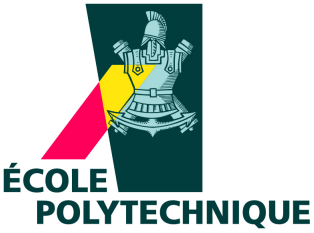
Polynomial approximation by means of the random discrete L2 projection and application to inverse problems for PDEs with stochastic data

Giovanni Migliorati

POLITECNICO DI MILANO - Department of Mathematics "F. Brioschi"
PhD School in Mathematical Models and Methods in Engineering - XXV° cycle

ÉCOLE POLYTECHNIQUE - Centre de Mathématiques Appliquées



Polynomial approximation by means of the random discrete L^2 projection and application to inverse problems for PDEs with stochastic data

presented by

Giovanni MIGLIORATI

graduated in Mathematical Engineering at Politecnico di Milano

Advisor at POLITECNICO DI MILANO: Prof. Fabio NOBILE
Advisor at ÉCOLE POLYTECHNIQUE: Prof. Houssem HADDAR

Tutor: Prof. Fabio NOBILE
Coordinator of the PhD Program: Prof. Roberto LUCCHETTI

April 3rd, 2013

Abstract

The main topic of this thesis concerns the polynomial approximation of aleatory functions by means of the random discrete L^2 projection, and its application to inverse problems for Partial Differential Equations (PDEs) with stochastic data. The motivations come from the parametric approximation of the solution to partial differential models and its application to Uncertainty Quantification in engineering. The thesis is arranged in two parts, with an introductory chapter which contains an overview of modern techniques for polynomial approximation of functions depending on random variables.

In the former part, from Chapter 1 to Chapter 4, the focus is on the theoretical analysis of the random discrete L^2 projection applied to solve the so-called *forward problem*, *e.g.* to approximate the moments of an aleatory function given its observations, or to compute the solution to a computational model with stochastic coefficients given initial and boundary data. In Chapter 1, the discrete L^2 projection on polynomial spaces with random evaluations is presented, as a tool to accurately approximate a multivariate function depending on a random variable distributed according to a given probability density. The stability and optimality of the approximation error evaluated in the L^2 weighted norm are addressed, under the assumption that the density is bounded away from zero. In this analysis, the main result achieved is a univariate probabilistic optimal convergence estimate with the uniform distribution, provided the number of evaluations scales as the square of the dimension of the polynomial space. Several numerical examples confirm the theoretical results, with aleatory functions defined on parameter spaces featuring low to moderately high dimension. The role of smoothness of the target function has been investigated as well.

In Chapter 2 the proof of the stability and optimality in expectation of the random discrete L^2 projection is extended to any monotone set of multi-indices identifying the polynomial space, and to any dimension of the parameter space. For a specific class of PDE models, that includes the elliptic model and the linear elasticity model, an exponential convergence estimate w.r.t. the number of sampling points has been proved, with an a priori optimal choice of the polynomial space. This estimate clarifies the dependence of the convergence rate on the dimension of the parameter space, and establishes a relation between the convergence rate of the random discrete projection and the convergence rate of the classical Stochastic Galerkin method.

Afterwards, in Chapter 3 the analysis of the random L^2 projection is extended to more general densities, focusing on how the choice of the density affects the optimal convergence rate. It is shown that the assumption on the density being bounded away from zero is strictly required in the proof of the optimal convergence theorem. The beta family, which includes the uniform and Chebyshev densities, is investigated. Some tests with the Gaussian and gamma densities have been performed.

The methodology based on the random L^2 projection is then applied in Chapter 4 to approximate Quantities of Interest related to the solution to PDE models with stochastic data. Several examples are presented, featuring the Darcy model with values of the coefficient and geometry of the inclusions governed by random variables. Hence, some examples with the linear elasticity model and with Navier-Stokes equations, both with stochastic data, are addressed.

In the latter part of the thesis, composed of Chapter 5 and Chapter 6, the methodology previously developed for the *forward problem* is applied to *inverse problems* for PDEs with stochastic coefficients. In Chapter 5 the problem of Electrical Impedance Tomography (EIT) is

introduced. The goal is to detect the presence and location of inclusions in the domain, when observing the solution to the associated PDE model only on the boundary of the domain. A numerical scheme to solve the dipole-like Neumann problem with inhomogeneous coefficient is proposed. Next, this scheme is employed in the Factorization Method in the framework of EIT, first in the case of inhomogeneous background, and then in the case of piecewise constant background, with values in each region affected by uncertainty. Several variants of the Factorization Method are proposed, and numerical results showing their effectiveness are presented.

Lastly, in Chapter 6 the variants of the Factorization Method proposed in the previous chapter are accelerated by means of the random discrete L^2 projection, exploiting the techniques that have been presented in the first part of the thesis.

Keywords: approximation theory; error analysis; Uncertainty Quantification; polynomial approximation; random discrete L^2 projection; nonparametric regression; PDE stochastic data; Inverse Problems; Electrical Impedance Tomography; Factorization Method; inhomogeneous background; uncertain background.

Sommario

Questa tesi verte sull'approssimazione polinomiale di funzioni aleatorie tramite proiezione L^2 discreta con valutazioni casuali, con applicazione a problemi inversi per Equazioni alle Derivate Parziali (EDP) con dati stocastici. Possibili utilizzi di questa tecnica di approssimazione consistono nella soluzione parametrica di modelli di EDP nell'ambito della Quantificazione dell'Incertezza. La tesi è strutturata in due parti, con un capitolo introduttivo che contiene una panoramica sulle moderne tecniche di approssimazione polinomiale di funzioni aleatorie.

Nella prima parte, dal Capitolo 1 al Capitolo 4, viene presentata l'analisi teorica della proiezione L^2 discreta e la sua applicazione per risolvere *problemi diretti*, *e.g.* per approssimare i momenti di una funzione aleatoria a partire da sue osservazioni puntuali, o per calcolare la soluzione numerica di modelli computazionali con coefficienti stocastici. Nel Capitolo 1 introduciamo la proiezione L^2 discreta su spazi polinomiali come strumento per approssimare accuratamente una funzione di una variabile aleatoria multidimensionale distribuita con densità nota. Analizziamo la stabilità e l'ottimalità dell'errore di approssimazione valutato nella norma L^2 pesata rispetto alla densità, nell'ipotesi che la densità sia strettamente positiva. Il risultato principale ottenuto è una stima di convergenza ottimale in probabilità nel caso di densità uniforme monodimensionale, a condizione che il numero di valutazioni sia proporzionale al quadrato della dimensione dello spazio polinomiale. I risultati teorici sono confermati da numerosi esempi numerici, con funzioni aleatorie di variabili univariate e multivariate con dimensione del supporto moderatamente alta. Il ruolo svolto dalla regolarità della funzione è stato considerato.

Nel Capitolo 2 estendiamo la dimostrazione della stabilità e ottimalità in valore atteso della proiezione L^2 discreta al caso di insiemi di multi-indici monotoni, e per ogni valore della dimensione dello spazio dei parametri. Per una classe specifica di modelli di EDP, che include il modello ellittico ed il modello di elasticità lineare, dimostriamo una stima di convergenza esponenziale rispetto al numero di valutazioni della funzione. Questa stima chiarisce la dipendenza del tasso di convergenza dalla dimensione dello spazio dei parametri, e stabilisce una relazione tra il tasso di convergenza della proiezione discreta ed il classico metodo di Galerkin Stocastico.

In seguito, nel Capitolo 3 estendiamo l'analisi della proiezione L^2 discreta a densità di probabilità più generali, ed esaminiamo la dipendenza del tasso di convergenza ottimale dalla particolare densità scelta. Consideriamo prima la famiglia di densità beta, che contiene in particolare la densità uniforme e le densità di Chebyshev, e successivamente le densità Gaussiana e gamma. Inoltre, mostriamo che l'ipotesi che la densità di probabilità sia strettamente positiva è strettamente necessaria nella dimostrazione del teorema di convergenza ottimale.

La tecnica di approssimazione basata sulla proiezione L^2 discreta viene poi applicata nel Capitolo 4 per approssimare Quantità di Interesse dipendenti dalla soluzione di modelli di EDP con dati stocastici. Presentiamo diversi esempi relativi al modello di Darcy con coefficiente di diffusione e geometria delle inclusioni descritti per mezzo di variabili aleatorie, ed alcuni esempi con il modello di elasticità lineare e le equazioni di Navier-Stokes, entrambi con dati stocastici.

Nella seconda parte della tesi, formata dai Capitoli 5 e 6, applichiamo a *problemi inversi* per EDP con coefficienti stocastici le tecniche di approssimazione sviluppate nella prima parte. Nel Capitolo 5 introduciamo il problema della Tomografia ad Impedenza Elettrica, in cui l'obiettivo è determinare l'eventuale presenza e posizione di inclusioni nel dominio, osservando la soluzione dell'EDP associata solo sul bordo del dominio. Costruiamo uno schema numerico per risolvere il problema di Neumann singolare con coefficiente di diffusione spazialmente

inomogeneo. Successivamente applichiamo questo schema al Metodo di Fattorizzazione nell'ambito della Tomografia ad Impedenza Elettrica, prima nel caso in cui il coefficiente di diffusione sia deterministico ma spazialmente inomogeneo, e poi nel caso il coefficiente sia costante a tratti con valori in ogni regione affetti da incertezza. Proponiamo diverse varianti del Metodo di Fattorizzazione, e ne mostriamo la loro efficacia tramite alcuni esempi numerici.

Infine, nel Capitolo 6 le varianti del Metodo di Fattorizzazione proposte nel capitolo precedente vengono accelerate per mezzo della proiezione L^2 discreta, utilizzando le tecniche di approssimazione sviluppate nella prima parte della tesi.

Parole chiave: teoria dell'approssimazione; analisi dell'errore; Quantificazione dell'Incertezza; approssimazione polinomiale; proiezione L^2 discreta a valutazioni casuali; regressione non parametrica; EDP con dati stocastici; Problemi Inversi; Tomografia ad Impedenza Elettrica; Metodo di Fattorizzazione; coefficiente di diffusione inomogeneo; coefficiente di diffusione a valori incerti.

Résumé

Le sujet principal de cette thèse porte sur l'approximation polynômiale des fonctions aléatoires au moyen de la projection L^2 aléatoire discrète, et son application aux problèmes inverses pour les équations aux dérivées partielles avec des données aléatoires. Les motivations proviennent de l'approximation paramétrique de la solution de modèles aux dérivées partielles et son application à la quantification des incertitudes en ingénierie. La thèse se compose de deux parties, avec un chapitre d'introduction qui résume les techniques modernes de l'approximation polynômiale des fonctions de variables aléatoires. Dans la première partie, du chapitre 1 au chapitre 4, l'analyse théorique de la projection L^2 aléatoire discrète est développée pour résoudre le problème direct, par exemple, pour rapprocher les moments d'une fonction aléatoire à partir de ses observations, ou pour calculer la solution à un modèle numérique avec des coefficients stochastiques.

Dans le chapitre 1, la projection L^2 discrète sur les espaces de polynômes avec des évaluations aléatoires est présenté comme un outil d'approche d'une fonction d'une variable aléatoire, qui est distribuée selon une densité de probabilité donnée. La stabilité et l'optimalité de l'erreur d'approximation évaluée dans la norme L^2 pondérée sont traités, en supposant que la densité est borné et n'approche pas zéro. Dans cette analyse, le principal résultat obtenu est une estimation en probabilité de convergence optimale avec la distribution uniforme, à condition que le nombre d'évaluations soit proportionnel au carré de la dimension de l'espace polynômial. Plusieurs exemples numériques confirment les résultats théoriques, avec des fonctions aléatoires définies sur des ensembles de paramètres présentant faible ou modérément élevé dimension. Le rôle de la régularité de la fonction a été étudiée aussi bien.

Dans le chapitre 2, la démonstration de la stabilité et l'optimalité de la projection L^2 aléatoire discrète est étendue à un ensemble monotone de multi-indices qui identifie l'espace polynômial, et à toute dimension de l'ensemble des paramètres. Pour une classe spécifique de modèles aux dérivées partielles, qui comprend le modèle elliptique et le modèle d'élasticité linéaire, une estimation de convergence exponentielle en fonction du nombre de points d'échantillonnage a été prouvé, avec une choix optimal a priori de l'espace polynômial. Cette estimation dépend de la dimension de l'ensemble des paramètres et établit une relation entre le taux de convergence de la projection aléatoire discrète et la vitesse de convergence de la méthode de Galerkin Stochastique.

Dans le chapitre 3 l'analyse de la projection aléatoire L^2 est étendu à d'autres densités généraux, en se concentrant sur la façon dont le choix de la densité influe sur la vitesse de convergence optimale. Il est montré que l'hypothèse sur la densité, étant borné et n'approchant pas zéro, est strictement nécessaire à la démonstration du théorème de convergence optimale. La famille bêta, qui comprend la densité uniforme et de Chebyshev, est étudiée. Quelques essais avec les densités Gaussiennes et gamma ont été effectuées.

La méthodologie basée sur la projection aléatoire L^2 est ensuite appliquée au chapitre 4, pour évaluer des Quantités d'Intérêt qui dépendent de la solution du équations aux dérivées partielles avec données stochastiques. Plusieurs exemples sont présentés, avec le modèle de Darcy aux valeurs du coefficient et de la géométrie des inclusions réglés par des variables aléatoires. De plus, quelques exemples avec le modèle d'élasticité linéaire et équations de Navier-Stokes, ainsi que avec des données aléatoires, sont pris en compte.

Dans la dernière partie de la thèse, composé des chapitres 5 et 6, la méthodologie développée précédemment pour le problème direct est appliqué aux problèmes inverses pour les équations aux dérivées partielles à coefficients stochastiques. Dans le chapitre 5, le problème de la tomographie

par impédance électrique (TIE) dans des milieux inhomogène est introduit. L'objectif est de détecter la présence, la position et la géométrie des inclusions dans le domaine, lorsque l'on observe la solution au modèle EDP associé uniquement sur le bord du domaine. Un schéma numérique est proposé pour résoudre le problème singulier de Neumann avec un coefficient inhomogène. Ce système est utilisé dans la méthode de factorisation dans le cadre de l'TIE, d'abord dans le cas de coefficient inhomogène, puis dans le cas de coefficient constante par morceaux, à valeurs dans chaque région affectée par l'incertitude. Plusieurs variantes de la méthode de factorisation sont proposées avec des résultats numériques qui montrent leur efficacité.

Enfin, dans le chapitre 6 les variantes de la méthode de factorisation proposées dans le chapitre précédent sont accélérés à l'aide de la projection L^2 aléatoire discrète, en utilisant les techniques qui ont été présentés dans la première partie de la thèse.

Mots-clés : théorie d'approximation ; analyse d'erreur ; quantification de l'incertitude ; approximation polynômiale ; projection L^2 discrète aléatoire ; régression non paramétrique ; équations aux dérivées partielles stochastiques ; problèmes inverses ; tomographie d'impédance électrique ; la méthode de factorisation ; milieux inhomogène ; milieux incertain.

Acknowledgments

At the end of this work I would like to thank some people for their valuable contributions and feedbacks. I wish to thank Fabio Nobile for the overall supervision, and Houssem Haddar for introducing me to qualitative methods. Then I wish to thank Raúl Tempone for the long-distance support and Erik von Schwerin for his contribution to the first and fourth chapter. I am also grateful to Prof. Albert Cohen for walking with us towards discrete least squares and to Abdellah Chkifa for his contribution to the second chapter, in particular to the proof of Lemma 2.2.

Thanks to Prof. Olivier Le Maître, Prof. Michele Piana and Prof. Bruno Sudret for the time dedicated to review the thesis, as well as to Prof. Luca Pavarino and Prof. Karim Ramdani for being members of the committee.

Contents

Introduction

Approximation theory plays a relevant role in the modern development of science, where functional relations are employed to describe dependencies and express connections among entities. Whether the framework is completely abstract, *e.g.* in pure mathematics, or trying to describe an underlying real context *e.g.* in physics and medicine, the matter of how to handle any quantifiable relation from a mathematical standpoint is of primary importance. Applied mathematics and engineering applications completely rely on accurate and efficient methodologies to approximate functional relations, leading to challenging issues as long as more complicated mathematical objects get involved. In recent years, the topics of uncertainty quantification [GS03, CDS10, CDS11, DNP$^+$04, LGH11, Eld11, SG04, MNvST11, Zha02, Eld09, EWC08, SG11, CCDS13, CCM$^+$13, MNvST13, LMRN$^+$02, LMKNG01, NT09, BE09, LMNGK04, LMKNG04, XK02b, DO11, Sud07, GZ07a], stochastic optimization and optimization under uncertainty [Che02, Cat04, Eld11, RW98, RW91] gained such a relevance to be considered as a new branch of applied mathematics itself, rather than minor subsections specifically devoted to address marginal effects. Several other major areas are directly related, risk or failure analysis, insurance and finance modeling to name just a few. The key point concerns the analysis of the presence of uncertainty in any of the steps that lead to the development of mathematical models trying to describe a given phenomenon. Uncertainty arises from the lack of knowledge (epistemic uncertainty) or when dealing with aleatory processes (aleatory uncertainty). It may concern the initial data or initial state of the system, but also be originated from controllable or uncontrollable external factors that can be treated as random variables. An approach that has been advocated in recent years, the Polynomial Chaos Expansion (PCE), consists in expanding over a spectral basis the random output parameters of the computational model, (see *e.g.* [XK02a, LK04, DNP$^+$04, DGS05, LM06, GG07, CLMM09, Sud07, BS08, BS11, EMSU12, BE09, DGRH07, SG09, Soi10]). This thesis presents an approximation methodology that relies on PCE, focusing on the approximation of aleatory functions and on the solution of computational models formulated by means of Partial Differential Equations (PDEs). The aforementioned methodology is based on the random discrete projection, and will be outlined in Section 0.1.

In the mathematical models, the identification of the forward and inverse directions to solve the problem is stated by the concept of well-posedness in the sense of Hadamard, and the forward problem is the one which formulation is well-posed [TA77]. According to this definition, the problems of polynomial approximation of the solution to PDE models with stochastic data, that are presented in the first part of the thesis, are forward problems. The identification problem for the coefficient of the PDE model presented in the second part of the thesis is an inverse problem.

Inverse problems are commonly difficult to solve, because they are ill-posed [TA77]. The field is very wide [MT02, Tar05, AK07, BBG$^+$11, MS12, Kir96, Vog02, KS05], and extends beyond the area of PDEs. In the present thesis we focus only on the inverse problem of Electrical Impedance Tomography (EIT) [MS12, BDG08, CIN99, HHP07], and more specifically on the application of the Factorization Method to the Continuous Model in the framework of EIT. The Factorization Method is an imaging technique belonging to the class of *Qualitative Methods*, see *e.g.* [CC06] for an introduction. The Continuous Model refers to the modeling of the boundary condition in the PDE associated with the EIT problem. It is named continuous due to the unrealistic assumption that any continuous pattern of current can be injected, not taking into account the finite number of electrodes that usually only partially cover the boundary of the

computational domain. See [MS12] for an overview of several models currently employed in EIT, including the Continuous Model and other more realistic models such as the Complete Electrode Model. In the Continuous Model of the inverse problem of EIT, the solution to the elliptic PDE model is observed on the whole boundary of the domain, and is used to recover information on the diffusion coefficient of the model inside the domain. The EIT problem is simple to formulate and delicate at the same time to handle from a mathematical and computational standpoint. It is a severely ill-posed inverse problem, demanding for appropriate regularization techniques, see *e.g.* [EHN96]. Despite of its low spatial resolution, EIT features a high contrast resolution which makes it interesting in many applications, *e.g.* geophysics, nondestructive testing, medicine, see *e.g.* the monography [MS12] and the references therein.

In reality, uncertainty affects all the inverse problems where physical measurements are collected: *e.g.* weather forecast, medical imaging, seismic inversion, image restoration, remote sensing. Dealing with uncertainty is itself a large dimensional problem, since in the modeling of real phenomena it is often the case that a huge number of parameters has to be taken into account. For these reasons several branches of applied mathematics, ranging from inverse problems, uncertainty quantification, model reduction [BTWG08, DGRH07, SG09, GR11], reduced basis [BMNP04, RHP08, GMNP07], are nowadays merging their efforts to come up with new efficient and more powerful methodologies to face one of the main challenges for the forthcoming years, *i.e.* the treatment of high dimensional objects, that arise when handling tensors, experimental measurements, parameters of large-scale models or clinical databases.

0.1 Polynomial approximation techniques for functions depending on random variables

The main topic of the thesis concerns the approximation in the L^2 probability sense of aleatory scalar functions $\phi : \Gamma \to \mathbb{R}$, or Banach-valued functions $\phi : \Gamma \to V$ where V is a Banach space, defined over a subset Γ of the d-dimensional euclidean space $\mathbb{R}^d$. The function ϕ depends on a random variable $\mathbf{Y}$ which is distributed according to a given probability density $\rho : \Gamma \to \mathbb{R}^+$. The least squares method is one of the most known approximation techniques, with applications ranging from functional identification to interpolation to data fitting.

Let $L^2_\rho(\Gamma)$ be the space of square integrable functions endowed with the norm $\|\cdot\|^2_{L^2_\rho(\Gamma)} = \langle \cdot, \cdot \rangle_{L^2_\rho(\Gamma)}$, induced by the inner product

$$\langle u, v \rangle_{L^2_\rho(\Gamma)} = \int_\Gamma u(\mathbf{Y}) v(\mathbf{Y}) \rho(\mathbf{Y}) \, d\mathbf{Y}.$$

Denote by $\mathcal{F}$ any finite dimensional subspace of $L^2_\rho(\Gamma)$. The continuous L^2_ρ projection $\Pi_\mathcal{F}$ of a scalar target function ϕ on $\mathcal{F}$ is defined as

$$\Pi_\mathcal{F} \phi = \operatorname*{argmin}_{v \in \mathcal{F}} \|\phi - v\|_{L^2_\rho(\Gamma)}. \tag{0.1.1}$$

This continuous projection (0.1.1) cannot be computed exactly, and in practice the minimization in the L^2_ρ norm is replaced by the minimization in some discrete norm $\|\cdot\|_M$ that uses only M pointwise evaluations of the target function. One possible choice, which will be thoroughly investigated in this thesis, consists in taking M independent realizations $\mathbf{y}_1, \dots, \mathbf{y}_M \overset{\text{i.i.d.}}{\sim} \rho$ of the random variable $\mathbf{Y}$. The random discrete inner product on Γ is defined as

$$\langle u, v \rangle_M = \frac{1}{M} \sum_{m=1}^{M} u(\mathbf{y}_m) v(\mathbf{y}_m), \tag{0.1.2}$$

and induces on Γ the corresponding discrete norm $\|\cdot\|^2_M := \langle \cdot, \cdot \rangle_M$. The random discrete L^2_ρ projection $\Pi^M_\mathcal{F}$ of a scalar target function ϕ on $\mathcal{F}$ is defined as

$$\Pi^M_\mathcal{F} \phi = \operatorname*{argmin}_{v \in \mathcal{F}} \|\phi(\mathbf{y}_m) - v(\mathbf{y}_m)\|_M. \tag{0.1.3}$$

The continuous (0.1.1) or discrete (0.1.3) L^2_ρ projection, when applied to a function of a random variable, falls in the field of nonparametric regression in the statistical literature.

Nowadays, the field of nonparametric regression is well established (see *e.g.* [GKKW02]), and a huge literature is available in the case of noisy measurements. In the noisy framework with (finite) variance σ^2, the observed function ϕ is contaminated by noise. A general convergence estimate is stated in [GKKW02, Theorem 11.3]: the error evaluated in the L^2_ρ norm, committed by the random discrete L^2 projection (0.1.3) when approximating the true function $\overline{\phi}$ (not contaminated by noise), $\overline{\phi} = \mathbb{E}[\phi|\mathbf{y}]$, reads

$$\mathbb{E}\Big[\|\overline{\phi} - \Pi^M_{\mathcal{F}}\overline{\phi}\|^2_{L^2_\rho}\Big] \leq \frac{C_1(\sigma^2)}{M} + C_2 \inf_{v\in\mathcal{F}} \|\overline{\phi} - v\|_{L^2_\rho}, \tag{0.1.4}$$

with M being the number of evaluations of the target function. The estimate (0.1.4) holds for M large enough compared to the dimension of $\mathcal{F}$. The best approximation error on the right side of (0.1.4) quantifies the best error that can be achieved when approximating the target function in the subspace $\mathcal{F}$. The convergence rate is slowed down by the presence of the term M^{-1}. The constant C_2 is independent of M, while C_1 can contain logarithmic factors in M, and is such that it does not vanish in the noiseless case, *i.e.*

$$C_1(\sigma^2) \not\to 0, \quad \text{if } \sigma \to 0. \tag{0.1.5}$$

An additional truncation operator is needed in (0.1.4) to make the expectation well defined. See [GKKW02] for further details.

The noiseless case is much less complete, since the estimates like (0.1.4) are useful only in the noisy framework due to (0.1.5). In the noiseless case one should not expect the convergence rate to contain the Monte Carlo-type term that prevents to achieve the optimal L^2_ρ-projection convergence rate, and we will show in this thesis that this is actually the case provided M is large enough compared to the dimension of $\mathcal{F}$. Apart from the approximation problem itself, the interest towards this framework relies in the application of this approximation technique to the field of parametric differential equations, and more generally to stochastic differential equations where stochastic terms could be present in the initial or boundary data, in the coefficients of the model or in the geometry of the domain. The presence of these sources of randomness takes into account the uncertainty that often is present in reality. Due to the randomness in the data, the solution to the differential model is random itself. However, the random input parameters can be sampled, and the solution to the differential model can be computed exactly without noise (up to eventual deterministic discretization errors) in the sampled realization of the input parameters.

The idea based on expanding the random output over a suitable spectral basis $\{\psi_j\}_j$ of orthogonal polynomials which are orthonormal w.r.t. the underlying density ρ (PCE) goes back to the work of [Wie38, CM47] for Gaussian densities and Hermite polynomials, and has lately been generalized to other discrete and continuous probability densities, see *e.g.* [XK02b, SG04, XH05]. This idea is motivated by the fact that the functional dependence between the solution of the differential problem and the random data is often very smooth. In the sequel $\mathbb{P}_\Lambda$ denotes the polynomial space generated by suitable combination/tensorization of the family of orthogonal polynomials $\{\psi_j\}_j$. The symbol Λ denotes a set of multi-indices and characterizes the elements of the space $\mathbb{P}_\Lambda$.

We can summarize the main following techniques, that are currently employed to compute truncated PCE of random functions in the space $\mathbb{P}_\Lambda$:

Standard Monte Carlo [Fis96, RC04]. This is the most naïve approach, to numerically compute the correlations $\langle\phi,\psi_j\rangle_{L^2_\rho(\Gamma)}$ according to the L^2_ρ inner product weighted w.r.t. the density ρ. Given M samples $\{\mathbf{y}_m\}_{m=1}^M$ of the random parameter, being independent and identically distributed according to the density ρ, the correlations are approximated by

$$\langle\phi,\psi_j\rangle_{L^2_\rho} \approx \langle\phi,\psi_j\rangle_M. \tag{0.1.6}$$

The error $\varepsilon(M) = |\langle \phi, \psi_j \rangle_{L^2_\rho} - \langle \phi, \psi_j \rangle_M|$ scales w.r.t. the sample size as $M^{-1/2}$. Nonetheless, the slow convergence rate of the standard Monte Carlo method exhibits the remarkable property of being independent of the dimension of the parameter set.

Stochastic Collocation [BNT07, NTW08a, NT09, BE09, GKW+08, GZ07b, XH05]. In general, the aim is to approximate or interpolate a function ϕ that depends on a finite set of independent random parameters which are associated with a probability density ρ. In the Stochastic Collocation approach, the evaluations of the target function are completely uncoupled among them, and can be performed in parallel. The target function can be replaced by any computational model, being evaluated in the "collocation points" of the parameter set. Then these evaluations are employed to build the lagrangian polynomial interpolation, that can later be used to compute statistics of the solution or to predict the value of the solution to the model in any point of the parameter set. To achieve this, the stochastic interpolant $\Pi_{\mathrm{SG}}\phi$ is built as

$$\Big(\Pi_{\mathrm{SG}}\phi\Big)(\mathbf{y}) = \sum_j \phi(\mathbf{y}_j)\varphi_j(\mathbf{y}), \quad \mathbf{y} \in \Gamma, \tag{0.1.7}$$

using the evaluations of the function ϕ in the points $\mathbf{y}_j$ belonging to a given design (or grid). Many works in the literature discuss the choice of the grid where the function is evaluated. The so-called Sparse Grids approach, based on the original work [Smo63] and then developed in [BG04, BNR00], represents the state of the art in this context, for moderately high dimension of the parameter set.

Stochastic Galerkin [Gha99, GS03, LMRN+02, LMKNG01, NT09, EEU07, BS09]. The well-known Stochastic Galerkin method has been proposed in [GS03] (see also the more recent monography [LMK10]) to solve problems described by means of differential equations with any source of uncertainty embedded. Theoretical results on convergence rates can be found *e.g.* in [BTZ04, CDS10, CDS11].

Consider a differential model in parametric form $\mathcal{M}(u, \mathbf{y}) = 0$, provided with a suitable variational formulation on a Hilbert space V for each value of the parameter $\mathbf{y} \in \Gamma$. For brevity, assume a smooth dependence of the parameter-to-solution map $\mathbf{y} \mapsto u(\mathbf{y})$. When the parameter $\mathbf{y}$ is distributed according to a given density ρ, then the solution u of the model $\mathcal{M}$ belongs to the space $L^2_\rho \otimes V$, which is the tensor product of the L^2_ρ probability space with the physical space V. The projection Π^{G}_Λ associated with the Stochastic Galerkin method is defined as

$$\text{find } \Pi^{\mathrm{G}}_\Lambda u \in L^2_\rho \otimes V \quad : \quad \mathbb{E}\left[\langle M(u, \mathbf{y}), v\rangle_V\right] = 0, \quad \forall\, v \in \mathbb{P}_\Lambda \otimes V. \tag{0.1.8}$$

This formulation (0.1.8) brings to a system of coupled PDEs. In practice it consists of a Galerkin projection on the tensor subspace $\mathbb{P}_\Lambda \otimes V$ of the space $L^2_\rho \otimes V$, which is composed by the tensorization of the usual functional space with the probability space. At this stage, V can already be considered as the discretized physical space (*e.g.* discretized by the standard Finite Element Method), rather than the original physical functional space.

Due to its coupled structure, the Stochastic Galerkin method requires some additional efforts to be performed in parallel, and the use of a black-box solver is possible only with remarkable difficulties in the implementation.

On the other hand, the direct projection of the solution u on the tensor space $\mathbb{P}_\Lambda \otimes V$ named Π_Λ (as in (2.4.7)-left in Chapter 2) and defined by

$$\Pi_\Lambda u = \operatorname*{argmin}_{v \in \mathbb{P}_\Lambda \otimes V} \|u - v\|_{L^2_\rho \otimes V}, \tag{0.1.9}$$

is of limited interest, because in practice it is not directly computable.

Random discrete L^2 projection. The methodology based on the random discrete L^2 projection [CDL13, MNvST11, MNvST13, BS08, BS11, LARH11, HWB10, Eld09] falls somehow between the aforementioned two methodologies Stochastic Galerkin and Stochastic Collocation. In the

case of scalar functions ϕ, the random discrete L^2 projection on a general finite dimensional subspace $\mathcal{F} \subset L^2_\rho(\Gamma)$ has been introduced in (0.1.3). When the dependence of the target function on the random parameters is smooth, then it is sound to accomplish the projection on polynomial spaces, and therefore to choose $\mathcal{F} = \mathbb{P}_\Lambda$ in (0.1.3). The random discrete projection is based on a random sampling of the parameter set Γ, and is presented as a nonparametric regression problem on a random design. It differs from other common techniques such as quasi-Monte Carlo methods [Lem09, GKN$^+$11] or Stochastic Collocation on Sparse Grids [BG04, BNT10] which are based on a deterministic design. In (0.1.3) the evaluations of the target function ϕ are still uncoupled, but a projection on a suitable polynomial space is performed afterward, rather than a simple interpolation as in the Stochastic Collocation method. In Chapter 2, the random discrete projection in the case of Hilbert-valued functions $\phi : \Gamma \to V$ is given in (2.4.7)-right, and reads

$$\Pi^M_\Lambda u = \underset{v \in \mathbb{P}_\Lambda \otimes V}{\operatorname{argmin}} \frac{1}{M} \sum_{m=1}^{M} \|u(\mathbf{y}_m) - v(\mathbf{y}_m)\|^2_V. \tag{0.1.10}$$

The difference with Stochastic Galerkin in (0.1.8) relies on the two-step projection, first on the physical space and then on the probability space, while in (0.1.8) a unique large system couples the discretizations both in the probability and in the physical space due to the random nature of the equation $\mathcal{M}(u, \mathbf{y}) = 0$.

Improved Monte Carlo techniques. Many improved strategies to sample the parameter set have been proposed in the literature, to considerably fasten the convergence rate of the standard Monte Carlo method. We mention the **quasi-Monte Carlo** [SJ94, Lem09, DP10, GKN$^+$11], **Latin Hypercube Sampling** [MBC79, Loh96, HD03], **Multi-level Monte Carlo** [Gil08, HSST12], **Markov Chain Monte Carlo**. See *e.g.* [SW98] where an analysis is presented on how quasi-Monte Carlo outperforms the standard Monte Carlo method.

The application of the aforementioned techniques can be to the approximation of the moments of scalar and vectorial random functions or functionals depending on the solution to differential models with stochastic data (also known as Quantities of Interests), or to the approximation of the solution to models with stochastic data that has to be accomplished both in the usual functional space and in the probability space.

In some applications a non-intrusive approach is advised, because the Stochastic Galerkin method would require a significant programming effort to be implemented. This is the case, in the application of polynomial approximation presented in Chapter 6, that is naturally prone to be treated with the random discrete projection or with Stochastic Collocation.

Despite our focus are expansions of the output over a spectral polynomial basis, the random discrete projection, the Stochastic Collocation and the Stochastic Galerkin can be intended in a broader sense, and can be performed on any subspace of the L^2_ρ probability space.

On the other hand, the Monte Carlo family of methods is not directly associated with an expansion over any basis of the space of random input parameters. For instance, the moments of the function ϕ could be computed in (0.1.6) using $\psi \equiv 1$, together with powers of the evaluations of the target function ϕ. However, the random discrete L^2_ρ projection can also be viewed as a post-process of a Monte Carlo simulation to construct an expansion on a basis and significantly improve the convergence rate.

0.2 Motivations and results achieved

Polynomial approximation by the random discrete L^2 projection. In the literature, a lot of contributions deal with the analysis of the random discrete L^2 projection in the noisy framework, and the field of nonparametric regression covers this subject in detail, see *e.g.* [GKKW02]. Needless to say, the regression estimation problem has countless applications towards applied mathematics and all the branches of science dealing with experimental measurements.

In the noiseless case, very few contributions were present in the literature on nonparametric regressions, and a complete theoretical analysis of the stability and convergence of the random L^2 projection was missing. The interest towards the noiseless framework arises from the development of efficient techniques to deal with computational models expressed in parametric form. In the applications, the uncertainty often affects the measurements of the input parameters, which can be treated as random variables. Then it is reasonable to assume that the solution to the model can be evaluated without any additional noise, since the effects of deterministic approximation-type errors such as round-off errors are negligible w.r.t. the effects due to measurement errors.

This motivates a deeper investigation of the impact that the presence of noise has on estimates like (0.1.4), in particular on the nature of the term (0.1.5) that prevents the exponential convergence of the estimate when approximating smooth target functions.

The present thesis is devoted to the theoretical study of the stability and optimality properties of the approximation methodology based on the random discrete L^2 projection, with a focus on polynomial-type approximations of smooth functions. A condition between the number of sampling points and the dimension of the polynomial space is introduced, and its key role in the stability and optimality of the random discrete projection is presented. Under this condition, even exponential convergence can be achieved for smooth functions. The effects of the dimension of the parameter space and of the smoothness of the target function are clarified as well. Then the methodology based on the random discrete projection is applied to approximate aleatory functions and the solution to PDE models with stochastic data.

The interest towards the methodology based on the random discrete projection in Uncertainty Quantification is related to the increasing attention dedicated to non-intrusive approaches like Stochastic Collocation, in contrast to intrusive methods like Stochastic Galerkin. The random discrete projection preserves the feature of being non-intrusive, still resorting to a global projection in the probability space. This approach is particularly suited to the application to PDE with stochastic data, being the evaluations of the computational model mutually uncoupled. In the Stochastic Collocation method based on Sparse Grids, the computational cost required to compute the optimal sparse grid becomes unaffordable as the dimension of the parameter set increases. To alleviate this effect, Adaptive Sparse Grids (see *e.g.* [GG03]) or greedy approaches to build quasi-optimal sparse grids might be considered, but the application of the Sparse Grids approach remains solely confined to moderately high dimensions of the parameter space. On the other hand, the analysis in this thesis reveals that the stability of the random discrete projection improves as the dimension of the parameter space increases, making it a tool naturally oriented towards high dimensional problems. A flexible and easy incorporation of new measurements to improve the current accuracy of the polynomial approximation is another point of strength of the random discrete projection versus the Stochastic Collocation method, since the new measurements do not have to satisfy any hierarchical compatibility. In the random discrete projection, the computational cost required to compute the sparse grid is replaced by the computational cost required to perform the projection (0.1.3). With the same underlying polynomial space, the accuracy of the projection (0.1.1) is superior to the Lagrangian interpolation (0.1.7). However, in the random discrete projection (0.1.3) replaces (0.1.1), and whether (0.1.3) is more accurate than (0.1.7) is still under investigation. In the end, the two approaches are quite different and a fair comparison among them is still an open topic of research.

Application of the Factorization Method to Electrical Impedance Tomography in inhomogeneous uncertain background. The present thesis addresses the application of the Factorization Method to the framework of EIT. The Factorization Method is an imaging technique belonging to the class of *Qualitative Methods*, see *e.g.* [CC06] for an introduction. These types of methods were initially developed in the context of inverse scattering [CPP97, CK98, CHP03, BBP$^+$09], and then extended to EIT [KG08]. Nowadays, the application of the Factorization Method to EIT with homogeneous background has been analyzed in many contributions proposed in the literature, see [CFMV98, Brü01, BHV03, GH07, AHOS07, HS08, LHH08]. In the case of deterministic inhomogeneous background the only contribution is [HS09], where a slightly inhomogeneous background is treated as a perturbation of the homogeneous case.

The contribution of this thesis concerns the numerical analysis of the Factorization Method applied to the Continuous Model in EIT, in the case of inhomogeneous uncertain background. The Factorization Method aims to detect the presence and location of inclusions inside the computational domain, exploiting the information obtained when observing the solution to the associated PDE model on the whole boundary of the domain. A numerical scheme to solve the dipole-like Neumann problem in inhomogeneous background has been proposed. This allowed to extend the range of application of the Factorization Method to inhomogeneous deterministic background, *i.e.* to background diffusion coefficient featuring a nonlinear spatial dependence on the spatial coordinates, as well as to the case of inhomogeneous uncertain background. More specific types of background, *i.e.* piecewise constant in a partition of the computational domain, are then considered. The range of variation of the background coefficient has been investigated up to two orders of magnitude.

In addition, a Tikhonov regularization technique embedding the Morozov principle has been compared with another regularization technique proposed in the literature and based on the Picard Criterion. The effectiveness of the proposed approach has been checked also in presence of artificial noise contaminating the measurements of the observation operator.

When uncertainty in the background coefficient is present, a classification has been proposed depending on how the measurements of the random observation operator are collected. Two situations are distinguished: the case of arbitrary measurements and the case of paired measurements. Accordingly, three variants of the Factorization Method have been proposed to cope with the uncertainty in the background, and their capabilities have been presented in an extensive set of numerical tests.

Acceleration of the Factorization Method by means of the random discrete projection. The methodology based on the random discrete projection has been applied to the polynomial approximation of the solution to a specific elliptic model, which is embedded in the variant of the Factorization Method proposed for paired measurements. In addition, the polynomial approximation has been applied to a specific Neumann-to-Dirichlet map embedded in the variant of the Factorization Method proposed for arbitrary measurement. The aforementioned elliptic model and Neumann-to-Dirichlet map contain both the same random background diffusion coefficient.

In the two cases, a convergence analysis of the approximation error committed by the random discrete projection has been proposed. The use of a suitable norm ensures to control the error between the spectrum of the original operator and the spectrum of the polynomial surrogate model, for each realization of the random variable describing the random background coefficient.

Therefore the evaluations of the original operator, each one requiring to numerically solve a PDE model, can be replaced by the costless evaluations of the polynomial surrogate model. This allows to accelerate the variants of the Factorization Method in the case of arbitrary measurements and in the case of paired measurements.

0.3 Thesis overview

This thesis includes the following five manuscripts, written in collaboration with A. Chkifa, A. Cohen, H. Haddar, F. Nobile, E. von Schwerin, R. Tempone:

- [MNvST11] "Analysis of the discrete L^2 projection on polynomial spaces with random evaluations" by G. Migliorati, F. Nobile, E. von Schwerin and R. Tempone, which forms Chapter 1.

- [CCM+13] "Discrete least squares polynomial approximation with random evaluations; application to parametric and stochastic elliptic PDEs" by A. Chkifa, A. Cohen, G. Migliorati, F. Nobile and R. Tempone, which forms Chapter 2.

- [MNvST13] "Approximation of Quantities of Interest for stochastic PDEs by the random discrete L^2 projection on polynomial spaces" by G. Migliorati, F. Nobile, E. von Schwerin and R. Tempone, which forms Chapter 4.

- [HM11] "Numerical analysis of the Factorization Method for Electrical Impedance Tomography in inhomogeneous background" by H. Haddar and G. Migliorati. The Sections 5.2.3 and 5.5 of Chapter 5 are extracted from here.

- [HM13] "Numerical analysis of the Factorization Method for Electrical Impedance Tomography with a piecewise constant uncertain background" by H. Haddar and G. Migliorati, which forms the remaining part of Chapter 5.

The main topic of the thesis concerns the polynomial approximation of aleatory functions, by means of the random discrete L^2 projection. The thesis is structured in two parts: the former is dedicated to the analysis of methods for the approximation of random aleatory functions, and for the approximation of the solution to computational models formulated in terms of Partial Differential Equations. This part is composed by the following four chapters:

- **Chapter 1.** The random discrete L^2 projection is presented as a tool to approximate functions which depend on random variables, distributed according to given probability densities. Under some assumptions, *e.g.* tensor structure of the density, density bounded away from zero and smoothness of the function, the target function is projected on a suitable space of polynomials of random variables. The projection is built using pointwise noise-free evaluations of the target function on randomly selected points of the parameter set. The stability and approximation properties of the random projection are analyzed. For any dimension of the parameter set, and any density of the multidimensional random variable (with a tensor structure) we proved in Proposition 1.1 the optimality of the L^2_ρ error committed by the random projection, with respect to the L^∞ best approximation error. This result holds for any realization of the random sample. Then in Theorem 1.2 a one-dimensional probabilistic optimality result is proved, in the case of the uniform density and under the condition (1.3.1) that the number of sampling points scales as the square of the dimension of the polynomial space. In this chapter the optimal convergence of RDP has to be understood in the sense that the L^2-error behaves like the best approximation error (measured in the L^∞ norm) of the target function in the chosen polynomial space, up to logarithmic factors. Moreover, it is shown in Proposition 1.4 that both the stability of the random projection and the optimal convergence rate are related to the supremum of the ratio between the continuous L^2 weighted and discrete norms of the functions in the polynomial space, *i.e.* the random variable C^ω defined in (1.2.15).

  Several numerical tests in the univariate and multivariate case confirm our theoretical results, and point out how the smoothness of the target function affects the optimal convergence rate. Many numerical examples show that, if the target function is smooth, then an optimal convergence up to a given threshold is obtained with a linear scaling of the number of sampling points w.r.t. the dimension of the polynomial space. In higher dimensions a linear scaling allows to achieve the quasi-optimal convergence rate even in cases where the target function features a low regularity. The stability of the random projection improves as well in higher dimension, making this tool naturally suited to high dimensional approximations.

- **Chapter 2.** This chapter begins with an overview of the theoretical results obtained in expectation and in probability on the random discrete L^2 projection. The condition (1.3.1) presented in Chapter 1, necessary to achieve stability and optimality of the random discrete L^2 projection, is extended in Section 2.3 to any multi-index monotone set identifying the polynomial space, and in any arbitrary dimension of the parameter set. Moreover, in Theorem 2.6 a convergence estimate (2.4.19) in expectation for the random discrete projection with an a priori chosen polynomial space is derived. This estimate holds for

any function $\phi : \Gamma \to \mathbb{R}$ in $L^2_\rho(\Gamma)$ analytic in a polydisc. In particular, its application is presented in Section 2.5 to approximate the solution to a class of PDE models with stochastic coefficients. For example, the aforementioned class contains elliptic PDEs and the linear elasticity model, both with stochastic coefficients described by means of random variables with a disjoint support in the physical domain. The estimate (2.4.19) shows that the random discrete L^2 projection on suitably chosen sequences of polynomial spaces converges at least with sub-exponential convergence rate w.r.t. the number of sampling points. In addition, the same estimate establishes a relation between the random discrete L^2 projection on a priori chosen polynomial spaces and the classical Stochastic Galerkin method.

- **Chapter 3.** The assumption on the density presented in Chapter 1 is relaxed, and the case of the beta family of probability densities is investigated, from the point of view of stability and optimality of the convergence rate. Then the case of Gaussian density is investigated, and it is shown that the Assumption 1.1 on the density being bounded away from zero is strictly required for the probabilistic optimality result to hold. Lastly, the robustness of the random projection w.r.t. perturbations in the density is investigated through some numerical tests.

- **Chapter 4.** The random discrete projection is applied to approximate Quantities of Interest (QOI), defined as integral functionals of the solution to PDE models with random coefficients and random domains. The theoretical results obtained with the analysis developed in Chapter 1 and Chapter 2 extend straightforwardly to the case of QOI. Throughout the chapter, the polynomial approximation is applied with the Total Degree polynomial space, as motivated in Chapter 2. Several examples with elliptic PDEs are presented (Darcy model), when the diffusion coefficient is parametrized by random variables featuring one and higher dimension. Then, the linear elasticity model with random Young's modulus and the Navier-Stokes model with random viscosity and random geometry of the computational domain have been tested. The role of smoothness in the dependence of the QOI on the random variables has been numerically investigated as well.

In the latter part of the thesis, some of the techniques developed in the first part for the forward problem are applied to solve more complicated inverse problems, involving PDEs with stochastic data. The type of inverse problems treated concerns the recovery of information on the coefficient of the PDE model, with a specific target towards the case where uncertainty is present.

- **Chapter 5.** The Factorization Method has been extended to the case of deterministic inhomogeneous background, and then of piecewise uncertain background. At first, a numerical scheme to solve the dipole-like singular Neumann problem has been proposed in Section 5.2.2, when the background coefficient is deterministic but spatially inhomogeneous. Then, this scheme has been incorporated in the Factorization Method to solve the EIT inverse problem. Several types of deterministic inhomogeneous background are addressed: piecewise constant background with two orders of magnitude jumps between neighboring regions and more general backgrounds with a nonlinear dependence on the spatial coordinates. It is shown that, in the inhomogeneous case, the inversion method achieves reconstructions of the same accuracy as those obtained in the homogeneous case. The method has been tested also when the measurements are contaminated by noise. Afterwards, the case of piecewise constant background with values of the coefficients affected by randomness is addressed, again with variations of the coefficient up to two orders of magnitude between neighboring regions. Depending on the procedure employed to collect the measurements, two situations are distinguished, *i.e.* arbitrary measurements and paired measurements. In the case of arbitrary measurements, two variants of the Factorization Method have been proposed. The first variant accomplishes a background reconstruction through an optimization procedure, and then performs the sampling step similarly to the standard

Factorization Method. The second variant avoids the optimization step, and uses a weighted linear combination of the indicator functions provided by the Factorization Method, as long as a sufficient number of measurements are available. In the case of paired measurements, a suitable variant of the Factorization Method based on a Monte Carlo type of indicator function has been proposed. The effectiveness of all the variants of the Factorization Method proposed is demonstrated by means of several numerical tests.

- **Chapter 6.** The polynomial approximation techniques presented in Chapter 1 and Chapter 2 are utilised to accelerate the variants of the Factorization Method proposed in Chapter 5. The proposed approach relies on the polynomial approximation of the inclusion-free Neumann-to-Dirichlet map Λ_0 recalled in (6.0.2), and allows to accelerate either the optimized variant of the Factorization Method (Algorithm 2) or its pure variant (Algorithm 3). The random discrete projection is applied to compute the polynomial approximation Λ_0^M of the operator Λ_0 with random background coefficient, *e.g.* the one defined in (6.0.1). The pointwise evaluations of the operator $\Lambda_0 = \Lambda_0(\mathbf{y})$, for each value of the parameter $\mathbf{y} \in \Gamma$, can then be replaced by the computationally costless pointwise evaluations of $\Lambda_0^M = \Lambda_0^M(\mathbf{y})$. In Section 6.3 some convergence estimates are derived, for the approximation error committed by the random discrete L^2 projection when approximating the solution to model (6.0.3), which defines the operator Λ_0. The error estimates are derived in a convenient norm that ensures pointwise accuracy, rather than accuracy in expectation. In Theorem 6.3 a convergence estimate in the Frobenius norm for the polynomial approximation Λ_0^M of the operator Λ_0 has been derived, and it ensures the convergence of all the Ky-Fan k-norms with $k < K$, with K being the highest frequency in the measurements of the EIT problem.

0.4 Mind map of the chapters

The reader of the thesis is advised to begin with either Chapter 1 or Chapter 2 which are both self-contained and introduce the random discrete projection. These chapters are prerequisites to the following Chapter 3 and Chapter 4 which deal with different aspects related to the random discrete projection. Another starting point to read the thesis could be Chapter 5, where the Factorization Method applied to the framework of Electrical Impedance Tomography is introduced. However, before moving to Chapter 6, the reader is advised to have a look at Chapter 1 and Chapter 2 as well. Fig. 1 resumes the suggested order to read the chapters.

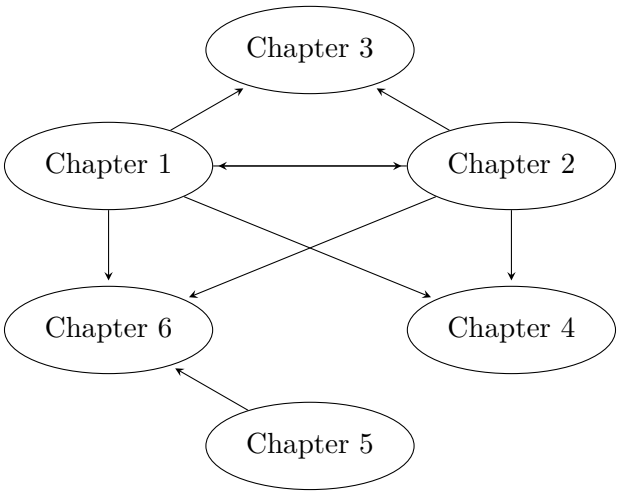


Figure 1: Structure of the chapters. The arrows indicate the advised reading order.

Part I

Polynomial approximation by the discrete L^2 projection with random evaluations

Chapter 1

Analysis of the discrete L^2 projection on polynomial spaces with random evaluations

The present chapter is based on the work [MNvST11] entitled "Analysis of the discrete L^2 projection on polynomial spaces with random evaluations" by G. Migliorati, F. Nobile, E. von Schwerin and R. Tempone.

Abstract

We analyze the problem of approximating a multivariate function by discrete least-squares projection on a polynomial space starting from *random, noise-free* observations. An area of possible application of such technique is Uncertainty Quantification (UQ) for computational models. We prove an optimal convergence estimate, up to a logarithmic factor, in the univariate case, when the observation points are sampled in a bounded domain from a probability density function bounded away from zero, provided the number of samples scales quadratically with the dimension of the polynomial space. Several numerical tests are presented both in the univariate and multivariate case, confirming our theoretical estimates. The numerical tests also clarify how the convergence rate depends on the number of sampling points, on the polynomial degree, and on the smoothness of the target function.

Key words: approximation theory, error analysis, noise-free data, multivariate polynomial approximation, point collocation, generalized polynomial chaos, nonparametric regression.

AMS subject classification: 41A10, 41A25, 65N12, 65N15, 65N35

1.1 Introduction

Given a smooth multivariate function $\phi = \phi(Y_1, \ldots, Y_N)$, where $Y_1, \ldots, Y_N$ are random variables, we consider in this work the classical problem of approximating ϕ in a multivariate polynomial space, starting from noise-free observations of ϕ on random samples of $Y_1, \ldots, Y_N$.

The motivations for such work come from the field of Uncertainty Quantification (UQ) in computational models [LMK10, Xiu09], where often uncertainty is present in many input parameters entering the mathematical model used to describe some problem in engineering, physics, biology, etc. and can be characterized in probabilistic terms by considering the input

parameters as random variables. The goal of the analysis is typically to compute statistics of the solution to the mathematical model or some output quantities of interest.

