquée industriellement.
- *Erreur du modèle.* Cela se rapporte à l'erreur faite lors de la construction du modèle numérique, mais également à l'erreur des équations physiques simplifiées qui reproduisent les phénomènes réels. De plus, lorsque le modèle est remplacé par un métamodèle, l'erreur d'approximation est une source supplémentaire d'erreur.
- *Incertitudes d'entrée.* Plusieurs sources d'incertitudes affectent les paramètres d'entrée. Cela peut provenir d'un manque de données, mais les données elles-mêmes peuvent être de mauvaise qualité, incomplètes ou entachées d'erreur. De plus, certaines informations sur les paramètres d'entrées proviennent de l'avis d'expert qui est potentiellement biaisé.

Dans cette thèse, nous étudions particulièrement l'incertitude qui affecte les données d'entrées. La méthode la plus classique pour prendre en compte ces incertitudes est d'utiliser un cadre probabiliste pour modéliser le manque d'information des variables d'entrée.

D.4 Propagation d'incertitudes

Les incertitudes qui affectent les variables d'entrée peuvent être prise en compte de différentes manières (intervalles, ensembles aléatoires, probabilités floues, info-gap, et autres... voir Ferson and Ginzburg (1996); Ferson and Oberkampf (2009); Paté-Cornell (1996); Qiu et al. (2008); Ben-Haim (2006)). Toutefois, le cadre probabiliste demeure le plus classique. Dans cette thèse, nous utilisons de manière importante la théorie de la mesure, c'est pourquoi certains prérequis sont présentés ci-dessous.

D.4.1 Modélisation probabiliste

Soit $\mathcal{X}$ un espace topologique, dans cette thèse il s'agira essentiellement d'un sous-espace de $\mathbb{R}^d$. On note $\mathcal{P}(\mathcal{X})$ l'ensemble des mesures de probabilité sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, où $\mathcal{B}(\mathcal{X})$ désigne la σ-algèbre borélienne sur $\mathcal{X}$. Un élement $\mu \in \mathcal{P}(\mathcal{X})$ induit un espace de probabilité $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mu)$. Un vecteur aléatoire $\mathbf{X} = (X_1, \ldots, X_d)$ sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ est une fonction (mesurable) telle que

$$\mathbf{X} : \quad \mathcal{B} \to \mathbb{R}^d \tag{D.2}$$

$$\omega \mapsto X(\omega) \ . \tag{D.3}$$

Si $\mathbf{X}$ est un vecteur aléatoire, alors pour tout ensemble Borélien B, $\mathbf{X}^{-1}(B) \in \mathcal{B}$. Il est alors possible de définir une mesure de probabilité $\mu_{\mathbf{X}}$ vérifiant

$$\mu_{\mathbf{X}}(B) = \mathbb{P}(\mathbf{X} \in B) = \mathbb{P}(\mathbf{X}^{-1}(B)).$$

De plus, pour toutes mesures de probabilité μ sur $\mathbf{R}^d$, il est possible d'associer a un vecteur aléatoire X_μ vérifiant pour tout $B \in \mathcal{B}(\mathbb{R}^d)$

$$\mathbb{P}(\mathbf{X}_\mu \in B) = \mu(B).$$

C'est pourquoi, il est d'usage de confondre indifféremment une mesure de probabilité et une variable aléatoire. Une mesure de probabilité μ sur $\mathbb{R}^d$ est souvent décrite par

sa fonction de répartition notée F_μ définie pour tout vecteur $\mathbf{h} = (h_1, \dots, h_d) \in \mathbb{R}^d$ par

$$F_\mu(\mathbf{h}) = \mathbb{P}_\mu(\mathbf{X} \leq \mathbf{h}) = \mathbb{P}_\mu(X_1 \leq h_1, \dots, X_d \leq h_d) = \mu\left(\prod_{1\leq i\leq d}]-\infty, h_i]\right) . \quad \text{(D.4)}$$

La traditionnelle séparation entre mesures discrètes et mesures continues, telle qu'enseignée dans l'enseignement supérieur n'est pas nécessaire ici. En effet, en théorie de la mesure, la continuité absolue est définie par rapport à une mesure de référence. En toute généralité, le théorème de décomposition de Lebesgue prouve que toute mesure de probabilité μ peut se décomposer de la manière suivante par rapport à une mesure de référence ν

$$\mu = \mu_{ac} + \mu_s \ ,$$

où μ_{ac} est une mesure absolument continue par rapport à ν et où μ_s et ν sont deux mesures singulières. Sur $\mathbb{R}^d$ le théorème de décomposition se précise de la manière suivante (Hewitt and Ross, 1979, Theorem 19.20)

$$\mu = \mu_{ac} + \mu_d + \mu_s \ ,$$

où μ_{ac} est absolument continue par rapport à la mesure de Lebesgue, μ_d est une mesure discrète supportée sur un ensemble dénombrable et μ_s une mesure singulièrement continue, c'est-à-dire supportée par un ensemble d'ahérence non-nulle pour la mesure de Lebesgue, mais où la probabilité de chaque point en cet ensemble est nulle (voir la fameuse distribution de Cantor par exemple).

Si la mesure μ est absolument continue par rapport à la mesure de Lebesgue, c'est-à-dire $\mu = \mu_{ac}$, alors sa densité de probabilité f_μ est définie comme la dérivée de Radon-Nikodym de la mesure μ par rapport à la mesure de Lebesgue. Si la mesure est discrète, alors sa densité est définie comme la dérivée de Radon-Nikodym par rapport à la mesure de comptage sur le support de μ.

Dans toute cette thèse, nous employons indifféremment les termes mesure, variable aléatoire, distribution, densité. Toutefois, une certaine préférence inscrit la thèse sous le formalisme de la mesure. En effet, soit $B \in \mathbb{R}^d$ un ensemble mesurable par rapport à une mesure μ, le calcul d'une probabilité sous la notation

$$\mathbb{P}(B) = \int_{\mathbb{R}^d} \mathbb{1}_B(\mathbf{x}) f(\mathbf{x}) \, \mathrm{d}\mathbf{x} \ ,$$

est utilisé davantage en présence d'une mesure absolument continue par rapport à la mesure de Lebesgue. Tandis que la notation plus générale faisant apparaître l'intégrale de Stieljes

$$\mathbb{P}(B) = \int_{\mathbb{R}^d} \mathbb{1}_B(\mathbf{x}) \, \mathrm{d}\mu(\mathbf{x}) \ ,$$

aussi connue sous le nom de transformée de Fourier-Stieljes, est toujours définie.

L'hypothèse d'indépendance des entrées sera également faite dans la quasi-totalité du manuscrit. Ainsi, la distribution d'entrée F_μ peut s'écrire comme un produit de distribution marginale $\prod_{1\leq i\leq d} F_{\mu_i}$ et $f_\mu = \prod_{1\leq i\leq d} f_{\mu_i}$. En d'autres termes, sous hypothèse d'indépendance, l'espace d'entrée $\mathcal{X}$ est un pavé s'écrivant $\prod_{1\leq i\leq d} \mathcal{X}_i$ et la mesure $\mu \in \mathcal{P}(\mathcal{X})$ se décompose $\mu = \otimes_{1\leq i\leq d}\mu_i$ où $\mu_i \in \mathcal{P}(\mathcal{X}_i)$ est la $i$ème projection canonique de μ.

Lorsque l'hypothèse d'indépendance n'est pas vérifiée, le vecteur aléatoire en entrée peut être défini aux travers de ses distributions marginales et d'une copule qui représente la dépendance entre chaque variable (Nelsen, 2006).

D.4.2 Choix des distributions d'entrée

Les ingénieurs rassemblent autant d'informations que possible afin d'améliorer leur connaissance des variables d'entrées. Cependant, ces informations pouvant être incomplètes ou même contradictoires, chaque paramètre d'entrée est modélisé par une variable aléatoire. La modélisation probabiliste reflète le caractère incertain et la variabilité des entrées du modèle.

Certaines informations proviennent d'avis d'experts, et peuvent aider à déterminer des contraintes spécifiques à chaque distribution d'entrée. Bien que cette source d'information soit naturellement biaisée, l'expérience humaine apporte de précieuses informations lorsque les données sont incomplètes. En effet, les données disponibles constituent la principale source d'information et proviennent en général de mesures expérimentales. L'analyste souhaite alors que la distribution de probabilité modélisant les entrées s'ajuste sur les données. De ce fait, les mesures des valeurs d'une entrée constituent un échantillon $\mathscr{X} = \{\mathbf{x}^{(i)}\}_{1\leq i\leq n}$ de taille n suivant une distribution μ qui doit être reconstruite. Il existe plusieurs façons de réaliser cette inférence.

- Une hypothèse est faite sur la nature de la distribution de probabilité à reconstruire, en particulier qu'elle appartienne à une famille de distribution paramétrique, telle que celle des lois Gaussienne, uniforme, log-normale, etc. Dans ce cas, reconstruire la loi de probabilité consiste à trouver le meilleur jeu de paramètres. Il est possible de procéder à une calibration bayésienne, ou une estimation directe des paramètres (Rohde, 2014). La calibration bayésienne repose sur l'élicitation d'une loi *a priori* qui repose sur l'avis d'expert. L'estimation directe peut être faite par exemple par maximum de vraisemblance.

- Lorsqu'aucune hypothèse n'est faite, la distribution de probabilité non-paramétrique est estimée, par exemple en utilisant une estimation à noyau ou l'approximant par histogramme.