It is assumed here that, for each value of the input parameters, the solution or output quantity can be accessed without errors. This is, of course, an idealization as deterministic approximation-type errors will typically be present whenever the model involves differential or integral operators. Also, round-off errors will be present as well. However, these sources of errors are quite different in nature from "measurement errors" appearing in an experimental setting, which are usually modeled as random and statistically independent. In the context of UQ in computational models, it is therefore reasonable to assume that the approximation errors can be kept under control by some careful a-posteriori error estimation and mesh refinement (see e.g. [AO00, BR03] and references therein).

A technique that has received considerable attention in the last few years is the so called generalized Polynomial Chaos expansion (gPC); see e.g. [GS03, XK02b]. It consists in expanding the solution in polynomials of the input random variables. The use of global polynomial spaces is sound in many situations, where the input/output (parameters-to-solution) map is smooth. This is the case, for instance, in elliptic PDEs with random diffusion coefficient [BNT10, CDS10, CDS11, BNTT11b].

Once a truncated gPC expansion has been computed by some means, it can be used later for inexpensive computations of solution statistics or as a reduced model of the input/output map for "global sensitivity analysis" [CLMM09, Sud07], or optimization under uncertainty [Eld11].

As a tool to build such a gPC approximation, we consider in this work an L^2 projection, starting from a random sample of the input parameters. Such an idea has already been proposed in the framework of UQ and has been given several names: Point Collocation [HWB10, DNP$^+$04, NTW08b], non intrusive gPC [LMK10, Eld09] regression [BS08, BS11]. As a practical recipe, the number of samples drawn from the input distribution is typically taken to be 2 to 3 times the dimension of the polynomial space.

The proposed approach falls within the classical topic of polynomial regression estimation, i.e. minimization of the empirical L^2 risk within the given polynomial space. We insist, however, that unlike the traditional statistical approach, here we consider noise-free data evaluated in random points.

A relevant question is whether such a minimizer is "optimal" in the sense that it achieves an approximation of the (unknown) function that is equivalent to the "best approximation" in the polynomial space.

A lot of literature is available on regression estimations in the noisy case. We recall here the book [GKKW02] that provides a general framework for analysis of regression estimators with random design, as well as the works [BCD$^+$05, BCDD07] that show optimality of the noisy regression when using piecewise constant or linear functions. The aforementioned works give estimates on the expected L^2 error under the assumption that the function is bounded in L^∞ by some a priori fixed constant. Other works in the field of distribution-free regression with noise have derived convergence rates for the L^2 risk which are optimal up to logarithmic factors, *e.g.* [Koh00, Koh03, BCK09].

The L^2-error in expectation is bounded by two terms: the (purely deterministic) best approximation error of the (unknown) function in the approximating space, and the statistical error due to the random sampling and the noise in the observations. The latter scales as $1/\sqrt{M}$ if M is the number of samples. In the aforementioned works, the statistical error does not vanish in the noise-free setting.

Hence, the main question that we address in this paper: in the noise-free polynomial approximation based on random samples, does the randomness of the evaluation points introduce a statistical error $\mathcal{O}(1/\sqrt{M})$ or not?

We study theoretically the problem for a univariate function $\phi(Y)$ where Y is a bounded random variable with probability density function bounded away from zero. Denoting by $\#\Lambda$ the dimension of the polynomial space, we prove that the L^2 projection on a polynomial space with random samples leads to quasi optimal convergence rates (up to logarithmic factors) provided

that the number of samples scales as $M \sim (\#\Lambda)^2$.

Similar results have been derived very recently in [CDL13]. However, the proof proposed therein uses different techniques than ours.

We also show, in the general multivariate setting, the relation between the optimality of the projection on random points and the condition number of the corresponding design matrix, when using an orthonormal basis.

We present several numerical tests, both on univariate and multivariate functions, that clearly show that a choice $M \sim (\#\Lambda)^2$ leads to a stable regression problem and an optimal approximation, whereas $M \sim \#\Lambda$ leads to an ill conditioned problem when $\#\Lambda$ is large and, eventually, to a divergent approximation. Moreover, our numerical tests show some significant differences between the one-dimensional and the multidimensional case.

The outline of the paper is the following: Section 1.2 introduces the approximation problem as an L^2 projection on a space of polynomials in N underlying variables; some common choices of polynomial spaces are described in Section 1.2.1. The optimality of the random L^2 projection, in terms of a best approximation constant, is shown in Section 1.2.2; the asymptotic behaviour of this best approximation constant, as the number of random evaluation points goes to infinity, is analyzed in Section 1.2.3. Next, Section 1.3 restricts the study to polynomial spaces in one variable, with uniformly distributed random points; in this case, a bound on the best approximation constant is derived and used to prove Theorem 1.2, which given the maximal polynomial degree provides a rule for the number of random points that makes the discrete random L^2 projection nearly optimal (up to a logarithmic factor) with any prescribed confidence level. Section 1.4 gives the algebraic formulation of the random projection problem, in view of its numerical discretization. In particular, Section 1.4 provides an analysis of how the condition number of the design matrix depends on the best approximation constant of the random L^2 projector. Finally, Section 1.5 complements the analysis in Sections 1.2–1.4 with numerical tests, both in the one-dimensional case and in higher dimensions.

1.2 Discrete L^2 projection with random points

Let $\mathbf{Y} = [Y_1, \dots, Y_N]$ be a vector of N random variables, taking values in a bounded set $\Gamma \subset \mathbb{R}^N$. We assume that Γ has a tensor structure $\Gamma = \Gamma_1 \times \cdots \times \Gamma_N$ and that the random vector $\mathbf{Y}$ has a probability density function $\rho : \Gamma \to \mathbb{R}^+$.

We consider a random variable $Z = \phi(\mathbf{Y})$, where $\phi : \Gamma \to \mathbb{R}$ is assumed to be a smooth function. The goal of the analysis is to compute statistical moments of Z. This will be achieved by first constructing a reduced model (approximate response function); i.e. we approximate the function $\phi(Y_1, \dots, Y_N)$ by a suitable multivariate polynomial $\phi_\Lambda(Y_1, \dots, Y_N)$. We then compute statistical moments using the approximate function ϕ_Λ.

We denote by

$$\mathbb{E}[Z] := \int_\Gamma \phi(\mathbf{Y})\rho(\mathbf{Y})d\mathbf{Y}$$

the expected value of the random variable Z and by

$$P(A) := \int_A \rho(\mathbf{Y})d\mathbf{Y}$$

the probability of the event $A \in \mathcal{B}(\Gamma)$, where $\mathcal{B}(\Gamma)$ is the Borel σ-algebra with respect to the measure $\rho(\mathbf{Y})d\mathbf{Y}$. Throughout the paper we also assume that

Assumption 1.1. $0 < \rho_{min} \le \rho(\mathbf{y}) \le \rho_{max} < \infty$, $\forall\, \mathbf{y} \in \Gamma$.

We introduce the space L^2_ρ of square integrable functions $f : \Gamma \to \mathbb{R}$, endowed with the norm

$$\|f\|_{L^2_\rho} = \left(\int_\Gamma f^2(\mathbf{Y})\rho(\mathbf{Y})d\mathbf{Y} \right)^{1/2}.$$

Observe that under Assumption 1.1, the norm $\|\cdot\|_{L^2_\rho}$ is equivalent to the standard L^2 norm (with Lebesgue measure), *i.e.*

$$\sqrt{\rho_{\min}} \leq \frac{\|v\|_{L^2_\rho(\Gamma)}}{\|v\|_{L^2(\Gamma)}} \leq \sqrt{\rho_{\max}}, \qquad \forall\, v \in L^2_\rho.$$

Let $\mathbf{p} = (p_1, \ldots, p_N)$ be a multi-index and $\Lambda \subset \mathbb{N}^N$ an index set which is monotonous in the following sense:

Property 1 (Monotonicity of Λ)**.** *Consider two multi-indices $\mathbf{p}', \mathbf{p}'' \in \mathbb{N}^N$ such that $p''_n \leq p'_n$, $\forall n = 1, \ldots, N$. The multi-index set Λ is monotonous if the following holds:*

$$\mathbf{p}' \in \Lambda \Longrightarrow \mathbf{p}'' \in \Lambda.$$

We denote by $\mathbb{P}_\Lambda(\Gamma)$ the multivariate polynomial space

$$\mathbb{P}_\Lambda(\Gamma) = \text{span}\left\{ \prod_{n=1}^{N} y_n^{p_n},\ \text{with } \mathbf{p} \in \Lambda \right\}, \tag{1.2.1}$$

and by $\#\Lambda = dim(\mathbb{P}_\Lambda)$ the dimension of the polynomial space, which corresponds to the cardinality of the index set Λ. For convenience, the set Λ can be arranged in lexicographical order:

Property 2 (Lexicographical order)**.** *Given $\mathbf{p}', \mathbf{p}'' \in \Lambda$,*

$$\mathbf{p}' \overset{L}{<} \mathbf{p}'' \quad \Longleftrightarrow \quad \exists\, \widetilde{n} > 0 \quad : \quad (p'_{\widetilde{n}} < p''_{\widetilde{n}}) \wedge (p'_n = p''_n \quad \forall\, n < \widetilde{n}).$$

According to this order, we can let $\mathbf{p}^j$ denote the jth multi-index of Λ. Sometimes we refer to the elements of Λ with the generic multi-index $\mathbf{p}$, rather than listing them by the lexicographical index.

Since the monomial basis in (1.2.1) is very ill-conditioned, in practice we use an orthonormal polynomial basis. A typical choice is to take orthogonal polynomials with respect to the measure $\rho(\mathbf{Y})d\mathbf{Y}$.

We introduce an N-dimensional orthonormal polynomial basis $\{l_j\}_{j=1}^{\#\Lambda}$ of $\mathbb{P}_\Lambda$ with respect to the weighted inner product

$$(u, v)_{L^2_\rho} = \int_\Gamma u(\mathbf{Y}) v(\mathbf{Y}) \rho(\mathbf{Y})\, d\mathbf{Y}, \tag{1.2.2}$$

i.e. $(l_i, l_j)_{L^2_\rho} = \delta_{ij}$. Assumption 1.1 guarantees that the orthonormal basis is complete in L^2_ρ when $\Lambda = \mathbb{N}^N$.

In the case where the density factorizes as $\rho(\mathbf{Y}) = \prod_{n=1}^N \rho_n(Y_n)$ the basis can be constructed by tensorizing 1D orthogonal polynomials with respect to each weight ρ_n separately. Given n, let $\{\varphi_j^n(\cdot)\}_j$ be the orthogonal polynomials with respect to ρ_n. Picking the jth (according to Property 2) multi-index $\mathbf{p}^j \in \Lambda$, the corresponding jth multidimensional basis function is given by

$$l_j(\mathbf{Y}) = \prod_{n=1}^{N} \varphi^n_{p_n^j}(Y_n). \tag{1.2.3}$$

Thus, using the basis function provided by (1.2.3), the definition (1.2.1) of $\mathbb{P}_\Lambda$ becomes

$$\mathbb{P}_\Lambda(\Gamma) = \text{span}\{l_j,\, j = 1, \ldots, \#\Lambda\}. \tag{1.2.4}$$

Observe that in general (1.2.1) and (1.2.4) are equivalent only if the index set Λ satisfies the Monotonicity Property 1.

To construct the polynomial approximation ϕ_Λ we first sample the exact function ϕ in $M > \#\Lambda$ independent points $\mathbf{y}_1, \dots, \mathbf{y}_M$, i.e. $\mathbf{y}_i$ are independent identically distributed (i.i.d.) random vectors with probability density function ρ. Then, we compute a discrete least square approximation of the values $\phi(\mathbf{y}_i)$ in the polynomial space $\mathbb{P}_\Lambda$, i.e.,

$$\phi_\Lambda = \Pi_M^{\Lambda,\omega}\phi = \underset{v\in\mathbb{P}_\Lambda(\Gamma)}{\operatorname{argmin}} \frac{1}{M}\sum_{i=1}^{M}(\phi(\mathbf{y}_i) - v(\mathbf{y}_i))^2. \tag{1.2.5}$$

We will use the superscript (or subscript) ω to denote a quantity that depends on the random sample $\mathbf{y}_1, \dots, \mathbf{y}_M$ (and therefore is random itself).

We now introduce the *random* discrete inner product

$$(u, v)_{M,\omega} = \frac{1}{M}\sum_{i=1}^{M} u(\mathbf{y}_i)v(\mathbf{y}_i), \tag{1.2.6}$$

and the corresponding norm $\|u\|_{M,\omega} = (u, u)_{M,\omega}^{1/2}$. Observe that this is actually a norm in $\mathbb{P}_\Lambda$ if for all $v \in \mathbb{P}_\Lambda$,

$$v(\mathbf{y}_i) = 0, \ \text{ for } i = 1, \dots, M \qquad \Leftrightarrow \qquad v = 0, \tag{1.2.7}$$

which by Assumption 1.1 happens with probability one provided that M is large enough. Then, the discrete least square problem can be written equivalently as

$$\text{find } \Pi_M^{\Lambda,\omega}\phi \in \mathbb{P}_\Lambda(\Gamma) \quad \text{s.t.} \quad (\Pi_M^{\Lambda,\omega}\phi, v)_{M,\omega} = (\phi, v)_{M,\omega}, \quad \forall v \in \mathbb{P}_\Lambda(\Gamma).$$

1.2.1 Common multivariate polynomial spaces

Some of the most common choices of function spaces are, Tensor Product, Total Degree, and Hyperbolic Cross, which are defined by the index sets below. We index the set Λ by the subscript w denoting the maximum polynomial degree used:

$$\textbf{Tensor Product (TP)}, \quad \Lambda_{\mathrm{w}} = \Big\{\mathbf{p} \in \mathbb{N}^N : \max_{n=1,\dots,N} p_n \le \mathrm{w}\Big\}, \tag{1.2.8}$$

$$\textbf{Total Degree (TD)}, \quad \Lambda_{\mathrm{w}} = \Big\{\mathbf{p} \in \mathbb{N}^N : \sum_{n=1}^{N} p_n \le \mathrm{w}\Big\}, \tag{1.2.9}$$

$$\textbf{Hyperbolic Cross (HC)}, \quad \Lambda_{\mathrm{w}} = \Big\{\mathbf{p} \in \mathbb{N}^N : \prod_{n=1}^{N} (p_n + 1) \le \mathrm{w} + 1\Big\}, \tag{1.2.10}$$

$$\textbf{Smolyak polynomial space (SM)}, \quad \Lambda_{\mathrm{w}} = \Big\{\mathbf{p} \in \mathbb{N}^N : \sum_{n=1}^{N} f(p_n) \le f(\mathrm{w})\Big\}, \tag{1.2.11}$$

$$\text{with } f(p) = \begin{cases} 0, & p = 0, \\ 1, & p = 1, \\ \lceil \log_2(p) \rceil, & p \ge 2. \end{cases}$$

These spaces are isotropic and the maximum polynomial degree w is the same in all spatial dimensions. Anisotropic versions could be considered as well, see *e.g.* Appendix A.1 and [BNTT11a]. The dimensions of TP and TD spaces are

$$\#TP(\mathrm{w}, N) = (\mathrm{w} + 1)^N, \tag{1.2.12}$$

$$\#TD(\mathrm{w}, N) = \binom{N + \mathrm{w}}{N}. \tag{1.2.13}$$

The dimension of the HC space is harder to quantify, so we report its exact dimension $\#HC(\mathrm{w}, N)$ in Fig. 1.1, computed for some values of w and N. A sharp upper bound, that holds when $N = 2$, is given by

$$\#HC(\mathrm{w}, 2) \leq \left\lfloor (\mathrm{w}+1) \cdot (1 + \log(\mathrm{w}+1)) \right\rfloor. \tag{1.2.14}$$

See Appendix A.2 for a proof.

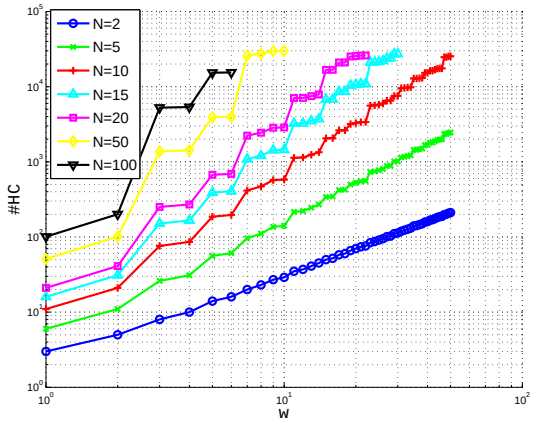


Figure 1.1: Dimension of the HC space, $N = 2, 5, 10, 15, 20, 50, 100$.

1.2.2 L^2_ρ optimality of the random discrete L^2 projection

Let us first introduce the following quantity

$$C^\omega(M, \Lambda) := \sup_{v \in \mathbb{P}_\Lambda \setminus \{v \equiv 0\}} \frac{\|v\|^2_{L^2_\rho}}{\|v\|^2_{M,\omega}}. \tag{1.2.15}$$

Notice that this quantity depends on the random sample and is therefore a random variable. The following result holds:

Proposition 1.1. *With $C^\omega(M, \Lambda)$ defined as in (1.2.15), it holds that*

$$\|\phi - \Pi^{\Lambda,\omega}_M \phi\|_{L^2_\rho} \leq \left(1 + \sqrt{C^\omega(M, \Lambda)}\right) \inf_{v \in \mathbb{P}_\Lambda(\Gamma)} \|\phi - v\|_{L^\infty}. \tag{1.2.16}$$

Proof. When $v = \Pi^{\Lambda,\omega}_M \phi \in \mathbb{P}_\Lambda$ the statement (1.2.16) trivially holds. Let $v \neq \Pi^{\Lambda,\omega}_M \phi$ be an arbitrary polynomial in $\mathbb{P}_\Lambda$. Then

$$\begin{aligned}
\|\phi - \Pi^{\Lambda,\omega}_M \phi\|_{L^2_\rho} &\leq \|\phi - v\|_{L^2_\rho} + \|v - \Pi^{\Lambda,\omega}_M \phi\|_{L^2_\rho} \\
&= \|\phi - v\|_{L^2_\rho} + \frac{\|v - \Pi^{\Lambda,\omega}_M \phi\|_{L^2_\rho}}{\|v - \Pi^{\Lambda,\omega}_M \phi\|_{M,\omega}} \|v - \Pi^{\Lambda,\omega}_M \phi\|_{M,\omega} \\
&\leq \|\phi - v\|_{L^2_\rho} + \sup_{v \in \mathbb{P}_\Lambda \setminus \{v \equiv 0\}} \frac{\|v\|_{L^2_\rho}}{\|v\|_{M,\omega}} \|v - \Pi^{\Lambda,\omega}_M \phi\|_{M,\omega} \\
&= \|\phi - v\|_{L^2_\rho} + \sqrt{C^\omega(M, \Lambda)} \left(\|\phi - v\|^2_{M,\omega} - \|\phi - \Pi^{\Lambda,\omega}_M \phi\|^2_{M,\omega} \right)^{\frac{1}{2}} \\
&\leq \left(1 + \sqrt{C^\omega(M, \Lambda)}\right) \|\phi - v\|_{L^\infty}.
\end{aligned}$$

The thesis follows from the arbitrariness of v. □

Clearly, the convergence property of the random discrete projection is strictly related to the properties of the quantity $C^\omega(M,\Lambda)$. In Section 1.3 we will characterize this quantity in the particular case of a one-dimensional problem ($N=1$) and the uniform distribution. The extension to multivariate functions is an ongoing work.

1.2.3 Asymptotic limit of $C^\omega(M,\Lambda)$

Here we show that when $M\to\infty$, while Λ is kept fixed, the limit of $C^\omega(M,\Lambda)$ is almost surely 1. For the purpose of this analysis we also introduce the constant

$$\widetilde{C}_\Lambda = \sup_{\varphi\in\mathbb{P}_\Lambda} \frac{\|\varphi\|^2_{L^\infty}}{\|\varphi\|^2_{L^2_\rho}} < +\infty. \tag{1.2.17}$$

Remark 1.1. *Given any Λ, the constant $\widetilde{C}_\Lambda$ is always finite in any dimension N since the space $\mathbb{P}_\Lambda$ is finite dimensional and the domain is bounded. For example in one dimension, any $v\in\mathbb{P}_{\mathrm{w}}(-1,1)$ can be expanded in Legendre series $v=\sum_{j=0}^{\mathrm{w}}\alpha_j\varphi_j$ with $\|v\|^2_{L^2}=\sum_{j=0}^{\mathrm{w}}\alpha_j^2$. Moreover, assuming a uniform distribution ρ on the interval $[-1,1]$,*

$$\begin{aligned}\|v\|_{L^\infty} &\le \sum_{j=0}^{\mathrm{w}}|\alpha_j|\|\varphi_j\|_{L^\infty} \le \sum_{j=0}^{\mathrm{w}}|\alpha_j|\sqrt{j+\frac{1}{2}}\\ &\le \sqrt{\sum_{j=0}^{\mathrm{w}}\alpha_j^2}\sqrt{\sum_{j=0}^{\mathrm{w}}\left(j+\frac{1}{2}\right)} = \frac{(\mathrm{w}+1)}{\sqrt{2}}\|v\|_{L^2} = (\mathrm{w}+1)\|v\|_{L^2_\rho},\end{aligned}$$

so $\sqrt{\widetilde{C}_\Lambda}\le \mathrm{w}+1$. One can tensorize the one-dimensional case, to show that $\widetilde{C}_\Lambda$ is bounded with respect to the maximum polynomial degree w *also in higher dimensions.*

We recall that a sequence of functions $\{f_n\}$, where $f_n: S\to\mathbb{R}$ converges uniformly to f if and only if

$$\lim_{n\to\infty}\sup_{s\in S}\left|f_n(s)-f(s)\right| = 0.$$

We denote the uniform convergence by $f_n \xrightarrow{\text{unif.}} f$. We also recall the definition of an equicontinuous sequence of functions:

Definition 1.1. *Consider the usual ε-δ definition of continuity for a real function f in the point x_0:*

$$\forall\varepsilon>0 \quad \exists\delta=\delta(x_0,f)>0 \quad \forall x\in dom(f)\,:\,|x-x_0|<\delta \Rightarrow |f(x)-f(x_0)|<\varepsilon. \tag{1.2.18}$$

We say that the family of continuous functions $F=\{f_i\}_i$ is

- *equicontinuous if δ in* (1.2.18) *is independent of f_i,*
- *uniformly equicontinuous if δ in* (1.2.18) *is independent of f_i and x_0.*

Now consider the set of functions $\widehat{\mathbb{P}}_\Lambda=\{v\in\mathbb{P}_\Lambda\,:\,\|v\|_{L^2_\rho}=1\}\subset L^2_\rho$. For any outcome $\{\mathbf{y}_j\}_{j=1}^\infty$ of the random variables $\{\mathbf{Y}_j\}_{j=1}^\infty$ we define the sequence $\{\tau_{M,\omega}\}_M$ of functionals, whose elements are defined as

$$\tau_{M,\omega}(\cdot) = \frac{\|\cdot\|^2_{M,\omega}}{\|\cdot\|^2_{L^2_\rho}} : \widehat{\mathbb{P}}_\Lambda\to\mathbb{R}^+. \tag{1.2.19}$$

Proposition 1.2. *For any M and Λ the function $\tau_{M,\omega}$ in* (1.2.19) *is Lipschitz continuous over $\widehat{\mathbb{P}}_\Lambda$.*

Proof. Consider the constant $\widetilde{C}_\Lambda$ defined in (1.2.17). Clearly $\|v\|_{L^\infty} \leq \sqrt{\widetilde{C}_\Lambda}$ for all $v \in \widehat{\mathbb{P}}_\Lambda$, so they are uniformly bounded. Taking arbitrary v_1 and v_2 in $\widehat{\mathbb{P}}_\Lambda$,

$$\begin{aligned}\Big|\tau_{M,\omega}(v_1) - \tau_{M,\omega}(v_2)\Big| &\leq \frac{1}{M}\sum_{j=1}^{M}\Big|v_1(y_j)^2 - v_2(y_j)^2\Big| \\ &= \frac{1}{M}\sum_{j=1}^{M}\Big|\Big(v_1(y_j)+v_2(y_j)\Big)\Big(v_1(y_j)-v_2(y_j)\Big)\Big| \\ &\leq \frac{1}{M}\Big(\|v_1\|_{L^\infty}+\|v_2\|_{L^\infty}\Big)\sum_{j=1}^{M}\Big|v_1(y_j)-v_2(y_j)\Big| \\ &\leq 2\sqrt{\widetilde{C}_\Lambda}\,\|v_1-v_2\|_{L^\infty} \leq 2\,\widetilde{C}_\Lambda\,\|v_1-v_2\|_{L^2_\rho}. \end{aligned} \tag{1.2.20}$$

□

From Proposition 1.2 it follows immediately that the sequence $\{\tau_{M,\omega}\}$ is uniformly equicontinuous. Moreover, from the Strong Law of Large Numbers, since $\mathbb{E}[v(y_j)^2] < \infty$, it follows that

$$\|v\|^2_{M,\omega} = \frac{1}{M}\sum_{j=1}^{M} v(x_j)^2 \overset{M\to+\infty}{\longrightarrow} \mathbb{E}\Big[(v(y_j))^2\Big] = \|v\|^2_{L^2_\rho},$$

almost surely; hence the sequence $\{\tau_{M,\omega}\}$ is also converging to 1 pointwise. We have then the following result:

Proposition 1.3. *Under the assumptions of Proposition 1.2 it holds*

$$\tau_{M,\omega}(v) \overset{unif.}{\longrightarrow} 1, \quad \forall\, v \in \widehat{\mathbb{P}}_\Lambda, \quad a.s. \tag{1.2.21}$$

Proof. For any outcome ω, the sequence $\{\tau_{M,\omega}\}_{M=1}^{\infty}$ is (uniformly) equicontinuous on $\widehat{\mathbb{P}}_\Lambda$ and converges almost surely on $\widehat{\mathbb{P}}_\Lambda$, therefore it converges uniformly in $\widehat{\mathbb{P}}_\Lambda$ (see *e.g.* [Rud76, Theorem 7.25]). □

Theorem 1.1. *Let $C^\omega(M,\Lambda)$ be the constant defined in* (1.2.15)*. Then*

$$\lim_{M\to\infty} C^\omega(M,\Lambda) = 1, \quad a.s.$$

Proof. We can extend the definition of $\tau_{M,\omega}$ in (1.2.19) to $\mathbb{P}_\Lambda \setminus \{v \equiv 0\}$ using the same formula. Then we rewrite the following supremum as

$$\lim_{M\to+\infty}\sup_{v\in\widehat{\mathbb{P}}_\Lambda}\left|\frac{\|v\|^2_{M,\omega}}{\|v\|^2_{L^2_\rho}} - 1\right| = \lim_{M\to+\infty}\sup_{v\in\mathbb{P}_\Lambda\setminus\{v\equiv 0\}}\left|\frac{\|v\|^2_{M,\omega}}{\|v\|^2_{L^2_\rho}} - 1\right|,$$

and, by the continuity of $\tau_{M,\omega}(v)$ in $\mathbb{P}_\Lambda$,

$$\tau_{M,\omega}(v) \overset{\text{unif.}}{\longrightarrow} 1, \quad \forall\, v \in \mathbb{P}_\Lambda \setminus \{v \equiv 0\}, \tag{1.2.22}$$

from which we also deduce that

$$\Big(\tau_{M,\omega}(v)\Big)^{-1} \overset{\text{unif.}}{\longrightarrow} 1, \quad \forall\, v \in \mathbb{P}_\Lambda \setminus \{v \equiv 0\}.$$

But this implies

$$\lim_{M\to+\infty}\sup_{v\in\mathbb{P}_\Lambda\setminus\{v\equiv 0\}}\left|\frac{\|v\|^2_{L^2_\rho}}{\|v\|^2_{M,\omega}} - 1\right| = 0,$$

which is the thesis. □

1.3 Error analysis in one dimension for the uniform distribution

We restrict the analysis to the case $N = 1$ and consider the polynomial space $\mathbb{P}_\Lambda = \text{span}\{y^p,\ p = 0, \dots, \mathrm{w}\}$. For convenience, we rename the polynomial space as $\mathbb{P}_\mathrm{w}$ and the random discrete projector as $\mathit{\Pi}_M^{\mathrm{w},\omega}$. The main result of this section is Theorem 1.2. Its proof is postponed until the end of the section as we will need several lemmas and intermediate results.

Theorem 1.2. *For any $\alpha \in (0,1)$, under the condition*

$$\frac{M}{3\log((M+1)/\alpha)} \geq 4\sqrt{3}\,\mathrm{w}^2 \tag{1.3.1}$$

holds

$$\mathbb{P}\left(\|\phi - \mathit{\Pi}_M^{\mathrm{w},\omega}\phi\|_{L^2_\rho} \leq \left(1 + \sqrt{3\log\frac{M+1}{\alpha}}\right) \inf_{v\in\mathbb{P}_\mathrm{w}} \|\phi - v\|_{L^\infty}\right) \geq 1 - \alpha. \tag{1.3.2}$$

The previous theorem states that, with confidence level $1-\alpha$, the discrete projection with random points is (nearly) optimal up to a logarithmic factor in M, provided M is large enough and satisfies the condition (1.3.1).

Now we proceed to derive some probabilistic results on the distribution of order statistics for the standard uniform distribution.

1.3.1 Useful results on order statistics of the uniform distribution

To study order statistics of the uniform distribution it is more convenient to consider standard uniform random variables in $[0,1]$ instead of $[-1,1]$. Therefore, we introduce a linear transformation $\mathcal{T} : [-1,1] \to [0,1]$ and define a new set of i.i.d. random variables

$$U_i = \mathcal{T}(Y_i) = \frac{Y_i + 1}{2}, \qquad i = 1, \dots, M.$$

Thus we continue working with a sample of M independent random variables from the standard uniform parent distribution,

$$U_1, \dots, U_M \overset{\text{i.i.d.}}{\sim} \mathbb{U}(0,1). \tag{1.3.3}$$

We know that the order statistics $U_{(1)} < U_{(2)} < \ldots < U_{(M)}$ are

$$U_{(k)} \sim \text{Beta}(k, M+1-k), \qquad k = 1, \dots, M,$$

where we recall that a $\text{Beta}(k, M+1-k)$ random variable has distribution

$$f(y) = \frac{M!}{(k-1)!(M-k)!} y^{k-1}(1-y)^{M-k}, \qquad y \in [0,1].$$

Let us define $\Xi_{(0)} = U_{(1)}$, $\Xi_{(M)} = 1 - U_{(M)}$, and $\Xi_{(k)} = U_{(k+1)} - U_{(k)}$ for $k = 1, \dots, M-1$. It can be shown that $\Xi_{(k)} \sim \text{Beta}(1, M)$ for each $k = 0, \dots, M$; see [DN03, page 14] where a more general result on the distribution of $U_j - U_i$ is proved, namely

$$U_j - U_i \sim \text{Beta}(j-i, M-j+i+1), \quad 1 \leq i < j \leq M.$$

In particular, the distribution is independent of the values i and j, and depends only on their difference $j - i$. The following result gives a bound on the probability distribution of $\max_{k=0,\dots,M} \Xi_{(k)}$.

Lemma 1.1. *For any $\alpha \in (0,1)$ and $M \in \mathbb{N}_+$*

$$Pr\left(\max_{k=0,\dots,M} \Xi_{(k)} > \frac{\log((M+1)/\alpha)}{M}\right) \leq \alpha.$$

Proof. Trivially, if $0 < \alpha \le \dfrac{M+1}{\exp(M)}$, then $\dfrac{\log\left((M+1)/\alpha\right)}{M} \ge 1$ and

$$Pr\left(\max_{k=0,\dots,M} \Xi_{(k)} > 1 \right) = 0 < \alpha.$$

Consider now $\dfrac{M+1}{\exp(M)} < \alpha < 1$. Rewriting the random event

$$\Big\{ \max_{k=0,\dots,M} \Xi_{(k)} > \delta \Big\} = \bigcup_{k=0}^{M} \Big\{ \Xi_{(k)} > \delta \Big\},$$

we have, for $0 < \delta < 1$,

$$\begin{aligned} Pr\Big(\max_{k=0,\dots,M} \Xi_{(k)} > \delta \Big) &= Pr\Big(\bigcup_{k=0}^{M} \Big\{ \Xi_{(k)} > \delta \Big\} \Big) \\ &\le \sum_{k=0}^{M} Pr\Big(\Xi_{(k)} > \delta \Big) = (M+1)(1-\delta)^M. \end{aligned} \tag{1.3.4}$$

Therefore

$$\begin{aligned} Pr\left(\max_{k=0,\dots,M} \Xi_{(k)} > \frac{\log((M+1)/\alpha)}{M} \right) &\le (M+1)\left(1 - \frac{\log((M+1)/\alpha)}{M} \right)^M \\ &= (M+1)\left(\left(1 - \frac{\log((M+1)/\alpha)}{M} \right)^{\frac{M}{\log((M+1)/\alpha)}} \right)^{\log((M+1)/\alpha)} \\ &\le (M+1)\, e^{-\log((M+1)/\alpha)} = \alpha. \end{aligned}$$

□

1.3.2 Relation between $\|\cdot\|_{L^2_\rho}$ and $\|\cdot\|_{M,\omega}$ on $\mathbb{P}_{\mathrm{w}}([-1,1])$

Here we go back to the uniform distribution in $\Gamma = [-1,1]$. We are interested in finding an inequality between the continuous norm $\|\cdot\|_{L^2_\rho}$ and the discrete norm $\|\cdot\|_{M,\omega}$, *i.e.* finding a suitable $C_M^{\mathrm{w},\omega}$ such that

$$\|v\|^2_{L^2_\rho} \le C_M^{\mathrm{w},\omega} \|v\|^2_{M,\omega}, \qquad \forall\, v \in \mathbb{P}_{\mathrm{w}}(\Gamma). \tag{1.3.5}$$

This will be the random constant appearing in Proposition 1.1. We will also need an inverse inequality for the derivative (see e.g. [CQ82]),

$$\left\| \frac{\partial v}{\partial y} \right\|_{L^2(\Gamma)} \le \widehat{C}\, \mathrm{w}^2 \|v\|_{L^2(\Gamma)}, \qquad \forall\, v \in \mathbb{P}_{\mathrm{w}}(\Gamma). \tag{1.3.6}$$

In this case $\widehat{C} = \sqrt{3}$, while in general $\widehat{C}(\Gamma)$ depends on the length of Γ. The same inequality holds replacing $L^2(\Gamma)$ with L^2_ρ.

The sampled points $\{y_j\}_{j=1}^M$ are distinct almost surely. To each point y_j we associate an open interval I_j satisfying the requirement that

$$\left(\bigcup_{j=1}^{M} \overline{I_j} \right) \cap \Gamma = \Gamma. \tag{1.3.7}$$

In other words, the family of intervals $\{I_j\}_j$ is a (finite) covering of the domain Γ. We order the points in increasing order

$$-1 \le y_1 < \ldots < y_M \le 1,$$

keeping the notation $\{y_j\}$; it will be clear from the context if we refer to the ordered points or to the originally sampled points.

In one dimension, it is easy to build a finite covering of mutually disjoint intervals

$$I_j = \Big(y_j - \Delta y_{j-1}, y_j + \Delta y_j\Big), \qquad j = 1, \ldots, M, \tag{1.3.8}$$

taking

$$\Delta y_j = \begin{cases} |-1-y_1|, & j = 0, \\ \dfrac{|y_{j+1} - y_j|}{2}, & j = 1, \ldots, M-1, \\ |1 - y_M|, & j = M. \end{cases} \tag{1.3.9}$$

In general, the sets I_j of this covering are not centered in the sample points. It is useful to split each interval I_j in its left and right part,

$$I_j^- = I_j \cap \{y \in \mathbb{R} : y \le y_j\}, \qquad I_j^+ = I_j \cap \{y \in \mathbb{R} : y \ge y_j\},$$

with measures $|I_j^-| = \Delta y_{j-1}$ and $|I_j^+| = \Delta y_j$, respectively.

We also define the random variable ΔY whose realizations are

$$\Delta y = \max_{j=1,\ldots,M} |I_j|. \tag{1.3.10}$$

In Appendix C we report its cumulative distribution function simulated numerically, for different values of the number of sampling points M.

The link between the random variable ΔY and the result given in the previous section is the following:

Lemma 1.2 (Corollary of Lemma 1.1)**.** *For any $\alpha \in (0,1)$ and $M \in \mathbb{N}_+$ s.t. $M \ge 2$*

$$Pr\left(\Delta Y > \frac{3\log((M+1)/\alpha)}{M}\right) \le \alpha.$$

Proof. It is enough to notice that, for each realization $(\xi_0, \ldots, \xi_M)$ of the random variables $(\Xi_{(0)}, \ldots, \Xi_{(M)})$, it holds

$$\max_{k=0,\ldots,M} \xi_k < \delta \qquad \Longrightarrow \qquad \Delta y < 3\delta,$$

and thus

$$Pr\left(\Delta Y > 3\delta\right) \le Pr\left(\max_{k=0,\ldots,M} \Xi_{(k)} > \delta\right).$$

□

We now define two events:

$$\Sigma_{\mathrm{w}} = \left\{\Delta Y \le \frac{1}{4\widehat{C}\mathrm{w}^2}\right\} \qquad \text{and} \qquad \Sigma_M = \left\{\Delta Y \le \frac{3\log((M+1)/\alpha)}{M}\right\}. \tag{1.3.11}$$

Observe that, taking M large enough, the probability of Σ_{w} can be made arbitrary close to 1. More precisely

Lemma 1.3. *For any $\alpha \in (0,1)$, under the condition*

$$\frac{M}{3\,\log\Big((M+1)/\alpha\Big)} \geq 4\,\widehat{C}\,\mathrm{w}^2, \tag{1.3.12}$$

the following inclusion holds

$$\Sigma_M \subset \Sigma_{\mathrm{w}}$$

and

$$Pr(\Sigma_{\mathrm{w}}) \geq Pr(\Sigma_M) \geq 1-\alpha. \tag{1.3.13}$$

Proof. Clearly, under (1.3.12)

$$\Delta Y \leq \frac{3\log((M+1)/\alpha)}{M} \quad \Longrightarrow \quad \Delta Y \leq \frac{1}{4\widehat{C}\mathrm{w}^2}$$

and the inclusion $\Sigma_M \subset \Sigma_{\mathrm{w}}$ is proved. The bound from below on the corresponding probabilities is an immediate consequence of Lemma 1.2. □

Now we are ready to formulate the main result of this subsection:

Theorem 1.3. *Define on Σ_{w} the random variable*

$$C_M^{\mathrm{w},\omega} = \frac{M\Delta Y/2}{1-2\Delta Y\,\widehat{C}\,\mathrm{w}^2}. \tag{1.3.14}$$

Then, in Σ_{w} it holds

$$\|v\|_{L^2_\rho}^2 \leq C_M^{\mathrm{w},\omega}\|v\|_{M,\omega}^2, \qquad \forall\, v \in \mathbb{P}_{\mathrm{w}}(\Gamma). \tag{1.3.15}$$

Moreover, under condition (1.3.12)*, in $\Sigma_M \subset \Sigma_{\mathrm{w}}$ it holds*

$$C_M^{\mathrm{w},\omega} \leq 3\log((M+1)/\alpha).$$

Proof. For convenience the proof uses the standard $L^2(\Gamma)$ norm instead of the weighted norm L^2_ρ. Remember that in this case $\|\cdot\|_{L^2_\rho}^2 = \frac{1}{2}\|\cdot\|_{L^2(\Gamma)}^2$. To lighten the notation we also introduce on each interval $I_j^\pm$ the short notation

$$\|\cdot\|_{I_j^-} := \|\cdot\|_{L^2(I_j^-)},$$
$$\|\cdot\|_{I_j^+} := \|\cdot\|_{L^2(I_j^+)}.$$

Take $v \in \mathbb{P}_{\mathrm{w}}(\Gamma)$ arbitrarily. The following standard inequalities are used in the sequel:

$$\begin{aligned}\Big|v(y)^2 - v(y_j)^2\Big| &= \left|\int_{y_j}^{y}(v^2)'(\xi)d\xi\right| = \left|\int_{y_j}^{y} 2v(\xi)v(\xi)'d\xi\right| \\ &\leq \begin{cases} 2\|v\|_{I_j^+}\|v'\|_{I_j^+}, & \forall y \in I_j^+, \\ 2\|v\|_{I_j^-}\|v'\|_{I_j^-}, & \forall y \in I_j^-. \end{cases}\end{aligned} \tag{1.3.16}$$

Integrating (1.3.16) over I_j^- and I_j^+ yields

$$\begin{aligned}\int_{I_j^-} v(y)^2dy - v(y_j)^2|I_j^-| &= \int_{I_j^-}\Big(v(y)^2 - v(y_j)^2\Big)\,dy \\ &\leq \int_{I_j^-}\Big|v(y)^2 - v(y_j)^2\Big|dy \\ &\leq 2\,|I_j^-|\,\|v\|_{I_j^-}\|v'\|_{I_j^-},\end{aligned} \tag{1.3.17}$$

and similarly for I_j^+

$$\int_{I_j^+} v(y)^2 dy - v(y_j)^2 |I_j^+| \leq 2\,|I_j^+|\,\|v\|_{I_j^+}\|v'\|_{I_j^+}. \tag{1.3.18}$$

Summing (1.3.17) and (1.3.18) we get

$$\begin{aligned}\int_{I_j} v(y)^2 dy - |I_j| v(y_j)^2 &\leq 2\Big(|I_j^-|\,\|v\|_{I_j^-}\|v'\|_{I_j^-} + |I_j^+|\,\|v\|_{I_j^+}\|v'\|_{I_j^+}\Big) \\ &\leq 2\,|I_j|\,\|v\|_{I_j}\|v'\|_{I_j}, \end{aligned}\tag{1.3.19}$$

which implies

$$\int_{I_j} v(y)^2 dy \leq |I_j|\Big(v(y_j)^2 + 2\,\|v\|_{I_j}\|v'\|_{I_j}\Big). \tag{1.3.20}$$

Summing over all the intervals we have

$$\begin{aligned}
\|v\|_{L^2}^2 &= \sum_{j=1}^M \int_{I_j} v(y)^2 dy && \Big[\text{substitute (1.3.20)}\Big] \\
&\leq \sum_{j=1}^M |I_j| v(y_j)^2 + 2\sum_{j=1}^M |I_j|\,\|v\|_{I_j}\|v'\|_{I_j} && \Big[\text{using (1.3.10)}\Big] \\
&\leq \Delta y \sum_{j=1}^M v(y_j)^2 + 2\Delta y \sum_{j=1}^M \|v\|_{I_j}\|v'\|_{I_j} && \Big[\text{definition of } \|\cdot\|_{M,\omega}\Big] \\
&= M\Delta y \|v\|_{M,\omega}^2 + 2\Delta y \sum_{j=1}^M \|v\|_{I_j}\|v'\|_{I_j} && \Big[\text{Cauchy-Schwarz's ineq.}\Big] \\
&\leq M\Delta y \|v\|_{M,\omega}^2 + 2\Delta y\,\|v\|_{L^2(\Gamma)}\|v'\|_{L^2(\Gamma)} && \Big[\text{using (1.3.6)}\Big] \\
&\leq M\Delta y \|v\|_{M,\omega}^2 + 2\Delta y\,\widehat{C}\,\mathrm{w}^2\,\|v\|_{L^2(\Gamma)}^2. && (1.3.21)
\end{aligned}$$

Rearranging the terms in (1.3.21) we obtain

$$(1 - 2\Delta y\,\widehat{C}\,\mathrm{w}^2)\|v\|_{L^2(\Gamma)}^2 \leq M\Delta y \|v\|_{M,\omega}^2. \tag{1.3.22}$$

The coefficient in the left hand side is non zero on Σ_{w} so we have proved (1.3.14)-(1.3.15). If we now restrict to Σ_M under condition (1.3.12), then from Lemma 1.3 we have

$$\Delta Y \leq \frac{1}{4\widehat{C}\mathrm{w}^2} \quad \text{and} \quad \Delta Y \leq \frac{3\log((M+1)/\alpha)}{M}$$

so that

$$C_M^{\mathrm{w},\omega} = \frac{M\Delta Y/2}{1 - 2\Delta Y\,\widehat{C}\,\mathrm{w}^2} \leq 3\log((M+1)/\alpha)$$

and this concludes the proof. □

Remark 1.2. *Theorem 1.3 states that on Σ_M, which is an event of probability at least $1-\alpha$, the discrete and continuous norms are equivalent up to a logarithmic factor if condition* (1.3.12) *is fulfilled, which, roughly speaking corresponds to $M \propto \mathrm{w}^2$.*

1.3.3 Proof of Theorem 1.2

The proof of this theorem is merely a collection of results from Proposition 1.1, Theorem 1.3, and Lemma 1.2.

Proof of Theorem 1.2. We consider the event Σ_M defined in (1.3.11). From Lemma 1.2 we know that under condition (1.3.1) the probability of this event is at least $1-\alpha$. From Theorem 1.3 it holds

$$\|v\|_{L^2_\rho}^2 \leq C_M^{\mathrm{w},\omega} \|v\|_{M,\omega}^2 \quad \text{in } \Sigma_M$$

uniformly with respect to $v \in \mathbb{P}_{\mathrm{w}}(\Gamma)$ and the constant $C_M^{\mathrm{w},\omega} \leq 3\log((M+1)/\alpha)$ *uniformly* in Σ_M. We then apply Proposition 1.1 in Σ_M to conclude that

$$\|\phi - \mathit{\Pi}_M^{\mathrm{w},\omega}\phi\|_{L^2_\rho} \leq \left(1+\sqrt{C_M^{\mathrm{w},\omega}}\right) \inf_{v\in\mathbb{P}_{\mathrm{w}}} \|\phi - v\|_{L^\infty}, \quad \forall \phi \in L^\infty(\Gamma), \quad \forall \omega \in \Sigma_M.$$

Since under condition (1.3.1) the probability of Σ_M is at least $1-\alpha$, this concludes the proof. □

1.4 Algebraic formulation

The value of $\#\Lambda$ depends on the particular polynomial space (TP, TD, HC, ...), the maximal polynomial degree used, w, and the dimension of the physical space, N. The number of points M must satisfy the constraint

$$M \geq \#\Lambda,$$

to have an overdetermined problem (more data than unknowns). We have shown in Section 1.2 that, for univariate functions, M should scale as $M \propto \mathrm{w}^2$ to have a stable discrete projection. As a general rule to choose M for multivariate functions and arbitrary polynomial spaces we consider the formula

$$M = c\,(\#\Lambda)^\alpha, \tag{1.4.1}$$

where c is a positive constant, and $\alpha \geq 1$ a real number. We restrict our numerical tests in Section 1.5 to $1 \leq \alpha \leq 2$.

Given the polynomial space, we define the design matrix $D^\omega \in \mathbb{R}^{M\times\#\Lambda}$. The element $D^\omega_{i,j}$ contains the jth L^2_ρ-orthonormal basis function l_j evaluated in the ith sample point $\mathbf{y}_i$, that is

$$[D^\omega]_{i,j} = l_j(\mathbf{y}_i). \tag{1.4.2}$$

The discrete random projection $\mathit{\Pi}_M^{\Lambda,\omega}\phi$ can be expressed in terms of the orthonormal basis $\{l_j\}_j$ as

$$\mathit{\Pi}_M^{\Lambda,\omega}\phi(\mathbf{Y}) = \sum_{j=1}^{\#\Lambda} x_j^\omega l_j(\mathbf{Y}). \tag{1.4.3}$$

Then the algebraic problem to determine the unknown coefficients $\{x_j^\omega\}$ can be formulated as:

$$x^\omega = \operatorname*{argmin}_{x\in\mathbb{R}^{\#\Lambda}} \|D^\omega x - b^\omega\|_2 \tag{1.4.4}$$

where $b^\omega \in \mathbb{R}^{M\times 1}$ contains the evaluations of the target function ϕ in the M sample points: $b^\omega(i) = \phi(\mathbf{y}_i)$. The Normal Equations allow us to rewrite the rectangular system embedded in (1.4.4) as a square system:

$$(D^\omega)^T D^\omega x^\omega = (D^\omega)^T b^\omega. \tag{1.4.5}$$

We use problem (1.4.4) to calculate the approximation error, computing its solution x^ω by QR factorization. On the other hand, formulation (1.4.5) will be useful to measure the ill-conditioning of the problem, through the condition number of the matrix $(D^\omega)^T D^\omega$. Alternatively, one can also solve (1.4.5) by Cholesky factorization.

To evaluate the approximation error, we have considered a cross-validation approach: a random set of 100 cross-validating points is chosen at the beginning, and the corresponding design matrix D_{cv} is computed. The evaluations of ϕ in these points are stored in b_{cv}. Then the cross-validated error in ∞-norm is defined as

$$\|\phi - \mathit{\Pi}_M^{\Lambda,\omega}\phi\|_{cv} = \|D_{cv}x^\omega - b_{cv}\|_\infty. \tag{1.4.6}$$

Note that $\|\cdot\|_{cv}$ is not a norm on the function space of ϕ; we abuse the norm notation in the figures with cross-validation errors below to emphasize the dependence on ϕ. To estimate the variability of (1.4.6) due to the random sampling of the M collocation points, we have repeated the calculation over R independent sets of points $\{\mathbf{y}_j^{\omega_k},\ j = 1,\dots,M\}$, with $k = 1,\dots,R$ and we have computed the average error

$$\overline{E}_{cv} = \frac{\sum_{k=1}^R \|D_{cv}x^{\omega_k} - b_{cv}\|_\infty}{R}, \tag{1.4.7}$$

as well as the sample standard deviation by

$$s_E = \sqrt{\frac{1}{R-1}\sum_{k=1}^R \left(\|D_{cv}x^{\omega_k} - b_{cv}\|_\infty - \overline{E}_{cv}\right)^2}. \tag{1.4.8}$$

We also aim to analyze the condition number of $(D^\omega)^T D^\omega$,

$$\mathrm{cond}\Big((D^\omega)^T D^\omega\Big) = \frac{\sigma_{\max}\Big((D^\omega)^T D^\omega\Big)}{\sigma_{\min}\Big((D^\omega)^T D^\omega\Big)}, \tag{1.4.9}$$

where $\sigma_{\max}(\cdot)$ and $\sigma_{\min}(\cdot)$ are the maximum and minimum singular values. Again, denoting by D^{ω_k} the design matrix built with the k-th set $\{\mathbf{y}_j^{\omega_k}\}_j$, we estimate the mean condition number $\overline{K}$ over the R repetitions as

$$\overline{K} = \frac{\sum_{k=1}^R \mathrm{cond}\Big((D^{\omega_k})^T D^{\omega_k}\Big)}{R}, \tag{1.4.10}$$

and the standard deviation as

$$s_K = \sqrt{\frac{1}{R-1}\sum_{k=1}^R \left(\mathrm{cond}\Big((D^{\omega_k})^T D^{\omega_k}\Big) - \overline{K}\right)^2}. \tag{1.4.11}$$

In the rest of the section we show how the condition number of problem (1.4.5) relates to some quantities that already appeared previously. All the contents of this section hold for a generic polynomial space, in any dimension. Accordingly, we refer to the polynomial space as $\mathbb{P}_\Lambda$.

In addition to the constant $C^\omega(M,\Lambda)$ already introduced in (1.2.15), we define the constant $c^\omega(M,\Lambda)$ as

$$c^\omega(M,\Lambda) = \sup_{\varphi\in\mathbb{P}_\Lambda\setminus\{\varphi\equiv 0\}} \frac{\|\varphi\|_{M,\omega}^2}{\|\varphi\|_{L^2_\rho}^2}. \tag{1.4.12}$$

Proposition 1.4. *The spectral condition number (2-norm) of the matrix $(D^\omega)^T D^\omega$, as defined in* (1.4.9) *is equal to*

$$K\Big((D^\omega)^T D^\omega\Big) = c^\omega(M,\Lambda)\, C^\omega(M,\Lambda). \tag{1.4.13}$$

Proof. Each realization of the random matrix $(D^\omega)^T D^\omega$ is almost surely a symmetric and positive definite matrix under Assumption 1.1 on ρ.