L'approche paramétrique est la plus utilisée en pratique. Le choix d'une famille paramétrique peut être fait à l'aide de test statistique comme le test de Kolmogorov-Smirnov ou de Cramér-von-Mises, pour ne citer que les plus connus (Nikolaidis et al., 2004). Dans cette thèse, une hypothèse est faite sur l'existence, d'un point de vue purement formel, d'une distribution parfaite qui modélise le comportement réel de la variable physique. Cette distribution a pour but d'être approchée au plus près.

D.4.3 Quantité d'intérêt

Puisque les variables d'entrées $\{X_i\}_{1\leq i\leq d}$ sont modélisées par des variables aléatoires, la sortie du modèle Y est également une variable aléatoire. Plus précisément, supposons que les X_i sont indépendamment distribuées selon des distributions marginales μ_i, pour $1 \leq i \leq d$, de sorte que le vecteur $\mathbf{X} = (X_1, \ldots, X_d)$ est distribué selon la loi produit $\mu = \mu_1 \times \cdots \times \mu_d$, alors la distribution F_μ de la sortie de modèle numérique G vérifie l'équation suivante :

$$F_\mu(y) = \int_{\mathbb{R}^d} \mathbb{1}_{\{G(x_1,\ldots,x_d)\leq y\}} \, \mathrm{d}\mu_1(x_1) \ldots \mathrm{d}\mu_d(x_d) \; . \tag{D.5}$$

Ce résultat est connu sous le nom de théorème de transport (Barbé and Ledoux, 2007). Puisque, sauf cas particulier, cette équation n'a pas de solution analytique, on estime en pratique la distribution de sortie par des simulations de Monte-Carlo (MC). Pour

cela, un échantillon d'entrée est généré $\mathscr{X} = \{\mathbf{x}_i\}_{1\leq i\leq n}$ suivant la loi d'entrée μ, puis il est propagé à travers le code de calcul, afin d'obtenir un échantillon de sortie $\mathscr{Y} = \{G(\mathbf{x_i})\}_{1\leq i\leq d}$ (De Rocquigny et al., 2008).

La sortie du code de calcul est appelée variable d'intérêt. Toutefois, le statisticien ne s'intéresse pas directement à cette variable mais à des indicateurs statistiques sur la sortie appelés quantités d'intérêt (QoI). Par exemple, considérons un code qui retourne la hauteur d'une rivière. Dans une étude de sûreté, l'analyste va s'intéresser à des QoIs comme la probabilité que la rivière ne dépasse pas une certaine hauteur, ou bien à des quantiles, dans le but de construire une digue adaptée par exemple. Il existe un grand nombre de quantités d'intérêt différentes suivant l'information que l'on souhaite extraire de la sortie du code :

- Il peut s'agir comme indiqué ci-dessus d'une probabilité de défaillance $\mathbb{P}(G(X) \leq h)$ par rapport à un seuil donné $h \in \mathbb{R}$. Plus généralement, on peut par le même moyen reconstruire toute la distribution de sortie F_μ pour caractériser sa variabilité.

- Afin de définir des marges de sûreté, les ingénieurs s'intéressent souvent à des quantités d'intérêt conservatives comme un quantile de niveau p $(0 \leq p \leq 1)$ $Q_p^L(\mu) = \inf\{h \in \mathbb{R} : F_\mu(h) \geq p\}$ (Oakley, 2004; Iooss and Marrel, 2019).

- Il est également possible d'estimer des moments, par exemple la moyenne et/ou la variance de la sortie du code, afin de synthétiser en quelques indicateurs le comportement de la sortie du code. Il existe également des moments généralisés, qui sont définis avec une fonction mesurable par rapport la une mesure d'entrée μ:
$$\mathbb{E}_\mu[\phi] = \int_{\mathbb{R}^d} \phi(\mathbf{x})\, \mathrm{d}\mu(\mathbf{x}) \ ; \tag{D.6}$$

- Une étude de quantification d'incertitude est complémentaire d'une étude d'analyse de sensibilité. Il s'agit d'étudier comment l'incertitude sur la sortie du modèle mathématique peut être divisée et allouée aux différentes sources d'incertitude des entrées. L'analyse de sensibilité a pour but de déterminer et d'ordonner l'importance de chaque entrée sur la variabilité de la distribution en sortie. Cela peut concerner la variation globale de la sortie ou bien se concentrer autour de la variabilité d'une valeur nominale, dans ce cas on parle d'analyse de sensibilité locale. Trois classes de méthodes existent : le *screening* (classement des variables les plus influentes sur la sortie par ordre d'importance, dans un but de reduction de dimension du problème) *les mesures d'importance* (des indices de sensibilité quantitatifs), et l'exploration profonde du comportement du modèle (mesure des effets sur l'ensemble du domaine de variation des entrées). L'analyse de sensibilité retourne comme quantité d'intérêt un ou plusieurs indices facilement interprétables. Les plus connus sont les indices de Sobol, qui décomposent la variance de la sortie en fractions attribuables à chaque entrée (sous hypothèse d'indépendance). D'autres indices comme ceux de Shapley, ou bien les HSIC sont utilisables.

D.4.4 Résumé

La méthodologie présentée ci-dessus est résumée dans la Figure D.1. Comme indiqué, l'étape A précise le problème en définissant le modèle numérique, possiblement son métamodèle, sa sortie, et la quantité d'intérêt en question. Le caractère aléatoire de

la sortie apparaît à l'étape B en définissant des distributions d'entrées. L'étape B' permet de calibrer et de valider le modèle afin qu'il corresponde au mieux aux données disponibles, par exemple en réduisant l'incertitude sur les entrées ou bien en ajustant les paramètres des distributions. Une étape C' de screening permet de simplifier le modèle en fixant certains paramètres non-influent à des valeurs par défaut. Cette boucle de retour permet de mettre à jour le métamodèle et les distributions au besoin. L'étape C qui propage les entrées aléatoires jusqu'à la quantité d'intérêt est connue sous le nom de propagation des incertitudes.

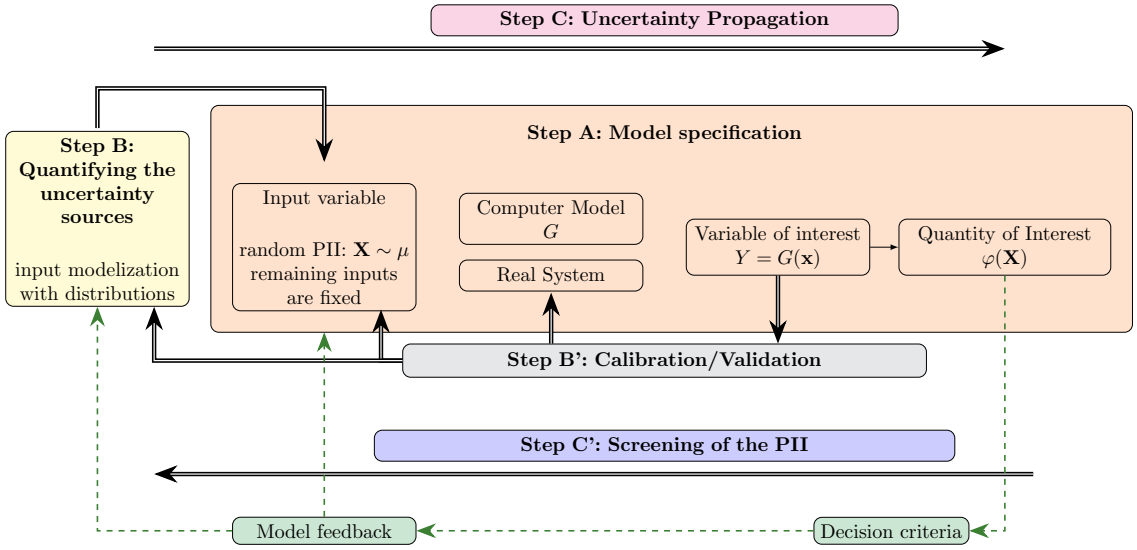


Figure D.1: Ce schéma représente les différentes étapes lors d'une étude de quantification d'incertitude.

D.5 Deuxième niveau d'incertitude

Les incertitudes sur les variables d'entrée sont prises en compte par la méthode présentée ci-dessus. Cependant, la distribution de probabilité qui modélise l'incertitude des variables d'entrée est elle-même incertaine. En effet, le biais naturel d'un avis d'expert et le manque d'information affectent le choix du modèle de probabilité. Par exemple, lorsque peu d'information est disponible et en l'absence d'avis d'expert, des tests statistiques de type test du χ^2 de Pearson ou G-test (Chernoff and Lehmann, 1954) peuvent accepter un grand nombre de modèles de probabilité différents, mais avec un faible pouvoir et une mauvaise estimation des paramètres.

Cette incertitude, qui affecte le choix même de la loi de probabilité, peut être vue comme une incertitude de deuxième niveau. Elle s'exprime de différentes manières :

- Le choix de la distribution est incertain, le choix d'une loi paramétrique peut être inadapté, qu'il s'agisse d'une uniforme, triangulaire, gaussienne, lognormale, etc. Le choix même d'une famille paramétrique est souvent critiquable, car s'il facilite l'approche numérique il n'est souvent qu'une interprétation simpliste de la réalité. Une unique distribution entièrement spécifiée reste cependant difficilement identifiable.