The ith element of the vector $D^\omega x$ is

$$[D^\omega x^\omega]_i = \sum_{j=1}^{\#\Lambda} l_j(y_i) x_j^\omega = \Pi_M^{\Lambda,\omega}\phi(y_i). \tag{1.4.14}$$

By definition (1.2.6) of the random discrete inner product

$$\|\Pi_M^{\Lambda,\omega}\phi\|_{M,\omega}^2 = \frac{1}{M}\sum_{i=1}^{M}\Big(\Pi_M^{\Lambda,\omega}\phi(y_i)\Big)^2 \tag{1.4.15}$$

and by (1.4.3) and the L^2_ρ-orthonormality of $\{l_j\}$

$$(x^\omega)^T x^\omega = \sum_{j=1}^{\#\Lambda}(x_j^\omega)^2 = \|\Pi_M^{\Lambda,\omega}\phi\|_{L^2_\rho}^2. \tag{1.4.16}$$

Using in sequence (1.4.16), (1.4.14) and (1.4.15), yields

$$\frac{(x^\omega)^T(D^\omega)^T D^\omega x^\omega}{(x^\omega)^T x^\omega} = \frac{\sum_{i=1}^M (\Pi_M^{\Lambda,\omega}\phi(y_i))^2}{\|\Pi_M^{\Lambda,\omega}\phi\|_{L^2_\rho}^2} = \frac{M\|\Pi_M^{\Lambda,\omega}\phi\|_{M,\omega}^2}{\|\Pi_M^{\Lambda,\omega}\phi\|_{L^2_\rho}^2}. \tag{1.4.17}$$

This shows that (1.4.17) is the Rayleigh quotient of the matrix $(D^\omega)^T D^\omega$. So the largest and smallest eigenvalues of $(D^\omega)^T D^\omega$ correspond to $c^\omega(M,\Lambda)$ and $C^\omega(M,\Lambda)^{-1}$, respectively. Since $(D^\omega)^T D^\omega$ is symmetric positive definite, its singular values are also eigenvalues. The conclusion follows from the definition (1.4.9) of the condition number. □

Remark 1.3. *In 1D and for the uniform distribution we can easily establish a norms equivalence between $\|\cdot\|_{M,\omega}$ and $\|\cdot\|_{L^2_\rho}$, collecting the results of Theorem 1.3 and Remark 1.1. Namely, under the condition* (1.3.1)*, with probability $1-\alpha$ we have*

$$\frac{1}{1+\mathrm{w}}\|v\|_{M,\omega}^2 \le \|v\|_{L^2_\rho}^2 \le 3\log\left(\frac{M+1}{\alpha}\right)\|v\|_{M,\omega}^2,$$

from which we get the bound on the condition number

$$cond\Big((D^\omega)^T D^\omega\Big) \le \frac{\log((M+1)/\alpha)}{\mathrm{w}+1}, \qquad in\ \Sigma_M. \tag{1.4.18}$$

where Σ_M is the event defined in (1.3.11) *that has probability at least $1-\alpha$. However, we have observed numerically that the bound* (1.4.18) *is very pessimistic as under condition* (1.3.1) *the condition number seems to be uniformly bounded with respect to M and* w.

A direct consequence of Theorem 1.1 is that

$$\mathrm{cond}\Big((D^\omega)^T D^\omega\Big) \overset{M\to+\infty}{\longrightarrow} 1, \quad \text{a.s.}$$

This is confirmed numerically. Fig. 1.2 shows the numerical results obtained for a 1D problem with an overkilling rule $M = 100\cdot(\#\Lambda)^4$, to simulate the asymptotic case $M \uparrow \infty$.

1.5 Numerical results

We present an illustrative selection of results from an extensive set of numerical tests. The aim is to seek the correct relation between the number of points to sample M and the dimension of the polynomial space to have a stable and accurate approximation. The following issues are addressed:

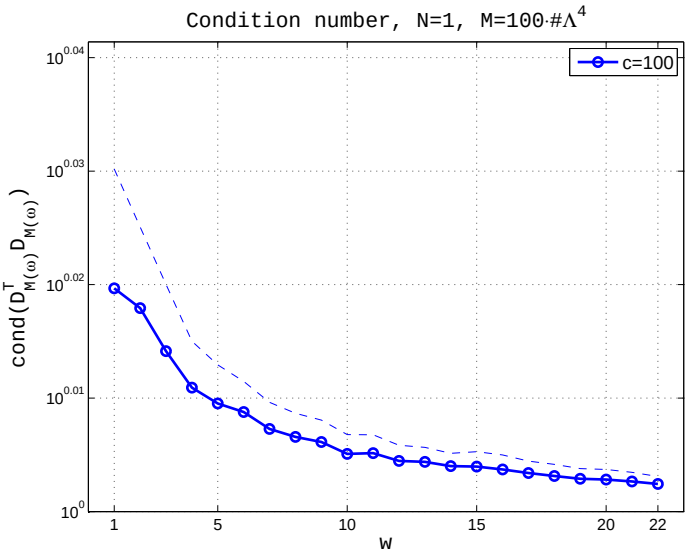


Figure 1.2: Condition number (1.4.9), $N = 1$, $M = 100 \cdot (\#\Lambda)^4$. The continuous, marked, lines show the mean condition number (1.4.10) over 200 repetitions. The dashed lines show the mean (1.4.10) plus one standard deviation (1.4.11). The scale on the y-axis ranges from $10^0 = 1$ to $10^{0.04} = 1.0965$

- how the condition number (1.4.9) depends on w, N, c, α and the choice of the polynomial space;
- analogously, how the cross-validation error (1.4.6) behaves.

In the convergence plots presented in the rest of this section we show the average error (1.4.7) and condition number (1.4.10) as well as their average plus one standard deviation.

1.5.1 The one-dimensional case

We first investigate the dependence of the condition number (1.4.9) on the rule (1.4.1) used to select the number $M(\mathrm{w})$ of sampling points. Observe that in the one-dimensional case $\#\Lambda = \mathrm{w}+1$.

As seen in Fig. 1.3, the condition number behaves differently depending on the rule chosen. In Fig. 1.3-Left we report results obtained with the linear rule $M = c \cdot \#\Lambda$, corresponding to $\alpha = 1$ in (1.4.1). We tested several values for c ranging from 2 to 20. All cases show an exponential growth of the condition number with respect to w, with rates decreasing with increasing c (as one would expect). Using $R = 10000$ repetitions the observed average condition number still shows a large variability. This is due to the large standard deviations of the condition number, as indicated in the figure in dashed line.

Note that the range of w goes up to 25, so in this range the choice of the largest c yields a linear rule which uses more sample points than some of the quadratic rules (shown in Fig. 1.3-Right).

In contrast to the exponential growth observed when using the linear rule, the results using the quadratic rule exhibit a condition number which is approximately constant for w ranging from 1 to 40. Fluctuations become smaller when c increases. This behaviour is consistent with the theoretical results in Section 1.3.

We now proceed to illustrate the convergence of the error for few functions of varying regularity in Figs. 1.4–1.7.

We focus on three target functions: an exponential function

$$\phi(Y) = \exp(Y), \qquad Y \in [-1, 1], \tag{1.5.1}$$

a meromorphic function

$$\phi(Y) = \frac{1}{1 + \beta Y}, \qquad Y \in [-1, 1], \tag{1.5.2}$$

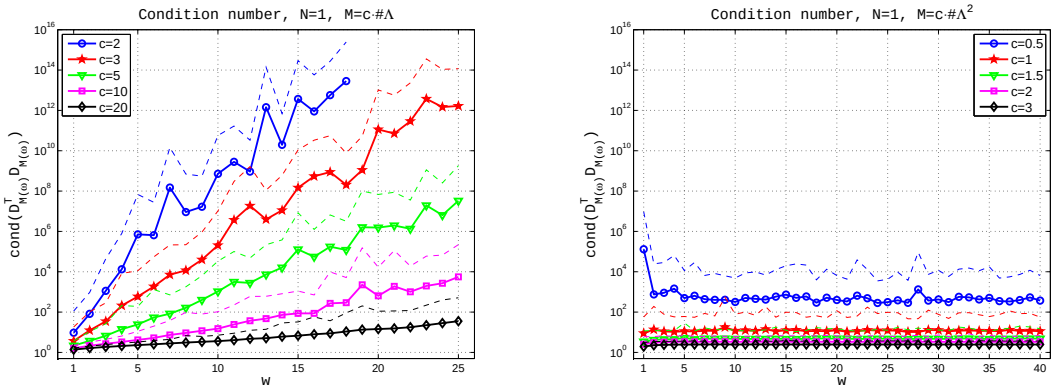


Figure 1.3: Condition number (1.4.9), $N = 1$. The continuous, marked, lines show the mean condition number (1.4.10) over 10000 repetitions. The dashed lines show the mean (1.4.10) plus one standard deviation (1.4.11). Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

that is a function which is analytic provided that $\beta \in (-1, 1)$, and a function with lower regularity

$$\phi(Y) = |Y|^3. \tag{1.5.3}$$

Fig. 1.4 shows the error computed as in (1.4.6), in approximating the exponential function (1.5.1) with different choices of c and α in the rule (1.4.1). The quadratic rule (to the right) displays the same exponential, optimal, convergence with respect to w independently of the constant c. The convergence is up to the machine precision.

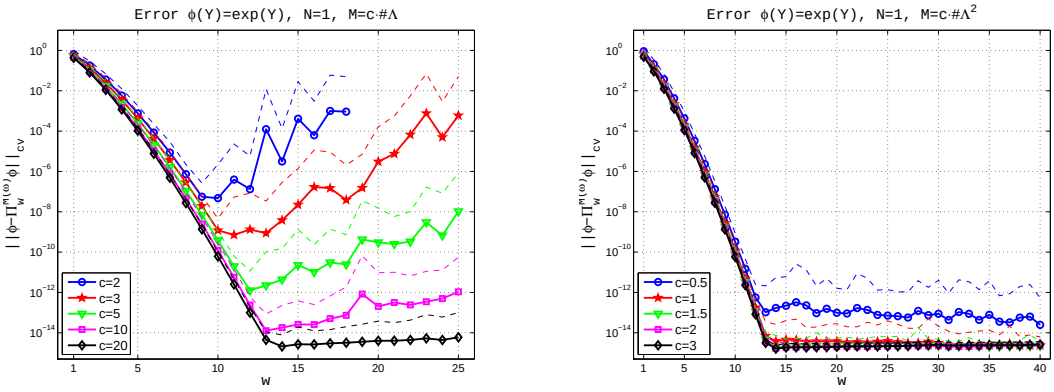


Figure 1.4: Error (1.4.6) for the function (1.5.1). The continuous marked lines show the mean error (1.4.7) over 10000 repetitions. The dashed lines show the mean (1.4.7) plus one standard deviation (1.4.8). Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

In contrast, the linear rule (on the left) displays a deterioration of the convergence using small values of c. The deterioration is due to the ill-conditioning of the matrix $D^T_{M(\omega)} D_{M(\omega)}$ when w increases. As noted before, the largest value of c yields at least as many sample points as the quadratic rule with the smallest value of c in the shown range and the errors behave accordingly. Again the fluctuations in the average error decrease with increasing c.

The use of the meromorphic function (1.5.2) with $\beta = 0.5$ (Fig. 1.5) and $\beta = 0.9$ (Fig. 1.6) yields analogous error graphs, but with a slower convergence rate. Unlike the function (1.5.2), which is analytic in $[-1, 1]$, the function (1.5.3) is only in $C^2([-1, 1])$, but not in $C^3([-1, 1])$. This decreased regularity manifests in the slower decay of the approximation error in Fig. 1.7. Note that the dependence of the error on the polynomial degree w is displayed in log-log scale, so that the error no longer decreases exponentially with respect to w.

When taking the number of sample points according to the quadratic rule (Fig. 1.7-Right), the error decreases like w^{-3}, and in this range of w the error shows no tendency to blow up for any of the studied values of c.

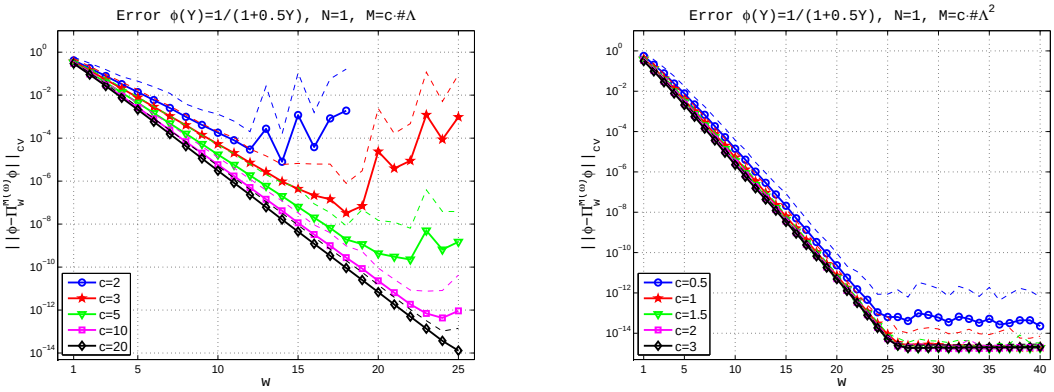


Figure 1.5: Error (1.4.6) for the function (1.5.2) with $\beta = 0.5$. The continuous marked lines show the mean error (1.4.7) over 10000 repetitions. The dashed lines show the mean (1.4.7) plus one standard deviation (1.4.8). Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

On the other hand, using the linear rule (Fig. 1.7-Left) yields a deterioration: the critical w, above which the error starts to grow, increases with increasing c.

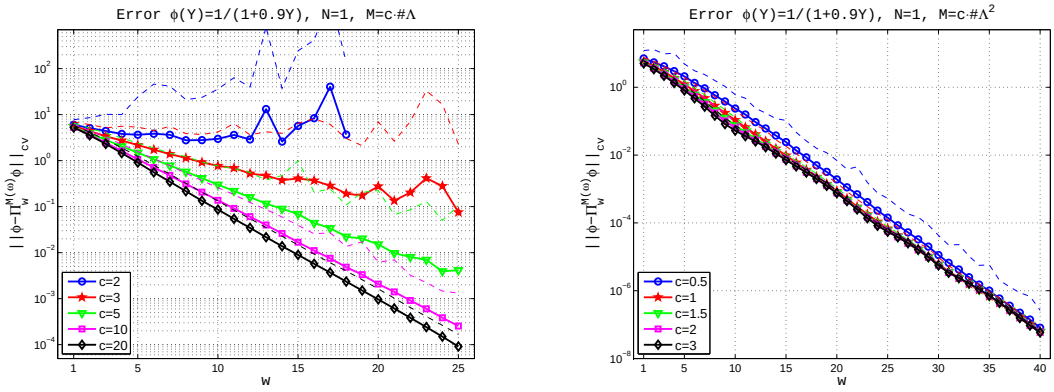


Figure 1.6: Error (1.4.6) for the function (1.5.2) with $\beta = 0.9$. The continuous marked lines show the mean error (1.4.7) over 10000 repetitions. The dashed lines show the mean (1.4.7) plus one standard deviation (1.4.8). Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

Note, in particular, that sooner or later the error starts to blow up for all shown constants. This is a clear indication that the linear rule does not lead to a stable and convergent approximation.

From a practical point of view, we are mainly interested in the error as a function of the computational work, not the polynomial degree itself. Fig. 1.8 shows how the error depends on the total number of sampling points, when we consider the function (1.5.2) with $\beta = 0.5$. Note that Fig. 1.8 shows the same errors as Fig. 1.5 but with M instead of w on abscissas.

In Fig. 1.8-Left we show the linear case: the error decays exponentially with increasing M in an initial phase, until the error starts to deteriorate. The convergence is faster for small values of c, but the deterioration also happens earlier, which prevents higher accuracies.

In Fig. 1.8-Right we show the quadratic case. In contrast to the linear case the convergence becomes subexponential with respect to M. On the other hand, all choices of $c \geq 1$ avoid the deterioration of the errors that we see using the linear rule, and the approximation remains stable and convergent. Fig. 1.9 compares the convergence of the error obtained with the linear and quadratic rules, with respect to M.

We remark that, even though the error deteriorates for high w when using the linear rule with a small c, we can still obtain an accuracy that suffices in many applications.

The plots in Fig. 1.8 show how the convergence speed is affected by the ratio between M and w. The convergence is fastest with the lowest constant $c = M/\mathrm{w}$ up to a certain point when blow up occurs.

A lower number of repetitions only turns out in an amplified variance in the results. In view

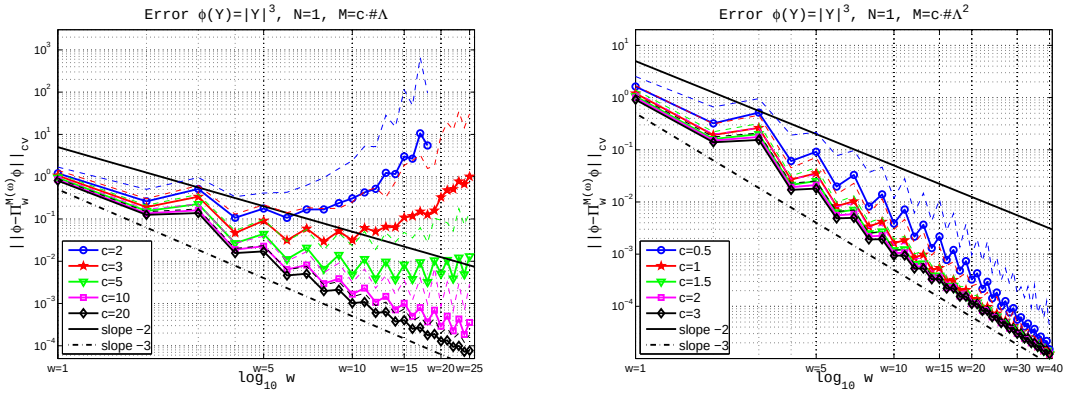


Figure 1.7: Error (1.4.6) for the function (1.5.3). The continuous marked lines show the mean error (1.4.7) over 10000 repetitions. The dashed lines show the mean (1.4.7) plus one standard deviation (1.4.8). Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

of the multiD section, where we choose $R = 100$, we report also in Fig. 1.10 the same graph of Fig. 1.3-Right, but with $R = 200$.

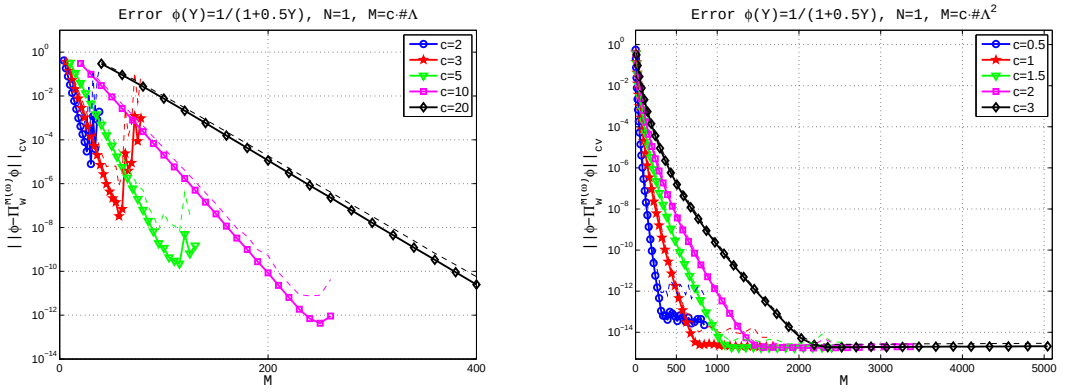


Figure 1.8: Dependence of the error (1.4.6) on the number of sample points M. The function is (1.5.2) with $\beta = 0.5$. The continuous marked lines show the mean error (1.4.7) over 10000 repetitions. The dashed lines show the mean (1.4.7) plus one standard deviation (1.4.8). Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

1.5.2 The multidimensional case

We now proceed to the multidimensional case, where we have an increased freedom to choose the space $\mathbb{P}_\Lambda(\Gamma)$. We will restrict our examples to isotropic versions of the TP, TD and HC spaces mentioned above. In this section we choose $R = 100$ repetitions to estimate the variability of the error and condition number. In the linear case, the values assumed by c are $1.1, 1.25, 2, 5, 20$. In the multidimensional case a constant c slightly larger than 1 is enough to have a good approximation. This is in contrast to the 1D case, where the linear rule with a constant $c = 2$ features a fast growth of the condition number and a large variability of the polynomial approximation.

Condition number

Fig. 1.11 shows the behaviour of the condition number for the 2D TP space. We see again an exponential growth of the condition number when M is chosen according to the linear rule. Note that the dimension of the PC space is equal to $(\mathrm{w} + 1)^2$ here.

As in the one-dimensional case, choosing the number of sample points M like $M \propto (\#\Lambda)^2$ yields condition numbers that are approximately constant in the studied range of w. Compared to the one-dimensional results of Fig. 1.3, the two-dimensional results exhibit a lower variability.

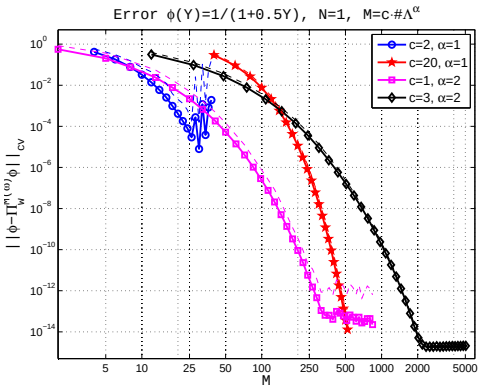


Figure 1.9: Dependence of the error (1.4.6) on the number of sample points M. Selected data from Fig. 1.8 corresponding to the rules $M = 2 \cdot \#\Lambda$, $M = 20 \cdot \#\Lambda$, $M = 1 \cdot (\#\Lambda)^2$, and $M = 3 \cdot (\#\Lambda)^2$.

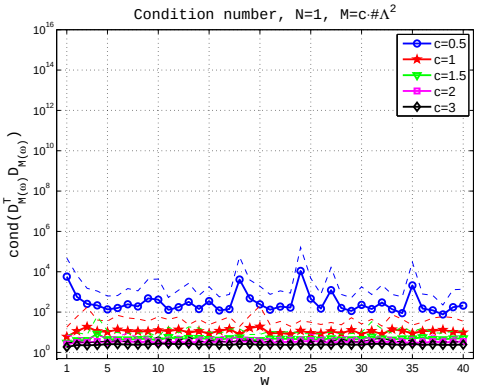


Figure 1.10: Condition number (1.4.9), $N = 1$. The continuous, marked, lines show the mean condition number (1.4.10) over 200 repetitions. The dashed lines show the mean (1.4.10) plus one standard deviation (1.4.11). $M = c \cdot (\#\Lambda)^2$.

Changing the polynomial space to TD we obtain Fig. 1.12, which looks similar to Fig. 1.11. The same holds with the HC space (Fig. 1.13). Therefore, the choice of the space does not seem to play a major role in the behaviour of the condition number (1.4.9). Note that the lowest value of c on the left of Fig. 1.13 is 1, instead of 1.1.

The situation is similar in higher dimensions; see Figs. 1.14–1.16. In addition, we observe a lower variability and a slower growth of the condition number with the linear rule, which, however, is still clearly exponential. Lastly, the HC space with the linear rule shows a very slow growth of the condition number also for very low values of c.

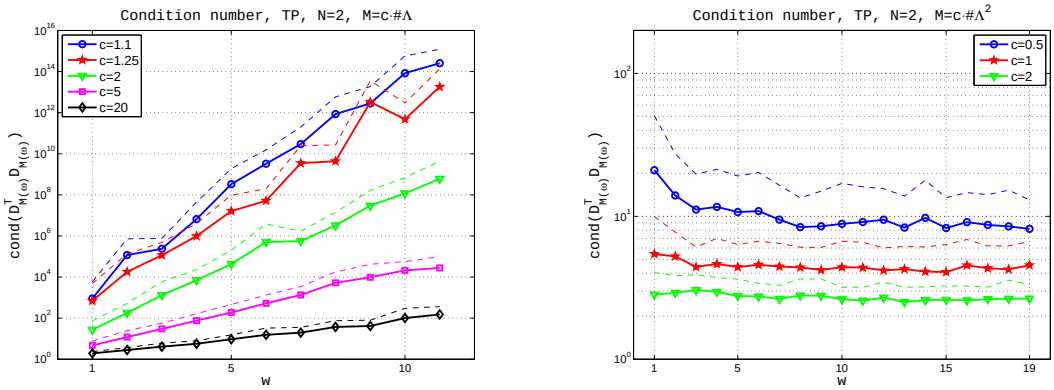


Figure 1.11: Condition number (1.4.9), TP, $N = 2$. The continuous, marked, lines show the mean condition number (1.4.10) over 100 repetitions. The dashed lines show the mean (1.4.10) plus one standard deviation (1.4.11). Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

Approximation error

Let us consider the error in approximating the target function

$$\phi(\mathbf{Y}) = \frac{1}{1 + \frac{\beta}{N} \sum_{i=1}^{N} Y_i}, \qquad \mathbf{Y} \in [-1, 1]^N, \tag{1.5.4}$$

which is a multidimensional generalization of (1.5.2), and inherits its regularity. We take $\beta = 0.5$ and start by considering the quadratic rule. Fig. 1.17 shows the optimal convergence rates. The TP space (Fig. 1.17-Left) seems to converge faster than the TD space (Fig. 1.17-Center),

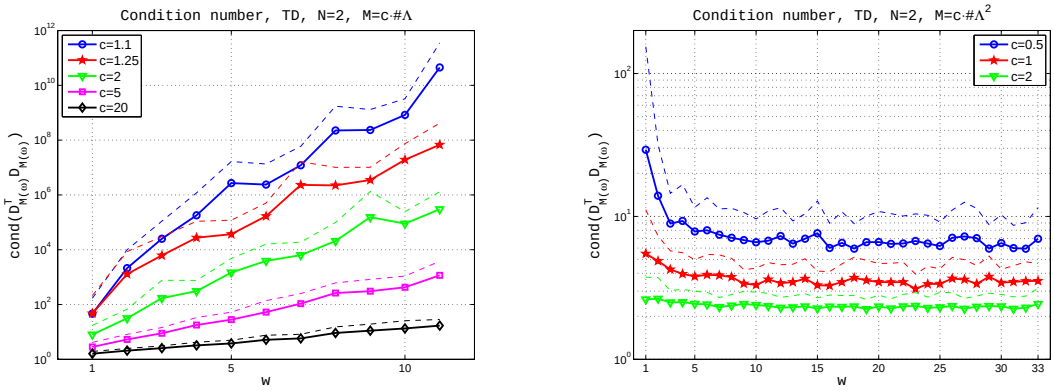


Figure 1.12: Condition number (1.4.9), TD, $N = 2$. The continuous, marked, lines show the mean condition number (1.4.10) over 100 repetitions. The dashed lines show the mean (1.4.10) plus one standard deviation (1.4.11). Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

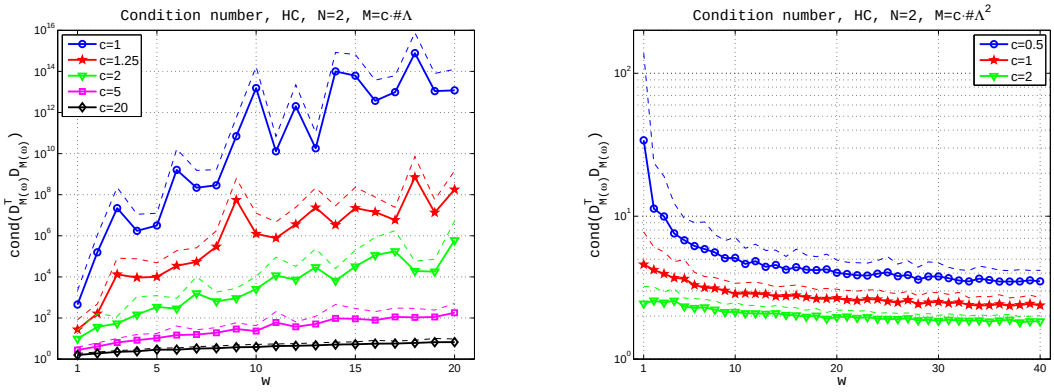


Figure 1.13: Condition number (1.4.9), HC, $N = 2$. The continuous, marked, lines show the mean condition number (1.4.10) over 100 repetitions. The dashed lines show the mean (1.4.10) plus one standard deviation (1.4.11). Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

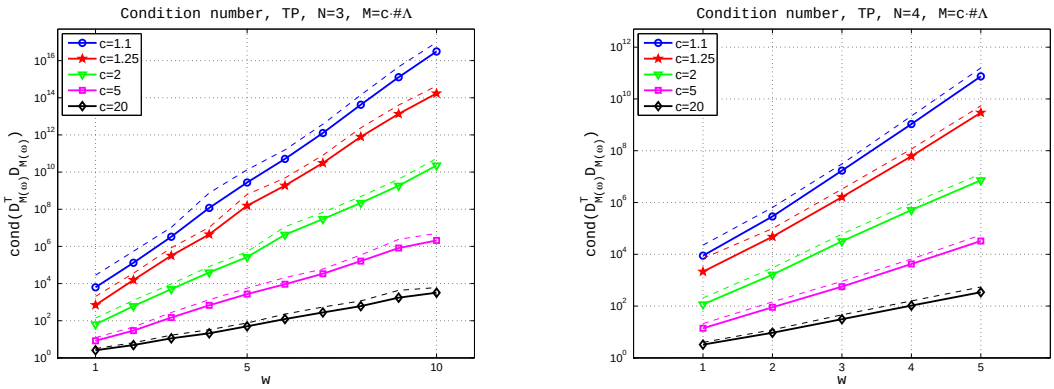


Figure 1.14: Condition number (1.4.9), TP, $M = c \cdot \#\Lambda$. The continuous, marked, lines show the mean condition number (1.4.10) over 100 repetitions. The dashed lines show the mean (1.4.10) plus one standard deviation (1.4.11). Left: $N = 3$. Right: $N = 4$.

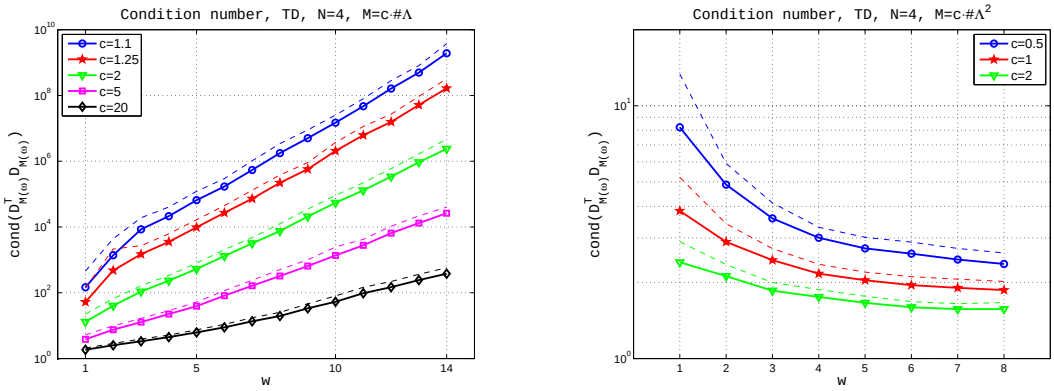


Figure 1.15: Condition number (1.4.9), TD, $N = 4$. The continuous, marked, lines show the mean condition number (1.4.10) over 100 repetitions. The dashed lines show the mean (1.4.10) plus one standard deviation (1.4.11). Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

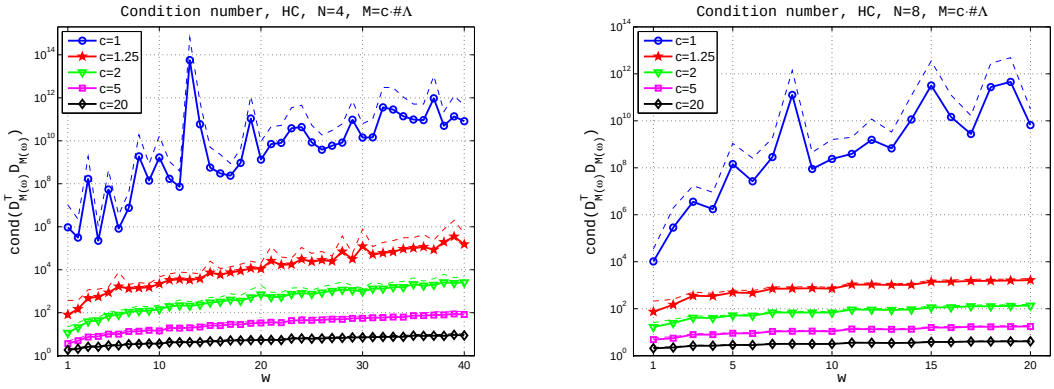


Figure 1.16: Condition number (1.4.9), HC, $M = c \cdot \#\Lambda$. The continuous, marked, lines show the mean condition number (1.4.10) over 100 repetitions. The dashed lines show the mean (1.4.10) plus one standard deviation (1.4.11). Left: $N = 4$. Right: $N = 8$.

but looking at the dimension $\#\Lambda$ of the space we recognize that spaces of the same dimension (TP or TD) introduce similar approximation errors (see Tables 1.1 and 1.2). Instead, we see in Fig. 1.17-Right that the convergence of the HC space is slower, also when looking at the effective dimension of the space, in Table 1.3.

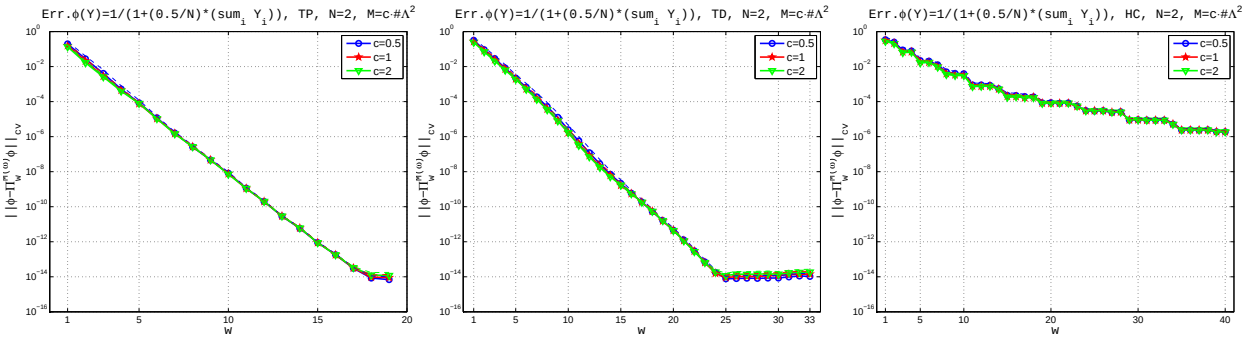


Figure 1.17: Error (1.4.6) with the function (1.5.4), $\beta = 0.5$, $N = 2$, $M = c \cdot (\#\Lambda)^2$. The continuous marked lines show the mean error (1.4.7) over 100 repetitions. The dashed lines show the mean (1.4.7) plus one standard deviation (1.4.8). Spaces: TP (left), TD (center), HC (right).

w	# TP	$\overline{E}_{cv}$
5	36	$0.790 \cdot 10^{-4}$
7	64	$1.456 \cdot 10^{-6}$
10	121	$0.724 \cdot 10^{-8}$
11	144	$1.108 \cdot 10^{-9}$
15	256	$0.876 \cdot 10^{-12}$

Table 1.1: Selected error values for TP and $M = 2\,\#\text{TP}$ from Fig. 1.17-Left.

w	# TD	$\overline{E}_{cv}$
7	36	$1.354 \cdot 10^{-4}$
10	66	$1.688 \cdot 10^{-6}$
14	120	$0.512 \cdot 10^{-8}$
16	153	$5.513 \cdot 10^{-10}$
21	253	$1.123 \cdot 10^{-12}$

Table 1.2: Selected error values for TD and $M = 2\,\#\text{TD}$ from Fig. 1.17-Center.

w	# HC	$\overline{E}_{cv}$
12	37	$7.505 \cdot 10^{-4}$
19	66	$7.971 \cdot 10^{-5}$
31	119	$8.788 \cdot 10^{-6}$
40	160	$1.889 \cdot 10^{-6}$

Table 1.3: Selected error values for HC and $M = 2\,\#\text{HC}$ from Fig. 1.17-Right.

Fig. 1.18 shows the error in approximating the function (1.5.4) in TP (on the left) or TD (on the right) space, for $N = 2$. We observe a lower variability in the error due to the reduced variability in the corresponding condition number. Despite the reduced variability we observe also in this case that the linear rule eventually leads to divergence when w increases if c is chosen too small. This effect will be much more dramatic for low regularity functions, as shown later on.

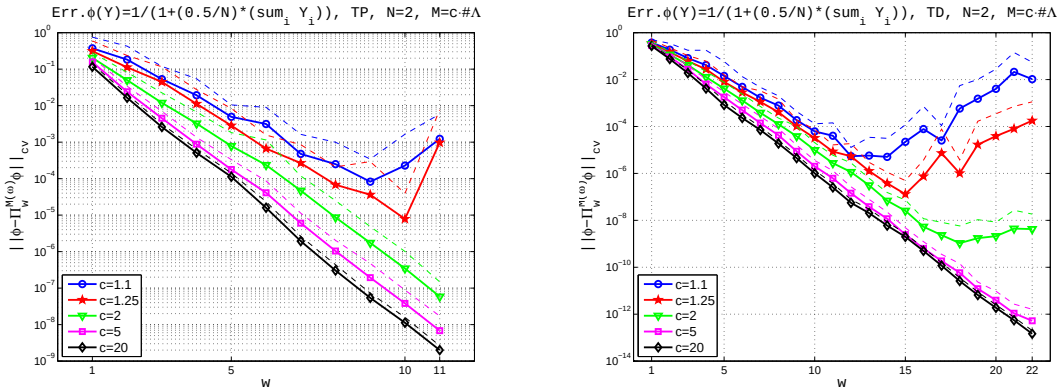


Figure 1.18: Error (1.4.6) with the function (1.5.4), $N = 2$, $M = c \cdot \#\Lambda$. The continuous marked lines show the mean error (1.4.7) over 100 repetitions. The dashed lines show the mean (1.4.7) plus one standard deviation (1.4.8). Left: TP space. Right: TD space.

A further comparison can be made between Fig. 1.18-Left and Fig. 1.19-Left. In the latter case N is increased to 4 (TP space), and this yields a faster convergence with respect to w. Note

however, that a larger amount of sample points is used, for corresponding w values.

Analogously, the comparison of Figs. 1.18-Right and Fig. 1.19-Right concerning the TD space reveals that in higher dimension the convergence is faster and more stable.

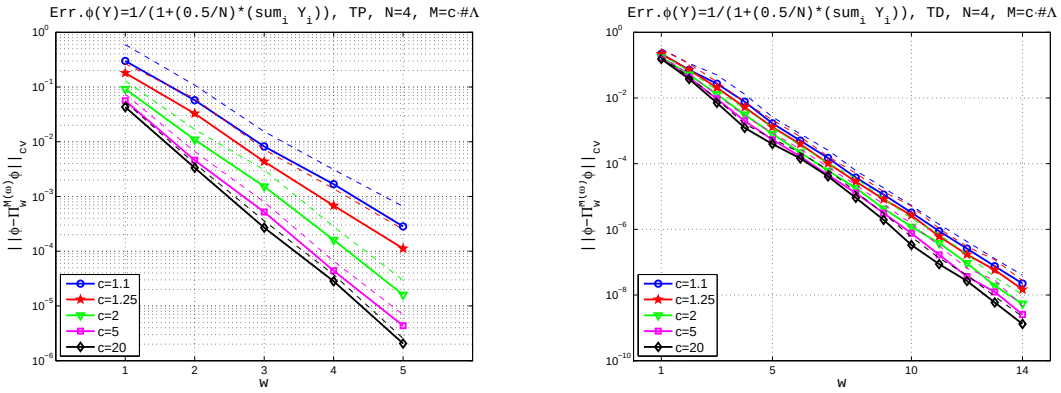


Figure 1.19: Error (1.4.6) with the function (1.5.4), $\beta = 0.5$, $N = 4$, $M = c \cdot \#\Lambda$. The continuous marked lines show the mean error (1.4.7) over 100 repetitions. The dashed lines show the mean (1.4.7) plus one standard deviation (1.4.8). Spaces: TP (left), TD (right).

We also consider the function,

$$\phi(\mathbf{Y}) = \sum_{i=1}^{N} |Y_i|^3, \qquad \mathbf{Y} \in [-1,1]^N, \tag{1.5.5}$$

which is a multidimensional extension of (1.5.3). Fig. 1.20 shows that the optimal convergence rate achieved by the quadratic rule does not depend on the choice of the space (*i.e.* TP, TD, or HC). In addition, for the linear rule Fig. 1.21 shows that the same convergence behaviour is obtained when increasing the dimension N, but with a significantly decreased variability. Note that we even allowed c to take the value 1 in the linear rule.

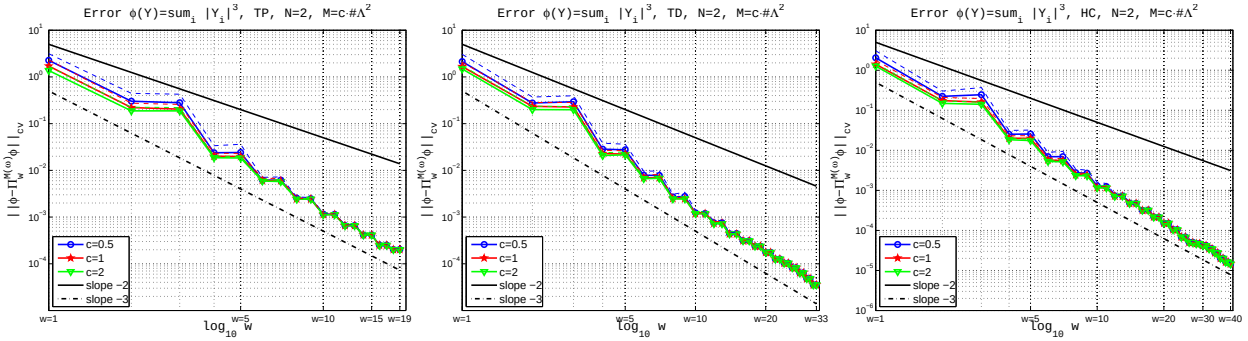


Figure 1.20: Error (1.4.6) with the function (1.5.5), $N = 2$, $M = c \cdot (\#\Lambda)^2$. The continuous marked lines show the mean error (1.4.7) over 100 repetitions. The dashed lines show the mean (1.4.7) plus one standard deviation (1.4.8). Spaces: TP (left), TD (center), HC (right).

A function with lower regularity across a circle

Now we give an example of a function which is hard to approximate in the TD spaces. When $N = 2$, we consider the target function

$$\phi(\mathbf{Y}) = \left| \sum_{i=1}^{2} Y_i^2 - 0.5 \right|^3, \qquad \mathbf{Y} \in [-1,1]^2, \tag{1.5.6}$$

which features a discontinuity in its derivatives over the circle with radius equal to $\sqrt{0.5}$ and centered in the origin. Note that (1.5.6) is a continuous function.

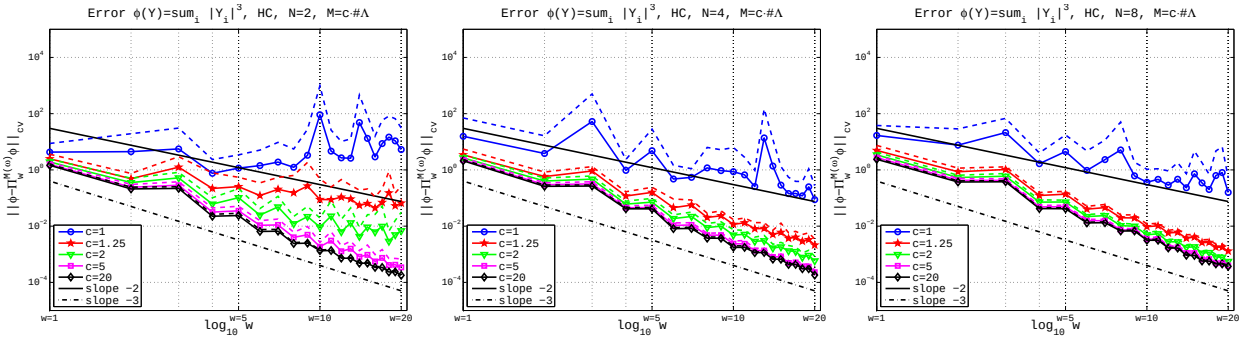


Figure 1.21: Error (1.4.6) with the function (1.5.5), HC space. The continuous marked lines show the mean error (1.4.7) over 100 repetitions. The dashed lines show the mean (1.4.7) plus one standard deviation (1.4.8). $M = c \cdot \#\Lambda$. $N = 2$ (left), $N = 4$ (center), $N = 8$ (right).

Choosing the quadratic rule leads to the expected theoretical convergence rates for both TP and TD spaces; see Fig. 1.22. However, the TD space exhibits a suboptimal convergence rate already when $w \leq 5$.

When choosing the linear rule (Fig. 1.23), the results obtained with the TP space slightly differ from those obtained with the TD space. In particular, the convergence rate is slower than the theoretically predicted one (Fig. 1.23-Right).

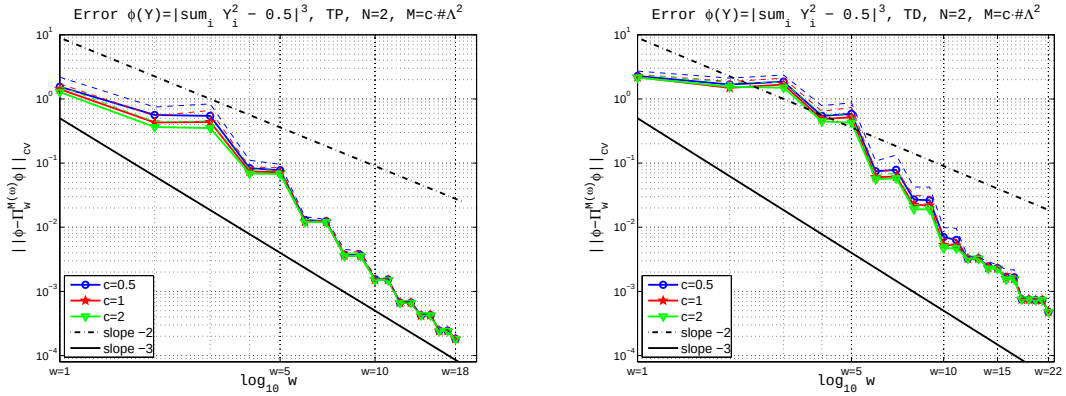


Figure 1.22: Error (1.4.6) with the function (1.5.6), $N = 2$, $M = c \cdot (\#\Lambda)^2$. The continuous marked lines show the mean error (1.4.7) over 100 repetitions. The dashed lines show the mean (1.4.7) plus one standard deviation (1.4.8). Left: TP. Right: TD.

1.6 Conclusions

In this work we have analyzed the problem of approximating a multivariate function by discrete L^2 projection on a polynomial space starting from random, noise-free observations.

In the 1D case with sampling points drawn from a bounded domain and a probability density function bounded away from zero, we have shown that the discrete L^2 projection leads to optimal convergence rates, equivalent to the best approximation error in L^∞, up to a logarithmic factor, *provided the number of samples M scales quadratically with the dimension of the polynomial space $\#\Lambda$*. We have also shown how this result reflects on the condition number of the design matrix.

The numerical tests we have performed confirm the theoretical results corresponding to a uniform distribution of sample points. In our 1D tests, we clearly see that the condition $M \sim (\#\Lambda)^2$ guarantees a condition number of the design matrix bounded independently of the polynomial degree and an optimal convergence rate. On the other hand, the relation $M \sim \#\Lambda$ leads to a condition number growing exponentially fast with the polynomial degree and a

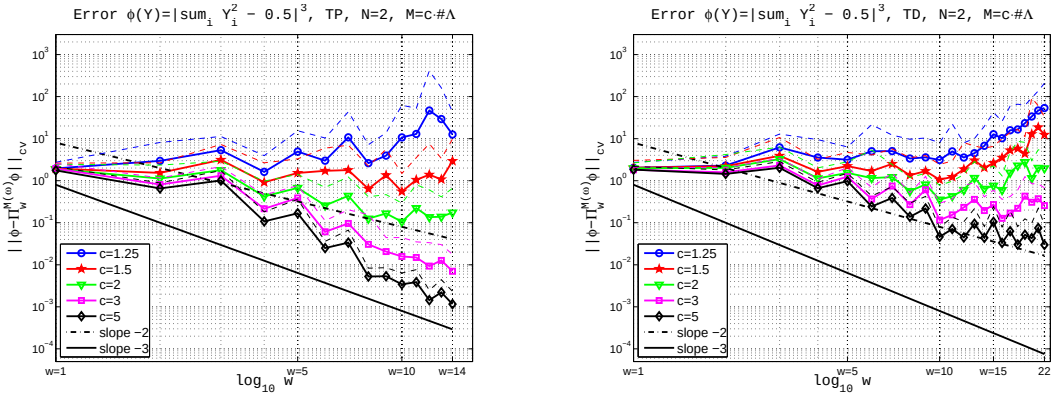


Figure 1.23: Error (1.4.6) with the function (1.5.6), $N = 2$, $M = c \cdot \#\Lambda$. The continuous marked lines show the mean error (1.4.7) over 100 repetitions. The dashed lines show the mean (1.4.7) plus one standard deviation (1.4.8). Left: TP. Right: TD.

convergence plot that features initially a suboptimal rate up to a critical polynomial degree beyond which divergence is observed. In addition, the sensitivity on the proportionality constant has been examined.

In high dimension we observe numerically in many cases that a choice $M \sim \#\Lambda$ does lead to optimal convergence rate within all reasonable tolerances (up to machine precision). Whether this is an indication that in high-D the relation $M \sim \#\Lambda$ is enough to have a stable and optimal approximation or just that the blow up of the error occurs much further (at tolerances below machine precision) is still an open question and a topic of current research.

In this work we have considered only functions with values in $\mathbb{R}$. In the field of UQ, one is often interested in functions with values in some Banach space, representing the solution of a (possibly nonlinear) differential or integral problem. Future research directions will include the extension of these results to Banach-valued functions.

Chapter 2

Discrete least squares polynomial approximation with random evaluations; application to parametric and stochastic elliptic PDEs

This chapter is based on the contents of the manuscript [CCM+13], in collaboration with A. Chkifa, A. Cohen, F. Nobile and R. Tempone.

2.1 Introduction

In this work we analyze the discrete least squares approximation based on random evaluations, to approximate functions depending on multivariate random variables distributed according to a given probability density. We first recall in Section 2.2 the main results achieved by two independent approaches that have been recently proposed in the univariate case: stability and optimality have been proved, in expectation in [CDL13] and in probability in [MNvST11], under the condition that the number of samples scales quadratically with respect to the dimension of the polynomial space. Then we focus on the extension to the multivariate case. In Section 2.3 the relation between the number of samples and the dimension of the polynomial space is extended to any monotone set, for any dimension of the parameter space.

The least squares methodology based on discrete random evaluations is then applied in Section 2.4 to the approximation of the solution to PDE models with stochastic coefficients, when the parameter space features moderately high dimensionality. Next we consider a specific application consisting of a linear elliptic PDE with a diffusion coefficient containing some inclusions with random diffusivity. For this specific model, results on the polynomial approximation by Stochastic Galerkin of the solution w.r.t. the random diffusion parameters have been obtained in [CCDS13, BNTT11b], and we recall them in Section 2.4.2. Afterwards, in Section 2.4.3 we focus on the approximation of the solution to the same model by the random discrete projection on Total Degree polynomial spaces, and show sub-exponential convergence w.r.t. the total number M of sampling points (hence deterministic PDEs to solve).

This estimate clarifies the dependence of the convergence rate on the number of sampling points and on the dimension of the parameter set. In addition, a relation has been established between the convergence rate of the least squares approximation and the convergence rate of the Stochastic Galerkin method derived in [BNTT11b]. The numerical results presented in Section 2.5 confirm our estimate and highlight a gap between the condition necessary to achieve

stability and optimality in the theory, and the condition that in practice yields the optimal convergence rate. Finally in Section 2.6 we draw some conclusions.

2.2 Least squares approximation with noise-free evaluations in random points

Let d be a non negative integer, Γ a bounded domain of $\mathbb{R}^d$ and ρ a probability measure on Γ. We consider $u : \Gamma \to \mathbb{R}$ an unknown function, depending smoothly on a random variable $\mathbf{Y}$ distributed according to a known probability density $\rho : \Gamma \to \mathbb{R}^+$. The goal of our analysis is to study the approximation of u by the discrete least squares estimation using noise-free samples $\mathbf{z}_i = u(\mathbf{y}_i)$, $i = 1, ..., M$ of u at the sample points $\mathbf{S}_M := (\mathbf{y}_i)_{i=1,...,M}$ where the $\mathbf{y}_i \in \Gamma$ are i.i.d samples from the measure ρ. The error between u and its estimator $\widetilde{u}$ is measured in the $L^2(\Gamma, d\rho)$ norm

$$\|v\| := \left(\int_\Gamma |v(x)|^2 d\rho(x) \right)^{\frac{1}{2}}. \tag{2.2.1}$$

Its associated inner product is denoted by $\langle \cdot, \cdot \rangle$. In general, the probability measure ρ is not known but we can sample according to ρ and we have access to observations of u at any sampled point $\mathbf{y} \in \Gamma$. A natural approach in this setting is to consider the solution of the discrete least squares problem

$$\widetilde{u} := \underset{v \in L^2(\Gamma, d\rho)}{\operatorname{argmin}} \sum_{i=1}^{M} |z_i - v(\mathbf{y}_i)|^2.$$

In general $\widetilde{u}$ is not unique. We consider instead the discrete least squares estimation of u on finite dimensional subspaces of $L^2(\Gamma, d\rho)$. Let $N \geq 1$ and $V_N \subset L^\infty(\Gamma)$ be a finite dimensional subspace of $L^2(\Gamma, d\rho)$ with $\dim V_N = N$, we introduce the best approximation of u in the space V_N in the continuous and discrete least squares sense

$$P_N u = \underset{v \in V_N}{\operatorname{argmin}} \|u - v\| \quad \text{and} \quad P_N^M u = \underset{v \in V_N}{\operatorname{argmin}} \|u - v\|_M, \tag{2.2.2}$$

where

$$\|v\|_M = \left(\frac{1}{M} \sum_{i=1}^{M} |v(\mathbf{y}_i)|^2 \right)^{\frac{1}{2}}, \tag{2.2.3}$$

is the L^2 norm with respect to the empirical measure. Analogously, $\langle ., . \rangle_M$ denotes the associated empirical inner product. We should emphasize that this discrete inner product depends on the sample $\mathbf{S}_M$ and that the projection $P_N u$ depends on the space V_N and the discrete projection $P_N^M u$ depends on both the space V_N and the sample $\mathbf{S}_M$.