- Une deuxième source d'incertitude provient de l'erreur dans le choix des paramètres de la loi, qui sont estimés sur un nombre limité de données.

Dans les deux cas la distribution d'entrée est imprécise. Ces incertitudes révèlent une limite de l'approche probabiliste classique. La méthodologie d'UQ qui se base sur ce modèle probabiliste est fragilisée à cause de ce deuxième niveau d'incertitude. En présence de cette incertitude, la modification de la distribution de probabilité d'entrée peut transformer significativement la valeur de la quantité d'intérêt. Il est donc primordial de quantifier l'impact de ce deuxième niveau d'incertitude, ce qui est l'objectif premier de cette thèse.

Plusieurs travaux s'inscrivent dans cette démarche sous différents contextes. Une liste non-exhaustive est présentée ci-dessous.

- Les premiers travaux remontent aux années 1980, alors connus sous le nom d'analyse bayésienne robuste (Berger, 1990; Ruggeri et al., 2005). Leur but est de quantifier l'amplitude de variation de la quantité d'intérêt lorsque la loi *a priori* n'est pas précisée. C'est pourquoi celle-ci est supposée appartenir à un certain ensemble de lois de probabilité qui tient compte de nos connaissances. La quantité d'intérêt est minimisée et maximisée sur cet espace de mesure de probabilité. Si l'amplitude de variation de la quantité d'intérêt est petite, alors le résultat est considéré comme robuste, dans le sens où le choix de la loi *a priori* a peu d'impact sur la quantité d'intérêt.

- Plus récemment, la théorie de l'optimal uncertainty quantification (OUQ) a été présentée par Owhadi et al. (2013). Il s'agit d'une analyse de robustesse sur les lois d'entrées. Il est fait l'hypothèse que la quantité d'intérêt est une fonction affine de la loi des entrées, comme par exemple une probabilité. Cette approche, qui s'apparente à l'analyse bayésienne robuste, suppose que la distribution d'entrée appartient à un espace de mesure défini par des contraintes de moments généralisés. Cet espace de mesure doit également être compatible avec nos connaissances sur les variables d'entrées. Des bornes minimales et maximales sont calculées sur la quantité d'intérêt, elles sont optimales dans le sens où ce sont les bornes les plus fines que l'on peut obtenir par rapport à l'information disponible sur les entrées. Cette thèse s'inspire de ces travaux et en reprend les notations.

- En analyse de sensibilité, un deuxième niveau d'incertitude est également nécessaire. Le but est alors de calculer comment l'incertitude sur la distribution d'entrée impacte la valeur des indices. Dans Meynaoui et al. (2019), les auteurs évaluent l'impact du deuxième niveau d'incertitude sur les indices HSIC. Dans Hart and Gremaud (2019), la robustesse des indices de Solob à une perturbation des lois d'entrée est étudiée. Pour ce faire, les auteurs étudient les variations locales des indices de Sobol en calculant leur dérivée de Fréchet par rapport à la distribution d'entrée. Dans Chabridon (2018) est introduit un ensemble d'indices de Sobol dédiés à la fiabilité, prenant en compte les deux niveaux d'incertitude dans un cadre paramétrique. Pour cela, l'auteur sépare l'incertitude aléatoire de celle épistémique en proposant une version désagrégée des variables aléatoires d'entrées. Enfin, les perturbed law indices (PLI) (Lemaître et al., 2015; Sueur et al., 2017; Gauchy et al., 2019) (voir également l'annexe A) sont des nouveaux indices de sensibilité dédié à la mesure de l'impact d'une distribution d'entrée imprécise. Ces indices quantifient la variation relative de la quantité d'intérêt sous perturbation de la distribution d'entrée. Une idée assez proche se trouve dans Pesenti et al. (2019), où la perturbation de la loi d'entrée (appelé stress)

mène ensuite au calcul de la dérivée de la quantité d'intérêt. Dans le même article, la dépendance entre les lois d'entrées et prise en compte de manière directe et indirecte.

- En statistique financière, la terminologie est quelque peu différente, même si la modélisation mathématique est similaire. Les lois de probabilité des entrées représentent des facteurs de risque qui sont agrégé à travers une fonction de perte G. La sortie scalaire représente les pertes d'un portefeuille. La quantité d'intérêt est calculée en utilisant une mesure de risque qui est donc une fonction ne dépendant que de la distribution d'entrée. Ces mesures de risques doivent vérifier certaines propriétés statistiques connues sous le nom d'axiomes de cohérence (Artzner et al., 1999). Les mesures de risques historiques sont la *value at risk* (VaR) équivalente au quantile, et la *conditional value at risk* (CVaR) aussi connue sous le nom d'*expercted shorfall* ou superquantile (Rockafellar and Uryasev, 2002). Un deuxième niveau d'incertitude est aussi intéressant dans ce contexte financier. Dans Cont et al. (2010), les auteurs étudient la sensibilité de la mesure de risque face à une perturbation de la distribution empirique (donc des données). Pour ce faire, ils définissent la robustesse d'une mesure risque par sa continuité par rapport à la métrique de Prokhorov sur l'espace des variables aléatoires intégrables.

- La théorie des probabilités imprécises permet de ne pas spécifier entièrement une distribution de probabilité. Le but est de distinguer les incertitudes épistémiques de celles aléatoires (Schöbi, 2017), ce qui n'est pas possible en présence d'une mesure de probabilité unique qui englobe les deux sources d'incertitude sous une même modélisation. Un grand nombre de théories existe dans la littérature, *evidence theory* (Dempster, 2008; Shafer, 1976), *probability-boxes* (Ferson and Ginzburg, 1996), *fuzzy variables* (Möller and Beer, 2004), *possibility theory* (Dubois and Prade, 1988), *info-gap theory* (Ben-Haim, 2006) en forment une liste non-exhaustive. Toutes considèrent que définir une unique loi de probabilité est trop contraignant. L'*evidence theory* consiste à déterminer des bornes sur des événements spécifiques de la variable aléatoire. Quand ces informations sont rassemblées, il est possible de reconstituer des distributions minimale et maximale appelées mesures de nécessité et mesures de possibilité qui correspondent à l'enveloppe inférieur et supérieur d'un ensemble de distribution vérifiant les données du problème. Une manière équivalente est de fournir directement ces enveloppes dans lesquelles la distribution inconnue doit se situer, c'est le principe des *probability boxes*. Les *probability boxes* sont donc un cas particulier de l'*evidence theory*, un résumé concis sur les probabilités imprécises peut être trouvé dans Schöbi (2017).

Tous ces exemples montrent la nécessité de prendre en compte le manque de confiance dans le choix de la distribution d'entrée.

D.6 Objectifs et déscription du manuscript

C'est dans ce contexte d'analyse de robustesse que prend place la thèse. Nous proposons donc une méthode pour prendre en compte ce deuxième niveau d'incertitude, en se basant sur les travaux de Owhadi et al. (2013). Les résultats théoriques étendus que nous proposons englobent à la fois les travaux en théorie bayésienne robuste et

l'*optimal uncertainty quantification*. Cela dit, l'interprétation de la robustesse et le but de ces théories étant différents, il est inutile de les confondre. Toutefois, le problème d'optimisation que l'on pose peut se résumer de la manière suivante "puisqu'il est difficile de spécifier une unique distribution d'entrée, nous considérons plutôt un espace de mesure". Cet espace doit être cohérent avec les incertitudes et les informations dont on dispose sur les variables d'entrée. C'est sur cet espace que nous tâcherons de calculer des bornes sur la quantité d'intérêt. Ces bornes sont intéressantes en ingénierie, par exemple dans un contexte d'analyse de risque, elles permettent de garantir la sécurité d'une structure puisqu'elles correspondent aux pires scénarios qui puissent survenir. Le but de la robustesse étant justement de garantir des normes de sûreté maximales quitte à surestimer le risque.

Le corps principal de cette thèse est constituée de 8 chapitres dont le contenu est détaillé ci-dessous. Ces chapitres ne sont pas indépendants, il est donc recommandé de lire la thèse dans un ordre chronologique.

Le **Chapitre 1** introduit les concepts fondamentaux de la quantification d'incertitude comme résumé ci-dessus. Nous introduisons la modélisation probabiliste, orientée vers le formalisme de la mesure. Nous introduisons également le concept d'analyse de robustesse liée au deuxième niveau d'incertitude. Un état de l'art est exposé sur ce sujet.

Le **Chapitre 2** présente les différents modèles numériques qui seront étudiés dans la thèse. Cette thèse a pour but de répondre à d'importantes problématiques industrielles et des problèmes d'ingénierie concrets. Deux modèles numériques sont introduits, le premier est un cas jouet suffisamment complexe pour illustrer la méthodologie. Le deuxième est une application réelle liée à des problématiques de sûreté nucléaire, il permettra de tester en profondeur la méthodologie et d'en exposer les limites.

Le **Chapitre 3** traite des aspects théoriques de la thèse. Nous démontrons une généralisation du théorème de réduction à la base de l'*optimal uncertainty quantification*. Les résultats sont essentiellement topologiques, nous démontrons que l'optimum d'une fonction quasi-convexe semi-continue inférieurement est atteint sur les points générateurs (sous réserve d'existence) d'un sous-ensemble convexe d'un espace topologique localement convexe. Nous étudions également des espaces de mesures qui satisfont les hypothèses du théorème qui seront utilisés dans la suite de la thèse. Ces espaces s'appellent la classe de moment et la classe de moment unimodale. La distribution d'entrée des modèles numérique du Chapitre 2 sera supposé appartenir à un produit de ces espaces de mesures.

Le **Chapitre 4** introduit la paramétrisation des deux classes de moments par les moments canoniques. Pour cela, nous faisons l'hypothèse que ces espaces de mesures sont contraints par des moments classiques. Puis nous exposons une paramétrisation nouvelle de la quantité d'intérêt. En particulier, nous prouvons l'existence d'une bijection entre les points extrémaux des classes de moments et un pavé de la forme $[0, 1]^n$, où n dépend du nombre de contraintes de moments. Nous exposons également dans ce chapitre l'algorithme principal de cette thèse servant à optimiser la quantité d'intérêt.

Le **Chapitre 5** présente de nombreuses quantités d'intérêt pratiques. Par exemple, une probabilité de défaillance, un quantile, un superquantile ou bien un indice de Sobol. Nous étudions à la fois leur intérêt théorique dans le cadre de l'OUQ et leur

intérêt pratique, en illustrant chacune d'entre elles sur le cas jouet du Chapitre 2. Nous démontrons également que l'analyse bayésienne robuste est un cas particulier du cadre théorique introduit.

Le **Chapitre 6** applique la méthodologie au cas d'étude réel présenté dans le Chapitre 2. Nous détaillons comment prendre en compte l'incertitude induite par le métamodèle dans cette situation. Ce chapitre peut servir de guide pratique pour une étude d'analyse de robustesse *via* l'OUQ.

Le **Chapitre 7** présente certaines pistes et leurs limites pour gérer la grande dimension et la dépendance des variables d'entrées. La première tente d'utiliser les moments canoniques dans un contexte de grande dimension des paramètres d'entrées, typiquement dans un réseau de neurones. La deuxième piste explore une méthode conservative en relaxant le problème d'optimisation sous forme d'un programme semi-défini. La dernière proposition se base sur les copules pour gérer la dépendance entre paramètres d'entrées. Toutes ces pistes montre la difficulté de gérer la dimension d'entrée du problème.

Le **Chapitre 8** conclut ces travaux. Nous y discutons les limites de l'OUQ appliqué à un contexte industriel. Nous proposons des perspectives d'évolutions de ces travaux, certaines étant actuellement à l'étude.

Table D.1: Table des mots-clés de la thèse

Keywords	**Chapter**
Modélisation des incertitudes / Théorie de la mesure / OUQ / Analyse de robustesse	Chapter 1
Modèle numérique / Expérience numérique / boîte noire	Chapter 2
Principe du maximum de Bauer / Quasi-convexité / Classe des moments / Classe des moments unimodale	Chapter 3
Moments canoniques / Optimisation globale / Calcul de la QoI	Chapter 4
Probabilité de défaillance / Quantile / Superquantile / Indices de Sobol / Analyse Bayesienne robuste	Chapter 5
Code thermaux-hydraulique / Quantile robuste	Chapter 6
Dépendence / Generalized Moment Problem / Programme semi-défini / Copule	Chapter 7

Publications et communications

Les publications liées à cette thèse sont présentées ci-dessous.

Stenger, J., Gamboa, F., Keller, M., and Iooss, B. 2020. Optimal Uncertainty Quantification of a Risk Measurement from a Thermal-Hydraulic Code using Canonical Moments. In *International Journal for Uncertainty Quantification 10*, 1, 35-53.

Stenger, J., Gamboa, F., Keller, M., and Iooss, B. 2019. Canonical Moments for Optimal Uncertainty Quantification on a Variety. In *Geometric Science of Information.* Springer, Cham, 571-578.

Stenger, J., Gamboa, F., and Keller, M. 2020. Optimization of a Quasi-Convex Function Over Product Measure Sets. *Preprint. arXiv: 1907.07934.*

Stenger, J., Gamboa, F., Keller, M., and Iooss, B. 2020. Quantification Robuste de l'Incertitude d'une Mesure de Risque Issue d'un Code de Calcul. *52èmes Journées de Statistiques de la Société Française de Statistique (SFdS).*

Gauchy, C., Stenger, J., Sueur, R., and Iooss, B. 2019. An Information Geometry Approach for Robustness Analysis in Uncertainty Quantification of Computer Codes. *Preprint. hal-02425477.*

Labopin-Richard, T., Gamboa, F., Garivier, A. and Stenger, J. 2019. Conditional Quantile Sequential Estimation for Stochastic Codes. In *Journal of Statistical Theory and Practice 13.*

Bibliography

Abramowitz, M. and Stegun, I., Eds. 1964. *Handbook of Mathematical Functions.* Dover Publications, Inc. New York. 148, 186

Aliprantis, C. D. and Border, K. C. 2007. *Infinite Dimensional Analysis: A Hitchhiker's Guide.* Springer Science & Business Media. 29, 75, 82, 85, 94

Amari, S. and Nagaoka, H. 2000. *Methods of Information Geometry.* Oxford University Press. 153

Amari, S.-i. 2016. *Information Geometry and Its Applications.* Springer Japan. 150

Arnold, V. I. 1997. *Mathematical Methods of Classical Mechanics (Graduate Texts in Mathematics, Vol. 60).* Springer. 152

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. 1999. Coherent Measures of Risk. *Mathematical Finance 9,* 3, 203–228. 16, 218

Baccou, J., Zhang, J., Fillion, P., Damblin, G., Petruzzi, A., Mendizábal, R., Reventós, F., Skorek, T., Couplet, M., Iooss, B., Oh, D.-Y., and Takeda, T. 2019. Development of good practice guidance for quantification of thermal-hydraulic code model input uncertainty. *Nuclear Engineering and Design 354*, 110173. 22, 162

Barbé, P. and Ledoux, M. 2007. *Probabilité.* EDP Sciences, Les Ulis, France. 12, 214

Baudin, M., Dutfoy, A., Iooss, B., and Popelin, A.-L. 2017. Title: Open TURNS: An industrial software for uncertainty quantification in simulation. In *Handbook of uncertainty quantification*, D. H. R. Ghanem and H. Owhadi, Eds. Springer. 120

Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. 2012. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing 22,* 3, 773–793. 168

Ben-Haim, Y. 2006. *Info-Gap Decision Theory: Decisions Under Severe Uncertainty.* Elsevier. 10, 16, 212, 218

Ben-Tal, A., Ghaoui, L. E., and Nemirovski, A. 2009. *Robust Optimization.* Princeton University Press. 105

Benes, V. and Štěpán, J. 1991. Extremal Solutions in the Marginal Problem. *Advances in Probability Distributions with Given Marginals*, 189–206. Publisher: Springer, Dordrecht. 128

Benoumechiara, N., Michel, B., Saint-Pierre, P., and Bousquet, N. 2018. Detecting and modeling worst-case dependence structures between random inputs of computational reliability models. *arXiv:1804.10527 [stat]*. arXiv: 1804.10527. 127

Bereanu, B. 1972. Quasi-convexity, strictly quasi-convexity and pseudo-convexity of composite objective functions. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique 6,* R1, 15–26. 31

Berger, J. O. 1990. Robust Bayesian analysis: sensitivity to the prior. *Journal of Statistical Planning and Inference 25,* 3 (July), 303–328. 15, 80, 217

Bertin, E. M. J., Cuculescu, I., and Theodorescu, R. 1997. *Unimodality of Probability Measures*. Mathematics and Its Applications. Springer Netherlands. 35

Bertin, E. M. J. and Theodorescu, R. 1984. Hinčin spaces and unimodal probability measures. *Mathematische Annalen 266,* 3 (Sept.), 357–367. 36

Bertsimas, D., Popescu, I., and Sethuraman, J. 2000. Moment Problems and Semidefinite Optimization. Tech. rep. type: dataset. 118, 127

Betro, B., Ruggeri, F., and Meczarski, M. 1994. Robust Bayesian analysis under generalized moments conditions. *Journal of Statistical Planning and Inference 41,* 3 (Oct.), 257–266. 97

Beyer, H.-G. and Sendhoff, B. 2007. Robust optimization – A comprehensive survey. *Computer Methods in Applied Mechanics and Engineering 196,* 33-34 (July), 3190–3218. 105

Bhattacharya, P. K. and Gangopadhyay, A. K. 1990. Kernel and nearest-neighbor estimation of a conditional quantile. *The Annals of Statistics*, 1400–1415. 170

Björck, A. and Pereyra, V. 1970. Solution of Vandermonde systems of equations. *Mathematics of Computation 24,* 112, 893–903. 61

Blatman, G. 2009. Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis. Ph.D. thesis, Université Blaise Pascal - Clermont-Ferrand II. 120

Blum, J. R. 1954. Approximation methods which converge with probability one. *The Annals of Mathematical Statistics*, 382–386. 169, 188

Bonnet, L., Akian, J.-L., Savin, E., and Sullivan, T. 2020. Adaptive reconstruction of imperfectly observed monotone functions, with applications to uncertainty quantification. *Algorithms 13,* 8 (Aug), 196. 142

Bovet, D. P. and Crescenzi, P. 1994. *Introduction to the Theory of Complexity*. Prentice Hall, New York. 132

Brockwell, P. J., Davis, R. A., and Fienberg, S. E. 1991. *Time Series: Theory and Methods: Theory and Methods*. Springer Science & Business Media. 141