We are interested in the case $N \leq M$ which is the regime where problem (2.2.2) may admit a unique solution. Let $\mathcal{B}_L := \{L_1, \ldots, L_N\}$ be any basis of V_N. The approximation $P_N^M u$ is then given by

$$P_N^M u = \sum_{j=1}^{N} w_j L_j,$$

where $\mathbf{w} := (w_1, \ldots, w_N)^t$ is the unique solution of

$$\mathbf{G}\mathbf{w} = \mathbf{J}\mathbf{u},$$

with $\mathbf{G} := (\langle L_i, L_j\rangle_M)_{1\le i,j\le N}$, $\mathbf{J} = \frac{1}{M}(L_i(y_j))_{\substack{1\le i\le N\\ 1\le j\le M}}$ and $\mathbf{u} := (\mathbf{z}_j)_{1\le j\le M}^t$. When $\mathbf{G}$ is not singular, then we have

$$P_N^M u = \sum_{j=1}^M u(\mathbf{y}_j)\pi_j,$$

where $\mathcal{B}_\pi := \{\pi_0, \dots, \pi_M\}$ is a family of elements in V_N given by

$$\mathcal{B}_\pi = \left(\mathbf{G}^{-1}\mathbf{J}\right)^t \mathcal{B}_L.$$

Indeed, for $\mathbf{y} \in \Gamma$ fixed, using the notation $\mathbf{L_y} := (L_j(\mathbf{y}))_{1\le j\le N}^t$, we have

$$P_N^M u(\mathbf{y}) = \sum_{j=1}^N w_j L_j(\mathbf{y}) = \mathbf{w}\cdot\mathbf{L_y} = (\mathbf{G}^{-1}\mathbf{J}\mathbf{u})\cdot\mathbf{L_y} = \mathbf{u}\cdot(\mathbf{G}^{-1}\mathbf{J})^t\mathbf{L_y}.$$

2.2.1 Stability of the projection P_N^M

We suppose now that the basis $\mathcal{B}_L$ is an orthonormal basis of V_N with respect to $\langle\cdot,\cdot\rangle$. We introduce the quantity

$$K(V_N) := \sup_{\mathbf{y}\in\Gamma}\left(\sum_{j=1}^N |L_j(\mathbf{y})|^2\right). \tag{2.2.4}$$

The quantity $K(V_N)$ does not depend on the orthonormal basis $(L_j)_{1\le j\le N}$ and depends only on V_N and on the measure ρ. See Proposition 2.1 below.

For brevity, in the sequel we denote by $\|v\|_\infty := \|v\|_{L^\infty(\Gamma)}$ the infinite norm of v. As in [MNvST11], we introduce the constant

$$\widetilde{C}(V_N) = \sup_{v\in V_N\,:\,v\neq 0}\frac{\|v\|_\infty^2}{\|v\|^2}. \tag{2.2.5}$$

Note that the values of $K(V_N)$ and $\widetilde{C}(V_N)$ are bounded, since we assumed $V_N \subset L^\infty(\Gamma)$. Moreover, we have the following result.

Proposition 2.1.

$$K(V_N) = \widetilde{C}(V_N).$$

Proof. To see this, we have by Cauchy-Schwartz inequality

$$\Big|\sum_{j=1}^N v_j L_j(\mathbf{y})\Big|^2 \le \sum_{j=1}^N |L_j(\mathbf{y})|^2 \sum_{j=1}^N |v_j|^2 \le K(V_N)\|v\|^2, \quad \mathbf{y}\in\Gamma,$$

with equality when $\mathbf{y} = \mathbf{y}^*$ is the point of Γ where the supremum in $K(V_N)$ is attained and the function v is defined by $v_j = L_j(\mathbf{y}^*)$. □

We also observe that $\mathbb{E}(\mathbf{G}) = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix. It has been proved in [CDL13] that $\mathbf{G}$ is well conditioned with high probability. We introduce the operator norm

$$|||M||| := \sup_{v\neq 0}\frac{\|Mv\|}{\|v\|} = \sup_{v,w\neq 0}\frac{|\langle Mv, w\rangle|}{\|v\|\|w\|}.$$

The following Theorem has been proved in [CDL13].

Theorem 2.1. *For $0 < \delta < 1$, then*

$$Pr\Big(|||\mathbf{G} - \mathbf{I}||| > \delta\Big) \leq 2N \exp\Big(-\frac{\beta_\delta M}{K(V_N)}\Big),$$

where $\beta_\delta := \delta + (1-\delta)\log(1-\delta)$.

Proof. See [CDL13, Theorem 1]. □

The theorem implies in particular that when $K(V_N)$ is smaller that M by a logarithmic factor, for example, if $K(V_N)$ is such that

$$K(V_N) \leq \frac{\beta_\delta}{1+\gamma}\frac{M}{\log M} \tag{2.2.6}$$

for some $\gamma > 0$, then

$$Pr\Big(|||\mathbf{G} - \mathbf{I}||| > \delta\Big) \leq \frac{2N}{M^{\gamma+1}} \leq 2M^{-\gamma}. \tag{2.2.7}$$

In the sequel, we pick $\delta = \frac{1}{2}$ and shorten the notation with $\beta = \beta_{\frac{1}{2}} \approx 0.15$. In Section 2.5 we will numerically analyze the effects of a different choice of the parameter δ.

Given M fixed, we denote $d\rho^M$ the probability measure of the draw. We denote by Ω^M the set of all possible draws and we divide it into Ω_+^M the set of draws such that

$$|||\mathbf{G} - \mathbf{I}||| \leq \frac{1}{2}.$$

Denote its complement by $\Omega_-^M = \Omega^M \setminus \Omega_+^M$. For any $\gamma > 0$, if $K(V_N)$ satisfies (2.2.6), then $Pr(\Omega_+^M) > 1 - 2M^{-\gamma}$. In addition, for any draw in Ω_+^M, we have by the definition of the spectral norms that,

$$|\langle v, w\rangle_M - \langle v, w\rangle| \leq \frac{1}{2}\|v\|\|w\|, \quad v, w \in V_N,$$

which implies in particular that

$$\frac{1}{2}\|v\|^2 \leq \|v\|_M^2 \leq \frac{3}{2}\|v\|^2, \quad v \in V_N. \tag{2.2.8}$$

Following [MNvST11], we introduce the random variable C^ω as

$$C^\omega(M, V_N) = \sup_{v \in V_N \setminus \{v \equiv 0\}} \frac{\|v\|^2}{\|v\|_M^2},$$

using the continuous weighted L^2 norm (2.2.1) and the discrete norm (2.2.3), and the random variable

$$c^\omega(M, V_N) := \sup_{v \in V_N \setminus \{v \equiv 0\}} \frac{\|v\|_M^2}{\|v\|^2}.$$

We can measure the stability of the discrete least squares approximation in terms of the condition number of the matrix $\mathbf{G}^T\mathbf{G}$: the following characterization has been proven in [MNvST11].

Proposition 2.2 ([MNvST11])**.** *The spectral condition number (2-norm) of the matrix $\mathbf{G}^T\mathbf{G}$ is equal to*

$$cond\Big(\mathbf{G}^T\mathbf{G}\Big) = c^\omega(M, V_N)\, C^\omega(M, V_N).$$

Proof. See [MNvST11, Proposition 4]. □

The previous considerations show that $\mathrm{cond}\Big(\mathbf{G}^T\mathbf{G}\Big) \leq 3$ with high probability under the condition (2.2.6). Moreover, $\mathrm{cond}\Big(\mathbf{G}^T\mathbf{G}\Big) \overset{M\to+\infty}{\longrightarrow} 1$ thanks to the following result proved in [MNvST11]:

Theorem 2.2 ([MNvST11])**.** *Given any finite dimensional set $V_N \subset L^\infty(\Gamma)$, it holds*

$$\lim_{M\to+\infty} c^\omega(M,V_N) = \lim_{M\to+\infty} C^\omega(M,V_N) = 1, \quad \textit{almost surely.}$$

Proof. See [MNvST11, Theorem 1]. □

2.2.2 Optimality of the projection P_N^M

In this section we analyze the optimality of the random discrete projection. The exact projection $P_N u$ of the function u on V_N and the discrete projection $P_N^M u$ of u on V_N have been introduced in (2.2.2). The corresponding approximation errors, both in the L^2 sense, are denoted as $\|u - P_N u\|$ and $\|u - P_N^M u\|$, respectively. Moreover, we denote by $\inf_{v\in V_N} \|u - v\|_\infty$ the best approximation error in L^∞ of u on V_N.

Assumption 2.1. *The function u is uniformly bounded and there exists $\tau > 0$ such that*

$$|u(\mathbf{y})| \leq \tau, \quad \forall\, \mathbf{y} \in \Gamma.$$

Under Assumption 2.1 on the target function u, we introduce the truncated projection

$$\widetilde{P}_N^M u = T_\tau(P_N^M u), \tag{2.2.9}$$

where $T_\tau(t) := \mathrm{sign}(t)\min\{\tau, |t|\}$.

Lemma 2.1 ([MNvST11])**.** *For any realization of the random sample $\mathbf{S}_M$ it holds*

$$\|u - P_N^M u\| \leq \Big(1 + \sqrt{C^\omega(M,V_N)}\Big) \inf_{v\in V_N} \|u - v\|_\infty.$$

Proof. See [MNvST11, Proposition 1]. □

Lemma 2.1 shows that, in any dimension d and for any density ρ, the approximation properties of the random least squares projection can be determined by the analysis of the random variable C^ω. In particular, Lemma 2.1 holds for any value of M. When M is large enough, from Theorem 2.2 the random least squares projection is stable and yields the same error as the the best approximation error achievable, up to a proportionality constant equal to two. The next theorem ensures the optimality of the random least squares projection, under the condition that M is large enough to satisfy (2.2.6).

Theorem 2.3. *For any $\gamma > 0$, if M is such that $K(V_N)$ satisfies (2.2.6), then for any $u \in L^\infty(\Gamma)$ with $\|u\|_\infty \leq \tau$, it holds*

$$\mathbb{E}(\|u - \widetilde{P}_N^M u\|^2) \leq (1 + \eta(M))\|u - P_N u\|^2 + 8\tau^2 M^{-\gamma}, \tag{2.2.10}$$

with $\eta(M) := \frac{4\beta}{(1+\gamma)\log M}$, and

$$Pr\Big(\|u - P_N^M u\| \leq \Big(1 + \sqrt{2}\Big) \inf_{v\in V_N} \|u - v\|_\infty\Big) \geq 1 - 2M^{-\gamma}. \tag{2.2.11}$$

Proof. The result in expectation given by (2.2.10) has been proven in [CDL13, Theorem 2].

To prove (2.2.11), let M be an integer satisfying the assumption of the theorem. According to (2.2.7), $Pr(\Omega_+^M) \geq 1 - 2M^{-\gamma}$. Now, given a draw in Ω_+^M. Using (2.2.8), we have for any $v \in V_N$

$$\|u - P_N^M u\| \leq \|u - v\| + \|v - P_N^M u\| \leq \|u - v\| + \sqrt{2}\|v - P_N^M u\|_M.$$

By Pythagoras identity $\|u - v\|_M^2 = \|u - P_N^M u\|_M^2 + \|P_N^M u - v\|_M^2$, we deduce

$$\|u - P_N^M u\| \leq \|u - v\| + \sqrt{2}\|u - v\|_M \leq (1 + \sqrt{2})\|u - v\|_\infty,$$

which completes the proof. □

In the specific case of polynomial approximation, the probabilistic optimality result expressed by (2.2.11) has been proven in [MNvST11, Theorem 2]. In particular, in the one-dimensional case ($d = 1$) and with the uniform density $\rho \sim \mathcal{U}(-1, 1)$, the relation (2.2.11) (up to a logarithmic factor) has been proven under the condition that $M \propto \big(\dim(\mathbb{P}_N)\big)^2$, using probability arguments and order statistics of the uniform density.

In many situations the evaluations of the target function u are contaminated by noise, and are denoted in the sequel by $\widehat{u}$. The nonparametric approach estimates then the conditional expectation $u(\mathbf{y}) = \mathbb{E}(\widehat{u}|\mathbf{y})$, with the maximal variance σ^2 of the noise given by

$$\sigma^2 := \max_{\mathbf{y}\in\Gamma} \mathbb{E}\Big[|\widehat{u} - u(\mathbf{y})|^2 \Big| \mathbf{y}\Big] < +\infty.$$

The noisy evaluations $\widehat{u}$ of the unknown function u are employed to compute the (truncated) least squares projection defined in (2.2.2). The following result has been proven in [CDL13]:

Theorem 2.4. *For any $\gamma > 0$, if M is such that $K(V_N)$ satisfies* (2.2.6)*, then for any $u \in L^\infty(\Gamma)$, with $\|u\|_\infty \le \tau$, it holds*

$$\mathbb{E}(\|u - \widetilde{P}_N^M u\|^2) \le (1 + 2\eta(M))\|u - P_N u\|^2 + 8\tau^2 M^{-\gamma} + 8\sigma^2 \frac{N}{M}, \tag{2.2.12}$$

with $\eta(M)$ as in Theorem 2.3.

Proof. See [CDL13, Theorem 3]. □

The next section focuses on the application of the random least squares projection on polynomial spaces.

2.3 Least squares approximation in multidimensional polynomial spaces

In this section we confine to the application of the random least squares projection on polynomial spaces, with either Legendre polynomials and uniform distribution or with Chebyshev polynomials and Chebyshev distribution.

We introduce the notation $\mathcal{F} = \mathbb{N}^d$ when $d < \infty$ and $\mathcal{F}$ the set of finitely supported multi-indices when $d = \infty$. Let $\boldsymbol{\nu} = (\nu_1, \dots, \nu_d)$ be a multi-index and $\Lambda \subset \mathcal{F}$ an index set which is monotonous in the following sense:

Property 3 (Monotonicity of Λ)**.** *Consider two multi-indices $\boldsymbol{\nu}', \boldsymbol{\nu}'' \in \mathcal{F}$ such that $\nu_j'' \le \nu_j'$, $\forall\, j = 1, \dots, d$. The multi-index set Λ is monotonous if the following holds:*

$$\boldsymbol{\nu}' \in \Lambda \Rightarrow \boldsymbol{\nu}'' \in \Lambda.$$

In the case where the density factorizes as $\rho(\mathbf{Y}) = \prod_{j=1}^d \rho_j(Y_j)$ then a basis of the polynomial space can be constructed by tensorizing one-dimensional orthogonal polynomials with respect to each weight ρ_j separately.

From now on, we confine to the case of $\rho_j = \mathcal{U}(-1, 1)$, $j = 1, \dots, d$. Therefore we introduce the univariate "*Probabilistic*" Legendre polynomials $(L_k)_{k\ge 0}$, that are orthonormal with respect to the measure $\frac{dy}{2}$ on $[-1, 1]$. The family $(L_k)_{k\ge 0}$ forms a complete orthonormal system in $L^2([-1, 1], \frac{dy}{2})$. We consider $\Gamma = [-1, 1]^d$, and then we introduce the tensorized Legendre polynomials $(\mathbf{L}_\nu)_{\nu\in\mathcal{F}}$ defined as

$$\mathbf{L}_\nu(\mathbf{y}) := \prod_{j:\nu_j\neq 0} L_{\nu_j}(y_j), \qquad \nu \in \mathcal{F}, \quad \mathbf{y} = (y_1, y_2, \dots) \in \Gamma. \tag{2.3.1}$$

The family $(\mathbf{L}_\nu)_{\nu\in\mathcal{F}}$ forms a complete orthonormal system in $L^2(\Gamma, d\rho)$. Next, we define the polynomial space $\mathbb{P}_\Lambda$ associated with any monotone multi-index set Λ as

$$\mathbb{P}_\Lambda = \operatorname{span}\{\mathbf{L}_\nu \,:\, \nu \in \Lambda\}.$$

In the sequel V_Λ will denote the polynomial space $\mathbb{P}_\Lambda$. Moreover, as in Section 2.2, $N = \dim(V_\Lambda) = \dim(\mathbb{P}_\Lambda) = \#\Lambda$, the last equality being justified by the one-to-one relation (2.3.1) between basis functions and multi-indices.

Possible choices of polynomial spaces are the Tensor Product and the Total Degree. We index the set Λ by the subscript w denoting the maximum polynomial degree retained in the space:

$$\begin{aligned} &\textbf{Tensor Product (TP)}, \qquad \Lambda(\mathrm{w}) = \Big\{\boldsymbol{\nu} \in \mathcal{F} : \|\boldsymbol{\nu}\|_\infty \leq \mathrm{w}\Big\}, \\ &\textbf{Total Degree (TD)}, \qquad \Lambda(\mathrm{w}) = \Big\{\boldsymbol{\nu} \in \mathcal{F} : \|\boldsymbol{\nu}\|_1 \leq \mathrm{w}\Big\}. \end{aligned} \tag{2.3.2}$$

The exact and discrete projections of the function u on the polynomial space V_Λ are denoted by $P_\Lambda u$ and $P_\Lambda^M u$, respectively. The truncation operator defined in (2.2.9) is denoted as $\widetilde{P}_\Lambda^M u$ when using the discrete projection on the polynomial space V_Λ.

With the polynomial basis (2.3.1) the definition of K given in (2.2.4) modifies as

$$K(\Lambda) = \sup_{\mathbf{y}\in\Gamma} \sum_{\boldsymbol{\nu}\in\Lambda} |\mathbf{L}_\nu(\mathbf{y})|^2.$$

This quantity can be bounded when the set Λ is monotone and when the basis is composed of the tensorized probabilistic Legendre or Chebyshev polynomials. The following result in Lemma 2.2 allows to quantify the number of sampling points M necessary to satisfy condition (2.2.6), in the specific case of polynomial approximation with Legendre polynomials. It is valid for any value of the dimension d, with $d = +\infty$ included.

The result in Lemma 2.4 concerns instead the case of polynomial approximation with Chebyshev polynomials and distribution.

In Lemma 2.5 we propose an alternative proof of the result expressed by Theorem 2.2, using multidimensional inverse inequalities.

Lemma 2.2. *For any monotone set $\Lambda \subset \mathcal{F}$, the quantity $K(\Lambda)$ with the tensorized Legendre polynomials satisfies*

$$K(\Lambda) \leq (\#\Lambda)^2. \tag{2.3.3}$$

Proof. For any $\boldsymbol{\nu} \in \mathcal{F}$ and any $k \in \mathbb{N}$, we denote by $\mathcal{F} \ni ((k, \boldsymbol{\nu})) := (k, 0, \dots) + \mathcal{S}(\boldsymbol{\nu})$ the addition of the multi-index $(k, 0, \dots)$ with all components equal to zero except the first one being equal to k, and the multi-index $\boldsymbol{\nu}$ shifted through the usual Right Shift Operator $\mathcal{S}$.

Since every polynomial L_k attains its maximum in $y = 1$ and $\|L_k\|_{L^\infty([-1,1])} = \sqrt{2k+1}$, then

$$K(\Lambda) = \sum_{\boldsymbol{\nu}\in\Lambda} \prod_{j:\nu_j\neq 0} (2\nu_j + 1).$$

We prove (2.3.11) using induction on $n_\Lambda := \#\Lambda$. When $n_\Lambda = 1$, then $\Lambda = \{0_\mathcal{F}\}$ and it is obviously true. Let $n \geq 1$ and Λ denote a monotone set with $n_\Lambda = n + 1$. Without loss of generality, we suppose that $\nu_1 \neq 0$ for some $\boldsymbol{\nu} \in \Lambda$, and denote by $J \geq 1$ the maximal value attained by the coordinate ν_1 when $\boldsymbol{\nu} \in \Lambda$. For $0 \leq k \leq J$, we introduce

$$\Lambda_k := \{\hat{\boldsymbol{\nu}} = (\nu_2, \nu_3, \dots) : ((k, \hat{\boldsymbol{\nu}})) \in \Lambda\}.$$

By monotonicity of Λ, every Λ_k is monotone. Also, since $J \geq 1$ then $\#\Lambda_k < \#\Lambda$ for any k, so that the induction hypothesis implies

$$K(\Lambda) = \sum_{k=0}^{J} (2k+1) K(\Lambda_k) \leq \sum_{k=0}^{J} (2k+1)(\#\Lambda_k)^2 \,.$$

In addition, we have

$$\Lambda_J \subset \cdots \subset \Lambda_1 \subset \Lambda_0,$$

since for $k \geq 1$, $\boldsymbol{\mu} \in \Lambda_k \Rightarrow ((k, \boldsymbol{\mu})) \in \Lambda \Rightarrow ((k-1, \boldsymbol{\mu})) \in \Lambda \Rightarrow \boldsymbol{\mu} \in \Lambda_{k-1}$. We deduce

$$k(\#\Lambda_k)^2 \leq \#\Lambda_k \#\Lambda_0 + ... + \#\Lambda_k \#\Lambda_{k-1},$$

and consequently

$$K(\Lambda) \leq \sum_{k=0}^{J} (\#\Lambda_k)^2 + 2\sum_{k=0}^{J} \Big(\#\Lambda_k \#\Lambda_0 + ... + \#\Lambda_k \#\Lambda_{k-1} \Big) \leq \Big(\sum_{k=0}^{J} \#\Lambda_k \Big)^2 = (\#\Lambda)^2,$$

which concludes the proof. □

Therefore, in the case of polynomial approximation, the optimality result of the least squares approximation stated in Theorem 2.3 holds under the condition

$$(\#\Lambda)^2 \leq \frac{\beta_\delta}{1+\gamma} \frac{M}{\log M}, \tag{2.3.4}$$

that thanks to (2.3.11) implies (2.2.6).

Lemma 2.3. *For any real positive numbers $a_0 \geq a_1 \geq ... \geq a_n$ and any $\alpha > \frac{\ln 3}{\ln 2}$ it holds*

$$(a_0 + ... + a_n)^\alpha \geq a_0^\alpha + 2(a_1^\alpha ... + a_n^\alpha). \tag{2.3.5}$$

Proof. The proof is by induction on n. Indeed, for $n = 0$ (2.3.5) is always true. For $n = 1$, since the function $x \mapsto (x + a_1)^\alpha - x^\alpha$ is increasing in $[a_1, \infty[$, then

$$(a_0 + a_1)^\alpha - a_0^\alpha \geq (2^\alpha - 1) a_1^\alpha \geq 2a_1^\alpha.$$

Now let $n \geq 1$ and $a_0 \geq a_1 \geq ... \geq a_{n+1}$ be positive number. By the induction assumption, we have

$$\begin{aligned} (a_0 + ... + a_{n+1})^\alpha &= \Big((a_0 + ... + a_n) + a_{n+1} \Big)^\alpha \\ &\geq (a_0 + ... + a_n)^\alpha + 2a_{n+1}^\alpha \\ &\geq a_0^\alpha + 2(a_1^\alpha ... + a_n^\alpha) + 2a_{n+1}^\alpha \\ &= a_0^\alpha + 2(a_1^\alpha ... + a_{n+1}^\alpha). \end{aligned}$$

□

Lemma 2.4. *For any monotone set $\Lambda \subset \mathcal{F}$, the quantity $K_T(\Lambda)$ with the tensorized Chebyshev polynomials satisfies*

$$K_T(\Lambda) \leq (\#\Lambda)^\beta, \quad \text{with} \quad \beta = \frac{\ln 3}{\ln 2}. \tag{2.3.6}$$

Proof. The univariate "Probabilistic" Chebyshev polynomials $(T_n)_{n \geq 0}$ are normalized according to

$$\max_{t \in [-1,1]} T_n(t) = T(1) = \begin{cases} 1, & \text{if } n = 0, \\ \sqrt{2}, & \text{if } n \neq 0. \end{cases}$$

Consequently, the multivariate "Probabilistic" Chebyshev polynomials $(T_\nu)_{\boldsymbol{\nu} \in \mathcal{F}}$ are normalized according to

$$\|T_\nu\|_\infty = T_\nu((1, 1, ..., 1)) = \Big(\sqrt{2} \Big)^{\#\mathrm{supp}(\boldsymbol{\nu})}, \tag{2.3.7}$$

and therefore

$$K_T(\Lambda) = \sum_{\boldsymbol{\nu} \in \Lambda} 2^{\#\mathrm{supp}(\boldsymbol{\nu})}. \tag{2.3.8}$$

We prove (2.3.6) using induction on $n_\Lambda := \#(\Lambda)$. When $n_\Lambda = 1$, then $\Lambda = \{0_\mathcal{F}\}$ and it is obviously true. Let $n \geq 1$ and Λ denote a monotone set with $n_\Lambda = n+1$. We define J and the monotone set Λ_k as in the proof of Lemma 2.2. We have by the induction assumption

$$K_T(\Lambda) = \sum_{k=0}^{J} \gamma(k) K_T(\Lambda_k) \leq \sum_{k=0}^{J} \gamma(k) (\#(\Lambda_k))^{\frac{\ln 3}{\ln 2}}.$$

with $\gamma(0) = 1$ and $\gamma(k) = 2$ for $k \geq 1$. Therefore

$$K_T(\Lambda) \leq (\#\Lambda_0)^{\frac{\ln 3}{\ln 2}} + 2 \sum_{k=1}^{J} (\#\Lambda_k)^{\frac{\ln 3}{\ln 2}}.$$

Since $0 < \#(\Lambda_J) \leq \cdots \#(\Lambda_1) \leq \#(\Lambda_0)$, then by the inequality (2.3.5) in Lemma 2.3 we deduce

$$K_T(\Lambda) \leq (\#\Lambda_0 + ... + \#\Lambda_J)^{\frac{\ln 3}{\ln 2}} = (\#\Lambda)^{\frac{\ln 3}{\ln 2}},$$

which completes the proof. □

Lemma 2.5. *In the cases of Tensor Product and Total Degree polynomial spaces, the quantity $K(\Lambda)$ with the tensorized Legendre polynomials satisfies*

$$K(\Lambda) \leq \begin{cases} (\#\Lambda)^2, & \textit{Tensor Product}, \\ \left(\dfrac{2\mathrm{w}}{d} + 1\right)^d \ \#\Lambda, & \textit{Total Degree}. \end{cases} \tag{2.3.9}$$

Proof. The univariate Legendre polynomials orthonormal in $L^2_\rho(-1,1)$, *i.e.* $\|L_k\| = 1$, are such that $\|L_k\|_{L^\infty(-1,1)} = \sqrt{2k+1}$. Given $v \in \mathbb{P}_\Lambda$, it can be expanded as $v = \sum_{\boldsymbol{\nu}\in\Lambda} v_\nu \boldsymbol{L}_\nu$. As in Lemma 2.2, it holds

$$K(\Lambda) = \left(\sum_{\boldsymbol{\nu}\in\Lambda} \prod_{j=1}^{d} (2\nu_j + 1) \right) \tag{2.3.10}$$

for any monotone Λ and any value of d. Then, in the case of TP we obtain

$$K(\Lambda) = \prod_{j=1}^{d} \sum_{i=0}^{\mathrm{w}} (2i+1) = \prod_{j=1}^{d} \left(\mathrm{w} + 1 + 2 \sum_{i=0}^{\mathrm{w}} i \right) = \prod_{j=1}^{d} \left(\mathrm{w} + 1 + \frac{2\mathrm{w}(\mathrm{w}+1)}{2} \right) = (\#\Lambda)^2.$$

In the case of TD it holds

$$K(\Lambda) = \sum_{\boldsymbol{\nu}\in\Lambda} \prod_{j=1}^{d} (2\nu_j + 1) \leq \sum_{\boldsymbol{\nu}\in\Lambda} \left(\frac{2\mathrm{w}}{d} + 1\right)^d \leq \left(\frac{2\mathrm{w}}{d} + 1\right)^d \sum_{\boldsymbol{\nu}\in\Lambda} 1 \leq \left(\frac{2\mathrm{w}}{d} + 1\right)^d \ \#\Lambda.$$

□

Remark 2.1. *In the case of Total Degree space, a direct calculation of $K(\Lambda)$ shows that the condition* (2.3.11) *expressed in Lemma 2.2 is conservative. In particular*

$$\#\Lambda \leq K(\Lambda) < (\#\Lambda)^2, \tag{2.3.11}$$

where the second inequality is strict for all $d > 1$. Fig. 2.1 shows that as the dimension d increases, the value of $K(\Lambda)$ departs more and more from the bound $(\#\Lambda)^2$. The bound in Lemma 2.5 is useful only for $d > 5$ where the factor $\left(1 + \frac{2\mathrm{w}}{d}\right)^d$ is smaller than $\#\Lambda$.

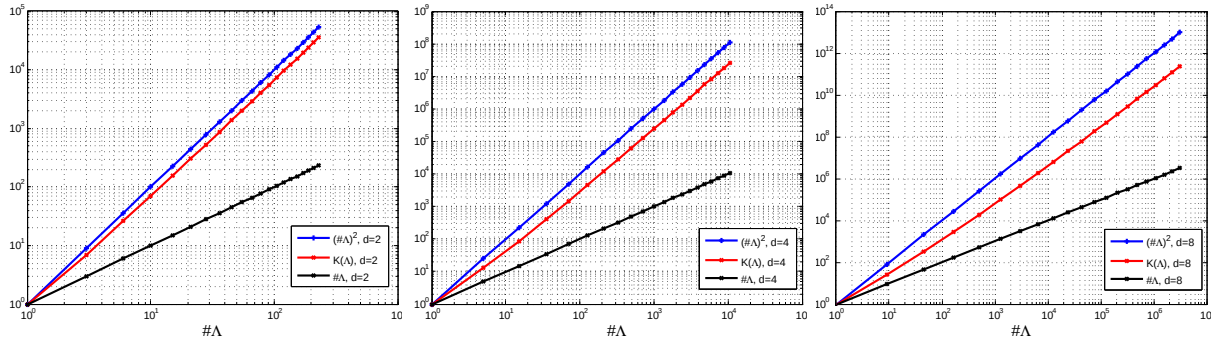


Figure 2.1: Comparison between $(\#\Lambda)^2$ and $K(\Lambda)$ in the case of Total Degree space. Left: $d = 2$. Center: $d = 4$. Right: $d = 8$.

2.4 Application to elliptic PDEs with random inclusions

We consider the stochastic elliptic boundary value problem

$$-\mathrm{div}(a\nabla u) = f \ \text{ in } \ D \subset \mathbb{R}^q, \ \ u = 0 \ \text{ on } \ \partial D, \tag{2.4.1}$$

where $f \in H^{-1}(D)$, q is a positive integer and

$$a(\mathbf{x}, \mathbf{y}) := \bar{a}(\mathbf{x}) + \sum_{j=1}^{d} y_j \psi_j(\mathbf{x}). \tag{2.4.2}$$

The $y_j, j = 1, \ldots, d$ are independent and uniformly distributed on $[-1, 1]$. Here d may be infinite; $\bar{a}$ and ψ_j are functions in $L^\infty(D)$ with nonoverlapping supports. We suppose that a satisfies the following uniform ellipticity assumption:

Assumption 2.2 (Uniform Ellipticity Assumption $\mathbf{UEA}(\mathbf{r}, \mathbf{R})$)**.**

$$0 < r \le a(x, y) \le R < \infty, \quad \forall\, x \in D, \quad \forall\, y \in \Gamma,$$

where $\Gamma := [-1, 1]^d$ *when* $d < \infty$ *and* $\Gamma := [-1, 1]^{\mathbb{N}}$ *when* $d = \infty$.

We introduce ρ the uniform measure on Γ defined by $d\rho := \otimes_{j\ge 1}^{d} \frac{dy_j}{2}$ and the Hilbert space $\mathcal{V} := L^2(\Gamma, V, d\rho)$ of square integrable functions taking value in V equipped with the inner product

$$\langle v, w\rangle := \int_\Gamma \Big(\int_D \nabla v(\mathbf{y}) \cdot \nabla w(\mathbf{y}) dx \Big) d\rho(\mathbf{y}). \tag{2.4.3}$$

The model (2.4.1) yields the variational problem: find $u \in \mathcal{V}$ such that

$$\mathcal{B}(u, v) = \mathcal{L}(v), \qquad \forall v \in \mathcal{V},$$

where the linear and bilinear form $\mathcal{L}$ and $\mathcal{B}$ are defined by,

$$\mathcal{B}(w, v) := \int_\Gamma \Big(\int_D a(x, \mathbf{y}) \nabla w(\mathbf{y}) \cdot \nabla v(\mathbf{y}) dx \Big) d\rho(\mathbf{y}), \qquad \mathcal{L}(v) := \int_\Gamma \Big(\int_D f(x) v(x, \mathbf{y}) dx \Big) d\rho(\mathbf{y}),$$

with the integrals over D to be understood as extensions by continuity from $L^2(D) \times L^2(D)$ to duality pairings between V^* and V. The Assumption 2.2 implies that there exists a unique solution $u \in \mathcal{V}$ to the variational problem, which satisfies

$$\|u(\mathbf{y})\|_V \le \frac{\|f\|_{V^*}}{r} =: \widetilde{\tau}, \quad \forall\, \mathbf{y} \in \Gamma. \tag{2.4.4}$$

As we did in Section 2.2.2 for the discrete projection P_N, we define the truncated discrete projection $\widetilde{P}_\Lambda$ on V_Λ,

$$\widetilde{P}_\Lambda^M u = T_{\widetilde{\tau}}(P_\Lambda^M u), \tag{2.4.5}$$

but using a truncation operator $T_{\widetilde{\tau}}$ defined as

$$T_{\widetilde{\tau}} v = \begin{cases} v & , \quad \text{if} \quad \|v\|_V \leq \widetilde{\tau}, \\ \dfrac{\widetilde{\tau}}{\|v\|_V}\, v & , \quad \text{if} \quad \|v\|_V > \widetilde{\tau}, \end{cases}, \tag{2.4.6}$$

with the threshold $\widetilde{\tau}$ instead of τ. It will be clear from the context if we use (2.2.9) or (2.4.6).

We have that $\mathcal{V}$ is isomorphic to $L^2(\Gamma, d\rho) \otimes V$ so that every v in $\mathcal{V}$ admits a unique expansion of the form

$$v = \sum_{\boldsymbol{\nu} \in \mathcal{F}} v_\nu \mathbf{L}_\nu, \quad \text{with} \quad v_\nu = \int_\Gamma v(\mathbf{y}) \mathbf{L}_\nu(\mathbf{y}) d\rho(\mathbf{y}) \in V, \quad \boldsymbol{\nu} \in \mathcal{F}.$$

We denote by

$$u = \sum_{\boldsymbol{\nu} \in \mathcal{F}} u_\nu \mathbf{L}_\nu,$$

the expansion of the solution u and introduce the sequence $(c_\nu)_{\boldsymbol{\nu} \in \mathcal{F}}$ with

$$c_\nu = \|u_\nu\|_V, \quad \boldsymbol{\nu} \in \mathcal{F}.$$

The summability properties of the sequence $(c_\nu)_{\boldsymbol{\nu} \in \mathcal{F}}$ are well understood. For $p > 0$, we introduce the space $\ell^p(\mathcal{F})$ of p-summable sequences indexed in $\mathcal{F}$. For a sequence $(a_\nu)_{\boldsymbol{\nu} \in \mathcal{F}}$, we introduce $\mathbf{a} := (\mathbf{a}_\nu)_{\boldsymbol{\nu} \in \mathcal{F}}$ the monotone envelope of $(a_\nu)_{\boldsymbol{\nu} \in \mathcal{F}}$ defined by

$$\mathbf{a}_\nu = \max_{\boldsymbol{\nu} \leq \boldsymbol{\mu}} |a_\mu|,$$

where $\leq$ is the order relation defined on $\mathcal{F}$ by $\boldsymbol{\nu} \leq \boldsymbol{\mu}$ iff $\nu_j \leq \mu_j$ for any j. The sequence $\mathbf{a}$ is monotone decreasing, in the sense that

$$\boldsymbol{\nu} \leq \boldsymbol{\mu} \Rightarrow \mathbf{a}_\mu \leq \mathbf{a}_\nu.$$

For $p > 0$, we introduce the space $\ell^p_m(\mathcal{F})$ of sequences that have their monotone envelopes in $\ell^p(\mathcal{F})$. The following result has been proved in [CDS11].

Theorem 2.5. *Under the uniform ellipticity assumption 2.2 and if the sequence $(\|\psi_j\|_{L^\infty(D)})_{j \geq 0}$ belongs to $\ell^p(\mathbb{N})$ for some $p < 1$, then the sequence $(c_\nu)_{\nu \in \mathcal{F}}$ belongs to $\ell^p_m(\mathcal{F})$.*

This result implies in particular that if Λ_N is the set of multi-indices in $\mathcal{F}$ corresponding to the N largest values of $\mathbf{c}_\nu$, then

$$\|u - \sum_{\boldsymbol{\nu} \in \Lambda_N} u_\nu L_\nu\|_{\mathcal{V}} = \Big(\sum_{\boldsymbol{\nu} \notin \Lambda_N} c_\nu^2 \Big)^{\frac{1}{2}} \leq \Big(\sum_{\boldsymbol{\nu} \notin \Lambda_N} \mathbf{c}_\nu^2 \Big)^{\frac{1}{2}} \leq \|\mathbf{c}\|_{\ell^p(\mathcal{F})} N^{-s}, \qquad s = \frac{1}{p} - \frac{1}{2}$$

where we have used Stechkin formula with the p-summable sequence $(\mathbf{c}_\nu)_{\boldsymbol{\nu} \in \mathcal{F}}$. Let us observe that since the sequence $(\mathbf{c}_\nu)_{\boldsymbol{\nu} \in \mathcal{F}}$ is monotone decreasing, then we can always choose the set Λ_N to be *monotone*, i.e.

$$\boldsymbol{\nu} \in \Lambda \ \text{ and } \ \boldsymbol{\mu} \leq \boldsymbol{\nu} \Rightarrow \boldsymbol{\mu} \in \Lambda.$$

This shows the existence of monotone sets that are quasi-optimal in the sense of best term approximation w.r.t. the monotone envelope. In practice, the construction of the sets Λ_N described earlier is impossible since the sequence $(c_\nu)_{\boldsymbol{\nu} \in \mathcal{F}}$ and thus $(\mathbf{c}_\nu)_{\boldsymbol{\nu} \in \mathcal{F}}$ is unknown. We know instead a sequence of computable a priori estimates $(q_\nu)_{\boldsymbol{\nu} \in \mathcal{F}}$ that is monotone decreasing and satisfies $|c_\nu| \leq q_\nu$. The sequence of nested monotone sets $(\Lambda_N)_{\boldsymbol{\nu} \in \mathcal{F}}$ corresponding each to the multi-indices of the N largest q_ν with ties broken arbitrarily, but with the condition that the monotonicity and nestedness hold, satisfies

$$\|u - \sum_{\boldsymbol{\nu} \in \Lambda_N} u_\nu L_\nu\|_{\mathcal{V}} \leq \|q\|_{\ell^p(\mathcal{F})} N^{-s}, \qquad s = \frac{1}{p} - \frac{1}{2}.$$

2.4.1 Discrete least squares approximation of parametric elliptic PDEs

We are also interested in the approximation of u using the discrete least squares approximation discussed in Section 2.2. Here the function u is not scalar but the definition of the least squares approximation introduced in (2.2.2) extends straightforwardly, with $\|\cdot\|_{L^2(\Gamma,V,d\rho)}$ replacing $\|\cdot\|$ and $\|\cdot\|_V$ replacing the absolute value in (2.2.3). We introduce the space $\mathcal{V}_\Lambda := \mathbb{P}_\Lambda \otimes V$, then

$$P_\Lambda u = \underset{v\in\mathcal{V}_\Lambda}{\operatorname{argmin}} \|u-v\|_{\mathcal{V}} \quad \text{and} \quad P_\Lambda^M u = \underset{v\in\mathcal{V}_\Lambda}{\operatorname{argmin}} \|u-v\|_M^V, \tag{2.4.7}$$

where

$$\|v\|_M^V = \Big(\frac{1}{M}\sum_{i=1}^M \|v(\mathbf{y}_i)\|_V^2\Big)^{\frac{1}{2}}, \tag{2.4.8}$$

and $\mathbf{y}_i$ are i.i.d. samples from the probability measure ρ. We have that

$$P_\Lambda u = \sum_{\boldsymbol{\nu}\in\Lambda} u_\nu \mathbf{L}_\nu.$$

The least squares projection is given by

$$P_\Lambda^M u = \sum_{j=1}^M u(\mathbf{y}_j)\pi_j, \tag{2.4.9}$$

where the π_i are polynomials in $\mathbb{P}_\Lambda$ given by the following argument. As in Section 2.2, we introduce the notation $\mathcal{B}_L := \{\mathbf{L}_{\nu^1},\dots,\mathbf{L}_{\nu^N}\}$. Then the family of polynomials $\mathcal{B}_\pi := \{\pi_1,\dots,\pi_M\}$ is given by

$$\mathcal{B}_\pi = \Big(\mathbf{G}^{-1}\mathbf{J}\Big)^t \mathcal{B}_L.$$

where this time $\mathbf{G} := \Big(\langle L_{\nu^i}, L_{\nu^j}\rangle_M\Big)_{1\le i,j\le N}$ and $\mathbf{J} = \frac{1}{M}(L_{\nu^i}(y_j))_{\substack{1\le i\le N\\ 1\le j\le M}}$. To prove (2.4.9), we fix $\alpha \in V$ and consider the scalar functions

$$U_\alpha : \mathbf{y} \mapsto \langle u(\mathbf{y}),\alpha\rangle_V \quad \text{and} \quad W_\alpha : \mathbf{y} \mapsto \langle P_\Lambda^M u(\mathbf{y}),\alpha\rangle_V,$$

that belongs to $L^2(\Gamma, d\rho)$ and $\mathbb{P}_\Lambda$ respectively. The projection $P_\Lambda^M u$ satisfies the orthogonality relation

$$\langle P_\Lambda^M u, v\rangle_M^V = \langle u, v\rangle_M^V, \qquad v \in \mathcal{V}_\Lambda,$$

therefore replacing v with any polynomial $P \in \mathbb{P}_\Lambda$, yields

$$\langle W_\alpha, P\rangle_M = \langle U_\alpha, P\rangle_M, \qquad P \in \mathbb{P}_\Lambda, \quad \forall \alpha \in V,$$

with $\langle\cdot,\cdot\rangle_M$ being the empirical inner product. Therefore, W_α is the least square approximation of U_α in $\mathbb{P}_\Lambda$. We deduce from the scalar case that

$$W_\alpha = \sum_{j=1}^M U_\alpha(\mathbf{y}_j)\pi_j.$$

This implies

$$\langle P_\Lambda^M u(\mathbf{y}),\alpha\rangle_V = \langle \sum_{j=1}^M u(\mathbf{y}_i)\pi_j(\mathbf{y}),\alpha\rangle_V, \quad \forall\, \mathbf{y}\in\Gamma, \quad \forall\, \alpha \in V,$$

which completes the justification. □

2.4.2 The case of nonoverlapping inclusions: approximation in Total Degree polynomial spaces

In [CDS10], it is shown that there exists a sequence of nested monotone sets $(\Lambda_N)_{N\geq 0}$ that can be built using a priori estimates on the Legendre coefficients. However, the sets in the sequence $(\Lambda_N)_{N\geq 0}$ are difficult to characterize, and might lead to approximations quite far from the best N-term due to the fact the a priori estimations on the Legendre coefficients are not sharp.

In this work, we focus on the subclass of stochastic PDEs (2.4.1), (2.4.2) characterized by the fact that the functions ψ_j have nonoverlapping support. This situation allows to model, for instance, the diffusion process in a medium with nonoverlapping inclusions of random conductivity (see *e.g.* Fig. 2.2). In this case, a priori estimates on the Legendre coefficients have been obtained *e.g.* in [BNTT12] and have been shown numerically to be quite sharp. They read:

$$\|u_\nu\|_V \leq C \prod_{j=1}^{d} \exp\{-\nu_j\, g_j\}, \quad \forall\, \boldsymbol{\nu} = (\nu_1, \ldots, \nu_d) \in \mathbb{N}^d.$$

Explicit expressions for the constant C can be found in [BNTT12, Corollary 8]. The coefficients $(g_j)_{1\leq j\leq d}$ can be tuned through an *a posteriori* procedure, that requires to solve only "one-dimensional" problems, *i.e.* analyzing the convergence when considering one random variable at a time and freezing all other variables to their expected value. In [BNTT12] it is also shown that the quasi-optimal sets associated with the problems in the aforementioned class are

$$\Lambda_{\mathrm{w}} = \{\boldsymbol{\nu} \in \mathbb{N}^d \,:\, \sum_{j=1}^{d} g_j \nu_j \leq \mathrm{w}\}, \qquad \mathrm{w} = 1, 2, \ldots$$

and correspond to anisotropic Total Degree spaces, *i.e.* the anisotropic variants of (2.3.2). Analogous estimates, showing the optimality of the TD sets, have been presented in [CCDS13].

For convenience we introduce the constants c_2, c_3 defined as:

$$c_2 := \frac{gd}{e}, \tag{2.4.10}$$

$$c_3 := \frac{\widehat{C}^2_{\mathrm{Leg}}}{(1-e^{-g})^d} \exp\left\{\frac{2\,e^2\,(1-e^{-1})\,c_2}{5}\right\}. \tag{2.4.11}$$

The expression of $\widehat{C}_{\mathrm{Leg}}$ can be recovered from [BNTT12, Corollary 8]. It depends on d, g_n and is related to the *Polydisc Analyticity Assumption* [BNTT12, Assumption 3].

Lemma 2.6. *In the isotropic case,* i.e. $g_n = g$ *for all* $n = 1, \ldots, d$*, the following estimate on the error of the* L^2 *projection* P_Λ *on the quasi-optimal TD sets holds*

$$\|u - P_\Lambda u\|^2_{\mathcal{V}} \leq c_3 \exp\left\{-c_2 N^{\frac{1}{d}}\right\}. \tag{2.4.12}$$

Proof. The following estimate has been obtained in [BNTT12, Theorem 22]:

$$\|u - P_\Lambda u\|^2_{\mathcal{V}} \leq \frac{\widehat{C}^2_{\mathrm{Leg}}}{(1-e^{-g})^d} \exp\left\{-c_2 \log\left(\left(1-\epsilon(N)\right)^{-1}\right) N^{\frac{1}{d}}\right\}, \tag{2.4.13}$$

with

$$\epsilon(N) = (1-e^{-1})\left(1 - \frac{2e}{5N^{\frac{1}{d}}}\right). \tag{2.4.14}$$

When $(2e/5)^d < N$ then $(1-\epsilon(N)) < 1$, and moreover

$$\lim_{N\to\infty} (1-\epsilon(N))^{\left(c_2 N^{\frac{1}{d}}\right)} = 0,$$

because the exponent is always positive and the basis is smaller than one. Introducing the change of variable z as

$$z = N^{\frac{1}{d}}, \tag{2.4.15}$$

using the definition of c_2 in (2.4.10) and replacing ϵ by (2.4.14), then the exponential term on the right side in (2.4.13) can be manipulated as

$$\begin{aligned}\left(e^{-1} + \frac{2\,e\,(1-e^{-1})}{5\,z}\right)^{c_2\,z} &= \left(1 + \frac{2\,e^2\,(1-e^{-1})}{5\,z}\right)^{c_2\,z} \cdot e^{-c_2\,z} \\ &< e^{\frac{2\,e^2\,(1-e^{-1})}{5}\,c_2} \cdot e^{-c_2\,z}. \end{aligned} \tag{2.4.16}$$

The last passage in (2.4.16) is justified observing that for any $a > 0$, $b > 0$, $z > 0$ it holds

$$\left(1+\frac{a}{z}\right)^{bz} < e^{ab} \quad \text{and} \quad \lim_{z\to+\infty}\left(1+\frac{a}{z}\right)^{bz} = e^{ab}.$$

Note from (2.4.15) that the limit $N \to +\infty$ is equivalent to $z \to +\infty$. Thanks to (2.4.16) we can bound the exponential term on the right of (2.4.13), and using the definition (2.4.11) of c_3 we obtain (2.4.12). □

2.4.3 Convergence of the least squares approximation

In this subsection we derive an estimate for the expected L^2 error $\mathbb{E}(\|u - \widetilde{P}_\Lambda^M u\|_\mathcal{V}^2)$ of the least squares approximation in terms of the number of sampling points M. To do this we rely on the estimates regarding the exact L^2 projection on Total Degree polynomial spaces that have been recalled in Section 2.4.2. To begin with, we will use the isotropic estimate (2.4.12). The extension to anisotropic problems will be presented in a forthcoming work.

To lighten the notation we introduce the constant

$$c_1 := 8\,\widetilde{\tau}^2, \tag{2.4.17}$$

with $\widetilde{\tau}$ coming from (2.4.4), and the factor

$$c_4 = c_4(M) := \big(1+\eta(M)\big)\,c_3, \tag{2.4.18}$$

that contains the proportionality constant appearing in (2.2.10). Note that $\eta = \eta(M)$ is a decreasing function and in practice its value is such that $c_4 \approx c_3$. Moreover, we name concisely as

$$l(u) := \|u - \widetilde{P}_\Lambda^M u\|_\mathcal{V}$$

the truncated error with the threshold $\widetilde{\tau}$ defined by means of the truncation operator in (2.4.5). We recall that $\widetilde{\tau}$ is related to the boundedness of the solution u by (2.4.4).

Theorem 2.6. *In the aforementioned PDE model class, the convergence rate of the least squares approximation with optimal choice of the polynomial space satisfies*

$$\mathbb{E}\Big[l(u)^2\Big] \le (c_4 + c_1)\,\exp\left\{-\left(c_2^{2d}\,\beta\,M\right)^{\frac{1}{2d+1}}\right\}. \tag{2.4.19}$$

Proof. The first step to characterize the optimal convergence rate w.r.t. M is to impose a relation between M, β, γ and N to have a stable least squares approximation. In the case of polynomial approximation, the relation (2.3.4) holds choosing the multi-index set Λ such that

$$N = \left(\frac{\beta}{\gamma+1}\frac{M}{\log M}\right)^{\frac{1}{2}}, \tag{2.4.20}$$

and this implies that (2.2.6) is satisfied. Therefore, the relation (2.4.20) prescribes how to enlarge the dimension of the polynomial space as M increases, to ensure stability and optimality of the random least squares projection thanks to Theorem 2.3. Of course the value of γ can be chosen optimally, and we will pursue this strategy in the sequel. Replacing N by (2.4.20) into the right side of (2.4.12) we have

$$\|u - P_\Lambda u\|_{\mathcal{V}}^2 \leq c_3 \exp\left\{-c_2\left(\frac{\beta M}{(\gamma+1)\log M}\right)^{\frac{1}{2d}}\right\} \lesssim c_3 \exp\left\{-c_2\left(\frac{\beta M}{\gamma \log M}\right)^{\frac{1}{2d}}\right\}. \tag{2.4.21}$$

Since we embedded the stability condition (2.2.6) as a constraint, then we can apply (2.2.10) (with the error evaluated in the $\mathcal{V}$-norm) and use (2.4.21) to bound the error on the right, obtaining

$$\mathbb{E}\Big[l(u)^2\Big] \leq c_4 \cdot e^{-\left(c_2\left(\frac{\beta M}{\gamma \log M}\right)^{\frac{1}{2d}}\right)} + c_1\, e^{-\gamma \log M}. \tag{2.4.22}$$

Note the factor $\gamma \log M$ in both the exponents of (2.4.22). Now we can choose $\gamma = \gamma(M, \beta)$ so that the exponents of the two exponential terms in (2.4.22) are equal, imposing

$$c_2\left(\frac{\beta M}{\gamma \log M}\right)^{\frac{1}{2d}} = \gamma \log M. \tag{2.4.23}$$

From (2.4.23) we obtain the explicit expression of γ as a function of M and β:

$$\gamma = \frac{\left(c_2^{2d}\,\beta\, M\right)^{\frac{1}{2d+1}}}{\log M}. \tag{2.4.24}$$

Finally, substituting γ with (2.4.24) in (2.4.22), we obtain the convergence rate (2.4.19) optimized w.r.t. γ. □

Looking at (2.4.19) we observe that:

- the error converges to zero exponentially fast as $\exp\Big\{-(c_2^{2d}M)^{\frac{1}{2d+1}}\Big\}$,
- the convergence rate depends on the dimension d, it is accelerated due to the presence of $d^{\frac{2d}{2d+1}}$ but also decelerated from the presence of d in the exponent of the term $M^{\frac{1}{2d+1}}$,
- an equivalence with Stochastic Galerkin is established. The error of the best N-term approximation converges to zero with the rate $\exp\{-c_2 N^{\frac{1}{d}}\}$ (see (2.4.12)). The random discrete projection converges to zero with the rate $\exp\Big\{-(c_2^{2d}M)^{\frac{1}{2d+1}}\Big\}$, with $M \sim N^2$.

2.5 Numerical results

In this section we present some numerical examples that confirm the theoretical findings presented in sections 2.2 and 2.4. In particular, we check that the convergence rate (2.4.19) is sharp when the number of sampling points M is chosen as

$$M = \frac{c_2}{\beta}\Big(N\Big)^{\left(2+\frac{1}{d}\right)}, \tag{2.5.1}$$

that comes from the optimal choice of γ prescribed by (2.4.24). To investigate the sensitivity w.r.t. to β, we denote again $\beta = \beta(\delta)$, for different choices of δ such that $\beta = 0.15$ or $\beta = 0.25$.

We consider the elliptic model (2.4.1) on the bounded domain $\Omega \subset \mathbb{R}^2$ shown in Fig. 2.2,

$$\begin{cases} -\nabla \cdot \Big(a(\mathbf{x},\mathbf{y})\,\nabla u\Big) = f(\mathbf{x}), & \mathbf{x} \text{ in } \Omega,\ \mathbf{y} \in \Gamma, \\ u = 0, & \mathbf{x} \text{ on } \partial\Omega,\ \mathbf{y} \in \Gamma, \end{cases} \tag{2.5.2}$$

with a being the random diffusion coefficient defined in (2.5.3) with the geometry displayed in Fig. 2.2. The eight inclusions $\Omega_1, \ldots, \Omega_8$ are circular with radii equal to 0.13, and are centered in the points $\mathbf{x} = (0.5, 0.5 \pm 0.3)$, $\mathbf{x} = (0.5 \pm 0.3, 0.5)$ and $\mathbf{x} = (0.5 \pm 0.3, 0.5 \pm 0.3)$. The 0.2-times-0.2 inner square Ω_0 lies in the center of Ω. The forcing term f is equal to 100 in Ω_0 and zero in $\Omega \setminus \Omega_0$. The random diffusion coefficient depends on a multivariate uniform random variable $\mathbf{Y} \sim \mathcal{U}([-1,1]^d)$, and is defined as

$$a(\mathbf{x},\mathbf{Y}) = \begin{cases} 0.395\Big(Y_i + 1\Big) + 0.01, & \mathbf{x} \in \Omega_i, \quad i = 1, \ldots, 8, \\ 1, & \mathbf{x} \in \Omega_0, \end{cases} \tag{2.5.3}$$

such that each random variable is associated with an inclusion. The range of variation of the coefficient in each inclusion is therefore $[0.01, 0.8]$. This test case has been used in [BNTT12], and allows a direct comparison of our results with those obtained when employing the classical Stochastic Galerkin method. The monodimensional convergence rate $g = 1.9$ of this example is has been estimated in [BNTT11b, Fig.7-left]. Note that the coefficient a in (2.5.3) satisfies Assumption 2.2.

We consider the following Quantity of Interest, related to the solution of the elliptic model (2.5.2):

$$\mathrm{QOI}_1(\mathbf{Y}) = \frac{1}{|\Omega|}\int_\Omega u(\mathbf{x},\mathbf{Y})\,d\mathbf{x},$$

and present the results obtained when approximating this function in the polynomial space. Similar results hold also with other Quantities of Interest, *e.g.*

$$\mathrm{QOI}_2(\mathbf{Y}) = \frac{1}{|\Omega|}\int_\Omega \Big|\nabla u(\mathbf{x},\mathbf{Y})\Big|^2\,d\mathbf{x}, \quad \mathrm{QOI}_3(\mathbf{Y}) = \frac{1}{|\Omega_0|}\int_{\Omega_0} u(\mathbf{x},\mathbf{Y})\,d\mathbf{x},$$

which will not be shown here. We consider three cases with $d = 2$, $d = 4$, $d = 8$ independent random variables. In the case $d = 2$, the first random variable describes the diffusion coefficient in the four inclusions at the top, bottom, left, right of the center square Ω_0. The second random variable describes the diffusion coefficient in the other four inclusions. In the case $d = 4$, each one of the four random variables is associated with two opposite inclusions w.r.t. the center of the domain. When $d = 8$ each of the random variables is associated with a different inclusion.

The Figs. 2.3, 2.4, 2.5 show the convergence plots obtained by the random least squares approximation using a number of samples as in (2.5.1) with two different choices of β (*i.e.* $\beta = 0.15$ and $\beta = 0.25$). The theoretical bound (2.4.19) is also shown as well as the reference slope $M^{-1/2}$ of a standard Monte Carlo method. In the same figures we also show the convergence plots obtained when using a simple linear rule $M = 3N$ or $M = 10N$.

The approximation error of the random least squares projection is approximated as

$$\mathbb{E}\left[\|\mathrm{QOI}_1(u) - \widetilde{P}_\Lambda^M \mathrm{QOI}_1(u)\|\right] \approx \mathbb{E}\left[\|\mathrm{QOI}_1(u) - \widetilde{P}_\Lambda^M \mathrm{QOI}_1(u)\|_{cv}\right],$$

employing the cross-validation procedure described in [MNvST11, Section 4].

The results presented in Figs. 2.3, 2.4, 2.5 show that the proposed bound (2.4.19) predicts very sharply the error $\mathbb{E}\Big[l(u)^2\Big]$, when the number of sampling points M is chosen according to (2.5.1). The bound accurately describes the effect of the dimension d as well, in the cases of moderately high dimensions.

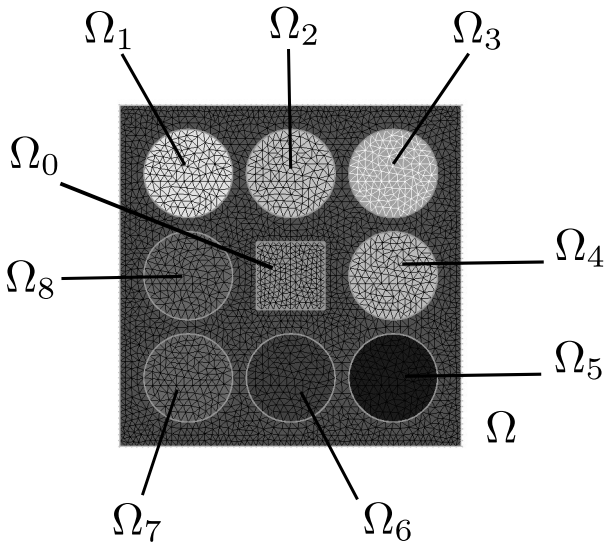


Figure 2.2: Mesh discretization and geometries of the inclusions. The domain Ω is the unitary square. The inner square is named Ω_0, the eight circular inclusions are $\Omega_1, \ldots, \Omega_8$.

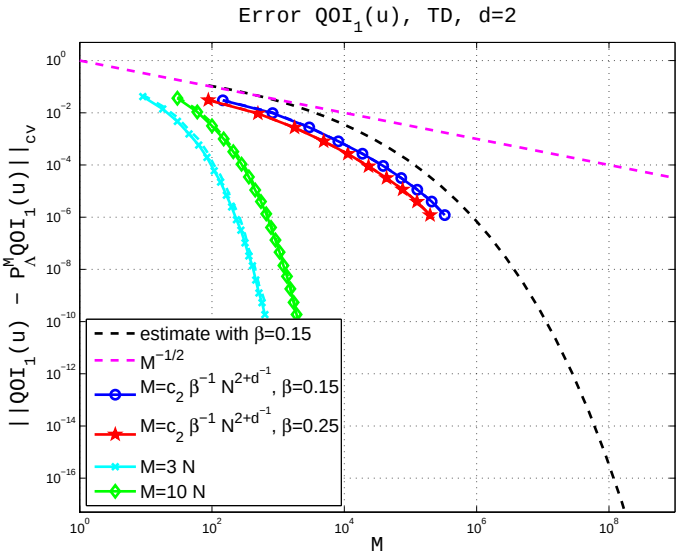


Figure 2.3: Error $\mathbb{E}\left[l(u)^2\right]$ in the case $d = 2$. Different relations between the number of samples M and the dimension of the polynomial space N are tested. The black dash line is the bound (2.4.19). The magenta dash line is the Monte Carlo convergence rate $M^{-1/2}$.

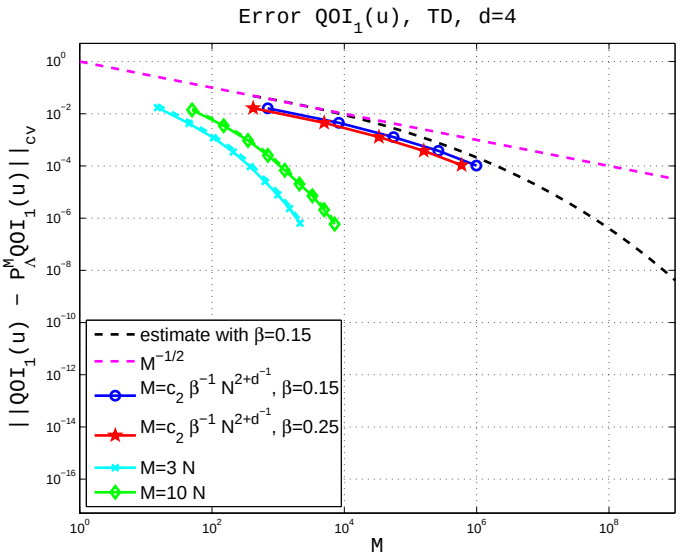


Figure 2.4: Error $\mathbb{E}\left[l(u)^2\right]$ in the case $d = 4$. Different relations between the number of samples M and the dimension of the polynomial space N are tested. The black dash line is the bound (2.4.19). The magenta dash line is the Monte Carlo convergence rate $M^{-1/2}$.

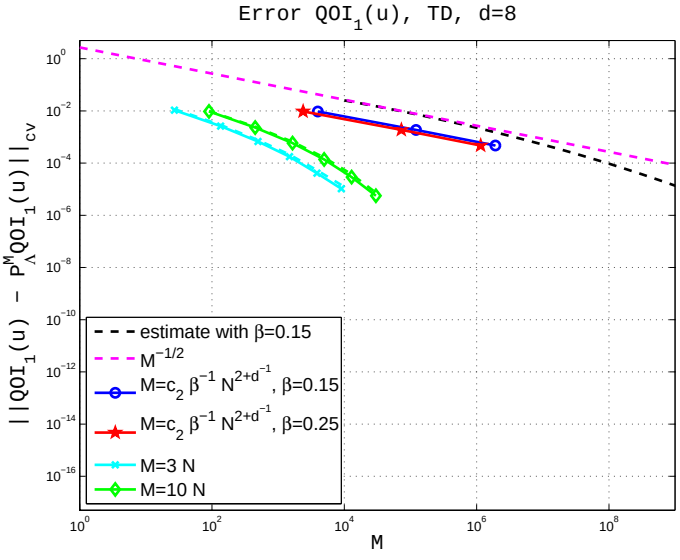


Figure 2.5: Error $\mathbb{E}\left[l(u)^2\right]$ in the case $d = 8$. Different relations between the number of samples M and the dimension of the polynomial space N are tested. The black dash line is the bound (2.4.19). The magenta dash line is the Monte Carlo convergence rate $M^{-1/2}$.

On the other hand, a faster convergence of the error $\mathbb{E}\Big[l(u)^2\Big]$ w.r.t. M is observed, with the linear scaling $M \propto N$ that yields a lower number of sampling points than (2.5.1), for a given set Λ. The efficiency of the linear scaling has been pointed out in [MNvST13], and its importance is motivated by the impossibility to employ the number of sampling points (2.5.1) when the dimension d is large. Fig. 2.5 shows that already when $d = 8$, the exponential gain of the bound (2.4.19) with respect to a Monte Carlo rate becomes perceivable only with an astronomical number of samples, making the choice (2.5.1) less attractive for the applications.

2.6 Conclusions

In this work the approximation technique based on least squares with random evaluations has been analyzed. The condition between the number of sampling points and the dimension of the polynomial space, which is necessary to achieve stability and optimality, has been extended to any monotone set of multi-indices identifying the polynomial space, in any dimension of the parameter set, and with the uniform and Chebyshev densities. When the density is uniform, this condition requires the number of sampling points to scale as the square of the dimension of the polynomial space.

Afterwards, this technique has been applied to a class of "inclusion-type" PDE models which includes the elliptic model and the linear elasticity model with stochastic coefficients, and an exponential convergence estimate in expectation has been derived. This estimate clarifies the dependence of the convergence rate on the number of sampling points and on the dimension of the parameter set. Moreover, this estimate establishes a relation between the convergence rate of the least squares approximation with random evaluations and the convergence rate of the classical Stochastic Galerkin method.

The numerical tests presented show that the proposed estimate is sharp, when the number of sampling points is chosen according to the condition that ensures stability and optimality. In addition, these results show that, in the aforementioned model class, a linear scaling of the number of sampling points w.r.t. the dimension of the polynomial space can yield a faster convergence rate.

Chapter 3

The cases of beta, Gaussian and gamma probability densities

The analysis developed in Chapter 1, concerning stability and optimality of the random discrete L^2 projection, is presented under the Assumption 1.1 on the probability density ρ:

Assumption 1.1. $0 < \rho_{min} \leq \rho(\mathbf{y}) \leq \rho_{max} < +\infty$, *for all* $\mathbf{y} \in \Gamma$.

Several numerical results featuring the uniform density are shown in Section 1.5. In this chapter the more general beta family of densities is considered, and it is shown that Assumption 1.1 is strictly required in our analysis of the discrete random projection, *i.e.* the stability and optimal convergence results do not hold if the probability density is not bounded away from zero. On the other hand, it is shown that the boundedness from above is not required, and an example with the *arcsine* density is given. Finally the Gaussian and gamma densities are considered, which have an unbounded support and therefore cannot be bounded away from zero.

3.1 The assumption on the density

The optimality proof in Theorem 1.2, Chapter 1, relies on the exact expression of order statistics. The main difficulty in extending this result to probability distributions other than the uniform one comes from the fact the distances of any two consecutive points are not identically distributed any more. If the probability density ρ of the sampling points satisfies Assumption 1.1, namely $\rho_{min} < \rho < \rho_{max}$, then the same relation between the discrete norm $\|\cdot\|_M$ and the continuous one $\|\cdot\|_{L^2_\rho}$ as in Theorem 1.3 holds up to a multiplicative factor involving the ratio ρ_{max}/ρ_{min} and the analysis can be recast to the case of uniform random variables, irrespectively of the exact shape of the density ρ.

However, an attempt to extend the same proof to any density different from the uniform gets stuck in Lemma 1.1, because of the additive probability argument used in its proof. Therefore, with any general probability density the optimality of the error evaluated in the $\|\cdot\|_{L^2_\rho}$ norm stated in Proposition 1.1 still holds, but the optimal convergence rate lacks of a probability quantification, *i.e.* a result like Theorem 1.2.

In general, the approximation properties of the random projection completely depend on the statistical distribution of the random variable C^ω defined in (1.2.15), that is not easily quantifiable.

3.2 Polynomial Chaos expansions

The same notation as in Chapter 1 is used. In particular, N denotes the dimension of the support Γ of the random variable $\mathbf{Y}$. Here Polynomial Chaos for different families of continuous and discrete probability densities have been introduced in [XK02b]. Our aim in the present section is to approximate, in a suitable polynomial space, a given univariate smooth function $\phi : \Gamma \to \mathbb{R}$,

that depends on a random variable $Y \sim \rho$, *i.e.* $\phi = \phi(Y)$. In the weighted scalar product introduced in (1.2.2) any probability density ρ can be used. When the density has been chosen, then a basis $\{\varphi_j\}_{j=1}^{\mathrm{w}}$ of the polynomial space $\mathbb{P}_{\mathrm{w}}$ with the property of being orthonormal with respect to the density ρ has to be built. The following cases for the density are considered:

- $\rho \sim \mathcal{N}(0,1)$, *i.e.* Gaussian density with zero mean and unitary variance,
- $\rho \sim \mathcal{G}(k,\theta)$, *i.e.* gamma density with the shape parameter $\theta = 1$, and $k > 0$,
- $\rho \sim \mathcal{B}(\alpha,\beta)$, *i.e.* beta density with $\alpha > 0$ and $\beta > 0$,

in addition to the case $\rho \sim \mathcal{U}([-1,1])$ that has been already analyzed in Chapter 1. For any ρ the family $\{\widetilde{\varphi}_j\}_j$ of one-dimensional orthogonal polynomials is built using the *Rodrigues' formula*

$$\widetilde{\varphi}_j(Y) = \Big(c_j \rho(Y)\Big) \frac{d^j}{dY^j}\Big(\rho(Y)(p(Y))^j\Big),$$

where c_j is a normalization factor and p is a polynomial in Y that does not depend on j. In this way, an orthogonal basis

$$\langle \widetilde{\varphi}_i, \widetilde{\varphi}_j \rangle_{L^2_\rho} = \delta_{ij} C^j_\rho$$

is built. Then the polynomials $\{\widetilde{\varphi}_j\}_j$ are orthonormalized by

$$\varphi_j = \frac{\widetilde{\varphi}_j}{\sqrt{C^j_\rho}}, \qquad j = 1, \ldots, \mathrm{w}.$$

The following constants C^j_ρ make orthonormal the family $\{\varphi_j\}_j$, for each choice of the density:

- *Gaussian* density: $\{\varphi_j\}_j$ Hermite polynomials with

  $$C^j_{\mathcal{N}} = \sqrt{2\pi}\,(j-1)!;$$

- *gamma* density: $\{\varphi_j\}_j$ (generalized) Laguerre polynomials with

  $$C^j_{\mathcal{G}}(k) = \frac{\Gamma(j+k+1)}{j!};$$

- *beta* density: $\{\varphi_j\}_j$ Jacobi polynomials with

  $$C^j_{\mathcal{B}}(\widetilde{\alpha},\widetilde{\beta}) = \frac{2^{\widetilde{\alpha}+\widetilde{\beta}+1}\Gamma(j+\widetilde{\alpha}+1)\Gamma(j+\widetilde{\beta}+1)}{(2j+\widetilde{\alpha}+\widetilde{\beta}+1)\Gamma(j+\widetilde{\alpha}+\widetilde{\beta}+1)j!};$$

  the Jacobi polynomials are orthonormal w.r.t. the weight $(Y-1)^{\widetilde{\alpha}}(Y+1)^{\widetilde{\beta}}$, and therefore $\widetilde{\alpha} = \alpha - 1 > 0$, $\widetilde{\beta} = \beta - 1 > 0$ since $\mathcal{B}(\alpha,\beta) \sim (Y-1)^{\alpha-1}(Y+1)^{\beta-1}$;

- *uniform* density: $\{\varphi_j\}_j$ Legendre polynomials with

  $$C^j_{\mathcal{U}} = \frac{2}{2j+1}.$$

From the context it is clear when the symbol Γ denotes the *Gamma Euler function* or the support of the density ρ. In the framework of Polynomial Chaos it is common to employ the gamma density with $k = 0$. In this case the generalized Laguerre polynomials are simply renamed Laguerre polynomials, and therefore $C^j_{\mathcal{G}} = 1$. In the rest of this chapter, we will present several univariate numerical tests with the aforementioned probability densities.

In the multivariate case, the multi-dimensional family $\{\psi_j\}_{j=1}^{\Lambda_{\mathrm{w}}}$ of polynomials can be built as in (1.2.3). Heterogeneous families of orthogonal polynomials, that differ in each direction, can be employed as well.

3.3 On the stability constraint with the beta density

In this section we show how the quantity K defined in equation (2.2.4) as

$$K^{\widetilde{\alpha}}(\mathrm{w}) = \sup_{y\in[-1,1]} \sum_{j=1}^{\mathrm{w}} \left(\varphi_j^{(\widetilde{\alpha},\widetilde{\alpha})}(y)\right)^2 \tag{3.3.1}$$

depends on the probability density ρ chosen in the beta family $\mathcal{B}(\alpha,\alpha)$, parametrized by the parameter $\alpha = \widetilde{\alpha}+1$, when $\widetilde{\alpha} \geq 0$. We remark that (3.3.1) states a relation between the number of sampling points and the dimension of the polynomial space, such that the random discrete projection is stable and accurate with high probability.

The evaluation of $\widetilde{\varphi}_j^{(\widetilde{\alpha},\widetilde{\alpha})}$ in the right endpoint of the interval $[-1,1]$ is equal to

$$\widetilde{\varphi}_j^{(\widetilde{\alpha},\widetilde{\alpha})}(1) = \binom{n+\widetilde{\alpha}+1}{n}, \tag{3.3.2}$$

see [Ask75, Equation (2.7)]. Moreover, the following symmetry relation [Sze39, Equation (4.1.3)] holds

$$\widetilde{\varphi}_j^{(\widetilde{\alpha},\widetilde{\alpha})}(-y) = (-1)^j\ \widetilde{\varphi}_j^{(\widetilde{\alpha},\widetilde{\alpha})}(y), \quad y\in[-1,1], \quad j\geq 0. \tag{3.3.3}$$

Using (3.3.2) and (3.3.3) we obtain the following expression for the evaluations of the polynomials $\varphi_j^{(\widetilde{\alpha},\widetilde{\alpha})}$ in the endpoints of the interval $[-1,1]$:

$$\left|\varphi_j^{(\widetilde{\alpha},\widetilde{\alpha})}(-1)\right| = \left|\varphi_j^{(\widetilde{\alpha},\widetilde{\alpha})}(1)\right| = \left|\frac{1}{\widetilde{\alpha}!}\sqrt{\frac{(j+2\widetilde{\alpha})!(2j+2\widetilde{\alpha}+1)}{j!2^{2\widetilde{\alpha}+1}}}\right|. \tag{3.3.4}$$

Therefore, the expression (3.3.4) can be used to get a lower bound of $K^{\widetilde{\alpha}}(\mathrm{w})$ as follows:

$$K^{\widetilde{\alpha}}(\mathrm{w}) \geq \frac{1}{\widetilde{\alpha}!2^{2\widetilde{\alpha}+1}} \sum_{j=1}^{\mathrm{w}} \sqrt{\frac{(j+2\widetilde{\alpha})!(2j+2\widetilde{\alpha}+1)}{j!}}. \tag{3.3.5}$$

Fig. 3.1 shows how the bound (3.3.5) performs for different values of the parameter $\alpha = \widetilde{\alpha}-1$ with $\alpha = 1, 1.05, 1.25, 1.5, 2, 3, 4$. In the case $\alpha = 1$ the bound is sharp, because the Legendre polynomials are associated with the uniform density, and these polynomials attain their maximum in the endpoints of the interval $[-1,1]$. As the value of α departs from one, the value of the lower bound of K increases. In the case of $\mathcal{B}(2,2)$, Fig. 3.1 shows that the condition (3.3.5) becomes very conservative, since K grows faster than the cubic power w^3. On the other hand, the results in the next section in Figs.3.17, 3.18, 3.19, 3.20, 3.21 show that a quadratic relation between the number of sampling points and w yields an optimally convergent error. The same holds in Figs.3.22, 3.23 with the density $\mathcal{B}(3,3)$, although from Fig.3.1 the lower bound of K in this case grows faster than the quartic power w^4.

3.4 Numerical tests with densities with a bounded support: the beta family

The beta family is useful to check the implications of Assumption 1.1 for a broad class of probability densities differing in some crucial features. This family allows to examine the case of a density bounded away from zero (on a bounded support) but unbounded from above, *e.g.* the density $\mathcal{B}(0.5,0.5)$ also known as *arcsine distribution* or *Chebyshev distribution of the 1st kind*, as well as the reversed case of a density not bounded away from zero but bounded from above, *e.g.* $\mathcal{B}(1,3)$ or $\mathcal{B}(2,5)$ or $\mathcal{B}(3/2,3/2)$. This last density $\mathcal{B}(3/2,3/2)$ is also known as *Wigner distribution* or *Chebyshev distribution of the 2nd kind*. The uniform density $\mathcal{B}(1,1)$ is also contained in the beta family. To illustrate the differences among the densities, refer to Fig. 3.2.

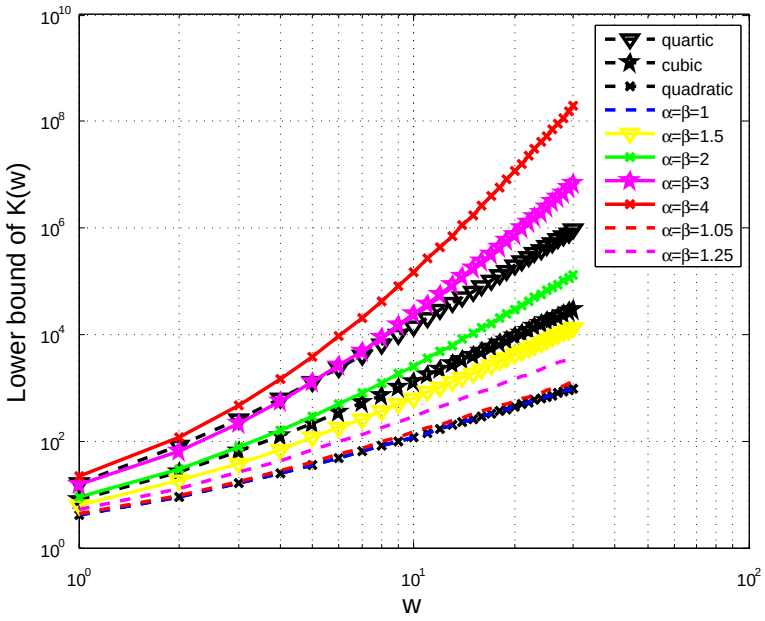


Figure 3.1: Lower bound of $K^{\widetilde{\alpha}}(\mathrm{w})$ in (3.3.5) over $\mathrm{w} = 1, \ldots, 25$, for different beta densities $\mathcal{B}(\alpha, \alpha)$ parametrized by $\alpha = \widetilde{\alpha} - 1$. Coloured lines: values of $K^{\widetilde{\alpha}}(\mathrm{w})$ for different values of $\widetilde{\alpha} = \alpha + 1$ with $\alpha = 1, 1.05, 1.25, 1.5, 2, 3, 4$. The black lines are the reference slopes w^2 (quadratic), w^3 (cubic), w^4 (quartic).

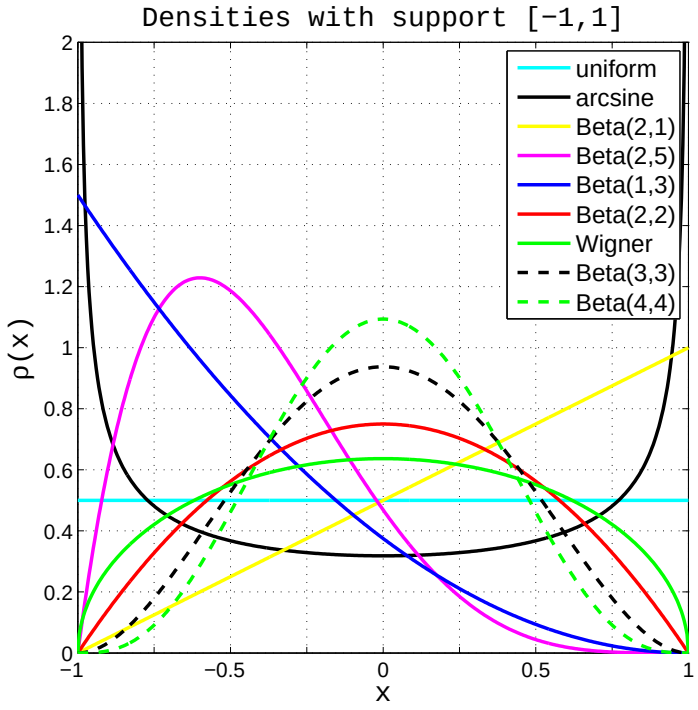


Figure 3.2: Some probability densities in the beta family with support in $[-1, 1]$.

To allow a comparison with the results presented in Section 1.5 featuring the uniform distribution, we focus on the same target functions but with the random variable distributed according to different distributions, always with support $\Gamma = [-1, 1]$. Again, we consider three smooth functions, *i.e.* an exponential function

$$\phi(y) = \exp(y), \qquad y \in [-1, 1], \tag{3.4.1}$$

and two meromorphic functions

$$\phi(y) = \frac{1}{1 + 0.5y}, \qquad y \in [-1, 1], \tag{3.4.2}$$

$$\phi(y) = \frac{1}{1 + 0.9y}, \qquad y \in [-1, 1], \tag{3.4.3}$$

and two functions with lower regularity

$$\phi(y) = |y|, \qquad y \in [-1, 1], \tag{3.4.4}$$

$$\phi(y) = |y|^3, \qquad y \in [-1, 1]. \tag{3.4.5}$$

3.4.1 An unbounded density bounded away from zero

Let us begin with the density $\mathcal{B}(0.5, 0.5)$, which is also known as *arcsine* distribution. This density is symmetric and promotes the points that are closer to the endpoints of the support Γ. Compared with the case with uniform density, this problem is much more stable.

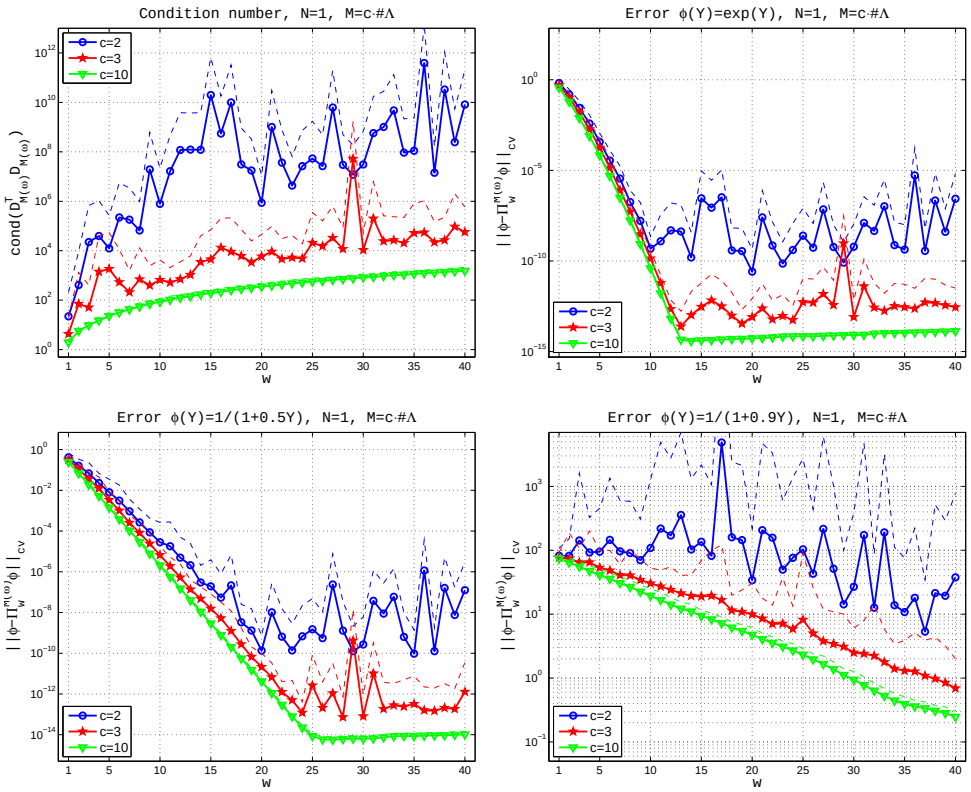


Figure 3.3: Condition number (1.4.9) and approximation error (1.4.6) with the smooth target functions (3.4.1), (3.4.2), (3.4.3), averaged over 1000 repetitions, $M = c\#\Lambda$. Beta density $\mathcal{B}(0.5, 0.5)$.

As shown in Fig. 3.3-top-left, a linear scaling of the number of sampling points $M = 3 \cdot \#\Lambda$ is enough to tame the condition number, and to avoid the (numerical) blow-up. Moreover, $M = 3 \cdot \#\Lambda$ converges like the optimal convergence rate, that in these plots is given by the

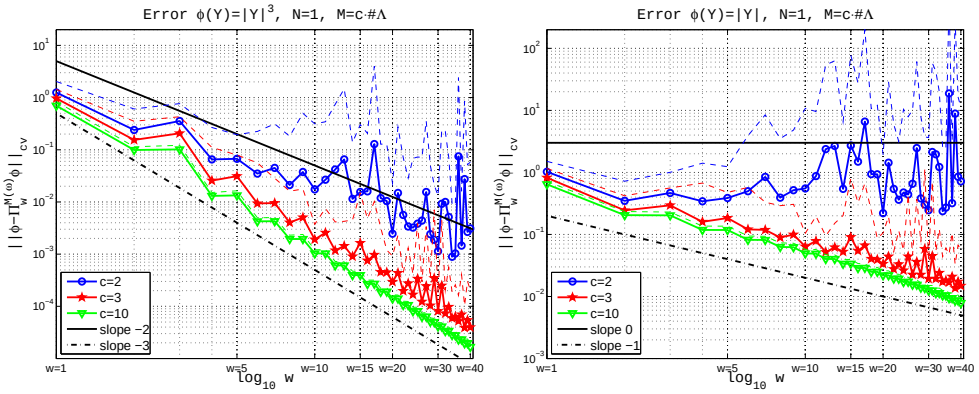


Figure 3.4: Approximation error (1.4.6) with nonsmooth target functions (3.4.5), (3.4.4) averaged over 1000 repetitions, $M = c\,\#\Lambda$. Beta density $\mathcal{B}(0.5, 0.5)$.

linear scaling $M = 10 \cdot \#\Lambda$ (that has even more points than the quadratic scaling $M = (\#\Lambda)^2$ in the range $\mathrm{w} = 1, \ldots, 10$). Fig. 3.3 shows the convergence obtained for three different analytic functions. Note that the optimal convergence obtained with $M = 3 \cdot \#\Lambda$ persists also when the function is not smooth (Fig. 3.4), catching exactly the correct slope according to the power of the modulus.

3.4.2 Symmetric densities not bounded away from zero

We investigate now the consequences of violating the condition that the density is bounded away from zero. We begin with the density $\mathcal{B}(1.05, 1.05)$ that slightly differs from the uniform density, for going to zero very slowly on the boundary of its support Γ. The graphs in Fig. 3.5 show that the linear scaling $M \sim c \cdot \#\Lambda$ and the quadratic scaling $M \sim c \cdot (\#\Lambda)^2$ with the density $\rho = \mathcal{B}(1.05, 1.05)$ yield results almost identical to those with the uniform density presented in Figs. 1.3. With the quadratic scaling $M \sim (\#\Lambda)^2$ the condition number slightly increases w.r.t. w, although the growth is very slow. The next examples with the density $\mathcal{B}(1.25, 1.25)$, $\mathcal{B}(2, 2)$ and $\mathcal{B}(3, 3)$ will show that this effect amplifies when the density gets flatter going to zero.

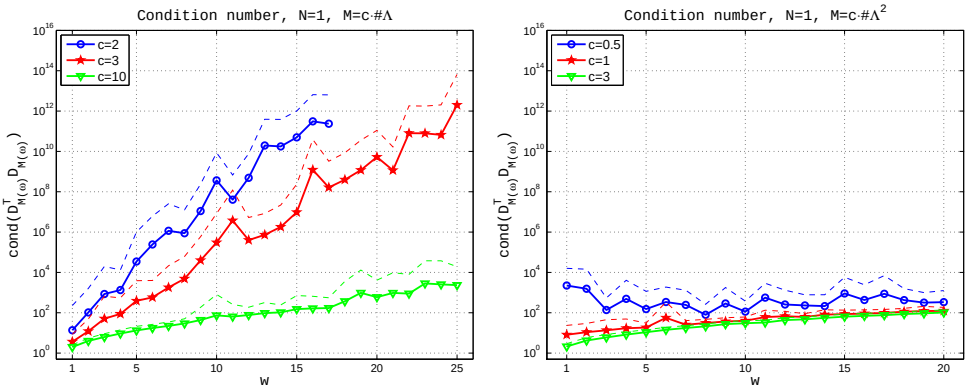


Figure 3.5: Condition number (1.4.9). Beta density $\mathcal{B}(1.05, 1.05)$. Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

Even though the condition number with the density $\mathcal{B}(1.05, 1.05)$ manifests some slight differences compared to the condition number of the uniform density, the plots concerning the convergence rate are completely identical with the two densities. Compare for examples the errors with the exponential function (1.5.1) in Fig. 1.4 and Fig. 3.6, the errors with the meromorphic function (1.5.2) in Fig. 1.5 and Fig. 3.7 with $\beta = 0.5$, or in Fig. 1.6 and Fig. 3.8 with $\beta = 0.9$. Observe that the convergence rate depends on the value of the parameter β, and a higher value produces a steeper function (1.5.2) that is also harder to approximate. As

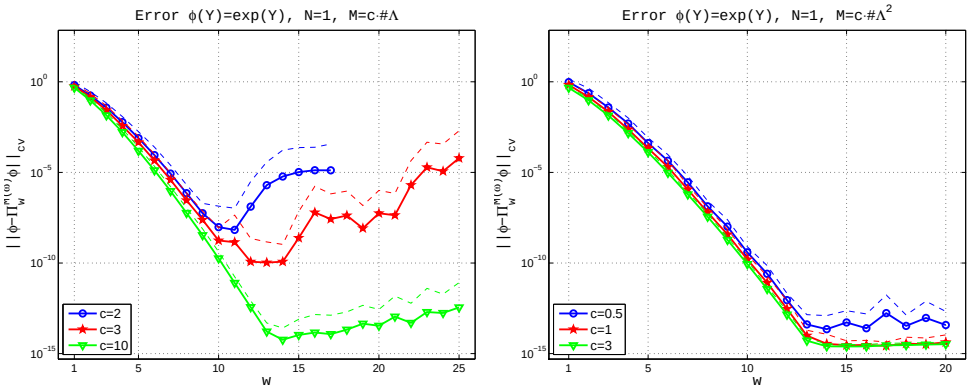


Figure 3.6: Approximation error (1.4.6) with the smooth target function (3.4.1). Beta density $\mathcal{B}(1.05, 1.05)$. Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

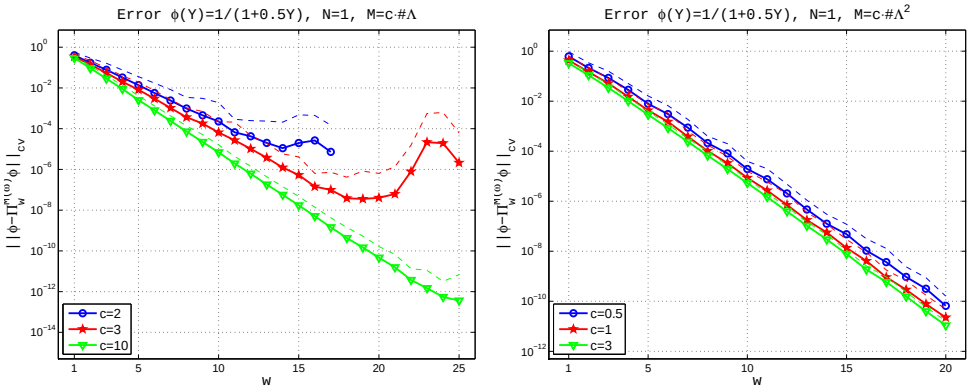


Figure 3.7: Approximation error (1.4.6) with the smooth target function (3.4.2). Beta density $\mathcal{B}(1.05, 1.05)$. Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

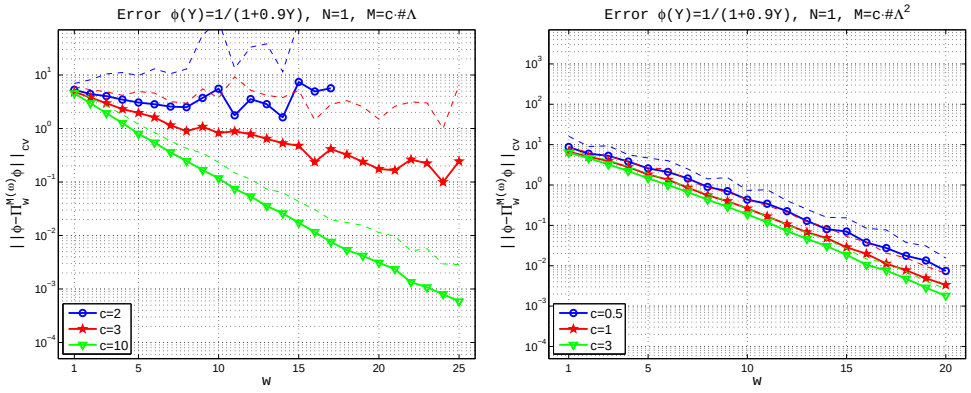


Figure 3.8: Approximation error (1.4.6) with the smooth target function (3.4.3). Beta density $\mathcal{B}(1.05, 1.05)$. Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

explained in Section 1.5.1 with the uniform density, when the target function is smooth the linear scaling optimally converges up to a certain value of w, and then the convergence degenerates, (Figs. 3.6-left, 3.7-left, 3.8-left).

Note in Figs. 3.9,3.10 that the linear scaling does not yield an optimal convergence rate when the function is not smooth. Conversely, the convergence with quadratic scaling is optimal also with nonsmooth target functions. Comparing the results with the two densities in Fig. 3.9 and Fig. 1.7 with the nonsmooth function (1.5.3), the rate of convergence with the uniform density is slightly faster, but still comparable to the rate of convergence of the density $\mathcal{B}(1.05, 1.05)$.

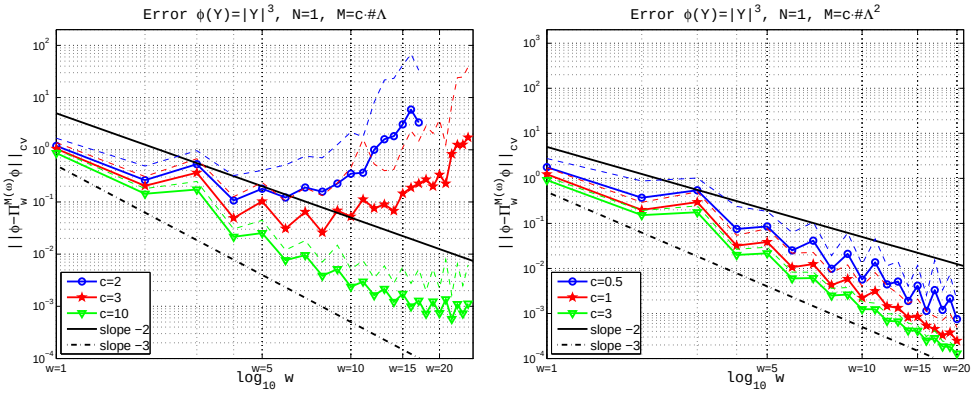


Figure 3.9: Approximation error (1.4.6) with the nonsmooth target function (3.4.5). Beta density $\mathcal{B}(1.05, 1.05)$. Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

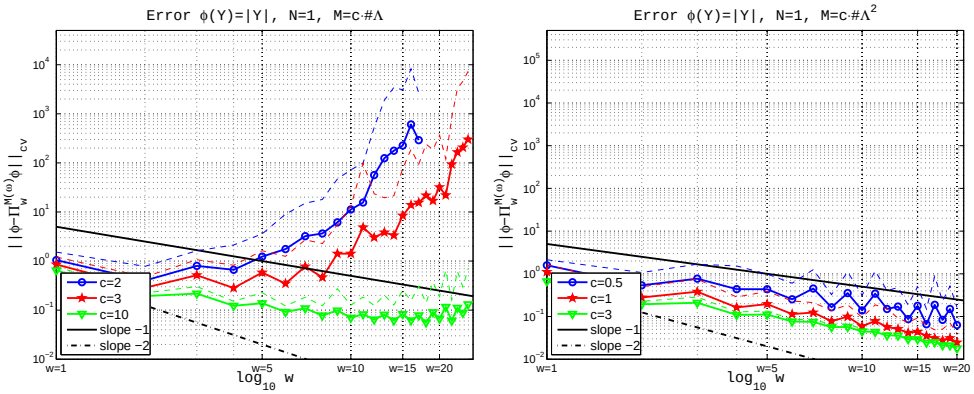


Figure 3.10: Approximation error (1.4.6) with the nonsmooth target function (3.4.5). Beta density $\mathcal{B}(1.05, 1.05)$. Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

Now let us consider the density $\mathcal{B}(1.25, 1.25)$, which goes to zero on the boundary of Γ faster than the density $\mathcal{B}(1.05, 1.05)$. In fact, the growth of the condition number with the quadratic scaling $M \sim (\#\Lambda)^2$ in Fig. 3.11 is even more accentuated than the growth in Fig. 3.5.

Again the approximation errors for the same smooth functions (3.4.1), (3.4.2) and (3.4.3) converge with the optimal rate independently of the linear or quadratic scaling, as confirmed by the results in Figs. 3.12, 3.13 and 3.14.

When approximating nonsmooth target functions (Figs. 3.15,3.16) the optimal convergence rate is achieved only with the quadratic scaling $M \sim (\#\Lambda)^2$.

Further increasing the values of the parameters of the Beta density, and keeping them equal to obtain a symmetric distribution, we can test the effect of an even flatter density on the boundary. Let us consider the density $\mathcal{B}(2, 2)$. From Fig. 3.17, it is clear that the quadratic scaling $M \sim (\#\Lambda)^2$ is not enough anymore to keep the condition number under control, but the blow-up is not yet dramatic, and the numerical computations are still feasible using for example a scaling $M = 10 \cdot (\#\Lambda)^2$.

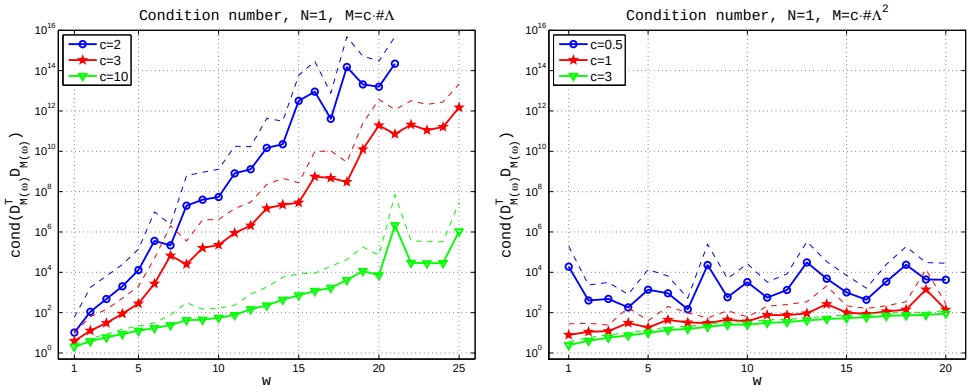


Figure 3.11: Condition number (1.4.9). Beta density $\mathcal{B}(1.25, 1.25)$. Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

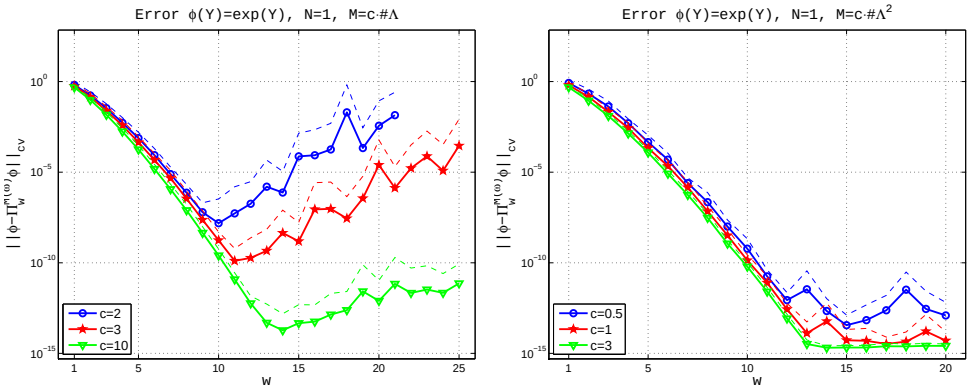


Figure 3.12: Approximation error (1.4.6) with the exponential function (3.4.1). Beta density $\mathcal{B}(1.25, 1.25)$. Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

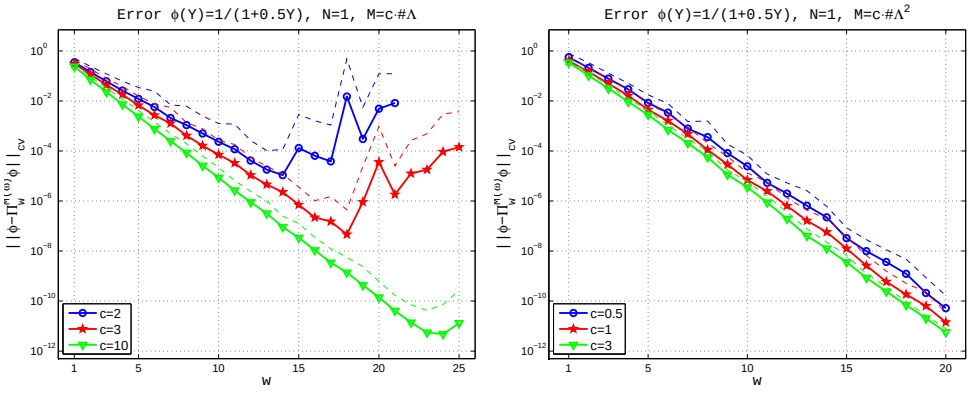


Figure 3.13: Approximation error (1.4.6) with the meromorphic function (3.4.2). Beta density $\mathcal{B}(1.25, 1.25)$. Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

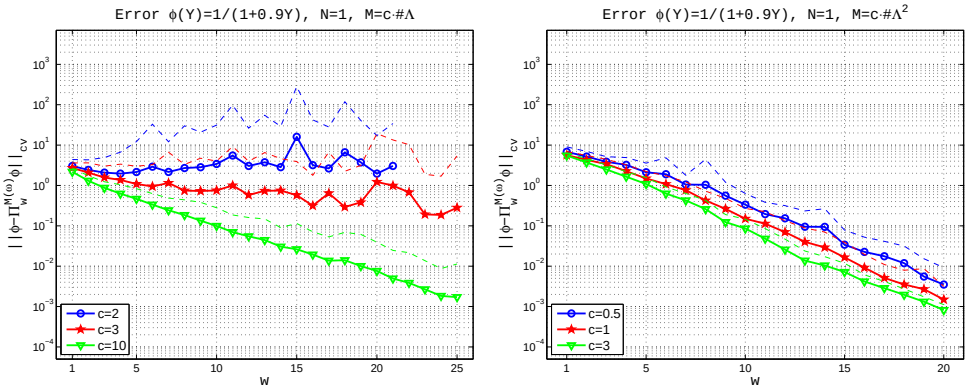


Figure 3.14: Approximation error (1.4.6) with the meromorphic function (3.4.3). Beta density $\mathcal{B}(1.25, 1.25)$. Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

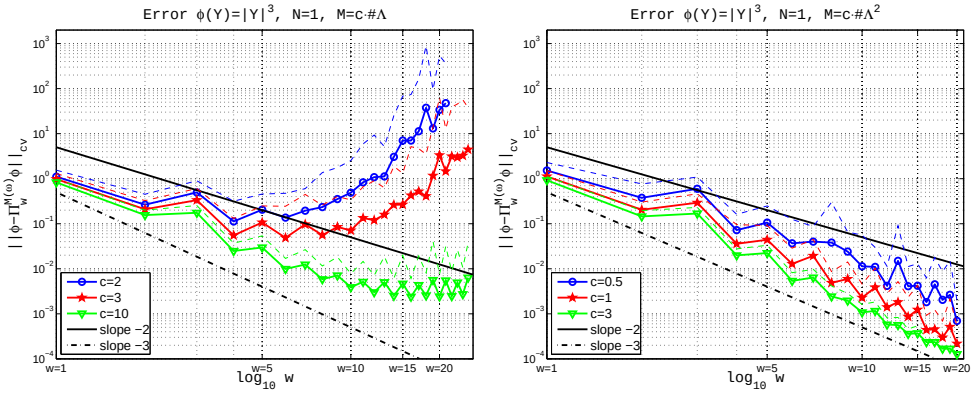


Figure 3.15: Approximation error (1.4.6) with the nonsmooth target function (3.4.5). Beta density $\mathcal{B}(1.25, 1.25)$. Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

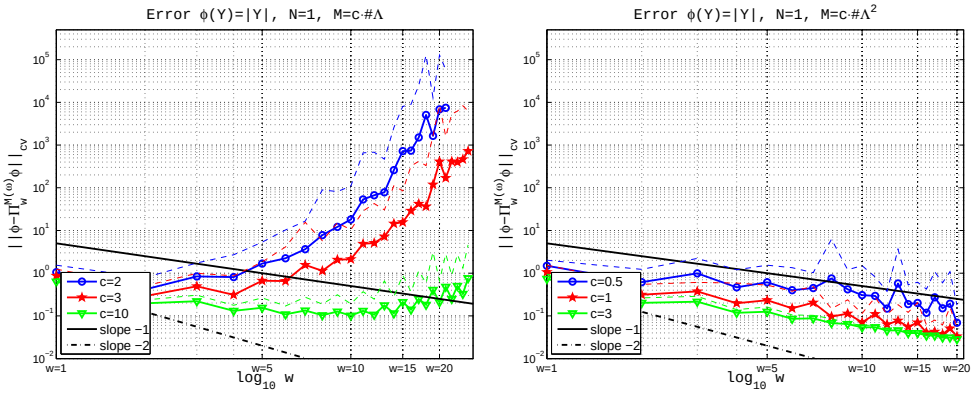


Figure 3.16: Approximation error (1.4.6) with the nonsmooth target function (3.4.4). Beta density $\mathcal{B}(1.25, 1.25)$. Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$.