Brown, J. R. 1965. Doubly stochastic measures and Markov operators. *The Michigan Mathematical Journal 12,* 3 (Sept.), 367–375. Publisher: The University of Michigan. 128

Bucalossi, A., Petruzzi, A., Kristof, M., and D'Auria, F. 2010. Comparison between Best-Estimate–Plus–Uncertainty Methods and Conservative Tools for Nuclear Power Plant Licensing. *Nuclear Technology 172,* 1 (Oct.), 29–47. 22, 145

Burkard, R. E., Dell'Amico, M., and Martello, S. 2009. *Assignment Problems, Revised Reprint.* SIAM. 131

Cannamela, C., Garnier, J., and Iooss, B. 2008. Controlled Stratification for Quantile Estimation. *The Annals of Applied Statistics 2,* 4, 1554–1580. 25

Cardot, H., Cénac, P., and Godichon, A. 2017. Online estimation of the geometric median in hilbert spaces: non asymptotic confidence balls. *The Annals of Statistics 45,* 2, 591–614. 170, 195

Chabridon, V. 2018. Reliability-oriented sensitivity analysis under probabilistic model uncertainty – Application to aerospace systems. Ph.D. thesis, Clermont-Auvergne. 15, 146, 217

Chakraborty, U., Ed. 2008. *Advances in Differential Evolution.* Studies in Computational Intelligence. Springer-Verlag, Berlin Heidelberg. 203

Chernoff, H. and Lehmann, E. L. 1954. The Use of Maximum Likelihood Estimates in χ^2 Tests for Goodness of Fit. *The Annals of Mathematical Statistics 25,* 3 (Sept.), 579–586. 14, 216

Choquet, G., Marsden, J., Lance, T., and Gelbart, S. S. 1969. *Lectures on analysis. Représentation theory ... Vol. 2. Vol. 2.* W.A. Benjamin, New York; Amsterdam. OCLC: 25191467. 29, 30, 31, 35, 36

Clausing, A. 1983. On Quotients of l_p-Means. *General Inequalities 3*, 43–68. 44

Coles, S. 2001. *An Introduction to Statistical Modeling of Extreme Values.* Springer Series in Statistics. Springer-Verlag, London. 20, 96

Cont, R., Deguest, R., and Scandolo, G. 2010. Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance 10,* 6 (June), 593–606. 16, 146, 218

Costa, S. I. R., Santos, S. A., and Strapasson, J. E. 2012. Fisher information distance: a geometrical reading. *Discrete Applied Mathematics 197*, 59–69. 146, 150, 151, 153

Croom, F. H. 2016. *Principles of Topology.* Courier Dover Publications. 94

Cybenko, G. 1989. Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals and Systems 2,* 4 (Dec.), 303–314. 112

Czado, C. 2010. Pair-Copula Constructions of Multivariate Copulas. *Copula Theory and Its Applications*, 93–109. Publisher: Springer, Berlin, Heidelberg. 127

Da Veiga, S. 2013. Global Sensitivity Analysis with Dependence Measures. *Journal of Statistical Computation and Simulation 85*, 1283–1305. 13, 23

Damblin, G. and Gaillard, P. 2020. Bayesian inference and non-linear extensions of the CIRCE method for quantifying the uncertainty of closure relationships integrated into thermal-hydraulic system codes. *Nuclear Engineering and Design 359*, 110391. 24, 100

David, H. A. and Nagaraja, H. N. 2003. *Order Statistics*. Wiley. 183

De Crécy, A. 2001. Determination of the uncertainties of the constitutive relationships of the CATHARE 2 code. In *M&C 2001*. Salt Lake City, Utah, USA. 24

De Lozzo, M. 2015. Substitution de modèle et approche multifidélité en expérimentation numérique. *Journal de la Société Française de Statistique 156,* 4. 8, 82, 210

De Lozzo, M. and Marrel, A. 2016. New improvements in the use of dependence measures for sensitivity analysis and screening. *Journal of Statistical Computation and Simulation 86,* 15 (Oct.), 3038–3058. 23

De Rocquigny, E., Devictor, N., and Tarantola, S. 2008. *Uncertainty in Industrial Practice: A Guide to Quantitative Uncertainty Management*. John Wiley & Sons. 8, 10, 12, 144, 210, 215

Delage, T., Sueur, R., and Iooss, B. 2018. Robustness analysis of epistemic uncertainties propagation studies in LOCA assessment thermal-hydraulic model. In *ANS Best Estimate Plus Uncertainty International Conference (BEPU)*. Lucca, Italy. 70, 155, 163, 164

Dempster, A. P. 2008. Upper and Lower Probabilities Induced by a Multivalued Mapping. *Classic Works of the Dempster-Shafer Theory of Belief Functions*, 57–72. Publisher: Springer, Berlin, Heidelberg. 16, 218

Dette, H. and Studden, W. J. 1997. *The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis*. Wiley-Blackwell, New York. 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 66

Devroye, L., Györfi, L., and Lugosi, G. 2013. *A probabilistic theory of pattern recognition*. Vol. 31. Springer Science & Business Media. 184

Dhillon, I. S. and Parlett, B. N. 2003. Orthogonal Eigenvectors and Relative Gaps. *SIAM Journal on Matrix Analysis and Applications 25,* 3 (Jan.), 858–899. 67

Drapeau, S. and Kupper, M. 2012. Risk Preferences and Their Robust Representation. *Mathematics of Operations Research 38,* 1 (Nov.), 28–62. 82

Dubois, D. and Prade, H. 1988. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Springer US. 16, 218

Duflo, M. and Wilson, S. S. 1997. *Random iterative models*. Vol. 22. Springer Berlin. 170

Durante, F. and Sempi, C. 2015. *Principles of Copula Theory*. CRC Press. 127, 128

Durante, F. and Sánchez, J. F. 2012. On the approximation of copulas via shuffles of Min. *Statistics & Probability Letters 82,* 10 (Oct.), 1761–1767. 129

Efron, B. 1979. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics 7,* 1 (Jan.), 1–26. Publisher: Institute of Mathematical Statistics. 157

Eguchi, S. and Copas, J. 2006. Interpreting Kullback-Leibler Divergence with the Neyman-Pearson Lemma. *Journal of Multivariate Analysis 97*, 2034–2040. 149

Fabian, V. 1968. On asymptotic normality in stochastic approximation. *The Annals of Mathematical Statistics*, 1327–1332. 169

Fan, K. 1953. Minimax Theorems. *Proceedings of the National Academy of Sciences of the United States of America 39,* 1, 42–47. 87

Fang, K.-T., Li, R., and Sudjianto, A. 2005. *Design and Modeling for Computer Experiments (Computer Science & Data Analysis)*. Chapman & Hall/CRC. 8, 23, 145, 163, 210

Feller, W. 1971. *An Introduction to Probability Theory and Its Applications, Vol. 2, 2nd Edition*, 2nd edition ed. John Wiley & Sons, Inc., New York. 50, 119

Ferson, S. and Ginzburg, L. R. 1996. Different methods are needed to propagate ignorance and variability. *Reliability Engineering & System Safety 54,* 2 (Nov.), 133–144. 10, 16, 212, 218

Ferson, S. and Oberkampf, W. 2009. Validation of imprecise probability models. *International Journal of Reliability and Safety*, 3–22. 10, 212

Freedman, D. A. 1965. On the Asymptotic Behavior of Bayes Estimates in the Discrete Case II. *Annals of Mathematical Statistics 36,* 2 (Apr.), 454–456. Publisher: Institute of Mathematical Statistics. 97

Frikha, N. and Menozzi, S. 2012. Concentration bounds for stochastic approximations. *Electronic Communications in Probability 17,* 47, 1–15. 169

Fu, S. 2012. Inverse problems occurring in uncertainty analysis. Ph.D. thesis, Paris 11. 9, 211

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. 1989. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology 61,* 2 (Feb.), 331–349. 112

Gadat, S., Klein, T., and Marteau, C. 2016. Classification with the nearest neighbor rule in general finite dimensional spaces: necessary and sufficient conditions. *The Annals of Statistics 44,* 3, 982–1009. 172

Gal, Y. and Ghahramani, Z. 2016. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. Vol. 48. W&CP, New York, NY, 10. 114

Gamboa, F. and Lozada-Chang, L.-V. 2004. Large Deviations for random power moment problem. *The Annals of Probability 32,* 3B (July), 2819–2837. Publisher: Institute of Mathematical Statistics. 52

Gassiat, E. 1990. Problèmes des moments et concentration de mesure. *Comptes rendus de l'Académie des sciences. Série 1, Mathématique 310,* 2, 41–44. 47

Gauchy, C., Stenger, J., Sueur, R., and Iooss, B. 2019. An information geometry approach for robustness analysis in uncertainty quantification of computer codes. *preprint*. 15, 70, 217

Gautschi, W. 1990. How (Un)stable Are Vandermonde Systems? *Asymptotic and computational analysis 124*, 193–210. 61

Gautschi, W. 2006. Orthogonal Polynomials, Quadrature, and Approximation: Computational Methods and Software (in Matlab). In *Orthogonal Polynomials and Special Functions*, J.-M. Morel, F. Takens, B. Teissier, F. Marcellán, and W. Van Assche, Eds. Vol. 1883. Springer Berlin Heidelberg, Berlin, Heidelberg, 1–77. Series Title: Lecture Notes in Mathematics. 67