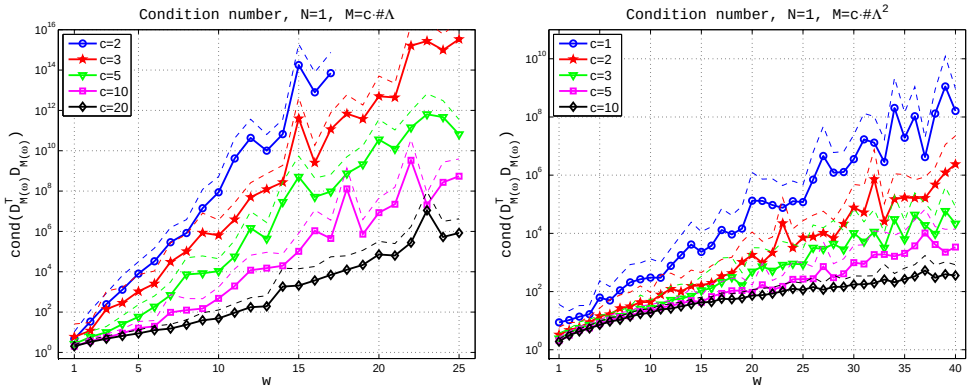


Figure 3.17: Condition number (1.4.9) averaged over 100 repetitions. Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$. Beta density $\mathcal{B}(2,2)$.

The error converges always at the optimal rate with the quadratic scaling, both with smooth (Figs. 3.18,3.19) and nonsmooth functions (Figs. 3.20,3.21). The linear scaling allow to approximate accurately only smooth functions, up to a certain threshold, while with nonsmooth functions the approximation is clearly divergent.

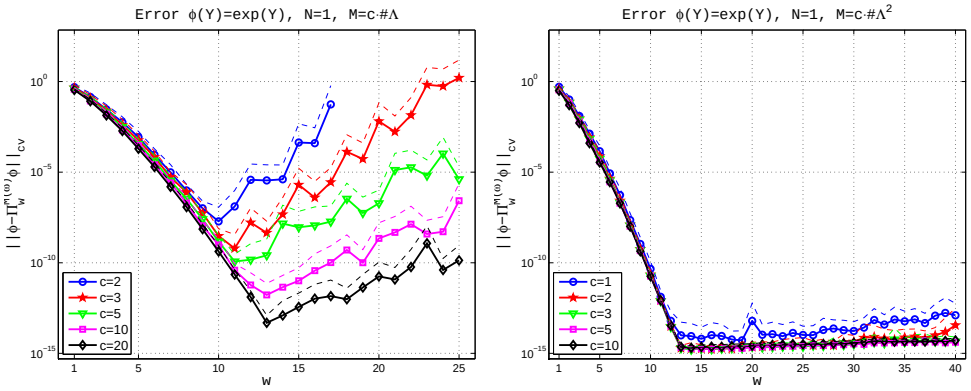


Figure 3.18: Approximation error (1.4.6) with the exponential function (3.4.1), averaged over 100 repetitions. Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$. Beta density $\mathcal{B}(2,2)$.

The density $\mathcal{B}(3,3)$ becomes convex close to the endpoints. Fig. 3.22-left shows that also a cubic scaling $M \sim (\#\Lambda)^3$ cannot avoid the growth of the condition number w.r.t. the parameter w. Nonetheless, Fig. 3.22-right shows that the error converges at the optimal convergence rate with $M = 2 \cdot (\#\Lambda)^2$ and the exponential target function (3.4.1). The same occurs in Fig. 3.23 with the nonsmooth functions (3.4.5), (3.4.4).

At this point we can make a general remark on the consequences of relaxing the assumption on the density ρ being bounded away from zero:

Remark 3.1. *When the density ρ is not bounded away from zero, the approximation problem with the random discrete projection* (1.2.5) *becomes more and more ill-conditioned, the faster the density approaches zero. Consequently, the condition number* (1.4.9) *increases and a larger sample size is necessay to preserve the numerical stability of the problem, see* e.g. *Figs. 3.17-right and 3.22-right. Despite of the ill-conditioning, the quadratic scaling $M \sim (\#\Lambda)^2$ provides an optimal convergent rate for smooth and nonsmooth functions,* e.g. *Figs. 3.20,3.21 with the density $\mathcal{B}(2,2)$, and Fig. 3.23 with the density $\mathcal{B}(3,3)$.*

3.4.3 Nonsymmetric densities not bounded away from zero

Remark 3.1 holds also for nonsymmetric beta densities. Some examples with the densities $\mathcal{B}(1,3)$ and $\mathcal{B}(2,5)$ follow. The density $\mathcal{B}(1,3)$ is monotone and convex approaching the right

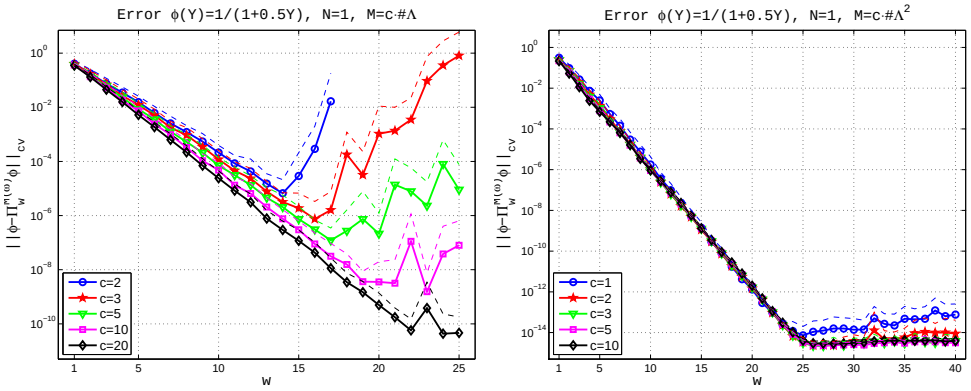


Figure 3.19: Approximation error (1.4.6) with the meromorphic function (3.4.2), averaged over 100 repetitions. Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$. Beta density $\mathcal{B}(2,2)$.

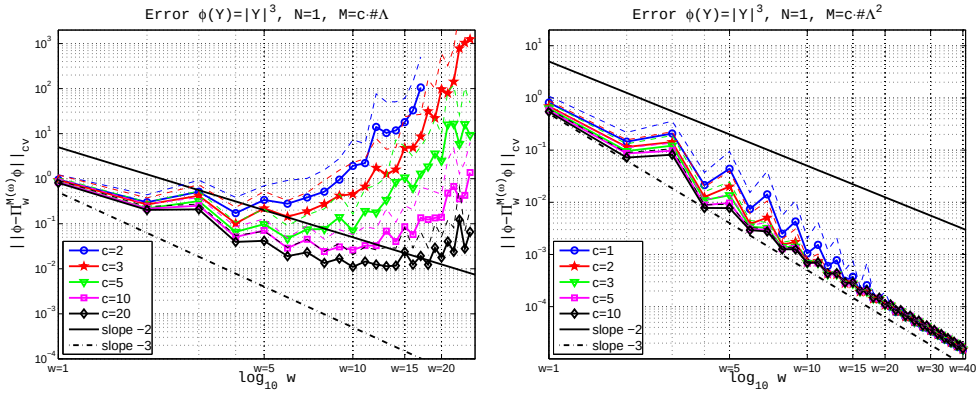


Figure 3.20: Approximation error (1.4.6) with the nonsmooth target function (3.4.5), averaged over 100 repetitions. Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$. Beta density $\mathcal{B}(2,2)$.

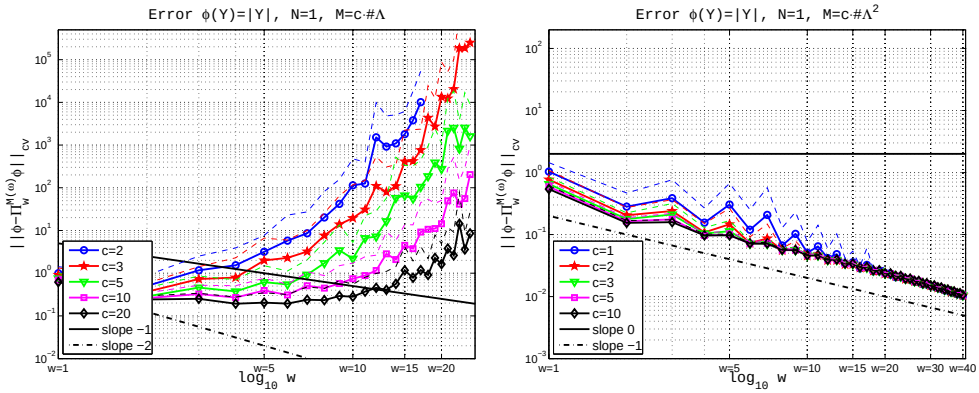


Figure 3.21: Approximation error (1.4.6) with the nonsmooth target function (3.4.4), averaged over 100 repetitions. Left: $M = c \cdot \#\Lambda$. Right: $M = c \cdot (\#\Lambda)^2$. Beta density $\mathcal{B}(2,2)$.

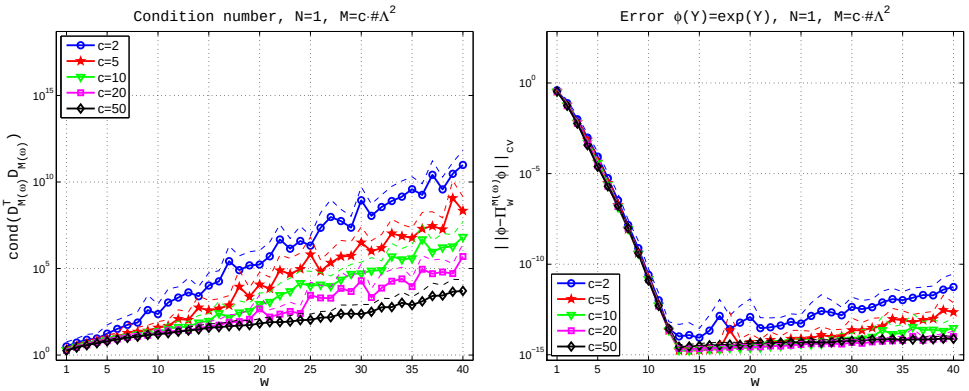


Figure 3.22: Condition number (1.4.9) and approximation error with the exponential function (3.4.1), averaged over 100 repetitions, $M = c \cdot (\#\Lambda)^2$. Beta density $\mathcal{B}(3,3)$.

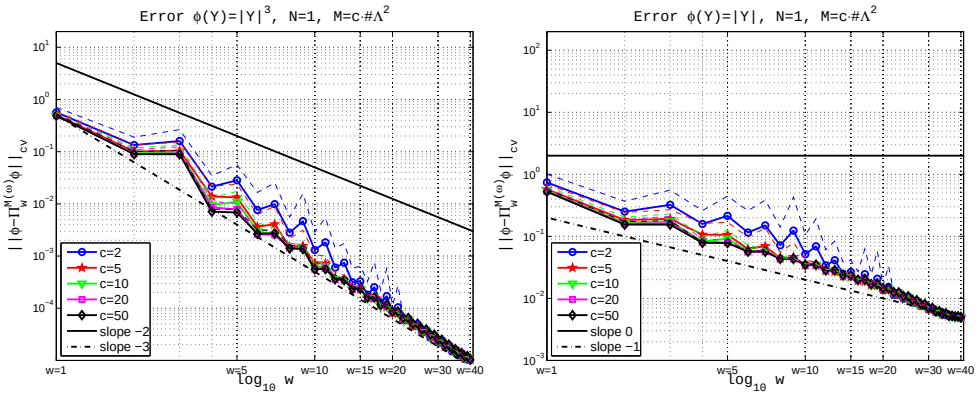


Figure 3.23: Approximation error (1.4.6) with the nonsmooth target functions (3.4.5), (3.4.4), averaged over 100 repetitions, $M = c \cdot (\#\Lambda)^2$. Beta density $\mathcal{B}(3,3)$.

endpoint of Γ. The shape of $\mathcal{B}(2,5)$ is even more flattened than $\mathcal{B}(1,3)$ approaching zero (see also Fig. 3.2). The results in Figs. 3.24-top-right, 3.25-top-right confirm that the linear scaling $M \sim \#\Lambda$ provides only a rough approximation of the exponential function (3.4.1). Concerning the function (3.4.3), the results in Fig. 3.24-bottom-left with the $\mathcal{B}(1,3)$ density show a slow convergence, and no convergence in Fig. 3.25-bottom-left for the same function but with the $\mathcal{B}(2,5)$ density.

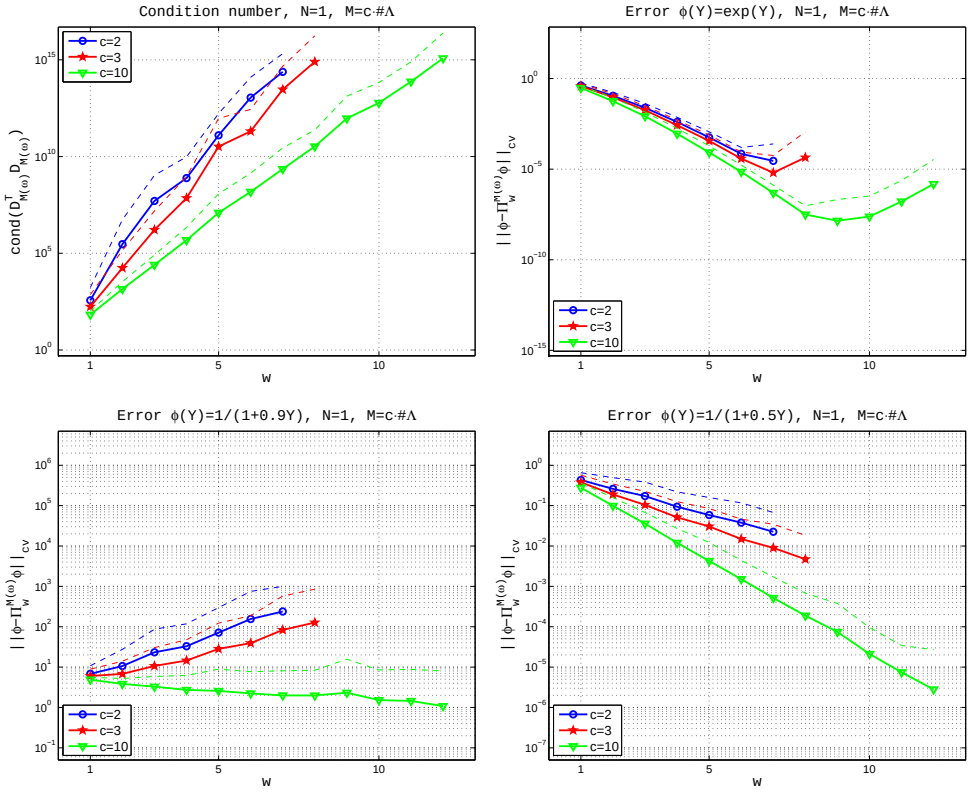


Figure 3.24: Condition number (1.4.9) and approximation error (1.4.6) with the smooth target functions (3.4.1), (3.4.2), (3.4.3), averaged over 1000 repetitions, $M = c \cdot \#\Lambda$. Beta density $\mathcal{B}(1,3)$.

Moreover, the results in Figs. 3.26,3.27 demonstrate that the linear scaling $M \sim \#\Lambda$ completely fails to approximate nonsmooth functions such as (1.5.3). The numerical stability of the approximation problem is lost as well and this contributes to deteriorate the accuracy of the approximation error.

On the other hand, the approximation obtained with the quadratic scaling $M \sim (\#\Lambda)^2$ always converges with the optimal convergent rate, until the numerical stability is lost (Figs. 3.28,3.29). Looking at the condition number in Fig. 3.28, the random projection can be considered stable until $\mathrm{w} = 8$ when $M = 2 \cdot (\#\Lambda)^2$, and until $\mathrm{w} = 10$ when $M = c \cdot (\#\Lambda)^2$ with $c \geq 5$.

A further test in Figs. 3.30,3.31 shows that the requirement on the density being bounded away from zero is crucial: renouncing to it the scalings $M \sim (\#\Lambda)^3$, $M \sim (\#\Lambda)^4$, $M \sim (\#\Lambda)^5$ do not avoid the growth of the condition number, and lead to an ill-posed approximation problem as the value of w increases. Note that in Figs. 3.30,3.31 the cubic scaling $M = c \cdot (\#\Lambda)^3$ with $c = 1, 10, 100$ yields more points than the scalings $M = (\#\Lambda)^4$ and $M = (\#\Lambda)^5$ on the range $\mathrm{w} = 1, \ldots, 10$.

At this stage we can summarize the following remark:

Remark 3.2. *The usefulness of the random projection is confined to situations where the number of sampling points M is large enough to ensure the numerical stability, quantified for example by the condition number* (1.4.9)*. In one dimension,* i.e. *$N = 1$, when the density ρ approaches zero with a flat slope the condition $M \sim (\#\Lambda)^2$ does not suffice anymore to ensure the stability.*

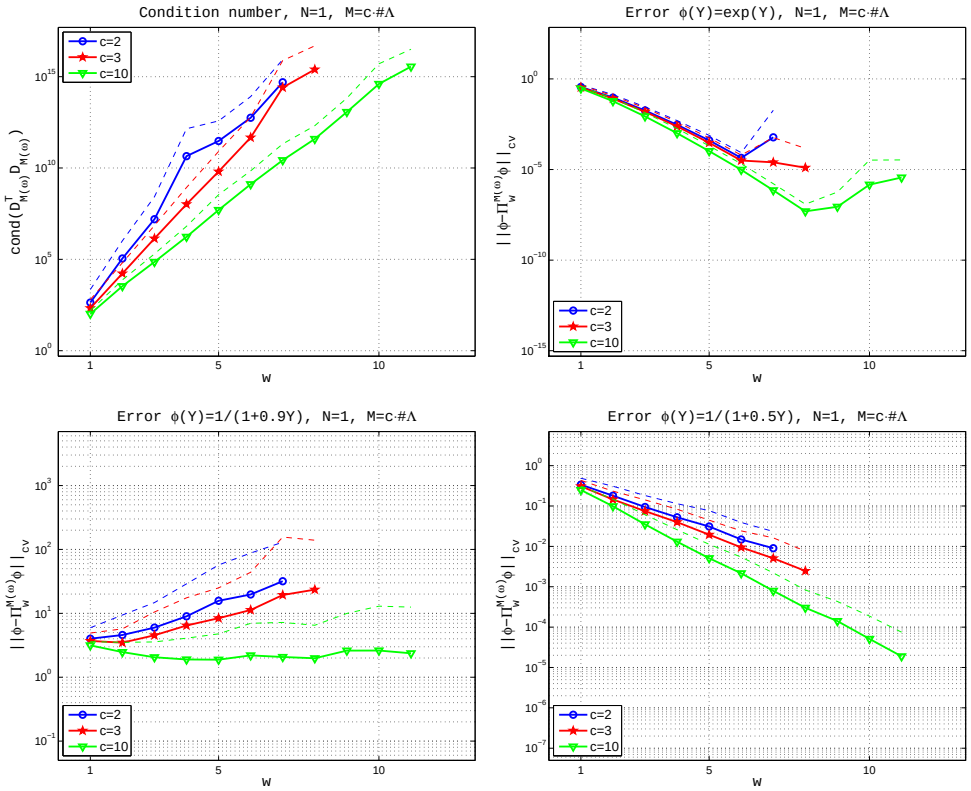


Figure 3.25: Condition number (1.4.9) and approximation error (1.4.6) with the smooth target functions (3.4.1), (3.4.2), (3.4.3), averaged over 1000 repetitions, $M = c \cdot \#\Lambda$. Beta density $\mathcal{B}(2, 5)$.

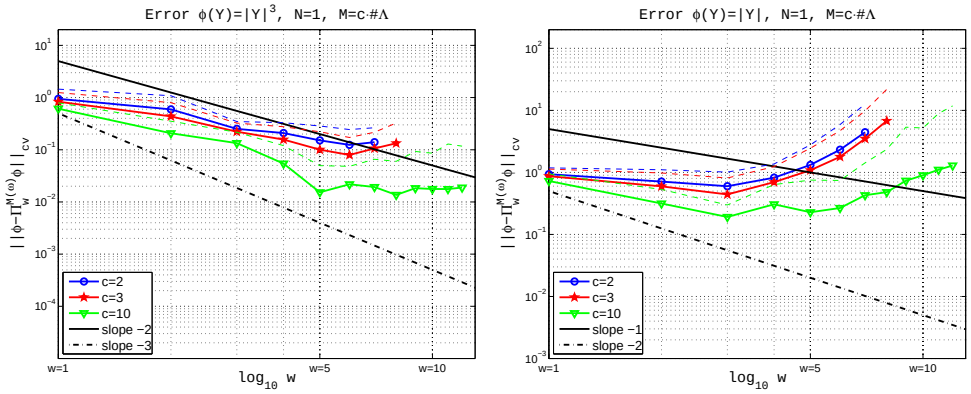


Figure 3.26: Approximation error (1.4.6) with the nonsmooth target functions (3.4.5), (3.4.4), averaged over 1000 repetitions, $M = c \cdot \#\Lambda$. Beta density $\mathcal{B}(1, 3)$.

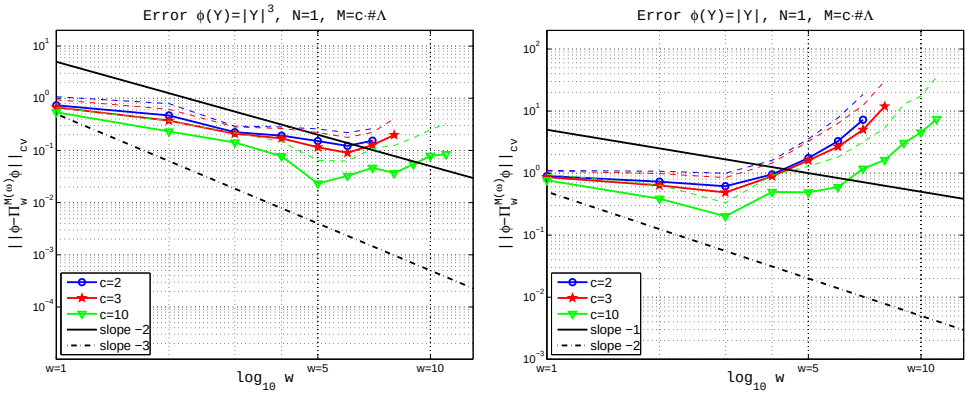


Figure 3.27: Approximation error (1.4.6) with the nonsmooth target functions (3.4.5), (3.4.4), averaged over 1000 repetitions, $M = c \cdot \#\Lambda$. Beta density $\mathcal{B}(2,5)$.

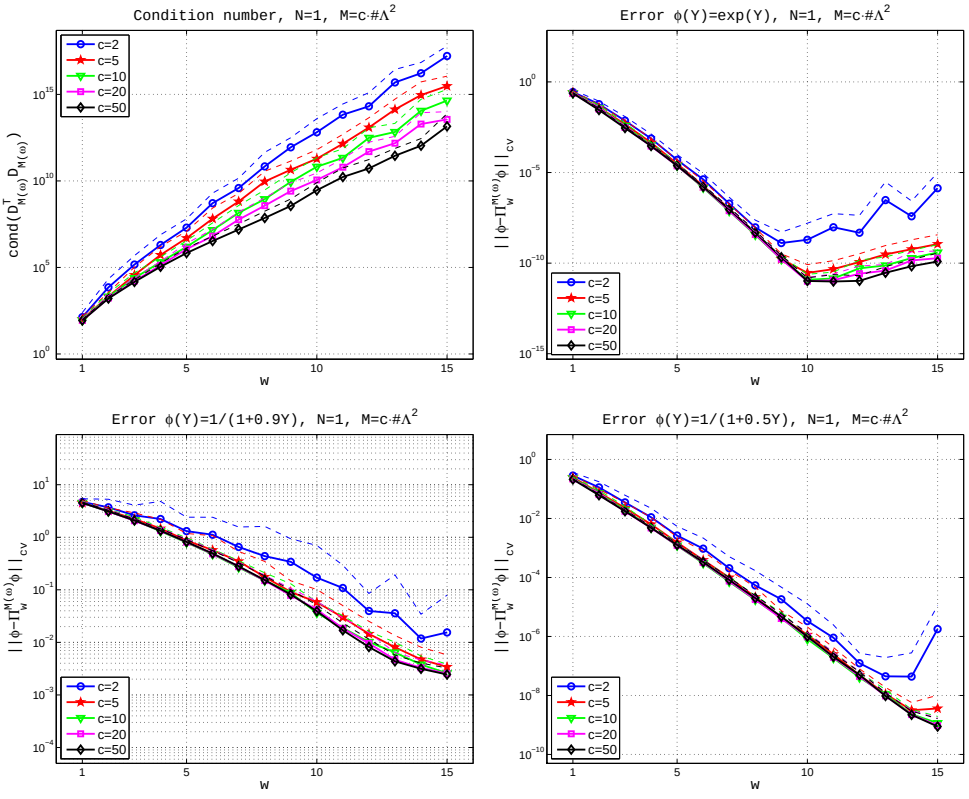


Figure 3.28: Condition number (1.4.9) and approximation error (1.4.6) with the smooth target functions (3.4.1), (3.4.2), (3.4.3), averaged over 100 repetitions, $M = c \cdot (\#\Lambda)^2$. Beta density $\mathcal{B}(2,5)$.

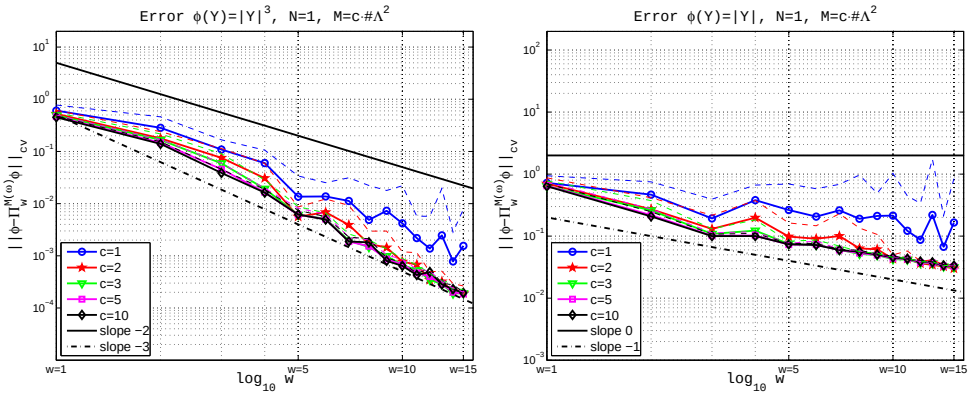


Figure 3.29: Approximation error (1.4.6) with the nonsmooth target functions (3.4.5), (3.4.4), averaged over 100 repetitions, $M = c \cdot (\#\Lambda)^2$. Beta density $\mathcal{B}(2,5)$.

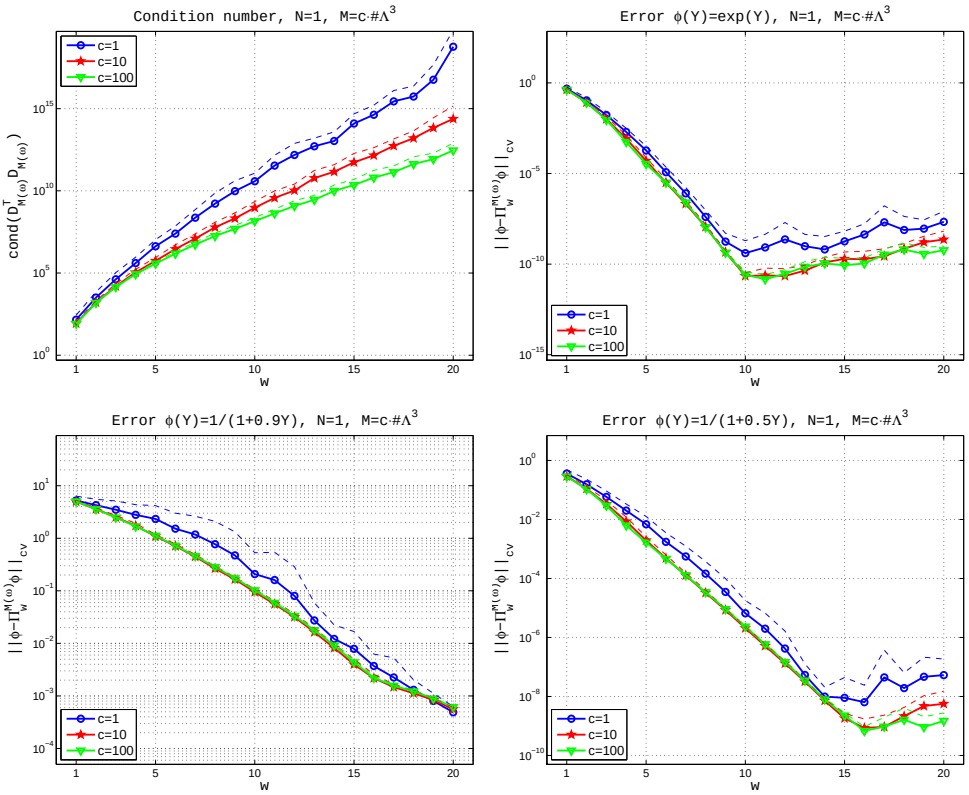


Figure 3.30: Condition number (1.4.9) and approximation error (1.4.6) with the smooth target functions (3.4.1), (3.4.2), (3.4.3), averaged over 100 repetitions, $M = c \cdot (\#\Lambda)^3$. Beta density $\mathcal{B}(2,5)$.

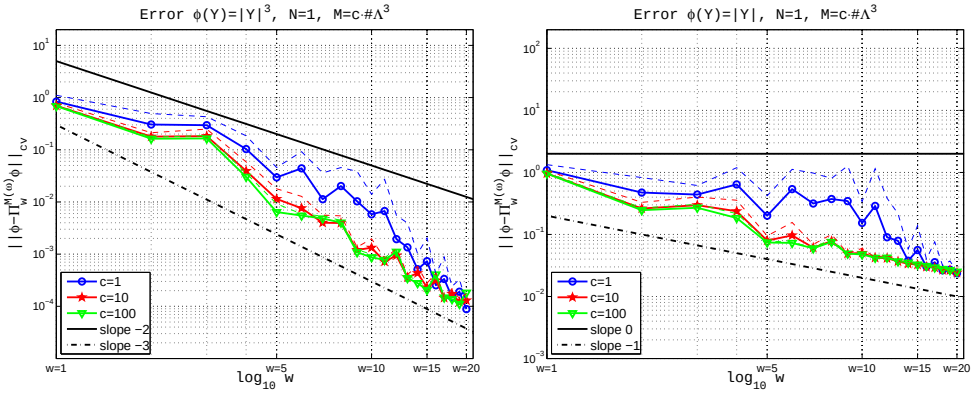


Figure 3.31: Approximation error (1.4.6) with the nonsmooth target functions (3.4.5), (3.4.4), averaged over 100 repetitions, $M = c \cdot (\#\Lambda)^3$. Beta density $\mathcal{B}(2,5)$.

However, as pointed out in Section 1.5.2 for the uniform distribution, in higher dimensions the condition for stability becomes less restrictive, and this holds more generally with any distribution ρ. The theoretical characterization of this behaviour in high dimensions demands a further analysis of the statistical distribution of the random variable $C^\omega = C^\omega(M, \Lambda)$ defined in (1.2.15)*, and is currently under investigation.*

3.5 Numerical results with densities with an unbounded support

In this section an overview of the numerical tests with densities on an unbounded support is given. The main focus is on the Gaussian density, and several numerical tests are presented. A test concerning the condition number with the gamma density is also presented.

3.5.1 The Gaussian case

We consider the following target functions, to test the random discrete projection (1.2.5) with the standard *Gaussian* density:

$$\phi(y) = \exp(y), \qquad y \in \mathbb{R}, \tag{3.5.1}$$

$$\phi(y) = \frac{1}{1+\exp(y)}, \qquad y \in \mathbb{R}, \tag{3.5.2}$$

$$\phi(y) = \frac{1}{1+y^2}, \qquad y \in \mathbb{R}. \tag{3.5.3}$$

Moreover, we test the following target functions featuring a lower regularity:

$$\phi(y) = |y|^3, \qquad y \in \mathbb{R}, \tag{3.5.4}$$

$$\phi(y) = |y|, \qquad y \in \mathbb{R}, \tag{3.5.5}$$

$$\phi(y) = \frac{|y|}{1+0.5|y|}, \qquad y \in \mathbb{R}. \tag{3.5.6}$$

In the Gaussian case, the cross-validated error (1.4.6) is evaluated in the 2-norm, because the evaluation in the ∞-norm naturally degenerates for extremal realizations of $\mathbf{Y}$, due to the asymptotic divergence of the Hermite polynomials. Therefore, the cross-validated error is evaluated as

$$\|\phi - \Pi_M^{\Lambda,\omega}\phi\|_{cv} = \|D_{cv}x^\omega - b_{cv}\|_2, \tag{3.5.7}$$

and accordingly, to estimate the variability of (3.5.7) due to the random sampling of the M collocation points, the calculation is repeated over R independent sets of points $\{\mathbf{y}_j^{\omega_k},\ j = 1,\dots,M\}$, with $k = 1,\dots,R$. The average error is computed by

$$\overline{E}_{cv} = \frac{\sum_{k=1}^{R} \|D_{cv}x^{\omega_k} - b_{cv}\|_2}{R}, \tag{3.5.8}$$

instead of (1.4.7). The sample standard deviation is computed by

$$s_E = \sqrt{\frac{1}{R-1}\sum_{k=1}^{R}\left(\|D_{cv}x^{\omega_k} - b_{cv}\|_2 - \overline{E}_{cv}\right)^2}, \tag{3.5.9}$$

instead of (1.4.8).

Let us begin to examine the stability of the least squares problem, in the case where we have to approximate a function that depends on a Gaussian random variable. Therefore we choose the Hermite family of orthogonal polynomials. As shown in Fig. 3.32, a quadratic scaling

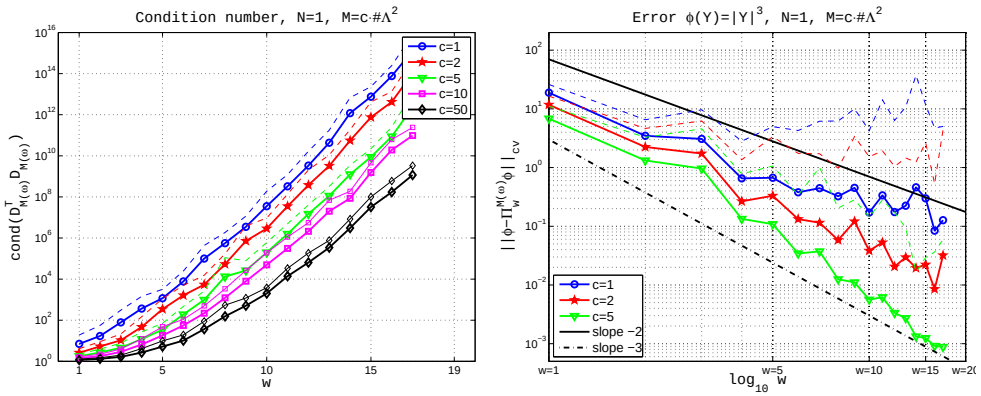


Figure 3.32: Gaussian density. $M = c\cdot(\#\Lambda)^2$. Left: condition number (1.4.9). Right: approximation error (1.4.6) for the function (3.5.4). The continuous marked lines show the mean error over 100 repetitions. The dashed lines show the mean error plus one standard deviation.

$M = c\cdot(\#\Lambda)^2$ is far from being stable. A power law $M = c\cdot(\#\Lambda)^L$ with exponent L exhibits the same behaviour. We tried $L = 3,\dots,10$, but we do not report the corresponding results.

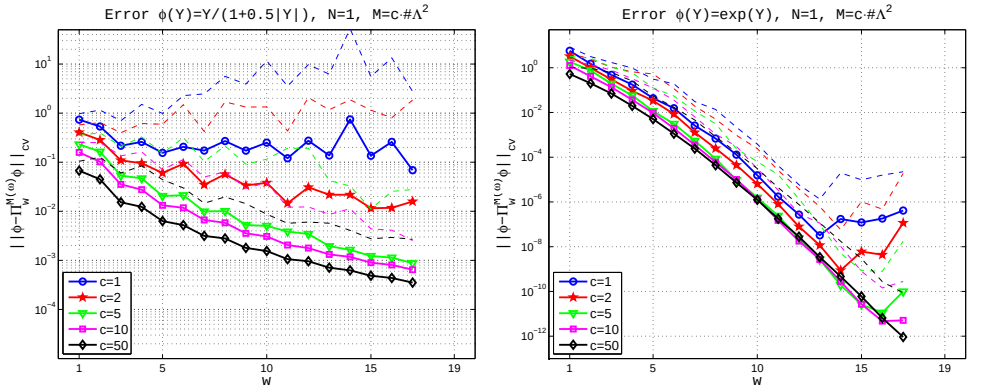


Figure 3.33: Gaussian density. $M = c\cdot(\#\Lambda)^2$. Left: approximation error (1.4.6) for the function (3.5.6). Right: approximation error (1.4.6) for the function (3.5.1). The continuous marked lines show the mean error over 100 repetitions. The dashed lines show the mean error plus one standard deviation.

A first numerical investigation shows that to bound the condition number independently of w, the scaling has to grow at least exponentially in the dimension, *i.e.* $M \propto (\#\Lambda)^{\#\Lambda}$.

Fig. 3.34 shows that a slightly less dramatic growth is allowed to tame the condition number, *e.g.* $M = c \cdot (\#\Lambda)^{\#\Lambda/2}$ but with $M = c \cdot (\#\Lambda)^{\#\Lambda/3}$ the condition number still grows significantly when w increases. However such an effort to maintain bounded the condition number can be unnecessary, as explained in Remark 3.3.

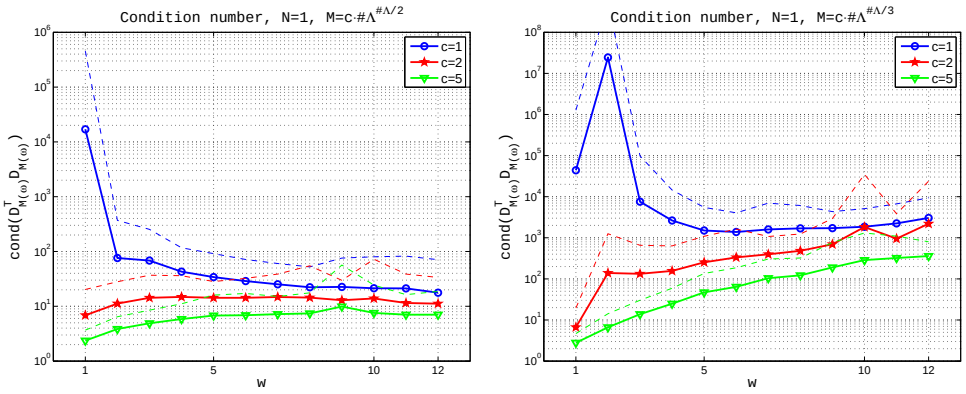


Figure 3.34: Gaussian density. Condition number (1.4.9). Left: $M = c \cdot (\#\Lambda)^{\#\Lambda/2}$. Right: $M = c \cdot (\#\Lambda)^{\#\Lambda/3}$. The continuous marked lines show the mean error over 1000 repetitions. The dashed lines show the mean error plus one standard deviation.

Fig. 3.35 shows the results obtained with the linear scaling $M \sim \#\Lambda$: of course the stability is lost soon, but surprisingly the error obtained approximating the exponential function is decaying exponentially fast, until the problem remains numerically stable.

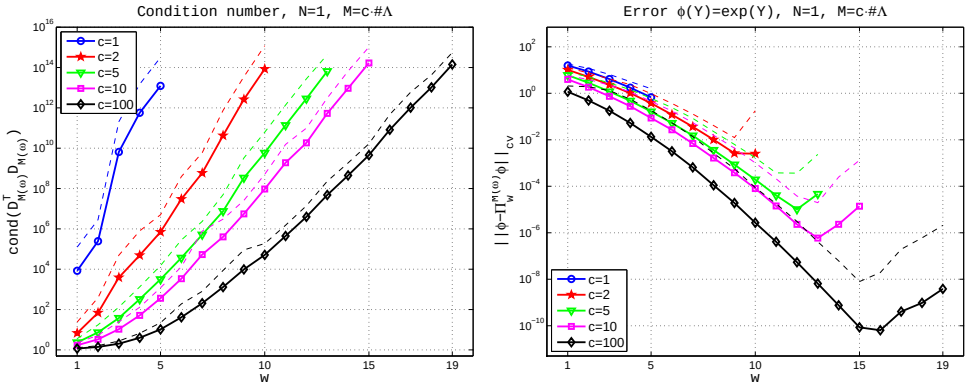


Figure 3.35: Gaussian density. $M = c \cdot \#\Lambda$. Left: condition number (1.4.9). Right: approximation error (1.4.6) for the function (3.5.1). The continuous marked lines show the mean error over 1000 repetitions. The dashed lines show the mean error plus one standard deviation.

The functions (3.5.2),(3.5.3) constitute an harder benchmark to test the random projection, and Fig. 3.36 shows that even the almost quadratical scaling $M = 10 \cdot \#\Lambda$ does not lead to a convergent approximation. The converge is achieved only with a very expensive scaling $M = 100 \cdot \#\Lambda$.

The same holds when approximating the nonsmooth functions (3.5.4), (3.5.5) (see Fig. 3.37).

The conclusions drawn on the behaviour of the random projection with Gaussian density in one dimension are summarized in Remark 3.3.

Remark 3.3. *In one dimension and when employing the Gaussian density:*

- *there is no hope to tame the condition number for practical applications, unless* w *has a very low value,*

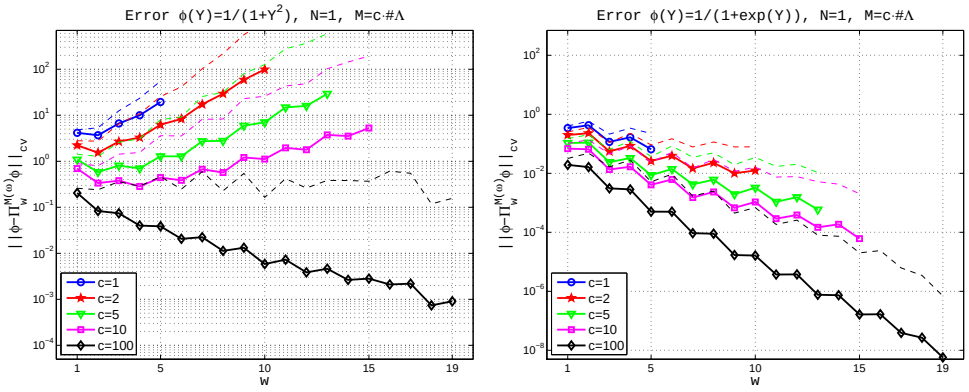


Figure 3.36: Gaussian density. $M = c \cdot \#\Lambda$. Approximation error (1.4.6) for the target functions (3.5.2) and (3.5.3). The continuous marked lines show the mean error over 1000 repetitions. The dashed lines show the mean error plus one standard deviation.

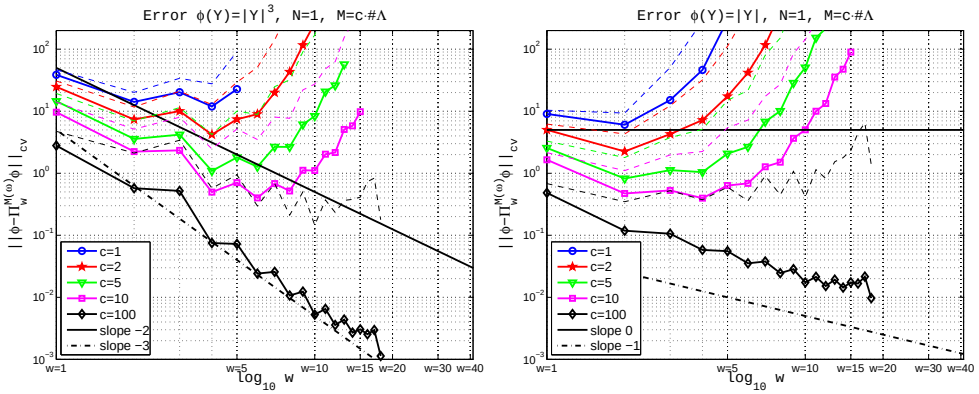


Figure 3.37: Gaussian density. $M = c \cdot \#\Lambda$. Approximation error (1.4.6) for nonsmooth target functions (3.5.4), (3.5.5). The continuous marked lines show the mean error over 1000 repetitions. The dashed lines show the mean error plus one standard deviation.

- *if the target function is smooth,* e.g. *exponential, the linear scaling yields optimal convergence up to a certain threshold on* w*, even if the least squares problem is very ill-posed. When this happens, one can simply use the (cheap) linear scaling and forget about the condition number (within certain limitations, of course),*
- *when the target function is not smooth, only a very expensive scaling (the linear one is not enough for sure) yields convergence, and the higher* w *the (exponentially) harder the optimality to achieve,*
- *in the Gaussian case the variability of the error is reduced, compared to the uniform case.*

3.5.2 The gamma case

The case of a random variable distributed according to a gamma density is even more ill-conditioned than the Gaussian case, as shown in Fig. 3.38.

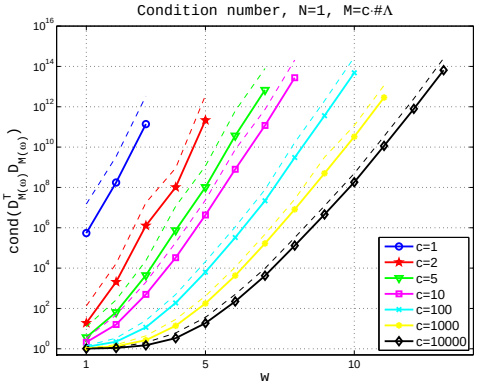


Figure 3.38: Gamma density. Condition number (1.4.9) averaged over 1000 repetitions. $M = c \cdot \#\Lambda$.

3.6 When the polynomial basis is not orthonormal w.r.t. the experimental density

In practical situations, the realizations of the random variable $\mathbf{Y}$ come from natural experiments, and the underlying experimental density $\widehat{\rho}$ is mostly unknown. Under some assumptions, a mathematical model can be proposed, and the experimental density $\widehat{\rho}$ can be assumed to be equal to a known density ρ, but the problem of investigating the effect of sampling from an experimental density $\widehat{\rho}$ that differs from the density ρ employed in the construction of the orthonormal polynomial basis (1.2.3) remains.

Therefore, in this section we build the polynomial basis (1.2.3) orthonormalized according to the *arcsine distribution* or *Wigner distribution*, and project over this basis a function that depends on a random variable distributed according to the uniform density.

The first test is to check the convergence of the random projection on the polynomial space. The *arcsine distribution* is used to orthonormalize the polynomials. Therefore, the polynomial basis is not orthonormal w.r.t. the uniform density, and we do not expect the condition number to converge to 1 as it happens in Fig. 1.2 when employing Legendre polynomials. Accordingly, Fig. 3.39 shows a slow but unavoidable growth of the mean condition number, when w increases. The error converges at the same optimal rates obtained in Figs. 1.5,1.7 when using the polynomial basis orthonormalized with the same density of the random variable $\mathbf{Y}$.

Then as shown in Fig. 3.40 the two graphs of the condition number exhibits the same growth, due to the lack of orthonormalization and to the linear scaling $M \sim \#\Lambda$. The basis is orthonormalized w.r.t. the *arcsine distribution* in Fig. 3.40-left, and w.r.t. to the *Wigner distribution* in Fig. 3.40-right.

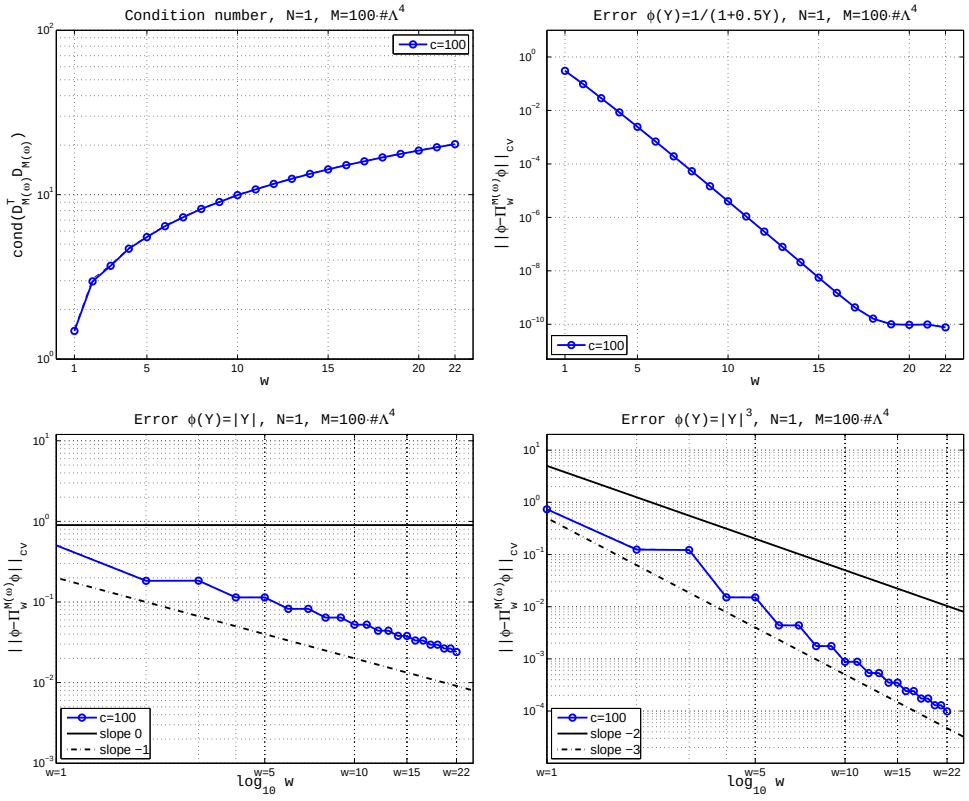


Figure 3.39: Chebyshev family of the first kind. Overkilling scaling $M = 100 \cdot (\#\Lambda)^4$. Top-left: mean condition number (1.4.9). Top-right: approximation error (1.4.6) for the function (3.4.2). Bottom-left: approximation error (1.4.6) for the function (3.4.5). Bottom-right: approximation error (1.4.6) for the function (3.4.4). The continuous marked lines show the mean error over 10 repetitions. The dashed lines show the mean error plus one standard deviation.

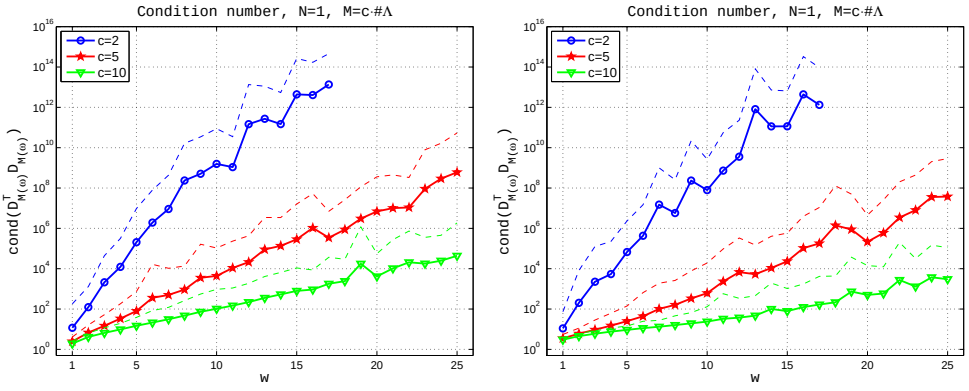


Figure 3.40: Condition number (1.4.9) averaged over 10000 repetitions, $M = c \cdot \#\Lambda$. Left: basis (1.2.3) orthonormalized according to the *arcsine distribution*. Right: basis (1.2.3) orthonormalized according to the *Wigner distribution*.

The theoretical analysis presented in Chapter 1 does not depend on the choice of the basis. Consequently, in Figs. 3.41,3.42 the error plots do not show any difference due to the basis, and the convergence rates are the same as those in Figs. 1.5,1.7 obtained using the polynomial basis orthonormalized with the same density of the random variable $\mathbf{Y}$.