Geffraye, G., Antoni, O., Farvacque, M., Kadri, D., Lavialle, G., Rameau, B., and Ruby, A. 2011. CATHARE 2 V2.5$_2$: A single version for various applications. *Nuclear Engineering and Design 241,* 11, 4456–4463. 22, 162

Gelfand, I. and Fomin, S. 2012. *Calculus of Variations*. Dover Publications. 152

Ghanem, R., Higdon, D., and Owhadi, H., Eds. 2017. *Springer Handbook on Uncertainty Quantification*. Springer. 145

Ghosh, P. P. and Bhandari, S. K. 2017. Characterization of Extreme Copulas. *arXiv:1709.02472 [math]*. arXiv: 1709.02472. 127, 128, 129, 132

Godichon, A. 2016. Estimating the geometric median in hilbert spaces with stochastic gradient algorithms. *Journal of Multivariate Analysis 146*, 209–222. 170, 175

Greenberg, H. J. and Pierskalla, W. P. 1971. A Review of Quasi-Convex Functions. *Operations Research 19,* 7, 1553–1570. 29

Han, S., Tao, M., Topcu, U., Owhadi, H., and Murray, R. M. 2015. Convex Optimal Uncertainty Quantification. *SIAM Journal on Optimization 25,* 3 (Jan.), 1368–1387. 142

Hart, J. and Gremaud, P. A. 2019. Robustness of the Sobol' Indices to Distributional Uncertainty. *International Journal for Uncertainty Quantification 9,* 5. Publisher: Begel House Inc. 15, 90, 146, 217

Hausdorff, F. 1923. Momentprobleme für ein endliches Intervall. *Mathematische Zeitschrift 16*, 220–248. 50

Henrici, P. 1993. *Applied and Computational Complex Analysis, Volume 3: Discrete Fourier Analysis, Cauchy Integrals, Construction of Conformal Maps, Univalent Functions*. John Wiley & Sons. 54

Henrion, D., Lasserre, J.-B., and Löfberg, J. 2009. GloptiPoly 3: moments, optimization and semidefinite programming. *Optimization Methods and Software 24,* 4-5 (Oct.), 761–779. 124, 125

Hesterberg, T. C. 1996. Estimates and confidence intervals for importance sampling sensitivity analysis. *Math. Comput. Modelling 23*, 79–85. 155

Hewitt, E. and Ross, K. A. 1979. *Abstract Harmonic Analysis*. Grundlehren der mathematischen Wissenschaften, vol. 115. Springer New York, New York, NY. 10, 213

Hofer, M. and Iacò, M. R. 2014. Optimal Bounds for Integrals with Respect to Copulas and Applications. *Journal of Optimization Theory and Applications 161,* 3 (June), 999–1011. 127, 131, 132

Holbrook, A., Lan, S., Streets, J., and Shahbaba, B. 2017. The nonparametric Fisher geometry and the chi-square process density prior. *arXiv e-prints*, arXiv:1707.03117. _eprint: 1707.03117. 165

Homma, T. and Saltelli, A. 1996. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety 52,* 1 (Apr.), 1–17. 91

Horn, R. A. and Johnson, C. R. 2012. *Matrix analysis*, 2nd ed ed. Cambridge University Press, Cambridge ; New York. 67

Hornik, K., Stinchcombe, M., and White, H. 1989. Multilayer feedforward networks are universal approximators. *Neural Networks 2,* 5 (Jan.), 359–366. 112

Iacca, G., Neri, F., and Mininno, E. 2012. Noise analysis compact differential evolution. *International Journal of Systems Science*. 206

Insua, D. R. and Ruggeri, F., Eds. 2000. *Robust Bayesian Analysis*. Lecture Notes in Statistics. Springer-Verlag, New York. 94, 118

Iooss, B. and Le Gratiet, L. 2019. Uncertainty and sensitivity analysis of functional risk curves based on Gaussian processes. *Reliability Engineering and System Safety 187*, 58–66. 146

Iooss, B. and Lemaître, P. 2015. A Review on Global Sensitivity Analysis Methods. *Uncertainty Management in Simulation-Optimization of Complex Systems 59*, 101–122. 19, 20, 90, 91, 145, 158, 160

Iooss, B. and Marrel, A. 2019. Advanced Methodology for Uncertainty Propagation in Computer Experiments with Large Number of Inputs. *Nuclear Technology* 237 (Mar.), 1–19. 8, 13, 22, 24, 25, 82, 101, 103, 145, 210, 215

Iooss, B., Vergès, V., and Larget, V. 2020. BEPU robustness analysis via perturbed-law based sensitivity indices. In *Accepted to the ANS Best Estimate Plus Uncertainty International Conference (BEPU 2020)*. Giardini Naxos, Italy. 146, 157

Ishigami, T. and Homma, T. 1990. An Importance Quantification Technique in Uncertainty Analysis for Computer Models. In *Proceedings. First International Symposium on Uncertainty Modeling and Analysis*. IEEE Comput. Soc. Press. 158

Isii, K. 1962. On sharpness of tchebycheff-type inequalities. *Annals of the Institute of Statistical Mathematics 14,* 1 (Dec.), 185–197. 118

Jala, M., Lévy-Leduc, C., Moulines, É., Conil, E., and Wiart, J. 2014. Sequential design of computer experiments for the assessment of fetus exposure to electromagnetic fields. *Technometrics 58,* 1, 30–42. 168

JCGM. 2008. Evaluation of measurement data — Guide to the expression of uncertainty in measurement. Tech. Rep. JCGM 100:2008. 9, 211

Jin, R., Chen, W., and Sudjianto, A. 2005. An Efficient Algorithm for Constructing Optimal Design of Computer Experiments. *Journal of Statistical Planning and Inference 134*, 268–287. 163

Joe, H. 1996. *Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters*. Institute of Mathematical Statistics. Pages: 120-141 Publication Title: Distributions with fixed marginals and related topics. 127

Johnson, N. L., Kotz, S., and Balakrishnan, N. 1995. *Continuous univariate distributions*. Wiley & Sons. 52

Kamga, P. H. T., Li, B., McKerns, M., Nguyen, L. H., Ortiz, M., Owhadi, H., and Sullivan, T. J. 2014. Optimal uncertainty quantification with model uncertainty and legacy data. *Journal of the Mechanics and Physics of Solids 72*, 1–19. 142

Kendall, D. G. 1962. Simplexes and Vector Lattices. *Journal of the London Mathematical Society s1-37,* 1, 365–371. 36

Kennedy, M. C. and O'Hagan, A. 2000. Predicting the output from a complex computer code when fast approximations are available. *Biometrika 87,* 1, 1–13. 168

Khintchine, A. Y. 1938. On Unimodal Distributions. *Izvestiya Nauchno-Issledovatel'skogo Instituta Matematiki i Mekhaniki 2*, 1–7. 35, 36

Kiureghian, A. D. and Ditlevsen, O. 2009. Aleatory or Epistemic? Does it Matter? *Structural Safety 31,* 2 (Mar.), 105–112. 9, 211

Kleijnen, J. P. C. and Sargent, R. G. 2000. A methodology for fitting and validating metamodels in simulation. *European Journal of Operational Research 120*, 14–29. 8, 210

Knight, F. H. 1921. Risk, Uncertainty and Profit. SSRN Scholarly Paper ID 1496192, Social Science Research Network, Rochester, NY. 15

Kuhn, H. W. 1955. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly 2,* 1-2 (Mar.), 83–97. 131

Larget, V. 2019. How to bring conservatism to a BEPU analysis. In *NURETH-18*. Portland, USA. 145, 146, 155, 163

Larget, V. and Gautier, M. 2020. Increasing conservatism in BEPU IB LOCA safety studies using complementary and industrially cost effective statistical tools. In *Accepted to the ANS Best Estimate Plus Uncertainty International Conference (BEPU 2020)*. Giardini Naxos, Italy. 146, 165

Lasserre, J. B. 2002. Bounds on measures satisfying moment conditions. *The Annals of Applied Probability 12,* 3 (Aug.), 1114–1137. 118

Lasserre, J.-B. 2010. *Moments, positive polynomials and their applications*. Number v. 1 in Imperial College Press optimization series. Imperial College Press ; Distributed by World Scientific Publishing Co, London : Signapore ; Hackensack, NJ. OCLC: ocn503631126. 13, 118, 119, 121, 122, 123, 124, 126, 127

Le Gratiet, L. 2013. Multi-fidelity gaussian process regression for computer experiments. Ph.D. thesis, Paris 7. 141

Le Gratiet, L., Marelli, S., and Sudret, B. 2017. Metamodel-Based Sensitivity Analysis: Polynomial Chaos Expansions and Gaussian Processes. *Handbook of Uncertainty Quantification*, 1289–1325. 8, 24, 120, 211

Leimkuhler, B. and Reich, S. 2005. *Simulating Hamiltonian Dynamics*. Cambridge University Press. 153

Lemaître, P. 2014. Sensitivity analysis in structural reliability. Ph.D. thesis, Université Bordeaux I. 146, 147, 149

Lemaître, P., Sergienko, E., Arnaud, A., Bousquet, N., Gamboa, F., and Iooss, B. 2015. Density modification-based reliability sensitivity analysis. *Journal of Statistical Computation and Simulation 85,* 6 (Apr.), 1200–1223. 15, 70, 146, 147, 149, 155, 165, 217

Loeppky, J. L., Sacks, J., and Welch, W. J. 2009. Choosing the Sample Size of a Computer Experiment: A Practical Guide. *Technometrics 51,* 4 (Nov.), 366–376. 23

Marcellán, F., Van Assche, W., Morel, J.-M., Takens, F., and Teissier, B., Eds. 2006. *Orthogonal Polynomials and Special Functions*. Lecture Notes in Mathematics, vol. 1883. Springer Berlin Heidelberg, Berlin, Heidelberg. 66

Marrel, A. and Chabridon, V. 2020. Statistical developments for target and conditional sensitivity analysis: application on safety studies for nuclear reactor. *Preprint, https://hal.archives-ouvertes.fr/hal-02541142*. 160

Mazgaj, P., Vacher, J.-L., and Carnevali, S. 2016. Comparison of CATHARE results with the experimental results of cold leg intermediate break LOCA obtained during ROSA-2/LSTF test 7. *EPJ Nuclear Sciences & Technologies 2*, 1. 22, 145, 162

McKerns, M., Owhadi, H., Scovel, C., Sullivan, T. J., and Ortiz, M. 2012. The Optimal Uncertainty Algorithm in the Mystic Framework. 45, 49, 65, 102

Meynaoui, A., Marrel, A., and Laurent, B. 2019. New statistical methodology for second level global sensitivity analysis. *arXiv:1902.07030 [math, stat]*. arXiv: 1902.07030. 15, 90, 146, 217

Mikusinski, P., Sherwood, H., and Taylor, M. D. 1992. Shuffles of Min. *Stochastica 13,* 1, 61–74. 129

Moreno, E. and Cano, J. 1991. Robust bayesian analysis with ϵ-contaminations partially known. *Journal of the Royal Statistical Society. Series B (Methodological) 53,* 1, 143–155. 140

Morio, J. and Balesdent, M. 2016. *Estimation of rare event probabilities in complex aerospace and other systems*. Woodhead Publishing. 145

Moulines, E. and Bach, F. R. 2011. Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In *Advances in Neural Information Processing Systems*. 451–459. 169, 174, 188

Mousseau, V. and Williams, B. J. 2017. Uncertainty quantification in a regulatory environment. In *Springer Handbook on Uncertainty Quantification*, R. Ghanem, D. Higdon, and H. Owhadi, Eds. Springer, 1613–1648. 145, 155

Möller, B. and Beer, M. 2004. *Fuzzy Randomness*. Springer Berlin Heidelberg, Berlin, Heidelberg. 16, 218

Neal, R. M. 1996. *Bayesian Learning for Neural Networks*. Lecture Notes in Statistics, vol. 118. Springer New York, New York, NY. 111, 112

Nelsen, R. B. 2006. *An Introduction to Copulas*, 2 ed. Springer Series in Statistics. Springer-Verlag, New York. 11, 127, 129, 213

Newey, W. K. and McFadden, D. 1994. Large sample estimation and hypothesis testing. In *Handbook of Econometrics*. Vol. 4. Elsevier, Chapter 36, 2111 – 2245. ISSN: 1573-4412. 150

Nielsen, F. 2013. Cramér-Rao Lower Bound and Information Geometry. In *Connected at Infinity II: A Selection of Mathematics by Indians*, R. Bhatia, C. S. Rajan, and A. I. Singh, Eds. Hindustan Book Agency, Gurgaon, 18–37. 146

Nikolaidis, E., Ghiocel, D. M., and Singhal, S. 2004. *Engineering Design Reliability Handbook*. CRC Press. 12, 214

Oakley, J. 2004. Estimating percentiles of uncertain computer code outputs. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 53,* 1, 83–93. 13, 25, 168, 215

Owen, A. B. 2014. Sobol' Indices and Shapley Value. *SIAM/ASA Journal on Uncertainty Quantification 2,* 1 (Jan.), 245–251. Publisher: Society for Industrial and Applied Mathematics. 13

Owhadi, H., Scovel, C., Sullivan, T. J., McKerns, M., and Ortiz, M. 2013. Optimal Uncertainty Quantification. *SIAM Review 55,* 2 (Jan.), 271–345. arXiv: 1009.0679. 15, 16, 17, 28, 33, 41, 42, 43, 102, 137, 146, 217, 218

Pasanisi, A., Keller, M., and Parent, E. 2012. Estimation of a quantity of interest in uncertainty analysis: Some help from Bayesian decision theory. *Reliability Engineering & System Safety 100*, 93–101. 19, 20

Paté-Cornell, M. E. 1996. Uncertainties in risk analysis: Six levels of treatment. *Reliability Engineering & System Safety 54,* 2 (Nov.), 95–111. 10, 212

Perrin, G. and Defaux, G. 2019. Efficient estimation of reliability-oriented sensitivity indices. *Journal of Scientific Computing 80,* 3. 146, 148

Pesenti, S. M., Millossovich, P., and Tsanakas, A. 2019. Cascade Sensitivity Measures. SSRN Scholarly Paper ID 3270839, Social Science Research Network, Rochester, NY. July. 15, 166, 217

Popescu, I. 2005. A Semidefinite Programming Approach to Optimal-Moment Bounds for Convex Classes of Distributions. *Mathematics of Operations Research 30,* 3 (Aug.), 632–657. Publisher: INFORMS. 118

Price, K., Storn, R. M., and Lampinen, J. A. 2005. *Differential Evolution: A Practical Approach to Global Optimization*. Natural Computing Series. Springer-Verlag, Berlin Heidelberg. 49, 202, 204, 205

Priestley, M. B. 1981. *Spectral Analysis and Time Series*. Academic Press. 141

Prieur, C. and Tarantola, S. 2015. Variance-Based Sensitivity Analysis: Theory and Estimation Algorithms. *Handbook of Uncertainty Quantification*, 1–23. 21, 93, 160, 161

Puccetti, G. and Wang, R. 2015. Extremal Dependence Concepts. *Statistical Science 30,* 4 (Nov.), 485–517. Publisher: Institute of Mathematical Statistics. 129, 131

Qiu, Z., Yang, D., and Elishakoff, I. 2008. Probabilistic interval reliability of structural systems. *International Journal of Solids and Structures 45,* 10 (May), 2850–2860. 10, 212

Rao, C. R. 1945. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc. 37*. 150

Rasmussen, C. E. and Williams, C. K. I. 2005. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press. 8, 24, 141, 211

Robbins, H. and Monro, S. 1951. A stochastic approximation method. *The Annals of Mathematical Statistics*, 400–407. 169, 173

Robert, C. P. and Casella, G. 2005. *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag, Berlin, Heidelberg. 103

Rockafellar, R. T. and Royset, J. O. 2014. Random Variables, Monotone Relations, and Convex Analysis. *Math. Program. 148,* 1-2 (Dec.), 297–331. 86, 87, 89

Rockafellar, R. T. and Uryasev, S. 2002. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance 26,* 7 (July), 1443–1471. 16, 86, 87, 218

Rohde, C. A. 2014. *Introductory Statistical Inference with the Likelihood Function*. Springer International Publishing. 12, 214

Royden, H. 1968. *Real analysis*. Macmillan. 59

Ruggeri, F., Ríos Insua, D., and Martín, J. 2005. Robust Bayesian Analysis. In *Handbook of Statistics*, D. K. Dey and C. R. Rao, Eds. Bayesian Thinking, vol. 25. Elsevier, 623–667. 15, 217

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. 1986. Learning representations by back-propagating errors. *Nature 323,* 6088 (Oct.), 533–536. Number: 6088 Publisher: Nature Publishing Group. 111

Ruppert, D. 1991. *Handbook of sequential analysis.* CRC Press. 169

Sacks, J. 1958. Asymptotic distribution of stochastic approximation procedures. *The Annals of Mathematical Statistics*, 373–405. 169

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. 1989. Design and Analysis of Computer Experiments. *Statistical Science 4,* 4 (Nov.), 409–423. Publisher: Institute of Mathematical Statistics. 8, 168, 211

Saltelli, A. and Tarantola, S. 2002. On the Relative Importance of Input Factors in Mathematical Models: Safety Assessment for Nuclear Waste Disposal. *Journal of the American Statistical Association 97,* 459, 702–709. 21, 160

Sanchez-Saez, F., Sánchez, A. I., Villanueva, J. F., Carlos, S., and Martorell, S. 2018. Uncertainty analysis of a large break loss of coolant accident in a pressurized water reactor using non-parametric methods. *Reliability Engineering & System Safety 174*, 19–28. 22, 145

Santner, T. J., Williams, B. J., and Notz, W. I. 2013. *The design and analysis of computer experiments.* Springer Science & Business Media. 168

Schreck, A., Fort, G., Moulines, E., and Vihola, M. 2016. Convergence of Markovian Stochastic Approximation with discontinuous dynamics. *SIAM J. Control Optim. 54,* 2, 866–893. 169

Schöbi, R. 2017. Surrogate models for uncertainty quantification in the context of imprecise probability modelling. Ph.D. thesis, ETH Zürich - Switzerland. 16, 218

Shafer, G. 1976. *A Mathematical Theory of Evidence.* Princeton University Press. Google-Books-ID: wug9DwAAQBAJ. 16, 218

Shapiro, A. 2001. On Duality Theory of Conic Linear Problems. In *Semi-Infinite Programming*, P. Pardalos, M. A. Goberna, and M. A. Lopez, Eds. Vol. 57. Springer US, Boston, MA, 135–165. Series Title: Nonconvex Optimization and Its Applications. 118