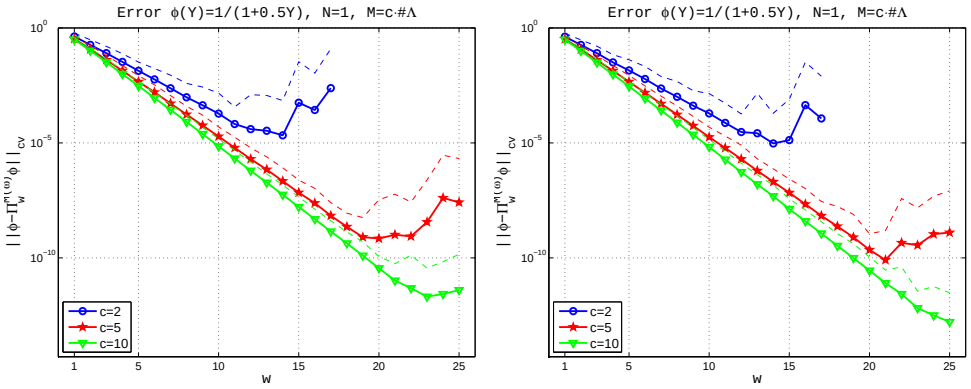


Figure 3.41: Approximation error (1.4.6) for the target function (3.4.2) with $\beta = 0.5$ averaged over 10000 repetitions, $M = c \cdot \#\Lambda$. Left: basis (1.2.3) orthonormalized according to the *arcsine distribution*. Right: basis (1.2.3) orthonormalized according to the *Wigner distribution*.

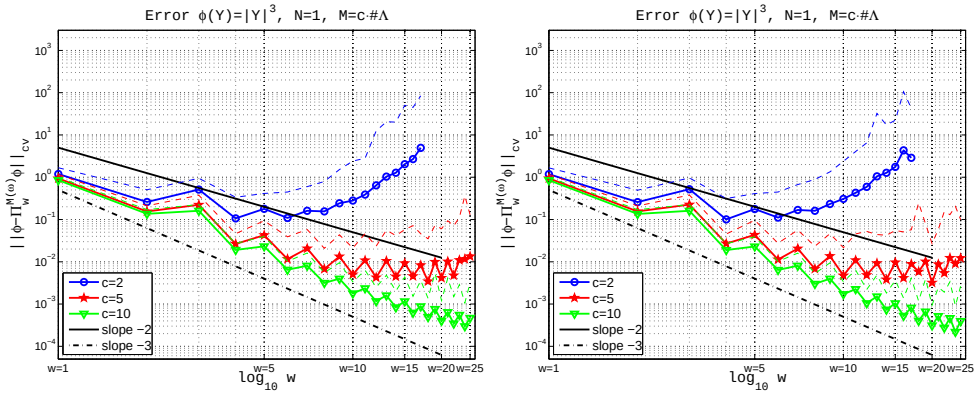


Figure 3.42: Approximation error (1.4.6) for the target function (3.4.5) averaged over 10000 repetitions, $M = c \cdot \#\Lambda$. Left: basis (1.2.3) orthonormalized according to the *arcsine distribution*. Right: basis (1.2.3) orthonormalized according to the *Wigner distribution*.

In Fig. 3.43 the results with the quadratic scaling $M \sim (\#\Lambda)^2$ are shown. The basis is orthonormalized w.r.t. the *arcsine distribution*. Again the same optimal convergence rates as in Figs. 1.5,1.7 are obtained. After the previous analysis we can wrap up the following conclusions, on the sensitivity analysis pursued orthonormalizing the basis (1.2.3) of the polynomial space w.r.t. densities ρ that differ from the experimental densities $\widehat{\rho}$ where the samples of the random variable are drawn:

- a worsening in the stability due to the choice of nonorthonormal basis of the polynomial space occurs,
- the errors obtained employing the basis of the polynomial space orthonormal w.r.t. the *arcsine distribution* and *Wigner distribution* and those obtained with Legendre polynomials are the same.

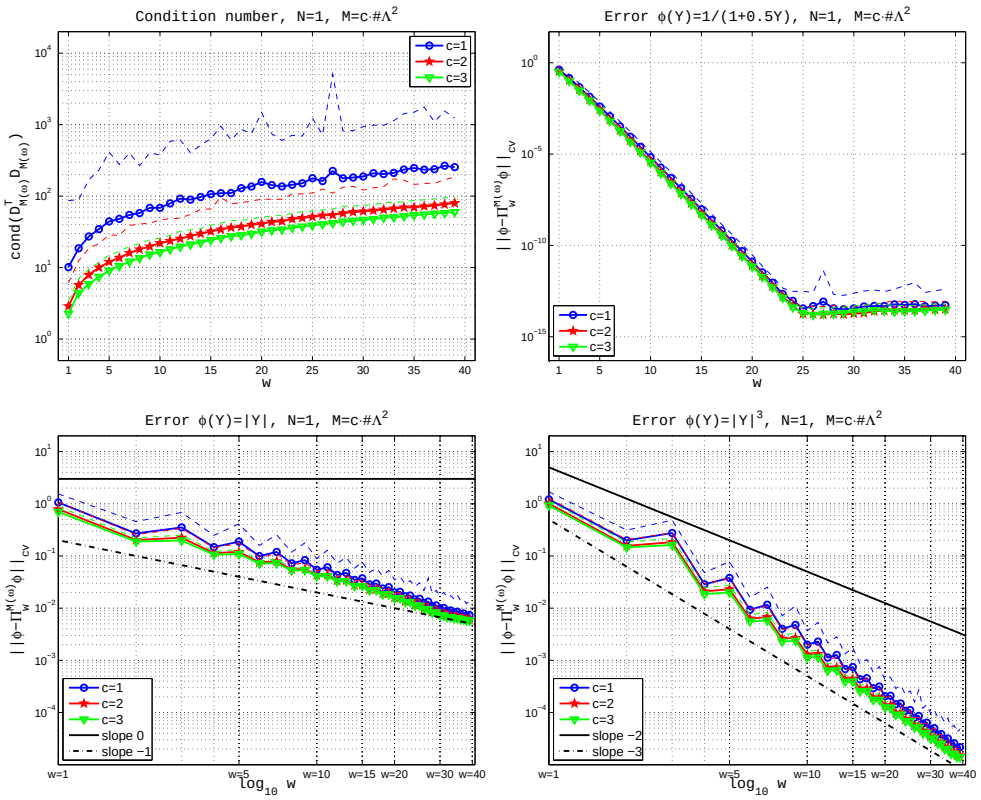


Figure 3.43: *Arcsine distribution.* $M = c \cdot (\#\Lambda)^2$. Top-left: mean condition number (1.4.9). Top-right: approximation error (1.4.6) for the function (3.4.2) with $\beta = 0.5$. Bottom-left: approximation error (1.4.6) for the function (3.4.5). Bottom-right: approximation error (1.4.6) for the function (3.4.4). The continuous marked lines show the mean error over 1000 repetitions. The dashed lines show the mean error plus one standard deviation.

Chapter 4

Approximation of Quantities of Interest in PDEs with stochastic coefficients and stochastic domain

This chapter is entirely based on the content of the paper [MNvST13] entitled *Approximation of Quantities of Interest in stochastic PDE by the random discrete L^2 projection on polynomial spaces*, by G. Migliorati, F. Nobile, E. von Schwerin and R. Tempone.

Abstract

In this work we consider the random discrete L^2 projection on polynomial spaces (hereafter RDP) for the approximation of scalar Quantities of Interest (QOIs) related to the solution of a Partial Differential Equation model with random input parameters. In the RDP technique the QOI is first computed for independent samples of the random input parameters, as in a standard Monte Carlo approach, and then the QOI is approximated by a multivariate polynomial function of the input parameters using a discrete least squares approach. We consider several examples including the Darcy equations with random permeability; the linear elasticity equations with random elastic coefficient; the Navier–Stokes equations in random geometries and with random fluid viscosity. We show that the RDP technique is well suited to QOIs that depend smoothly on a moderate number of random parameters. Our numerical tests confirm the theoretical findings in [MNvST11], which have shown that, in the case of a single uniformly distributed random parameter, the RDP technique is stable and optimally convergent if the number of sampling points is proportional to the square of the dimension of the polynomial space. Here optimality means that the weighted L^2 norm of the RDP error is bounded from above by the best L^∞ error achievable in the given polynomial space, up to logarithmic factors. In the case of several random input parameters, the numerical evidence indicates that the condition on quadratic growth of the number of sampling points could be relaxed to a linear growth and still achieve stable and optimal convergence. This makes the RDP technique very promising for moderately high dimensional uncertainty quantification.

Key words: PDE stochastic data, discrete least squares, polynomial approximation

AMS subject classification: 41A10, 65N35

4.1 Introduction

In recent years, the modeling of uncertainty in mathematical models has attracted a lot of attention in the scientific community. When a probabilistic framework is considered, uncertainty in the model parameters is modeled in terms of random variables. The underlying challenge concerns the accurate and efficient approximation of the model outcome in presence of many random input parameters. In the context of Partial Differential Equations (PDEs) with stochastic data, a well-established technique that has been used in many engineering applications [Eld11, HWB10, GZ07a] consists in the use of a spectral expansion to represent the input/output dependence; see e.g. [GS03, XK02a, LMK10]. Once such an expansion has been computed by some means, statistics of the model output can easily be recovered. The *random discrete* L^2 *projection* (RDP), also known as regression or point collocation approach, has been proposed in the context of applications devoted to uncertainty analysis in [HWB10, BS08, EWC08] as a tool for computing the spectral expansion of the model response. The regression approach is based on the evaluation of the target output function on randomly selected points, and aims to improve the slow convergence of the classical Monte Carlo method by performing a discrete projection onto a multivariate polynomial space. It differs from other techniques based on a deterministic choice of the points at which the function is evaluated, also known as *Collocation methods on Sparse Grids* [BG04].

In [MNvST11, CDS11] the RDP was analyzed in the context of approximating a smooth aleatory function in the L^2 probability sense. This approximation problem falls in the field of nonparametric regression with random design, and when noise is present there exist well-known estimates for the approximation error [GKKW02]. The RDP of a given aleatory function, consists in its discrete L^2 projection over a polynomial space, and is computed by evaluating the target function pointwise in randomly selected points of the parameter space. Unlike the usual nonparametric regression in statistics, the evaluations are assumed to be noise-free. The stability and convergence properties of RDP are analyzed in [MNvST11, CDS11]. Under some assumptions on the probability density, it is proved in these works that, in one dimension and when the number of sampling points is proportional to the square of the dimension of the polynomial space, the RDP converges optimally in the sense that the L^2error behaves like the best approximation error (measured in the "sup" norm) of the target function in the chosen polynomial space, up to logarithmic factors. It is reasonable to compare with the L^∞ best approximation error since the RDP is based on pointwise evaluations. In [CDS11] a bound in terms of the best approximation in L^2 norm is obtained by taking expectations. In contrast the bound in [MNvST11] is in terms of the L^∞ norm and holds with high probability. The above-mentioned quadratic relation between number of sampling points and dimension of the polynomial space actually guarantees the stability of the RDP. Several numerical tests in [MNvST11] show the capabilities of the method, and highlight the influence of the dimension of the parameter space and of the smoothness of the target function on the achieved convergence rate.

The present work focuses on the application of RDP to the approximation of QOIs related to the solution of PDEs with stochastic data. The aleatory function is defined as a functional of the solution of the PDE, *e.g.* an integral over a portion of the physical domain. We begin by considering the Darcy model with a random diffusion coefficient, parametrized in a one-dimensional parameter space. The randomness affects the value of the coefficient in a particular region of the physical domain, for instance an inclusion. Then we investigate a variation of the Darcy example where the randomness affects the location where the diffusion coefficient is discontinuous. Thus the example treats an inclusion with a random domain in the sense that the geometrical shape of the domain is parametrized in terms of a random variable.

Next, we move to higher dimensional parameter spaces, always using the isotropic Total Degree polynomial space (defined in Section 4.2). This choice is well motivated in cases where the target function depends analytically on each input parameter with analyticity region not affected by the other parameters [BNTT12]. We consider a Darcy model with a random diffusion coefficient parametrized in a five-dimensional parameter space, increasing to five the number of

nonoverlapping inclusions that are displaced in the physical domain.

Finally, we consider two more complex vectorial problems: the Navier–Lamé equations that govern the bending of a cantilever beam in the regime of linear elasticity, and the incompressible Navier–Stokes equations that govern the motion of a viscous fluid in a pipe. Both examples contain uncertainty in the model parameters: the Young modulus in the former, and the viscosity and the geometry of the domain in the latter.

All examples of the elliptic class presented in this paper, with one-dimensional parameter space, validate the theoretical results outlined in [MNvST11, CDS11]. In particular, a number of sampling points proportional to the square of the dimension of the polynomial space always yields optimal convergence, as predicted by the theory. Moreover, when approximating a smooth function an optimal convergence is observed up to a certain threshold, even if the number of sampling points is only linearly proportional to the dimension of the polynomial space. Beyond such a threshold, the error starts diverging. On the other hand, in all examples with higher dimensional parameter space, optimal convergence rate for smooth functions has always been achieved, even with a linear proportionality between the number of sampling points and the dimension of the polynomial space. No blow-up has been observed in all ranges of polynomial space dimensions tested. This is consistent with the observation in [MNvST11]: when the dimension gets higher the RDP seems to become stable even when the number of sample points is only linearly proportional to the dimension of the polynomial space. In this sense, the one-dimensional case is the most ill-conditioned case, making the RDP more promising in moderately high dimensional approximation problems.

The outline of the paper is the following: in Section 4.2 we introduce the formulation of the random discrete L^2 projection and recall the main results obtained in [MNvST11] concerning stability and optimality. In Section 4.3 the numerical examples are presented. Examples 1, 2, and 3 are based on the Darcy model with random permeability. Examples 4 and 5 address the linear elasticity equations with random elastic coefficient, and Navier–Stokes equations in random geometries and with random fluid viscosity. Finally, Section 4.4 contains the concluding discussion.

4.2 The random discrete L^2 projection on polynomial spaces

In this section we review the formulation of the random discrete L^2 projection in an abstract setting and recall the main results obtained in [MNvST11]. The technique will then be applied to PDEs with random data in Section 4.3.

Let $\Gamma \subseteq \mathbb{R}^N$ be an N-dimensional subset of the N-dimensional Euclidean space, with a tensor structure form $\Gamma = \Gamma^1 \times \cdots \times \Gamma^N$. Denote by $\rho : \Gamma \to \mathbb{R}^+$ a probability density function over Γ, and by $\mathbf{Y} = (Y_1, \ldots, Y_N)$ a vector of N random variables, taking values in Γ and distributed according to the density ρ.

We consider a random variable $Z = \phi(\mathbf{Y})$, where $\phi : \Gamma \to \mathbb{R}$ is assumed to be a smooth function, and we are interested in computing statistical moments of Z. This will be achieved by first constructing a reduced model; i.e. we approximate the function $\phi(Y_1, \ldots, Y_N)$ by a suitable multivariate polynomial $\phi_\Lambda(Y_1, \ldots, Y_N)$. We then compute statistical moments using the approximate function ϕ_Λ.

We denote by

$$\mathbb{E}[Z] := \int_\Gamma \phi(\mathbf{y})\rho(\mathbf{y})d\mathbf{y}$$

the expected value of the random variable $Z = \phi(\mathbf{Y})$ and by

$$Pr(A) := \int_A \rho(\mathbf{y})d\mathbf{y}$$

the probability of the event $A \in \mathcal{B}(\Gamma)$, where $\mathcal{B}(\Gamma)$ is the Borel σ-algebra with respect to the measure $\rho(\mathbf{y})d\mathbf{y}$. Throughout the paper we also assume the followin.

Assumption 1.1. Γ *is a bounded set, and* $0 < \rho_{min} \leq \rho(\mathbf{y}) \leq \rho_{max} < \infty$, *for all* $\mathbf{y} \in \Gamma$.

Remark 4.1. *This assumption excludes* e.g. *the normal and lognormal cases, since these densities are not bounded away from zero due to their unbounded support. In the case of densities with unbounded support,* e.g. *normal and lognormal, derivations of results corresponding to Propositions 4.1, 4.2 and Theorem 4.1 would have to rely on the use of suitable weighted* L^∞ *norms (see* e.g. *[BNT10]), but this is out of the scope of the present paper.*

We introduce the space L^2_ρ of square integrable functions $f : \Gamma \to \mathbb{R}$, endowed with the norm

$$\|f\|_{L^2_\rho} = \left(\int_\Gamma f^2(\mathbf{y})\rho(\mathbf{y})d\mathbf{y} \right)^{1/2}.$$

Let $\mathbf{p} = (p_1, \ldots, p_N)$ be a multi-index and $\Lambda \subset \mathbb{N}^N$ a set of multi-indices. In what follows we consider only sets Λ that are monotone in the following sense.

Property 1 (Monotonicity of Λ)**.** *Consider two multi-indices* $\mathbf{p}', \mathbf{p}'' \in \mathbb{N}^N$ *such that* $p''_n \leq p'_n$ *for all* $n = 1, \ldots, N$. *The multi-index set* Λ *is monotone if the following holds:*

$$\mathbf{p}' \in \Lambda \Longrightarrow \mathbf{p}'' \in \Lambda.$$

We denote by $\mathbb{P}_\Lambda(\Gamma)$ the multivariate polynomial space

$$\mathbb{P}_\Lambda(\Gamma) = \text{span}\left\{ \prod_{n=1}^N y_n^{p_n}, \text{ with } \mathbf{p} \in \Lambda \right\}, \tag{4.2.1}$$

and by $\#\Lambda = dim(\mathbb{P}_\Lambda)$ the dimension of the polynomial space, which corresponds to the cardinality of the multi-index set Λ. For convenience, the set Λ can be arranged in lexicographical order, and according to this order, we can denote by $\mathbf{p}^j$ the jth multi-index of Λ. Sometimes we refer to the elements of Λ with the generic multi-index $\mathbf{p}$, rather than listing them by the lexicographical index.

Since the monomial basis in (4.2.1) is very ill-conditioned, in practice we use an orthonormal polynomial basis. A typical choice is to take orthogonal polynomials with respect to the measure $\rho(\mathbf{y})d\mathbf{y}$. We introduce an N-dimensional orthonormal polynomial basis $\{l_j\}_{j=1}^{\#\Lambda}$ of $\mathbb{P}_\Lambda$ with respect to the weighted inner product

$$(u, v)_{L^2_\rho} = \int_\Gamma u(\mathbf{y})v(\mathbf{y})\rho(\mathbf{y})\, d\mathbf{y},$$

i.e. $(l_i, l_j)_{L^2_\rho} = \delta_{ij}$. Assumption 1.1 ensures that the orthonormal basis is complete in L^2_ρ when $\Lambda = \mathbb{N}^N$, applying Theorems 3.3 and 3.5 of [EMSU12].

In the particular case where the density factorizes as $\rho(\mathbf{y}) = \prod_{n=1}^N \rho_n(y_n)$ the basis can be constructed by tensorizing one-dimensional orthogonal polynomials with respect to each weight ρ_n separately. Given n, let $\{\varphi_j^n(\cdot)\}_j$ be the orthogonal polynomials with respect to ρ_n. The jth multi-index $\mathbf{p}^j \in \Lambda$ is associated with the corresponding jth multidimensional basis function by

$$l_j(\mathbf{y}) = \prod_{n=1}^N \varphi_{p_n^j}^n(y_n). \tag{4.2.2}$$

Thus, using the basis functions provided by (4.2.2), the definition (4.2.1) of $\mathbb{P}_\Lambda$ becomes

$$\mathbb{P}_\Lambda(\Gamma) = \text{span}\{l_j,\, j = 1, \ldots, \#\Lambda\}, \tag{4.2.3}$$

and of course $\dim(\mathbb{P}_\Lambda) = \#\Lambda$. Observe that in general (4.2.1) and (4.2.3) are equivalent only if the index set Λ satisfies the Monotonicity Property 1.

We consider a sample $\mathbf{y}_1, \dots, \mathbf{y}_M$ of size $M \geq \#\Lambda$ of independent random variables identically distributed according to the density ρ, and then evaluate the function ϕ pointwise at each value $\mathbf{y}_i$, $i = 1, \dots, M$.

Finally, we compute a discrete least square approximation of the values $\phi(\mathbf{y}_i)$ in the polynomial space $\mathbb{P}_\Lambda$, i.e.

$$\phi_\Lambda = \Pi_M^{\Lambda,\omega}\phi = \operatorname*{argmin}_{v \in \mathbb{P}_\Lambda(\Gamma)} \frac{1}{M}\sum_{i=1}^{M}(\phi(\mathbf{y}_i) - v(\mathbf{y}_i))^2. \tag{4.2.4}$$

We will use the superscript (or subscript) ω to denote a quantity that depends on the random sample $\mathbf{y}_1, \dots, \mathbf{y}_M$ (and therefore is random itself).

We now introduce the *random* discrete inner product

$$(u, v)_{M,\omega} = \frac{1}{M}\sum_{i=1}^{M} u(\mathbf{y}_i)v(\mathbf{y}_i) \tag{4.2.5}$$

on $\mathbb{P}_\Lambda(\Gamma)$, and the corresponding discrete norm $\|u\|_{M,\omega} = (u, u)_{M,\omega}^{1/2}$. Note that for $M \geq \#\Lambda$ the bilinear form (4.2.5) is an inner product on $\mathbb{P}_\Lambda(\Gamma)$ with probability one, by Assumption 1.1. With this notation we can write (4.2.4) as

$$\text{find } \Pi_M^{\Lambda,\omega}\phi \in \mathbb{P}_\Lambda(\Gamma) \quad \text{such that} \quad (\phi - \Pi_M^{\Lambda,\omega}\phi, v)_{M,\omega} = 0, \quad \forall\, v \in \mathbb{P}_\Lambda(\Gamma).$$

Some of the most common choices of function spaces are Tensor Product, Total Degree, and Hyperbolic Cross, which are defined by the index sets below. We index the set Λ by the subscript w, which denotes the maximum polynomial degree used:

$$\textbf{Tensor Product (TP)}, \quad \Lambda_{\mathrm{w}} = \Big\{\mathbf{p} \in \mathbb{N}^N : \max_{n=1,\dots,N} p_n \leq \mathrm{w}\Big\}, \tag{4.2.6}$$

$$\textbf{Total Degree (TD)}, \quad \Lambda_{\mathrm{w}} = \Big\{\mathbf{p} \in \mathbb{N}^N : \sum_{n=1}^{N} p_n \leq \mathrm{w}\Big\}, \tag{4.2.7}$$

$$\textbf{Hyperbolic Cross (HC)}, \quad \Lambda_{\mathrm{w}} = \Big\{\mathbf{p} \in \mathbb{N}^N : \prod_{n=1}^{N} (p_n + 1) \leq \mathrm{w} + 1\Big\}. \tag{4.2.8}$$

These spaces are isotropic in the sense that the maximum polynomial degree w is the same in all variables $Y_1, \dots, Y_N$. The dimensions of the TP and TD spaces are

$$\#TP(\mathrm{w}, N) = (\mathrm{w} + 1)^N, \tag{4.2.9}$$

$$\#TD(\mathrm{w}, N) = \binom{N + \mathrm{w}}{N}. \tag{4.2.10}$$

The dimension of the HC space is harder to quantify. An upper bound is given by

$$\#HC(\mathrm{w}, N) \leq \left\lfloor (\mathrm{w} + 1) \cdot (1 + \log(\mathrm{w} + 1))^{N-1} \right\rfloor, \tag{4.2.11}$$

where $\lfloor \cdot \rfloor$ denotes the operation of rounding a real number downwards to the nearest integer. This bound is sharp for $N = 2$ and becomes very conservative as N increases.

4.2.1 Stability and convergence rate of the random discrete L^2 projection

Here we recall some theoretical results concerning the discrete L^2 projection, derived in [MNvST11] and [CDS11]. Let us first introduce the quantity

$$C^\omega(M, \Lambda) := \sup_{v \in \mathbb{P}_\Lambda \setminus \{v \equiv 0\}} \frac{\|v\|_{L^2_\rho}^2}{\|v\|_{M,\omega}^2}, \tag{4.2.12}$$

which depends on the random sample and is therefore a random variable. The following proposition states the optimality of the discrete L^2 projection with respect to the L^∞ norm, when the error is evaluated in the L^2_ρ norm, that is:

Proposition 4.1 (see [MNvST11]). *With $C^\omega(M,\Lambda)$ defined as in* (4.2.12)*, it holds that*

$$\|\phi - \Pi_\Lambda^{M,\omega}\phi\|_{L^2_\rho} \leq \Big(1 + \sqrt{C^\omega(M,\Lambda)}\Big) \inf_{v\in\mathbb{P}_\Lambda(\Gamma)} \|\phi - v\|_{L^\infty}. \tag{4.2.13}$$

As a consequence, the convergence properties of the RDP are strictly related to the properties of the quantity $C^\omega(M,\Lambda)$. The next theorem quantifies the asymptotic behavior of the random variable $C^\omega(M,\Lambda)$.

Theorem 4.1 (see [MNvST11]). *Let $C^\omega(M,\Lambda)$ be the random variable defined in* (4.2.12)*. Then, for any given Λ we have*

$$\lim_{M\to\infty} C^\omega(M,\Lambda) = 1, \quad a.s.$$

The previous proposition and theorem are general results on the discrete L^2 projection: they hold in any dimension N, for any arbitrary N-dimensional monotone multi-index set Λ, and for any density ρ satisfying Assumption 1.1.

When $N = 1$ and $\rho = \mathbb{U}([-1,1])$ a probability estimate has been proved in [MNvST11]. In this particular case the polynomial space is denoted by $\mathbb{P}_{\mathrm{w}}$ rather than $\mathbb{P}_\Lambda$, because the multi-index set Λ is just $\{0,1,\dots,\mathrm{w}\}$ and its dimension $\#\Lambda = 1 + \mathrm{w}$. Accordingly, the projector $\Pi_\Lambda^{M,\omega}$ is denoted by $\Pi_{\mathrm{w}}^{M,\omega}$. The following theorem ensures the stability and accuracy of the discrete L^2 projection, under the condition $M \propto (\#\Lambda)^2$.

Theorem 4.2 ([MNvST11]). *For any $\delta \in (0,1)$, under the condition*

$$\frac{M}{3\log((M+1)/\delta)} \geq 4\sqrt{3}\,\mathrm{w}^2 \tag{4.2.14}$$

it holds

$$Pr\left(\|\phi - \Pi_{\mathrm{w}}^{M,\omega}\phi\|_{L^2_\rho} \leq \left(1 + \sqrt{3\log\frac{M+1}{\delta}}\right) \inf_{v\in\mathbb{P}_{\mathrm{w}}} \|\phi - v\|_{L^\infty}\right) \geq 1 - \delta. \tag{4.2.15}$$

We remark that in practice condition (4.2.14) is equivalent to $M \propto (\#\Lambda)^2$, since the effect due to the presence of the nonoptimal logarithmic factor is often negligible. In [CDS11] an estimate in expectation of the error $\|\phi - \Pi_{\mathrm{w}}^{M,\omega}\phi\|_{L^2_\rho}$ has also been derived, showing that it behaves as the L^2 best approximation error under the same condition $M \propto (\#\Lambda)^2$. Moreover, for analytic functions, the best approximation error converges exponentially as w increases. Note that, to keep condition (4.2.14) satisfied, increasing w requires increasing M as well.

There also exists a general relation between the optimal convergence rate of the random projection and its stability, which holds again in any dimension N and for any Λ and ρ. The same random variable C^ω, besides entering into the error estimate (4.2.13), plays a role in the stability of the random projection, as stated in the next proposition. As in [MNvST11], we denote by D^ω the random design matrix associated with problem (4.2.4); its elements are defined as $[D^\omega]_{i,j} = l_j(\mathbf{y}_i)$.

Proposition 4.2 (from [MNvST11]). *The spectral condition number (2 norm) of the matrix $(D^\omega)^T D^\omega$ is equal to*

$$cond\Big((D^\omega)^T D^\omega\Big) = c^\omega(M,\Lambda)\, C^\omega(M,\Lambda), \tag{4.2.16}$$

where

$$c^\omega(M,\Lambda) := \sup_{v\in\mathbb{P}_\Lambda\setminus\{v\equiv 0\}} \frac{\|v\|^2_{M,\omega}}{\|v\|^2_{L^2_\rho}}. \tag{4.2.17}$$

Remark 4.2. *It is also possible to construct a RDP approximation $\widetilde{\Pi}_{\mathrm{w}}^{M,\omega}\phi$ in which the sampling points are drawn from a density $\widetilde{\rho}$ different than ρ. Assuming that $\rho(\mathbf{y})/\widetilde{\rho}(\mathbf{y}) < +\infty$, $\forall \mathbf{y} \in \Gamma$, the following error bound will hold:*

$$\|\phi - \widetilde{\Pi}_{\mathrm{w}}^{M,\omega}\phi\|_{L^2_\rho(\Gamma)} \leq \left\|\frac{\rho}{\widetilde{\rho}}\right\|^{\frac{1}{2}}_{L^\infty(\Gamma)} \|\phi - \widetilde{\Pi}_{\mathrm{w}}^{M,\omega}\phi\|_{L^2_{\widetilde{\rho}}(\Gamma)}. \tag{4.2.18}$$

We propose then to choose the density $\widetilde{\rho}$ so as to improve the stability of RDP, while preserving its accuracy. A good candidate is given by the Chebyshev distribution. Indeed, it has been shown in [CDS11] that when drawing samples from the monovariate Chebyshev distribution, the RDP is stable and optimally convergent under the assumption $M \propto \#\Lambda$ which is much less demanding than (4.2.14)*.*

4.3 Parametric PDEs

In [MNvST11] we presented some numerical examples of RDP to approximate monovariate and multivariate target functions $\phi = \phi(\mathbf{y}) : \Gamma \to \mathbb{R}$ on polynomial spaces. The role of smoothness was investigated. When using a relation $M \propto (\#\Lambda)^2$, an optimal (in the sense described in the introduction) convergence rate was always observed. On the other hand, when using a linear relation $M \propto \#\Lambda$, the optimal convergence rate was observed up to a certain threshold, after which the error started increasing and eventually diverged. This effect was clearly observed in the monovariate case. However, in higher dimensions the blow-up of the error was not observed in the range of practical polynomial degrees explored. As pointed out in [MNvST11], the linear relation $M \propto \#\Lambda$ therefore seems to be sufficient for all practical purposes to achieve an accurate and stable approximation.

The aim of this paper is to test the RDP when the target function ϕ is related to the solution of a stochastic PDE model. We focus on QOIs of integral type over the spatial domain of the PDE model, such as the mean of the solution or its gradient in portions of the domain, or pointwise quantities such as the maximum or minimum of the solution in the domain.

Consider a steady state PDE model,

$$\mathcal{L}(\mathbf{x}, \mathbf{y}, u, f, g, \Omega) = 0, \tag{4.3.1}$$

defined on a bounded domain $\Omega \subset \mathbb{R}^d$ and parametrized by $\mathbf{y} \in \Gamma \subseteq \mathbb{R}^N$, with $u : \Omega \times \Gamma \to \mathbb{R}$ its solution, $f : \Omega \times \Gamma \to \mathbb{R}$ the forcing term, and $g : \partial\Omega \times \Gamma \to \mathbb{R}$ a suitable boundary condition. In general, the operator $\mathcal{L}$ can be nonlinear w.r.t. the solution u of the model. The following examples feature only two-dimensional spatial domains Ω, *i.e.* $d = 2$. However, extensions to three-dimensional problems are straightforward. The domain may also depend on the parameter $\mathbf{y}$, *i.e.* $\Omega = \Omega(\mathbf{y})$. Moreover, we assume that f and g satisfy proper conditions to make the whole model well-posed in the sense of Hadamard.

Considering $\mathbf{y} \in \Gamma$ as a realization of the random variable $\mathbf{Y}$ distributed according to the density $\rho : \Gamma \to \mathbb{R}^+$, the parametric model (4.3.1) can also be considered as a PDE model with stochastic data.

We now proceed with some examples to illustrate the application of the RDP to approximate QOIs depending on the solution of stochastic PDEs. First we focus on models in the elliptic class, where the solution typically depends smoothly on the random variable, as proved in [CDS11, BNT10].

The first three examples concern the Darcy flow in a medium containing some inclusions. The first and second examples have a one-dimensional stochastic parameter space: in Example 1 the value of the diffusion coefficient is stochastic, while in Example 2 the geometrical shape of the inclusion is stochastic. In Example 3 we increase the dimension of the parameter space to five.

Then we analyze the linear elasticity model and the incompressible Navier–Stokes model. The former exhibits a highly regular dependence of the solution on the random parameter affecting *Young's modulus*, while the latter shows some nonsmooth QOIs.

To quantify the error $\|\phi - I\!\!\Pi_\Lambda^{M,\omega}\phi\|_{L^2_\rho}$ committed by RDP we employ the cross-validation procedure described in [MNvST11, Section 4]. A random set of 100 cross-validating points is chosen at the beginning. Then the cross-validated error in the ∞ norm is defined as

$$\|\phi - I\!\!\Pi_\Lambda^{M,\omega}\phi\|_{cv} = \max_{\mathbf{z}_j} |\phi(\mathbf{z}_j) - I\!\!\Pi_\Lambda^{M,\omega}\phi(\mathbf{z}_j)|, \tag{4.3.2}$$

where $\mathbf{z}_j$, $j = 1, \ldots, 100$, is the set of cross-validation points. To estimate the variability of (4.3.2) due to the random sampling of the M collocation points, we have repeated the calculation over R independent sets of points $y_j^{\omega_k}, j = 1, \ldots, M$, with $k = 1, \ldots, R$, and we have computed the average error

$$\overline{E}_{cv} = \frac{\sum_{k=1}^R \|\phi - I\!\!\Pi_\Lambda^{M,\omega}\phi\|_{cv}}{R}, \tag{4.3.3}$$

as well as the sample standard deviation by

$$s_E = \sqrt{\frac{1}{R-1}\sum_{k=1}^R \left(\|\phi - I\!\!\Pi_\Lambda^{M,\omega}\phi\|_{cv} - \overline{E}_{cv}\right)^2}. \tag{4.3.4}$$

We also aim to analyze the condition number of $(D^\omega)^T D^\omega$ that has been introduced in (4.2.16). Following [MNvST11, Section 4], it can be quantified as

$$\mathrm{cond}\Big((D^\omega)^T D^\omega\Big) = \frac{\sigma_{\max}\Big((D^\omega)^T D^\omega\Big)}{\sigma_{\min}\Big((D^\omega)^T D^\omega\Big)}, \tag{4.3.5}$$

where $\sigma_{\max}(\cdot)$ and $\sigma_{\min}(\cdot)$ are the maximum and minimum singular values. Again, denoting by D^{ω_k} the design matrix built with the kth set $\{\mathbf{y}_j^{\omega_k}\}_j$, we estimate the mean condition number $\overline{K}$ over the R repetitions as

$$\overline{K} = \frac{\sum_{k=1}^R \mathrm{cond}\left((D^{\omega_k})^T D^{\omega_k}\right)}{R}, \tag{4.3.6}$$

and the standard deviation as

$$s_K = \sqrt{\frac{1}{R-1}\sum_{k=1}^R \left(\mathrm{cond}\Big((D^{\omega_k})^T D^{\omega_k}\Big) - \overline{K}\right)^2}. \tag{4.3.7}$$

In the convergence plots, the continuous lines mark the mean of the error $\|\phi - I\!\!\Pi_\Lambda^{M,\omega}\phi\|_{L^2_\rho}$, or the mean of the condition number of the random design matrix, while the dashed line mark the mean plus one standard deviation. The discretization of the PDE model over the spatial domain Ω is obtained by means of the Finite Element method. In Examples 1,2,3,4 the P1 finite elements are used. In Example 5 the *inf-sup* compatible P2-P1 finite elements are used.

4.3.1 Example 1: the Darcy flow in a domain with one inclusion

To start, we consider an elliptic model on a bounded deterministic domain $\Omega \subset \mathbb{R}^2$, with a random diffusion coefficient μ:

$$\begin{cases} -\nabla \cdot (\mu(\mathbf{x}, \mathbf{y})\, \nabla u(\mathbf{x}, \mathbf{y})) = 0, & \mathbf{x} \text{ in } \Omega,\ \mathbf{y} \in \Gamma, \\ u(\mathbf{x}, \mathbf{y}) = g_1(\mathbf{x}), & \mathbf{x} \text{ on } \partial\Gamma_1 \cup \partial\Gamma_3,\ \mathbf{y} \in \Gamma, \\ \partial_\nu u(\mathbf{x}, \mathbf{y}) = g_2(\mathbf{x}), & \mathbf{x} \text{ on } \partial\Gamma_2 \cup \partial\Gamma_4,\ \mathbf{y} \in \Gamma. \end{cases} \tag{4.3.8}$$

We set problem (4.3.8) in a unitary square domain, with a circular inclusion Ω_I with radius 0.2, as shown in Fig. 4.1. The edges are labeled clockwise as $\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4$ starting from the left. We

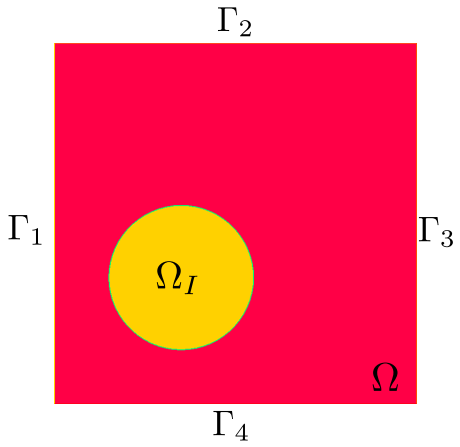


Figure 4.1: Square domain Ω with a circular inclusion Ω_I.

impose nonhomogeneous Dirichlet conditions on the vertical edges and homogeneous Neumann conditions on the horizontal ones, to force a steady state flow from left to right. The random diffusion coefficient depends on a uniform random variable $Y \sim \mathcal{U}(-1, 1)$ and is defined as

$$\mu(\mathbf{x}, Y) = \begin{cases} \exp(5\,Y), & \Omega_I, \\ 10^{-4}, & \Omega \setminus \Omega_I. \end{cases} \tag{4.3.9}$$

Such a model for the coefficient can be employed in practical situations where the value of the diffusion properties of the material are not accurately determined in a given region of the physical domain, or when the value is a function of the outcome of some stochastic process with a known underlying probability law. Notice that the aforementioned diffusion coefficient may jump by more than four orders of magnitude from the bulk to the inclusion.

The QOIs we analyze are defined by integrals of the solution over the physical domain. We consider the mean of the solution u in Ω,

$$\mathrm{QOI}_1(u) = \frac{1}{|\Omega|} \int_\Omega u \, d\Omega; \tag{4.3.10}$$

the mean of $|\nabla u|^2$ in Ω

$$\mathrm{QOI}_2(u) = \frac{1}{|\Omega|} \int_\Omega |\nabla u|^2 \, d\Omega; \tag{4.3.11}$$

and the mean of the solution on the left boundary segment Γ_1,

$$\mathrm{QOI}_3(u) = \frac{1}{|\Gamma_1|} \int_{\Gamma 1} u \, d\Omega. \tag{4.3.12}$$

All these QOIs are analytic functions of the parameter $\mathbf{y}$ (see *e.g.* [BNT10, CDS11]) so we expect the best approximation error to decay exponentially fast w.r.t. the polynomial degree w. The numerical results obtained with a sample size chosen as $M \propto \#\Lambda$ are reported in Fig. 4.2, where we plot the condition number (4.3.5) of the matrix $(D^\omega)^T D^\omega$ (top-left) and the error in the QOI (measured using (4.3.2)) versus the polynomial degree.

When the value of the proportionality constant c decreases too much, the convergence rate achieved by the random projection is reduced, and the variability of the error amplifies as well. Analogous results were shown in [MNvST11] for scalar target functions.

Fig. 4.3 shows the results for QOI_1 in the case of a sample size $M \propto (\#\Lambda)^2$. We see that, as the proportionality constant increases, the convergence curve "stabilizes" onto a limit curve that corresponds to the optimal convergence, *i.e.* the convergence of the exact L^2 projection. Note that Figs. 4.2 and 4.3 give convergence plots in terms of the dimension of the polynomial

space. However, the actual cost of RDP is proportional to the sample size M; therefore it is also useful to look at convergence plots of error versus M. Note that Fig. 4.2 (top-right) and Fig. 4.3 were based on the same data sets, but including only values of $\mathrm{w} \leq 10$. This is shown in Fig. 4.4, where error plots for QOI_1 are shown for different choices of proportionality relation $M \propto (\#\Lambda)^\alpha$.

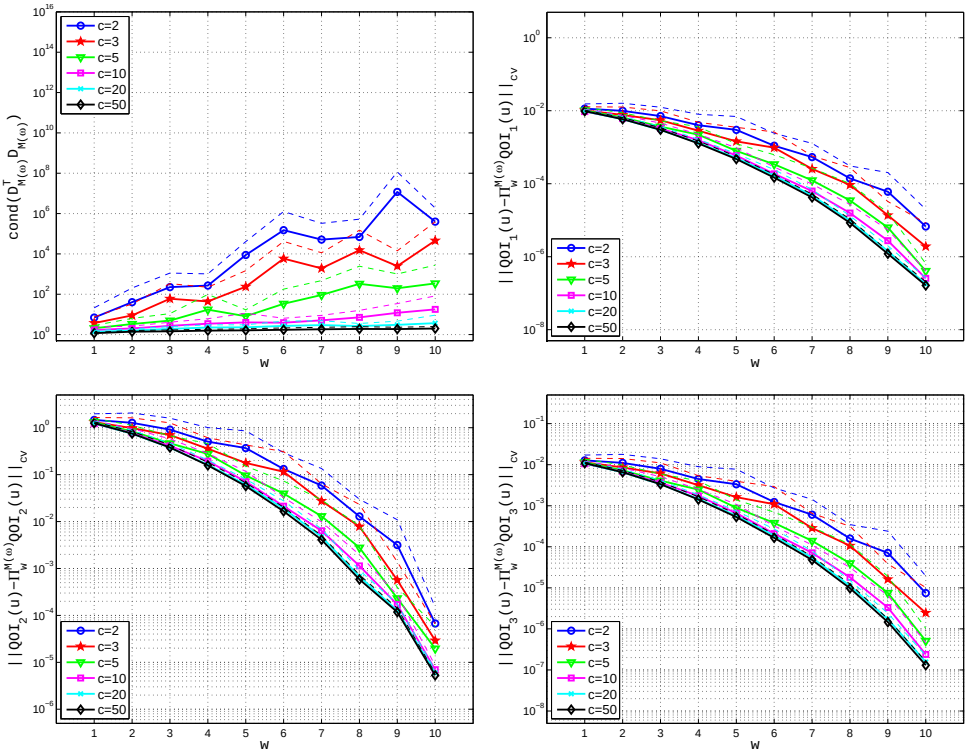


Figure 4.2: Example 1. Condition number (4.3.5) (top-left) and approximation errors (4.3.2) in the QOI_i, $i = 1, 2, 3$, using $M = c \cdot \#\Lambda = c \cdot (\mathrm{w}+1)$. Continuous lines: sample mean averaged over $R = 100$ repetitions. Dashed lines: sample mean plus one sample standard deviation averaged over $R = 100$ repetitions.

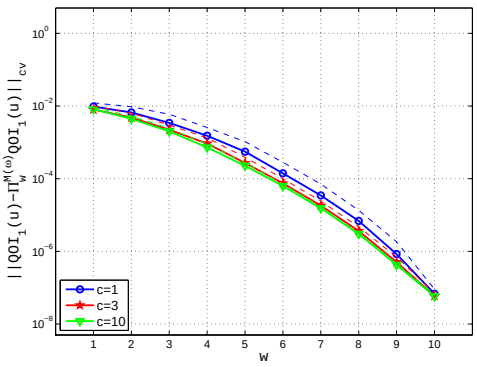


Figure 4.3: Example 1. Approximation errors (4.3.2), using $M = c \cdot (\#\Lambda)^2 = c \cdot (\mathrm{w}+1)^2$. Continuous lines: sample mean averaged over $R = 100$ repetitions. Dashed lines: sample mean plus one sample standard deviation averaged over $R = 100$ repetitions.

Observe that the choice $M \propto (\#\Lambda)^2$, which according to the theoretical estimates is the minimal relation that guarantees stability of the projection, leads to exponential convergence. Choosing a smaller sample size, *e.g.* $M = 2 \cdot \#\Lambda$, yields a faster convergence w.r.t. M up to a certain threshold, after which the solution deteriorates. This effect is also seen in [MNvST11, Fig.8].

As an example of a statistical moment obtained from the RDP reduced model we computed $\mathbb{E}[\mathrm{QOI}_1]$. The error versus number of samples M is reported in Fig. 4.4 (right) and its decay is very similar to that of the cross-validated error. We see that in this example RDP is clearly superior to a simple Monte Carlo approximation of the same expected value, where the error decays with the rate $M^{-1/2}$.

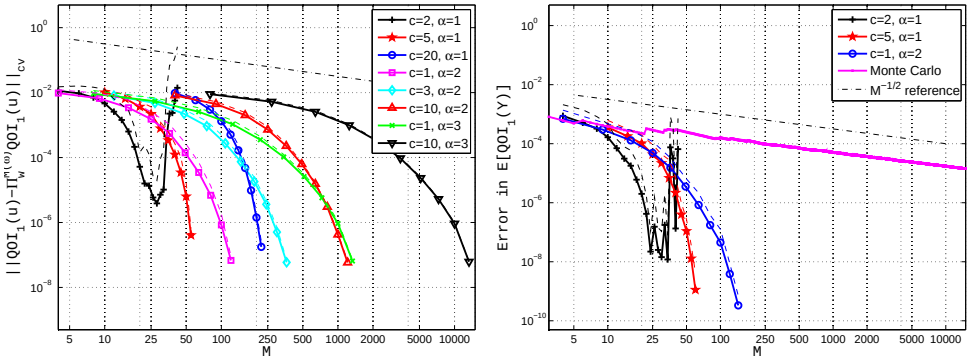


Figure 4.4: Example 1. Left: approximation errors (4.3.2) in the QOI_1 versus M for different choices of the sample size, $M = c \cdot (\#\Lambda)^\alpha = c \cdot (\mathrm{w}+1)^\alpha$. The markers are used to indicate progressive values of w, starting from $\mathrm{w} = 1$. Right: errors in the scalar quantity $\mathbb{E}[\mathrm{QOI}_1]$ versus M. Here $\mathbb{E}[\mathrm{QOI}_1]$ was approximated from the RDP reduced model, and the error was evaluated using a highly accurate reference solution computed with $\mathrm{w} = 20$ and $M = 5 \cdot (\mathrm{w}+1)^2$; the continuous curves show the mean error over $R = 100$ repetitions, and the dashed lines show the mean error plus one sample standard deviation. The estimated standard deviation of the sample mean in a simple Monte Carlo approximation of $\mathbb{E}[\mathrm{QOI}_1]$ is also included.

4.3.2 Example 2: the Darcy model in a domain with an inclusion of random shape

The second example we consider is based on problem (4.3.8), but now we choose a deterministic value of the diffusion coefficient μ as

$$\mu(\mathbf{x}, Y) = \begin{cases} 1, & \Omega_I, \\ 10^{-6}, & \Omega \setminus \Omega_I, \end{cases} \tag{4.3.13}$$

and the randomness is in the radius of the circular inclusion Ω_I, which is centered in $(0.4, 0.4)$ and has a radius R which is determined by the uniform random variable $Y \sim \mathcal{U}(-1, 1)$ as $R = (Y+2)/10$. For each realization of the random variable we remesh the whole domain. The motivation for this example is to investigate the effect of a discontinuity of the diffusion coefficient with random location in the physical domain. The discontinuity consists in the boundary of the inclusion, where the diffusion coefficient jumps by six orders of magnitude. The QOIs considered are the same as those in Example 1.

Fig. 4.5 shows the condition number (4.3.5) and the approximation error (4.3.2) in the QOIs in this example. In Fig. 4.5 we observe an exponential convergence w.r.t. w up to $\mathrm{w} = 6$ for all three QOIs, and this example shows that these are smooth QOIs even if the problem presents discontinuities with a random location (hence the solution itself, measured in the $L^2(\Omega)$ or $H^1(\Omega)$ norm, is not smooth with respect to $\mathbf{Y}$). For larger values than $\mathrm{w} = 6$ the error levels out due to nonnegligible contributions of the finite element error.

Notice again that, as the value of the proportionality constant c increases, the curves converge to the optimal curve corresponding to the exact L^2 projection error curve. In this case, already $M = 3 \cdot (\#\Lambda)^2$ yields the optimal convergence rate.

Example 2: a less smooth QOI We now build a QOI that exhibits a less regular dependence on the random variable $\mathbf{Y}$. We consider again problem (4.3.8) with the value of the diffusion

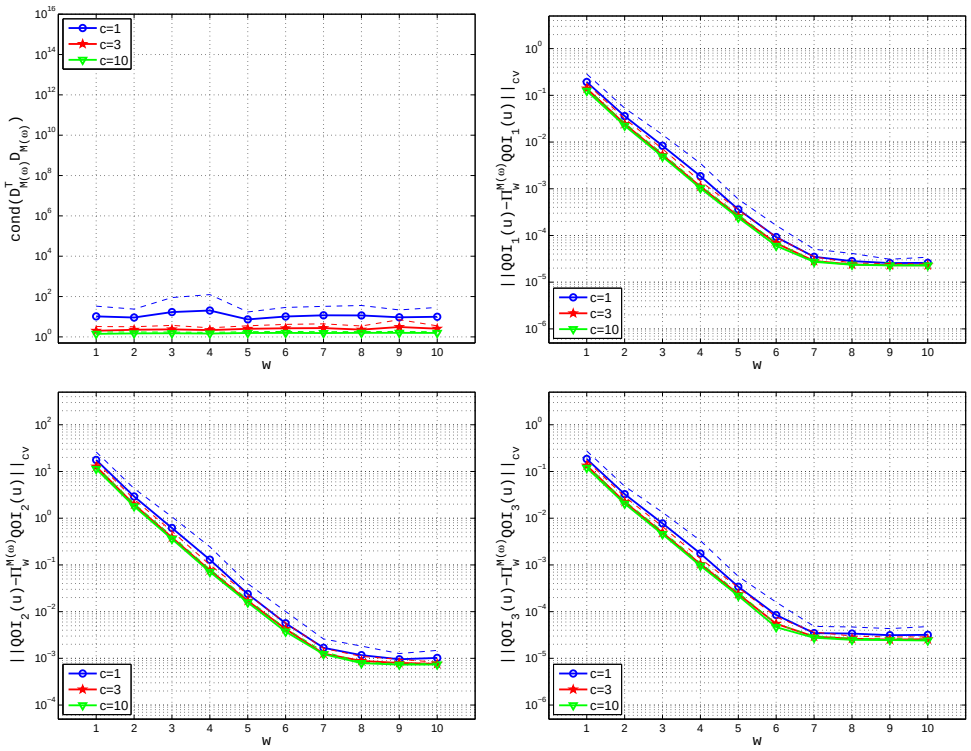


Figure 4.5: Example 2. Condition number (4.3.5) (top-left) and approximation errors (4.3.2) in the QOI_i, $i = 1, 2, 3$, using $M = c \cdot (\#\Lambda)^2 = c \cdot (\mathrm{w}+1)^2$. Continuous lines: sample mean averaged over $R = 100$ repetitions. Dashed lines: sample mean plus one sample standard deviation averaged over $R = 100$ repetitions.

coefficient μ given by

$$\mu(\mathbf{x}, Y) = \begin{cases} 1, & \Omega_I, \\ 10^{-2}, & \Omega \setminus \Omega_I, \end{cases} \tag{4.3.14}$$

and the circular inclusion Ω_I with random radius as in the previous section. Note that the discontinuity of the coefficient is a jump of two orders of magnitude across the boundary of the inclusion. Now we consider the following quantities of interest:

$$\mathrm{QOI}_4(Y) = u(\widetilde{\mathbf{x}}, Y)\Big|_{\widetilde{\mathbf{x}}=(0.4,0.4)}, \qquad \mathrm{QOI}_5(Y) = u(\widetilde{\mathbf{x}}, Y)\Big|_{\widetilde{\mathbf{x}}=(0.55,0.55)}, \tag{4.3.15}$$

which are pointwise evaluations of the solution u of problem (4.3.8) in two fixed positions $\widetilde{\mathbf{x}} = (0.4, 0.4)$ and $\widetilde{\mathbf{x}} = (0.55, 0.55)$. The former coincides with the center of the random shape inclusion and therefore always lies inside it. The latter point may or may not belong to the inclusion, depending on the outcome of the random variable Y that determines the radius of the inclusion. The corresponding results are displayed in Fig. 4.6. The QOI associated with the point $\widetilde{\mathbf{x}} = (0.4, 0.4)$ exhibits a faster convergence than the one associated with $\widetilde{\mathbf{x}} = (0.55, 0.55)$, since the discontinuity in the coefficient affects the regularity of the solution exactly in the point where it is evaluated. In this case, the point $\widetilde{\mathbf{x}} = (0.55, 0.55)$ is such that the probability of falling inside the inclusion is approximately twice the probability of falling outside it. Of course there are also QOIs that are hard to approximate, *e.g.* the one associated with the point $\widetilde{\mathbf{x}} = (0.6, 0.6)$ that falls inside the inclusion with a probability larger than 98%. In this case the use of *importance sampling* techniques (see *e.g.* [RC04]) should be considered.