Shohat, J. A. and Tamarkin, J. D. 1950. *The Problem of Moments.* American Mathematical Society. 47

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis.* CRC Press. 12

Sivaganesan, S. and Berger, J. O. 1989. Ranges of Posterior Measures for Priors with Unimodal Contaminations. *The Annals of Statistics 17,* 2 (June), 868–889. 36, 43, 96

Skibinsky, M. 1967. The range of the $(n+1)$th moment for distributions on $[0, 1]$. *Journal of Applied Probability 4,* 3 (Nov.), 543–552. 50

Skibinsky, M. 1977. The Maximum Probability on an Interval When the Mean and Variance Are Known. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002) 39,* 2, 144–159. 50

Skibinsky, M. 1986. Principal representations and canonical moment sequences for distributions on an interval. *Journal of Mathematical Analysis and Applications 120,* 1 (Nov.), 95–118. 50

Smith, R. C. 2014. *Uncertainty quantification*. SIAM. 145

Sobol', I. M. 1993. Sensitivity estimates for non linear mathematical models. *Mathematical Modelling and Computational Experiments 1*, 407–414. 13, 90, 146

Sobol', I. M. 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation 55,* 1, 271 – 280. 160

Song, E., Nelson, B. L., and Staum, J. 2016. Shapley Effects for Global Sensitivity Analysis: Theory and Computation. *SIAM/ASA Journal on Uncertainty Quantification 4,* 1 (Jan.), 1060–1083. 13

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research 15*, 1929–1958. 114

Stenger, J., Gamboa, F., and Keller, M. 2019a. Optimization Of Quasi-convex Function Over Product Measure Sets. *arXiv:1907.07934 [math]*. arXiv: 1907.07934. 40, 94

Stenger, J., Gamboa, F., Keller, M., and Iooss, B. 2019b. Canonical Moments for Optimal Uncertainty Quantification on a Variety. In *Geometric Science of Information*. Springer, Cham, 571–578. 61

Stenger, J., Gamboa, F., Keller, M., and Iooss, B. 2020. Optimal Uncertainty Quantification of a risk measurement from a thermal-hydraulic code using Canonical Moments. *International Journal for Uncertainty Quantification 10,* 1, 35–53. 19, 146

Stillwell, J. 1997. *Numbers and Geometry (Undergraduate Texts in Mathematics)*. Springer. 151

Stirzaker, D. 2003. *Elementary Probability*. Cambridge University Press. 148

Stone, C. J. 1976. Nearest neighbour estimators of a nonlinear regression function. *Proc. Comp. Sci. Statis. 8th Annual Symposium on the Interface*, 413–418. 170

Stone, C. J. 1977. Consistent nonparametric regression. *The Annals of Statistics*, 595–620. 170

Sueur, R., Bousquet, N., Iooss, B., and Bect, J. 2016. Perturbed-Law based sensitivity indices for sensitivity analysis in structural reliability. In *Proceedings of the 8th International Conference on Sensitivity Analysis of Model Output (SAMO 2016)*. Le Tampon, Réunion Island, France. 146, 147, 155

Sueur, R., Iooss, B., and Delage, T. 2017. Sensitivity analysis using perturbed-law based indices for quantiles and application to an industrial case. In *Proceedings of the 10th International Conference on Mathematical Methods in Reliability (MMR 2017)*. Grenoble, France. 15, 146, 155, 217

Sullivan, T. J. 2015. *Introduction to Uncertainty Quantification*. Texts in Applied Mathematics. Springer International Publishing. 8, 10, 37, 210

Sullivan, T. J., McKerns, M., Meyer, D., Theil, F., Owhadi, H., and Ortiz, M. 2013. Optimal uncertainty quantification for legacy data observations of Lipschitz functions. *ESAIM: Mathematical Modelling and Numerical Analysis 47,* 6 (Nov.), 1657–1689. Number: 6 Publisher: EDP Sciences. 142

Van der Vaart, A. W. 2000. *Asymptotic Statistics*. Cambridge University Press. 166, 167

Vandenberghe, L., Boyd, S., and Comanor, K. 2007. Generalized Chebyshev Bounds via Semidefinite Programming. *SIAM Review 49,* 1 (Jan.), 52–64. 118

Villani, C. 2008. Optimal transport – Old and new. Vol. 338. 129, 131

Walley, P. 2000. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning 24,* 2 (May), 125–148. 16

Wallis, G. B. 2007. Uncertainties and probabilities in nuclear reactor regulation. *Nuclear Engineering and Design 237,* 15-17, 1586–1592. 82

Weizsäcker, H. v. and Winkler, G. 1979. Integral Representation in the Set of Solutions of a Generalized Moment Problem. *Mathematische Annalen 246*, 23–32. 36

Wilks, S. S. 1941. Determination of sample sizes for setting tolerance limits. *Ann. Math. Statist. 12,* 1 (03), 91–96. 103

Winkler, G. 1985. *Choquet Order and Simplices: With Applications in Probabilistic Models*. Lecture Notes in Mathematics. Springer-Verlag, Berlin Heidelberg. 35

Winkler, G. 1988. Extreme Points of Moment Sets. *Mathematics of Operations Research 13,* 4, 581–587. 33, 36, 42

Woodroofe, M. 1972. Normal approximation and large deviations for the Robbins-Monro process. *Probability Theory and Related Fields 21,* 4, 329–338. 169

Zwillinger, D., Ed. 2002. *CRC Standard Mathematical Tables and Formulae, 31st Edition*, 31 edition ed. Chapman and Hall/CRC, Boca Raton, Fla. 63

Štěpán, J. 1993. Simplicial Measures and Sets of Uniqueness in the Marginal Problem. *Statistics & Risk Modeling 11,* 3, 289–300. 128

Résumé

La quantification des incertitudes lors d'une étude de sûreté peut être réalisée en modélisant les paramètres d'entrée du système physique par des variables aléatoires. Afin de propager les incertitudes affectant les entrées, un modèle de simulation numérique reproduisant la physique du système est exécuté avec différentes combinaisons des paramètres d'entrée, générées suivant leur loi de probabilité jointe. Il est alors possible d'étudier la variabilité de la sortie du code, ou d'estimer certaines quantités d'intérêt spécifiques. Le code étant considéré comme une boîte noire déterministe, la quantité d'intérêt dépend uniquement du choix de la loi de probabilité des entrées. Toutefois, cette distribution de probabilité est elle-même incertaine. En général, elle est choisie grâce aux avis d'experts, qui sont subjectifs et parfois contradictoires, mais aussi grâce à des données expérimentales souvent en nombre insuffisant et entachées d'erreurs. Cette variabilité dans le choix de la distribution se propage jusqu'à la quantité d'intérêt. Cette thèse traite de la prise en compte de cette incertitude dite de deuxième niveau. L'approche proposée, connue sous le nom d'Optimal Uncertainty Quantification (OUQ) consiste à évaluer des bornes sur la quantité d'intérêt. De ce fait on ne considère plus une distribution fixée, mais un ensemble de mesures de probabilité sous contraintes de moments sur lequel la quantité d'intérêt est optimisée. Après avoir exposé des résultats théoriques visant à réduire l'optimisation de la quantité d'intérêt aux point extrémaux de l'espace de mesures de probabilité, nous présentons différentes quantités d'intérêt vérifiant les hypothèses du problème. Cette thèse illustre l'ensemble de la méthodologie sur plusieurs cas d'applications, l'un d'eux étant un cas réel étudiant l'évolution de la température de gaine du combustible nucléaire en cas de perte du réfrigérant

Mots-clés: Quantification d'incertitude - Analyse de robustesse - Probabilité non paramétrique - Optimisation sur des espaces de mesures.

Abstract

Uncertainty quantification in a safety analysis study can be conducted by considering the uncertain inputs of a physical system as a vector of random variables. The most widespread approach consists in running a computer model reproducing the physical phenomenon with different combinations of inputs in accordance with their probability distribution. Then, one can study the related uncertainty on the output or estimate a specific quantity of interest (QoI). Because the computer model is assumed to be a deterministic black-box function, the QoI only depends on the choice of the input probability measure. It is formally represented as a scalar function defined on a measure space. We propose to gain robustness on the quantification of this QoI. Indeed, the probability distributions characterizing the uncertain input may themselves be uncertain. For instance, contradictory expert opinion may make it difficult to select a single probability distribution, and the lack of information in the input variables affects inevitably the choice of the distribution. As the uncertainty on the input distributions propagates to the QoI, an important consequence is that different choices of input distributions will lead to different values of the QoI. The purpose of this thesis is to account for this second level uncertainty. We propose to evaluate the maximum of the QoI over a space of probability measures, in an approach known as optimal uncertainty quantification (OUQ). Therefore, we do not specify a single precise input distribution, but rather a set of admissible probability measures defined through moment constraints. The QoI is then optimized over this measure space. After exposing theoretical results showing that the optimization domain of the QoI can be reduced to the extreme points of the measure space, we present several interesting quantities of interest satisfying the assumption of the problem. This thesis illustrates the methodology in several application cases, one of them being a real nuclear engineering case that study the evolution of the peak cladding temperature of fuel rods in case of an intermediate break loss of coolant accident.

Key-words: Uncertainty Quantification - Robustness analysis - Non parametric probability - Optimization over measure space.