4.3.3 Example 3: the Darcy flow in a domain with five inclusions

In the next test we again use problem (4.3.8) and increase the dimension of the parameter space Γ to $N = 5$ by adding some inclusions, as shown in Fig. 4.7. The inclusions are circular with

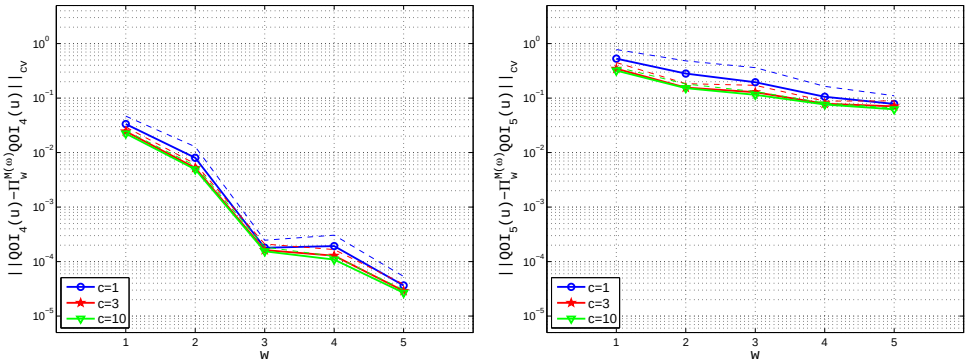


Figure 4.6: Example 2. Approximation errors (4.3.2) in the QOI_i, $i = 4, 5$, using $M = c \cdot (\#\Lambda)^2 = c \cdot (\mathrm{w}+1)^2$. Continuous lines: sample mean averaged over $R = 100$ repetitions. Dashed lines: sample mean plus one sample standard deviation averaged over $R = 100$ repetitions.

radius equal to 0.1 and are centered in the points $\mathbf{x} = (0.5, 0.5)$ and $\mathbf{x} = (0.5 \pm 0.25, 0.5 \pm 0.25)$. We denote by Ω_i, $i = 1, \ldots, 5$, the inclusion domains and $\Omega_0 = \Omega \setminus \cup_{i=1}^5 \Omega_i$ the bulk. Therefore $\Omega = \left(\bigcup_{i=0}^5 \Omega_i\right)$, and the sets Ω_i are not overlapping each other. The random diffusion coefficient now depends on a multivariate uniform random variable $\mathbf{Y} \sim \mathcal{U}([-1,1]^5)$ and is defined as

$$\mu(\mathbf{x}, \mathbf{Y}) = \begin{cases} \exp(\beta\, Y_i), & \Omega_i, \quad i = 1, \ldots, 5, \\ 10^{-4}, & \Omega_0, \end{cases} \tag{4.3.16}$$

such that each random variable is associated with an inclusion. We consider the same QOIs as in Example 1. The chosen polynomial space $\mathbb{P}_\Lambda$ is the isotropic TD space. As mentioned in the introduction, this choice is motivated by the analysis in [BNTT12]. We set $\beta = 2$, so that the coefficient variations in the inclusions are of two orders of magnitude, and report the results in Fig. 4.8. The convergence rate is exponential whenever the value of c is larger than 1. A number of points $M = 3 \cdot \#\Lambda$ is enough to achieve the optimal convergence rate, and no deterioration is observed up to the maximal polynomial degree $\mathrm{w} = 10$ considered. Then we set $\beta = 5$ and obtain the results in Fig. 4.9. Note that this case yields a variation of more than four orders of magnitude in the coefficient inside the inclusions. As a consequence we observe that the convergence remains exponential but with a slower rate.

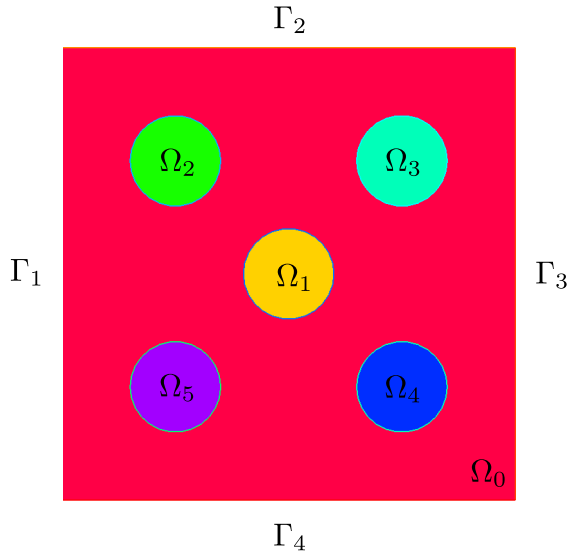


Figure 4.7: Domain with five inclusions $\Omega_1, \Omega_2, \Omega_3, \Omega_4, \Omega_5$ with random diffusivity for problem (4.3.8).

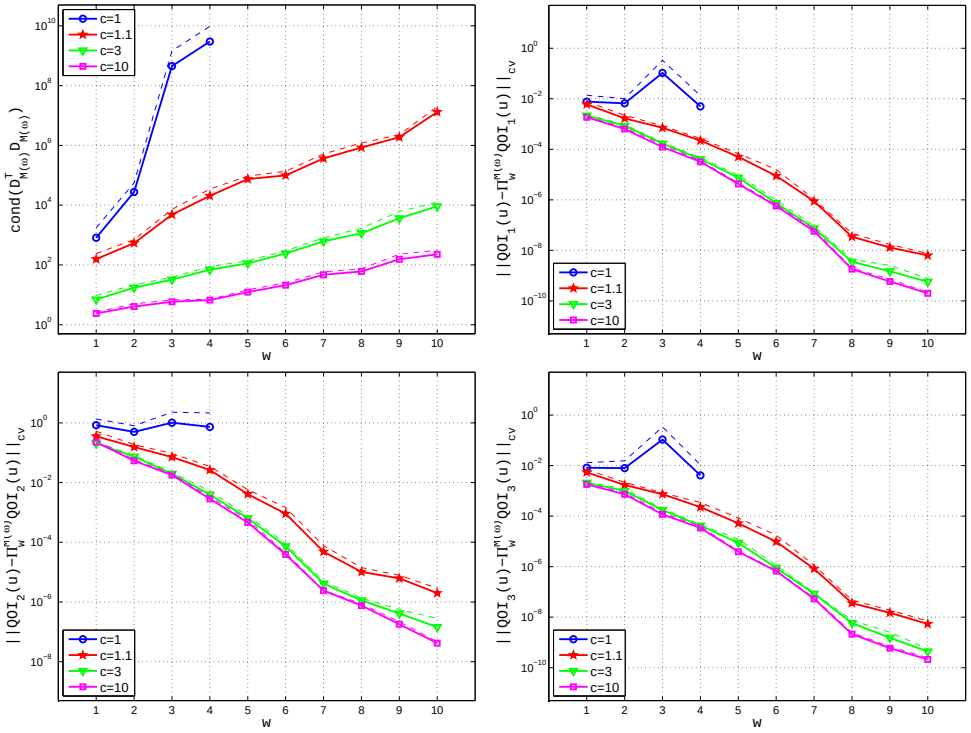


Figure 4.8: Example 3: the Darcy model. $\beta = 2$. Condition number (4.3.5) (top-left) and approximation errors (4.3.2) in the QOI_i, $i = 1, 2, 3$, using $M = c \cdot \#\Lambda = c \cdot (\mathrm{w} + 1)$. Continuous lines: sample mean averaged over $R = 100$ repetitions. Dashed lines: sample mean plus one sample standard deviation averaged over $R = 100$ repetitions.

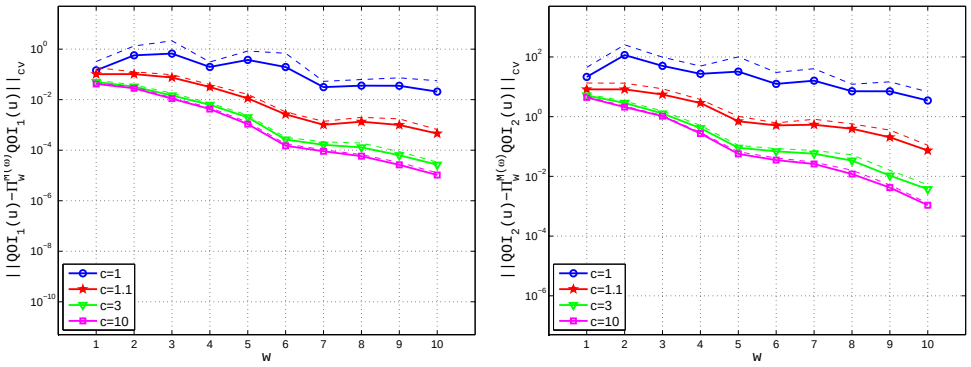


Figure 4.9: Example 3: the Darcy model. $\beta = 5$. Approximation errors (4.3.2) in the QOI_i, $i = 1, 2$, using $M = c \cdot \#\Lambda = c \cdot (\mathrm{w} + 1)$. Continuous lines: sample mean averaged over $R = 100$ repetitions. Dashed lines: sample mean plus one sample standard deviation averaged over $R = 100$ repetitions.

4.3.4 Example 4: Navier–Lamé equations with "inclusion type" coefficient

We consider the *Navier–Lamé* equations written in displacement form on the domain $\Omega = \cup_{i=1}^{7}\Omega_i$ depicted in Fig. 4.10:

$$\begin{cases} -(\lambda(\mathbf{x},\mathbf{y})+\mu(\mathbf{x},\mathbf{y}))\nabla(\nabla\cdot\mathbf{u})+\mu(\mathbf{x},\mathbf{y})\nabla^2\mathbf{u} = -f(\mathbf{x},\mathbf{y}), & \mathbf{x}\in\Omega,\ \mathbf{y}\in\Gamma, \\ \sigma(\mathbf{u})\cdot\mathbf{n} = 0, & \mathbf{x} \text{ on } \partial\Omega\setminus\Gamma_{wall},\ \mathbf{y}\in\Gamma, \\ \mathbf{u} = \mathbf{0}, & \mathbf{x} \text{ on } \Gamma_{wall},\ \mathbf{y}\in\Gamma, \end{cases} \tag{4.3.17}$$

with

$$\mu(\mathbf{x},\mathbf{Y}) = \frac{E(\mathbf{x},\mathbf{Y})}{2(1+\nu)}, \qquad \lambda(\mathbf{x},\mathbf{Y}) = \frac{\nu\, E(\mathbf{x},\mathbf{Y})}{(1+\nu)(1-2\,\nu)},$$

and with σ the usual stress tensor

$$\sigma(\mathbf{u}) = \lambda(\nabla\cdot\mathbf{u})\mathbf{I} + 2\,\mu\,\frac{\nabla\mathbf{u}+\nabla^T\mathbf{u}}{2}.$$

Young's modulus E is affected by uncertainty, and it depends on the random variable $\mathbf{Y}\sim\mathcal{U}([-1,1]^7)$ in the following way:

$$E(\mathbf{x},\mathbf{Y}) = \exp(7+Y_i),\ \text{in } \Omega_i, \qquad i=1,\dots,7.$$

Poisson's ratio ν is deterministic and equal to 0.28. The prescribed boundary conditions are null displacement on Γ_{wall} and null stress on $\partial\Omega\setminus\Gamma_{wall}$. The forcing term $f\equiv -1$ models the distributed action of the gravity force. The reference configuration of the cantilever is a 1×7 rectangle. Further details about the geometry are given in Fig. 4.10. As in Example 3, we choose $\mathbb{P}_\Lambda$ to be the isotropic TD space. We are interested in the following QOIs:

$$\mathrm{QOI}_6(\mathbf{u}) = \int_\Omega |\nabla u_1|^2 + |\nabla u_2|^2\, d\Omega,$$

$$\mathrm{QOI}_7(u_2) = \min_{\mathbf{x}\in\Omega} u_2(\mathbf{x}), \qquad \mathrm{QOI}_8(\mathbf{u}) = \int_\Omega \sigma_{12}(\mathbf{u})\, d\Omega.$$

In Fig. 4.11 we report the corresponding results. The convergence is exponential even with $M = 1.1\cdot\#\Lambda$, which is very close to the minimal number of points required to have an overdetermined problem. The red line corresponds to the choice $M = 3\cdot\#\Lambda$, and can be considered the optimal convergence rate, since no significant improvement is observed when going to $M = 10\cdot\#\Lambda$ (green line).

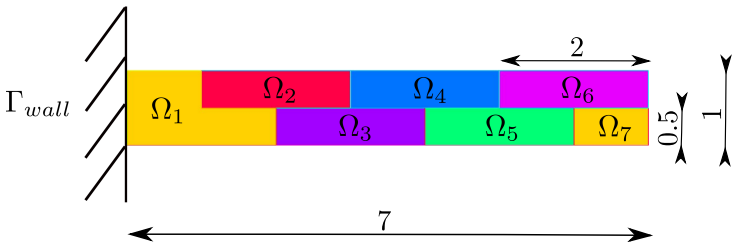


Figure 4.10: The cantilever beam.

4.3.5 Example 5: Navier–Stokes equations in a random domain

In the last example, we consider the steady state incompressible Navier–Stokes equations which govern the motion of a fluid in a pipe:

$$\begin{cases} -\nu\Delta\mathbf{u} + \big(\mathbf{u}\cdot\nabla\big)\mathbf{u} + \nabla p = 0, & \text{in } \Omega, \\ \nabla\cdot\mathbf{u} = 0, & \text{in } \Omega, \\ + \text{ boundary conditions} & \text{on } \partial\Omega. \end{cases} \tag{4.3.18}$$

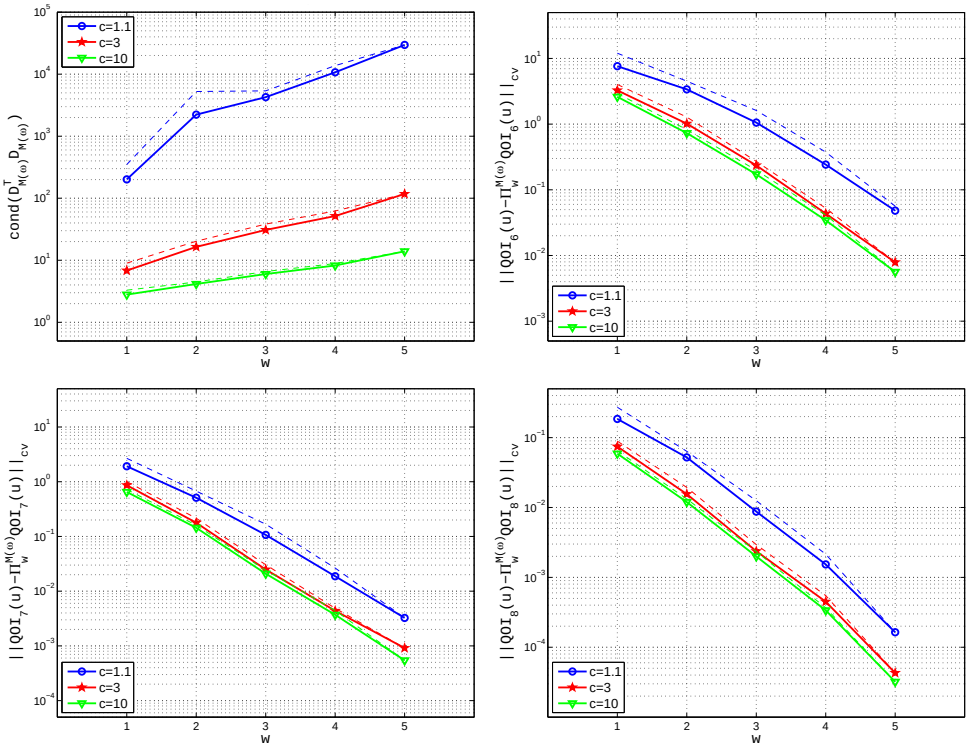


Figure 4.11: Example 4. Condition number (4.3.5) (top-left) and approximation errors (4.3.2) in the QOI_i, $i = 6, 7, 8$, using $M = c \cdot \#\Lambda = c \cdot (\mathrm{w}+1)$. Continuous lines: sample mean averaged over $R = 100$ repetitions. Dashed lines: sample mean plus one sample standard deviation averaged over $R = 100$ repetitions.

The presence of uncertainty in the model is described by a two-dimensional uniform random variable $\mathbf{Y} \sim \mathcal{U}([-1,1]^2)$. The first component Y_1 models the uncertainty in the kinematic viscosity $\nu(\mathbf{Y}) = 10^{-Y_1}$, while the second component determines the geometrical parameter r_1 by

$$r_1 = \frac{1.5 + Y_2}{5}.$$

The parameter r_1 defines the curvature in the innermost part of the elbow of the pipe. The parameter $r_2 = 0.3$ is kept fixed, since it has a minor influence on the solution of the model. The size of the inflow and outflow sections of the pipe and other geometrical details are reported in Fig. 4.12. We again choose $\mathbb{P}_\Lambda$ to be the isotropic TD space, although the two random variables

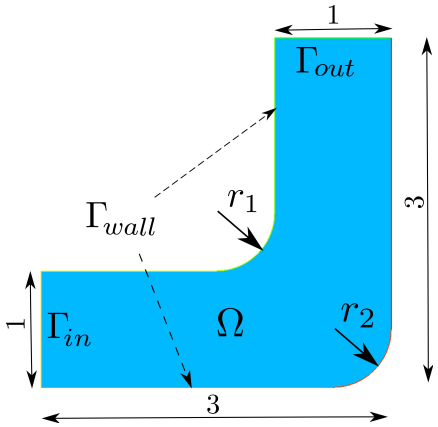


Figure 4.12: Geometry of the domain Ω in Example 5.

have clearly different roles. We impose a Poiseuille velocity profile on Γ_{in} with maximal velocity

equal to 4, no-slip conditions on Γ_{wall}, null tangential velocity,z and null pressure on Γ_{out}. The QOIs that we address are given by the pressure as

$$\mathrm{QOI}_9(\mathbf{Y}) = \frac{1}{|\Gamma_{\mathrm{in}}|}\int_{\Gamma_{\mathrm{in}}} p(\mathbf{x},\mathbf{Y})\,d\Omega, \qquad \mathrm{QOI}_{10}(\mathbf{Y}) = p(\widetilde{\mathbf{x}},\mathbf{Y})\Big|_{\widetilde{\mathbf{x}}=(2.5,1)},$$

and by the vorticity $\mathbf{v}(\mathbf{x},\mathbf{Y}) = \nabla \times \mathbf{u}(\mathbf{x},\mathbf{Y})$ as

$$\mathrm{QOI}_{11}(\mathbf{Y}) = \int_{\Omega} |\mathbf{v}(\mathbf{x},\mathbf{Y})|\,d\Omega.$$

The point $\widetilde{\mathbf{x}} = (2.5, 1)$ lies in a central region of the domain where the pressure is largely affected by the values of the random parameters. The Reynolds number ranges from 0.4 to 40, depending on the realizations of the random variable $\mathbf{Y}$. The flow of the fluid is always in the laminar regime.

We report the numerical results obtained in Fig. 4.13. The QOIs associated with the pressure converge exponentially fast. On the other hand, the QOI with the vorticity is very sensitive to the input parameters. As a consequence, the corresponding QOI exhibits a slow convergence, and the use of a high order polynomial approximation seems ineffective.

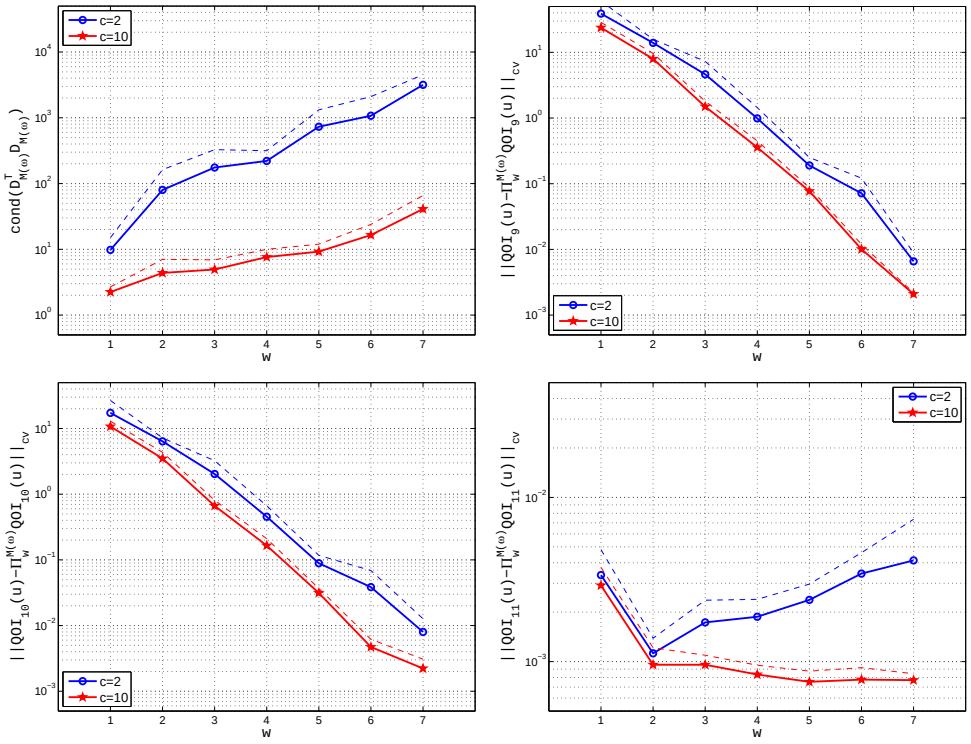


Figure 4.13: Example 5. Condition number (4.3.5) (top-left) and approximation errors (4.3.2) in the QOI_i, $i = 9, 10, 11$, using $M = c \cdot \#\Lambda = c \cdot (\mathrm{w}+1)$. Continuous lines: sample mean averaged over $R = 5$ repetitions. Dashed lines: sample mean plus one sample standard deviation averaged over $R = 5$ repetitions.

4.4 Conclusions

In this work we have presented the use of RDP to approximate QOIs related to the solution of PDEs with stochastic data. We have proposed several examples of QOIs related to the solution of the Darcy model, the linear elasticity model, and the Navier–Stokes equations, containing few random parameters, and we have shown numerically how the sample size affects the convergence rate when using a linear proportionality relation $M \propto \#\Lambda$ or a quadratic relation $M \propto (\#\Lambda)^2$.

For a one-dimensional parameter space, the theoretical analysis proposed in [MNvST11] ensures that when a quadratic proportionality $M \propto (\#\Lambda)^2$ is employed, the error committed by the random projection behaves as the best approximation error in the L^∞ norm. The first two numerical examples proposed, concerning the Darcy model with one random parameter describing either the permeability in an inclusion contained in the domain or the radius of the inclusion, confirm the theoretical results. The QOIs analyzed are all analytic functions of the random parameter, and exponential convergence has been obtained when employing a quadratic proportionality relation $M \propto (\#\Lambda)^2$. On the other hand, with a linear proportionality relation $M \propto \#\Lambda$, the error decays initially exponentially but a deterioration of the convergence is observed for high polynomial degrees (above w = 11 in Fig. 4.4).

The situation seems to be more favorable in higher dimension. Indeed, in all the proposed numerical tests with smooth (analytic) QOIs and more than one parameter, a linear proportionality relation $M \propto \#\Lambda$ yielded an almost-optimal exponential convergence rate, and no deterioration of the convergence due to an insufficient sample size has been observed in the range of polynomial degrees and tolerances tested. The design matrix $(D^\omega)^T D^\omega$ is also better conditioned than in the monovariate case. Although we can not exclude a blow-up of the error for high polynomial degrees, our conclusion is that for practical engineering tolerances, a linear proportionality $M = c \cdot \#\Lambda$ is acceptable, and the higher the dimension, the smaller the constant can be taken, which makes RDP an attractive technique for moderately high dimensional problems. However a theoretical result supporting this statement is still missing.

The overall efficiency of RDP has to be compared with classical methods as Stochastic Galerkin and Stochastic Collocation on Sparse Grids. The RDP is more suited to complex applications than Stochastic Galerkin, since the evaluations of the target function are completely uncoupled and one might use a black box deterministic solver. In addition, RDP is very promising for intermediate to large dimensions and could be competitive with or even better than Stochastic Collocation on Sparse Grids in terms of accuracy versus number of evaluations of the target function. A fair comparison between the two methods is out of the scope of the present paper and will be addressed in a forthcoming work. Meanwhile, in this paper, we have shown one example of the approximation of the expected value of a QOI, where RDP clearly outperforms a basic Monte Carlo method.

Finally, we remark that a potential great improvement in the stability and consequent efficiency of RDP could be achieved by sampling from a distribution different from the underlying one, as mentioned in Remark 4.2. This aspect will be investigated in a future work.

Part II

Application to Inverse Problems for Partial Differential Equations with stochastic data

Chapter 5

The Factorization Method for Electrical Impedance Tomography in piecewise constant uncertain backgrounds

The contents of this chapter are mainly rearranged from the manuscript [HM13], by H. Haddar and G. Migliorati, where the Factorization Method is applied to the Continuous Model in EIT with piecewise constant uncertain backgrounds. The Sections 5.2.3 and 5.5 are extracted from the technical report [HM11], where more general deterministic inhomogeneous backgrounds are addressed, also in presence of artificial noise.

Abstract

We extend the Factorization Method for Electrical Impedance Tomography to the case of background featuring uncertainty. We first describe the algorithm for known but inhomogeneous backgrounds and indicate expected accuracy from the inversion method through some numerical tests. Then we develop three methodologies to apply the Factorization Method to the more difficult case of piecewise constant but uncertain background. The first one is based on a recovery of the background through an optimization scheme and is well adapted to relatively low-dimensional random variables describing the background. The second one is based on a weighted combination of the indicator functions provided by the Factorization Method for different realizations of the random variables describing the uncertain background. We show through numerical experiments that this procedure is well suited to the case where many realizations of the measurement operators are available. The third strategy is a variant of the previous one when measurements for the inclusion-free background are available. In this case, a single pair of measurements is sufficient to achieve comparable accuracy to the deterministic case.

Key words: Inverse Problems, Electrical Impedance Tomography, Factorization Method, uncertain background.

AMS subject classification: 35R60, 35R30, 65M32

5.1 Introduction

The problem of *Electrical Impedance Tomography* (EIT) arises in many applied contexts. It leads to well-established operative procedures, *e.g.* in geophysics, nondestructive testing or imaging applications, while in other fields its use is still experimental *e.g.* medicine, [CIN99, Bor02, MS12, SVV$^+$01, GZ03].

For some key theoretical results related to uniqueness and conditional stability of the EIT inverse problem, we refer to [AP06, BT03, Ale88, LU89] and references therein. On the numerical side, many mathematical models and inversion algorithms for EIT were proposed in the literature, trying to take into account as many physical phenomena as possible, see [MS12] and the references therein for a complete overview. In this work we address the Factorization Method (FM) as introduced in [KG08] applied to the Continuous Model (in contrast to the Complete Electrode Model [LHH08, HHP07]) in the context of EIT featuring an inhomogeneous isotropic background. We shall in particular address the case where the inhomogeneous background cannot be considered as a small perturbation of the homogeneous one (as in [HS09]) and more importantly consider the case where this background is piecewise constant but with uncertain values on the conductivity (and not on the spatial distribution). The latter configuration is motivated by medical imaging experiments since the conductivity support may be known from the use of other imaging modalities, while the physical electrical parameters cannot be exactly determined.

Most of the works on the FM in the literature treat the case of the homogeneous background, *e.g.* [AHOS07, HS08]. Our work is mostly related to numerical issues associated with EIT in an inhomogeneous and uncertain background. To begin with, we propose a numerical scheme to solve the dipole-like Neumann boundary-value problem when the background coefficient is inhomogeneous and deterministic, and use this scheme to design an efficient implementation of the FM algorithm for inhomogeneous but deterministic background. We then discuss the case where the background is piecewise constant with known spatial distribution but unknown parameter values. We propose three variants of the FM to cope with this configuration. In the first algorithm, by means of an optimization scheme motivated by the structure of the sampling operator and the outcome of the FM, we simultaneously recover the background parameters and the shape of the inclusion. This algorithm is well suited to low-dimensional configurations of the parameter space, and can be applied to moderately high-dimensional configurations by means of acceleration with polynomial approximation techniques [BNT10, CDS11, MNvST11]. In addition, an alternative approach is proposed, where the optimization scheme is replaced by a weighting of the FM indicator function with a misfit indicator for the background. This procedure requires an extensive sampling of the parameter space which would be very expensive. However, in the case where many independent realizations of the measurement operator are available, we describe how this strategy can be made efficient. The third algorithm deals with the case where paired measurements are available, namely measurements for the inclusion-free background and measurements for the medium with inclusion. In this case, a uniform weight for the FM associated with random sampling of the parameter space provides an effective indicator function.

Let us recall that the FM is an imaging technique belonging to the class of *Qualitative Methods* [KG08, CC06]. In this sense the inversion output is an approximate image of a characteristic function of the inclusion domain and the method does not allow for a quantification of the inversion error in terms of the measurement or discretization errors. Only qualitative statements on the convergence of regularizing schemes as measurement error goes to zero can be obtained. The FM is also a deterministic approach to EIT: it relies on a contrast with an inclusion-free reference and no prior information is needed to apply the method, but additional information might be useful to interpret the imaging reconstruction. For statistical approaches to Inverse Problems and EIT based on Bayesian inference, where prior information is needed but the inclusion-free reference is not required, we refer to [KKSV00, Stu10] and references therein.

The outline of the paper is the following: in Section 5.2.1 we present the problem of EIT and

the notation. In Sections 5.2.2 and 5.2.3 we introduce the FM in deterministic inhomogeneous background and the numerical scheme proposed to solve the dipole-like singular problem. Then in Section 5.2.4 we describe the regularization technique employed. In Section 5.3 we introduce the problem of EIT in random media, starting from the parametrization of the background coefficient by random variables. A justification of how randomness affects the spectrum of the measurement operator is provided. Moreover, two situations are identified depending on how the measurements are collected, *i.e.* arbitrary or paired measurements. In Section 5.3.1 two variants of the FM are proposed, in the case of arbitrary measurements. In Section 5.3.2 a variant of the FM is proposed in the case of paired measurements. Sections 5.4, 5.5, 5.6, 5.7 illustrate the performance of the variants of the FM proposed through several test cases. Finally, in Section 5.8 we draw some conclusions.

5.2 The EIT problem and the Factorization Method

5.2.1 Mathematical formulation of the EIT problem

Consider a bounded Lipschitz domain $B \subset \mathbb{R}^2$ and its subset $D \subset B$. We assume that D is a union of possibly disjoint Lipschitz domains, each one with positive measure, and that $B \setminus \overline{D}$ is connected. The domain B represents the background medium, modeled by means of the background diffusion coefficient σ_B. The domain D represents an inclusion, displaced somewhere inside B (see Fig. 5.1). The inclusion is characterized by unknown shape and unknown value of its diffusion coefficient σ_D. We shall assume that σ_B and σ_D are real-valued functions. Now denote by $\sigma \in L^\infty(B)$ the diffusion coefficient

$$\sigma(\mathbf{x}) = \begin{cases} \sigma_B(\mathbf{x}), & \text{in } B \setminus D, \\ \sigma_D(\mathbf{x}), & \text{in } D, \end{cases} \tag{5.2.1}$$

and define the functional spaces $\mathring{L}^2(\partial B)$, $\mathring{H}^1(B)$, $\mathring{H}^{\frac{1}{2}}(\partial B)$ and $\mathring{H}^{-\frac{1}{2}}(\partial B)$ as the subspaces of respectively $L^2(\partial B)$, $H^1(B)$, $H^{\frac{1}{2}}(\partial B)$ and $H^{-\frac{1}{2}}(\partial B)$ of functions with zero (boundary) mean value on ∂B. Throughout the paper the differential operator is denoted by ∇, or by ∇_j and ∇_x to specify the differentiation with respect to the jth argument and to the variable $\mathbf{x}$, respectively.

Given $g \in \mathring{H}^{-\frac{1}{2}}(\partial B)$, consider the Neumann boundary-value problem,

$$\text{find } u \in \mathring{H}^1(B) \text{ s.t. } \begin{cases} \nabla \cdot (\sigma(\mathbf{x}) \nabla u) = 0, & \text{in } B, \\ \sigma(\mathbf{x}) \nabla u \cdot \boldsymbol{\nu} = g, & \text{on } \partial B, \end{cases} \tag{5.2.2}$$

where $\boldsymbol{\nu}$ denotes the outward normal to ∂B. It is well known that problem (5.2.2) has a unique solution if $\sigma \in L^\infty(B)$ and is positive definite. In the case of the continuous EIT model, one prescribes a current pattern g and measures the generated potential $f = u|_{\partial B}$ on the boundary ∂B. The final aim is to retrieve some information about D, *e.g.* its location inside B and its shape.

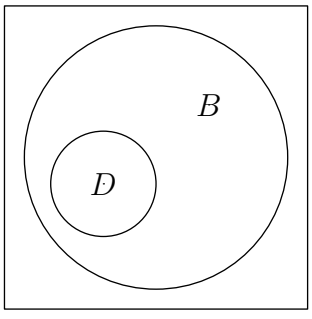


Figure 5.1: An inclusion D which lies in a medium B.

The operator Λ that maps currents g into potentials f is the Neumann-to-Dirichlet map associated with problem (5.2.2) with coefficient (5.2.1). It is a continuous operator from $\mathring{H}^{-\frac{1}{2}}(\partial B)$ to $\mathring{H}^{\frac{1}{2}}(\partial B)$. Denote also by Λ_0 the Neumann-to-Dirichlet map for the problem (5.2.2) with the coefficient

$$\sigma(\mathbf{x}) = \sigma_B(\mathbf{x}), \text{ in } B, \tag{5.2.3}$$

and by u_0 the solution to the same problem. In this way u_0 represents the potential generated by the incoming current g in the domain B, when the inclusion D is not present. Let f_0 be the corresponding measured potential $f_0 = u_0\big|_{\partial B} = \Lambda_0 g$. Both the operators Λ and Λ_0 are weakly compact, when restricted on $\mathring{L}^2(\partial B)$, *i.e.* $\Lambda : \mathring{L}^2(\partial B) \to \mathring{L}^2(\partial B)$ and $\Lambda_0 : \mathring{L}^2(\partial B) \to \mathring{L}^2(\partial B)$. Moreover, we define the operator $\widetilde{\Lambda}$

$$\widetilde{\Lambda} := \Lambda - \Lambda_0 : \mathring{L}^2(\partial B) \to \mathring{L}^2(\partial B), \tag{5.2.4}$$

that has the remarkable property of being strongly compact (eigenvalues decay exponentially fast: see for instance (5.3.8)).

We will need also the Green function $N(\cdot, \boldsymbol{\xi}) \in \mathring{L}^2(B)$, which is a solution to the following Neumann boundary-value problem with a Dirac source δ_ξ centered in $\boldsymbol{\xi} \in B$:

$$\begin{cases} \nabla_1 \cdot (\sigma_B(\mathbf{x}) \nabla_1 N(\mathbf{x}, \boldsymbol{\xi})) = -\delta_\xi(\mathbf{x}), & \mathbf{x} \in B, \\ \boldsymbol{\nu} \cdot \sigma_B(\mathbf{x}) \nabla_1 N(\mathbf{x}, \boldsymbol{\xi}) = -\frac{1}{|B|}, & \mathbf{x} \text{ on } \partial B. \end{cases} \tag{5.2.5}$$

Notice that (5.2.5) does not embed any information about the inclusion D. Denote by $\mathbf{p}$ a two-dimensional versor (*i.e.* $|\mathbf{p}| = 1$). In the sequel we will often need the test function

$$\psi(\mathbf{x}, \boldsymbol{\xi}, \mathbf{p}) := \mathbf{p} \cdot \nabla_2 N(\mathbf{x}, \boldsymbol{\xi}), \quad \mathbf{x} \in B, \tag{5.2.6}$$

as well as its restriction on ∂B

$$l_\xi^p = \psi(\mathbf{x}, \boldsymbol{\xi}, \mathbf{p})\Big|_{\mathbf{x} \text{ on } \partial B}. \tag{5.2.7}$$

Moreover, denoting by $\{\mathbf{e}_1, \mathbf{e}_2\}$ an orthonormal basis of $\mathbb{R}^2$, we define

$$l_\xi^k = l_\xi^{e_k}, \quad k = 1, 2.$$

Remark 5.1. *When the domain B is a circle with radius R there is an explicit formula for the solution $N(\mathbf{x}, \boldsymbol{\xi})$ of the Neumann problem* (5.2.5) *with $\sigma_B \equiv 1$ (see* e.g. *[Brü01]):*

$$N(\mathbf{x}, \boldsymbol{\xi}) = -\frac{1}{2\pi}\left(\log|\mathbf{x} - \boldsymbol{\xi}| + \log\left| \frac{R}{|\mathbf{x}|}\mathbf{x} - \frac{|\mathbf{x}|}{R}\boldsymbol{\xi} \right| \right) + \frac{\log R}{\pi}. \tag{5.2.8}$$

On the boundary of the unitary circle the evaluation of ψ defined in (5.2.6) *is given by*

$$\psi(\mathbf{x}, \boldsymbol{\xi}, \mathbf{p})\Big|_{\partial B} = \frac{1}{\pi} \frac{\mathbf{p} \cdot (\boldsymbol{\xi} - \mathbf{x})}{|\boldsymbol{\xi} - \mathbf{x}|^2}, \qquad \textit{for } |\mathbf{x}| = 1. \tag{5.2.9}$$

5.2.2 The Factorization Method with deterministic inhomogeneous background

In this section we outline the basis of the FM applied to EIT and some numerical issues related to the case of inhomogeneous background. The main result that motivates the FM algorithm is Theorem 5.1. It provides a necessary and sufficient criterion to localize a point inside the support of the inclusions. Remark 5.2 shows how this criterion becomes operative. This theorem can be proved following the lines of the proof of [KG08, Theorem 6.8] that treats the case of constant σ_B (see also [Brü01]). The only point that requires attention is the use of unique continuation arguments for equations of type (5.2.2), and this is why a regularity assumption on σ_B is required.

Assumption 5.1. *The coefficients σ and σ_B are bounded and positive definite functions on B. The coefficient σ_B is piecewise Lipschitz continuous and with interfaces with Lipschitz boundaries. Moreover, denoting by*

$$\sigma_B^m = \min_{x\in D} \sigma_B(\mathbf{x}) \quad \text{and} \quad \sigma_B^M = \max_{x\in D} \sigma_B(\mathbf{x}),$$

then σ_B and σ_D satisfy

$$\sigma_D(\mathbf{x}) < \sigma_B^m - \tau \quad \text{or} \quad \sigma_B^M < \sigma_D(\mathbf{x}) - \tau$$

a.e. in D, for some positive constant τ.

Theorem 5.1. *Under Assumption 5.1, the operator $\widetilde{\Lambda} : \mathring{L}^2(\partial B) \to \mathring{L}^2(\partial B)$ is self-adjoint and injective. Moreover, for any versor $\mathbf{p}$ it holds*

$$l_{\boldsymbol{\xi}}^p \in \mathit{Range}(|\widetilde{\Lambda}|^{1/2}) \Longleftrightarrow \boldsymbol{\xi} \in D.$$

Remark 5.2. *Denote by $\{\lambda_j, w_j : j \in J\}$ the spectral system of the self-adjoint and compact operator $\widetilde{\Lambda}$. Then*

$$w \in \mathit{Range}(|\widetilde{\Lambda}|^{1/2}) \quad \Longleftrightarrow \quad \sum_{j=1}^{\infty} \frac{|(w, w_j)_{L^2(\partial B)}|^2}{|\lambda_j|} < \infty. \tag{5.2.10}$$

The main difficulty in implementing the FM for the inhomogeneous background is the precise evaluation of $l_{\boldsymbol{\xi}}^p$ that can only be done in the general case by numerically solving problem (5.2.5). To overcome the singularity in the forcing term, we resort to the fundamental solution

$$\phi(\mathbf{x}, \boldsymbol{\xi}) = -\frac{1}{2\pi\sigma_B(\boldsymbol{\xi})} \log\left|\mathbf{x} - \boldsymbol{\xi}\right|, \quad \mathbf{x} \in \mathbb{R}^2, \tag{5.2.11}$$

of the problem

$$-\nabla \cdot \Big(\sigma_B(\boldsymbol{\xi}) \nabla\phi(\mathbf{x}, \boldsymbol{\xi})\Big) = \delta_{\xi}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^2. \tag{5.2.12}$$

Since the singularity at $\boldsymbol{\xi}$ in problem (5.2.5) is of the same kind as in problem (5.2.12), we can restrict problem (5.2.5) in a small neighborhood $\mathcal{O}(\boldsymbol{\xi}) \subset \mathbb{R}^2$ of $\boldsymbol{\xi}$ where

$$\sigma_B(\mathbf{x}) \approx \sigma_B(\boldsymbol{\xi}), \qquad \forall\, \mathbf{x} \in \mathcal{O}(\boldsymbol{\xi}),$$

and approximate the solution $N(\mathbf{x}, \boldsymbol{\xi})$ to problem (5.2.5) near the singularity in $\boldsymbol{\xi}$ as

$$N(\mathbf{x}, \boldsymbol{\xi}) \approx \phi(\mathbf{x}, \boldsymbol{\xi}), \qquad \forall\, \mathbf{x} \in \mathcal{O}(\boldsymbol{\xi}).$$

Then it is possible to write a nonsingular problem for the difference

$$\varphi_N(\cdot, \boldsymbol{\xi}) = N(\cdot, \boldsymbol{\xi}) - \phi(\cdot, \boldsymbol{\xi}), \tag{5.2.13}$$

plugging (5.2.13) in (5.2.5). The quantity $\psi(\cdot, \boldsymbol{\xi}, \mathbf{p})$ defined in (5.2.6) is a solution to problem (5.2.5) for any $\mathbf{p}$. Moreover, in ψ the dependence of $N(\cdot, \cdot)$ on the second argument is smooth. So we can exploit this regularity to derive a numerical scheme that given $\boldsymbol{\xi} \in B$ directly computes $\nabla_2 N(\cdot, \boldsymbol{\xi})$, that is needed to obtain $l_{\boldsymbol{\xi}}^p$ through ψ and apply Theorem 5.1. To this aim we differentiate (5.2.13) with respect to the second argument, and obtain

$$\nabla_2 \varphi_N(\cdot, \boldsymbol{\xi}) = \nabla_2 N(\cdot, \boldsymbol{\xi}) - \nabla_2 \phi(\cdot, \boldsymbol{\xi}). \tag{5.2.14}$$

The derivatives of ϕ w.r.t. $\boldsymbol{\xi}$ are

$$\frac{\partial \phi(\cdot, \boldsymbol{\xi})}{\partial \xi_k} = \frac{1}{2\pi\sigma_B(\boldsymbol{\xi})^2} \frac{\partial \sigma_B(\boldsymbol{\xi})}{\partial \xi_k} \log\Big(|\mathbf{x} - \boldsymbol{\xi}|\Big) + \frac{x_k - \xi_k}{2\pi\sigma_B(\boldsymbol{\xi})|\mathbf{x} - \boldsymbol{\xi}|^2}, \quad k = 1, 2,$$

and since ϕ is analytical when $\mathbf{x} \neq \boldsymbol{\xi}$ the mixed derivatives coincide:

$$\frac{\partial^2 \phi(\mathbf{x},\boldsymbol{\xi})}{\partial x_k \partial \xi_k} = \frac{1}{2\pi\sigma_B(\boldsymbol{\xi})^2}\frac{\partial \sigma_B(\boldsymbol{\xi})}{\partial \xi_1}\frac{x_k-\xi_k}{|\mathbf{x}-\boldsymbol{\xi}|^2} + \frac{|\mathbf{x}-\boldsymbol{\xi}|^2 - 2(x_k-\xi_k)^2}{2\pi\sigma_B(\boldsymbol{\xi})|\mathbf{x}-\boldsymbol{\xi}|^4},$$

$$\frac{\partial^2 \phi(\mathbf{x},\boldsymbol{\xi})}{\partial x_\ell \partial \xi_k} = \frac{1}{2\pi\sigma_B(\boldsymbol{\xi})^2}\frac{\partial \sigma_B(\boldsymbol{\xi})}{\partial \xi_1}\frac{x_\ell-\xi_\ell}{|\mathbf{x}-\boldsymbol{\xi}|^2} - \frac{(x_k-\xi_k)\,(x_\ell-\xi_\ell)}{\pi\sigma_B(\boldsymbol{\xi})|\mathbf{x}-\boldsymbol{\xi}|^4},$$

for $k,\ell = 1,2$ and $\ell \neq k$. Denoting by superscript the partial derivative, we have two problems for the unknowns $\varphi_N^k = \partial\varphi_N/\partial\xi_k$, $k = 1,2$:

$$\begin{cases} -\nabla_x \cdot \Big(\sigma_B(\mathbf{x})\nabla_1 \varphi_N^k(\mathbf{x},\boldsymbol{\xi})\Big) = -\nabla_x \cdot \Big(\Big(\sigma_B(\boldsymbol{\xi}) - \sigma_B(\mathbf{x})\Big)\boldsymbol{\Phi}^k(\mathbf{x},\boldsymbol{\xi})\Big), & \mathbf{x} \text{ in } B, \\ \sigma_B(\mathbf{x})\,\nabla_1\varphi_N(\mathbf{x},\boldsymbol{\xi})\cdot\boldsymbol{\nu} = -\sigma_B(\mathbf{x})\,\boldsymbol{\Phi}^k(\mathbf{x},\boldsymbol{\xi})\cdot\boldsymbol{\nu} - \frac{1}{|\partial B|}, & \mathbf{x} \text{ on } \partial B, \end{cases} \tag{5.2.15}$$

with

$$\boldsymbol{\Phi}^k(\mathbf{x},\boldsymbol{\xi}) = \left(\frac{\partial^2\phi(\mathbf{x},\boldsymbol{\xi})}{\partial x_1\partial\xi_k}, \frac{\partial^2\phi(\mathbf{x},\boldsymbol{\xi})}{\partial x_2\partial\xi_k}\right), \qquad \forall\,(\mathbf{x},\boldsymbol{\xi}) \in B\times B \,:\, \mathbf{x}\neq\boldsymbol{\xi}.$$

All the second derivatives of ϕ are singular in $\mathbf{x} = \boldsymbol{\xi}$, and behave like $1/|\mathbf{x}-\boldsymbol{\xi}|^2$. Therefore if $\sigma_B(\cdot)$ is piecewise Lipschitz then the solution of (5.2.15) is in $H^1(B)$. Hence the variational formulation of (5.2.15) is suitable for standard finite element discretizations. Once the solution of (5.2.15) has been computed, then $\nabla_2 N(\cdot,\boldsymbol{\xi})$ can be recovered from (5.2.14). Notice that when σ_B is piecewise constant and $\boldsymbol{\xi}$ does not fall where it jumps, the function $\zeta(\mathbf{x},\boldsymbol{\xi}) := \sigma_B(\boldsymbol{\xi}) - \sigma_B(\mathbf{x})$ vanishes as $\mathbf{x}$ approaches $\boldsymbol{\xi}$.

5.2.3 A closer look to the theory behind the Factorization Method

Theorem 5.1 merges many results in one, and introduces directly the *Range* criterion. To illustrate the theory behind the Factorization Method, we provide a more articulate presentation as well. The key result is the factorization of the operator $\widetilde{\Lambda}$ given in Theorem 5.2, (see [Sch10, Chap.12,Thm.4] for a proof). In [KG08] a more general three-term factorization is derived for $\widetilde{\Lambda} : \mathring{H}^{-1/2}(\partial B) \to \mathring{H}^{1/2}(\partial B)$. Also another simpler two-term factorization can be chosen, see *e.g.* [Sch10]. Although many factorizations are possible, each one of them provides a necessary and sufficient criterion to determine the inclusions. Lemma 5.1 and Theorem 5.5 show how this criterion becomes operative. See [Sch10, Chap.12] for the proof of Theorems 5.2, 5.4 and 5.5.

Let us introduce the operator $\widetilde{G} : \mathring{H}^{-1/2}(\partial D) \to \mathring{L}^2(\partial B)$. It acts as a Neumann-to-Dirichlet map of a suitable virtual problem,

$$\begin{cases} \nabla\cdot(\sigma_B(\mathbf{x})\nabla w) = 0, & \text{in } B\setminus\overline{D}, \\ \frac{\partial w}{\partial\nu} = 0, & \text{on } \partial B, \\ -\frac{\partial w}{\partial\nu} = \varphi, & \text{on } \partial D, \end{cases} \tag{5.2.16}$$

where $\varphi \in \mathring{H}^{-1/2}(\partial D)$ is the input, and $w|_{\partial B} \in \mathring{L}^2(\partial B)$ is the output. Moreover, define the space $W = \{w \in \mathring{H}^1(B\setminus\overline{D}) \,:\, w \text{ solves (5.2.16)}\}$, and denote by $\widetilde{G}^* : \mathring{L}^2(\partial B) \to \mathring{H}^{1/2}(\partial D)$ the adjoint of $\widetilde{G}$. To define the operator T we introduce the problems

$$\begin{cases} \nabla\cdot(\sigma(\mathbf{x})\nabla\varpi) = 0, & \text{in } B\setminus\partial D, \\ \frac{\partial\varpi}{\partial\nu} = 0, & \text{on } \partial B, \\ [\varpi]_{\partial D} = \varphi, & \text{on } \partial D, \\ [\boldsymbol{\nu}\cdot\sigma(\mathbf{x})\nabla\varpi]_{\partial D} = 0, & \text{on } \partial D, \\ \int_{\partial B}\varpi\,ds = 0, \end{cases} \qquad \begin{cases} \nabla\cdot(\sigma_B(\mathbf{x})\nabla\varpi_0) = 0, & \text{in } B\setminus\partial D, \\ \frac{\partial\varpi_0}{\partial\nu} = 0, & \text{on } \partial B, \\ [\varpi_0]_{\partial D} = \varphi, & \text{on } \partial D, \\ [\boldsymbol{\nu}\cdot\sigma_B(\mathbf{x})\nabla\varpi_0]_{\partial D} = 0, & \text{on } \partial D, \\ \int_{\partial B}\varpi_0\,ds = 0. \end{cases} \tag{5.2.17}$$

The operator T is defined as

$$T : \mathring{H}^{1/2}(\partial D) \to \mathring{H}^{-1/2}(\partial D) : \varphi \mapsto \frac{\partial(\varpi^+ - \varpi_0^+)}{\partial \boldsymbol{\nu}}\Big|_{\partial D},$$

where ϖ^+ and ϖ_0^+ are the restrictions of ϖ and ϖ_0 on $B \setminus \overline{D}$.

Theorem 5.2. *The operator $\widetilde{\Lambda} : \mathring{L}^2(\partial B) \to \mathring{L}^2(\partial B)$ introduced in* (5.2.4) *can be factorized as*

$$\widetilde{\Lambda} = \widetilde{G} T \widetilde{G}^*. \tag{5.2.18}$$

Theorem 5.3. *A point $\boldsymbol{\xi} \in B$ belongs to D if and only if l^p_ξ coincides with the trace of some potentials $w \in W$.*

Proof. Let $\boldsymbol{\xi} \in D$. From the definition (5.2.7) we know that l^p_ξ is the trace of $\psi(\cdot, \boldsymbol{\xi}, \mathbf{p}) \in \mathring{H}^1(B \setminus \overline{D})$. For every $\boldsymbol{\xi}$ the function $N(\cdot, \boldsymbol{\xi})$ has the same Neumann data on the boundary. Thus its normal derivative w.r.t. $\boldsymbol{\xi}$ is null on ∂B. Moreover, on each connected component of D, the *Green's formula* gives

$$\int_{\partial D_j} \nabla_2 \psi(\mathbf{x}, \boldsymbol{\xi}, \mathbf{p}) \cdot \boldsymbol{\nu}\, ds = 0, \tag{5.2.19}$$

for every component $D_j \subset D$ such that $\boldsymbol{\xi} \notin D_j$. But also the flux of $\psi(\cdot, \boldsymbol{\xi}, \mathbf{p})$ across $\partial(B \setminus \overline{D})$ vanishes, so (5.2.19) holds also for the component D_j of D such that $\boldsymbol{\xi} \in D_j$. Therefore $\psi(\cdot, \boldsymbol{\xi}, \mathbf{p}) \in W$.

In the case $\boldsymbol{\xi} \notin \overline{D}$, if there exists $w \in W$ such that l^p_ξ is the trace of w then we obtain a contraddiction. In fact, w and $\psi(\cdot, \boldsymbol{\xi}, \mathbf{p})$ (related to l^p_ξ by (5.2.7)) satisfy the same Neumann problem in $B \setminus (\overline{D} \cup \{\boldsymbol{\xi}\})$ with the same boundary conditions. The uniqueness of the solution implies that w and $\psi(\cdot, \boldsymbol{\xi}, \mathbf{p})$ coincide in $B \setminus (\overline{D} \cup \{\boldsymbol{\xi}\})$. But now w extends harmonically into a neighborhood $\mathcal{O}(\boldsymbol{\xi})$ of $\boldsymbol{\xi}$, so that w is bounded in $\mathcal{O}(\boldsymbol{\xi})$ while $\psi(\cdot, \boldsymbol{\xi}, \mathbf{p})$ is not.

If $\boldsymbol{\xi} \in \partial D$ with the same argument w and $\psi(\cdot, \boldsymbol{\xi}, \mathbf{p})$ coincide in $B \setminus (\overline{D})$. But this is a contraddiction because $w \in \mathring{H}^1(\mathcal{O}(\boldsymbol{\xi}) \cap (B \setminus (\overline{D}))$, but $\psi(\cdot, \boldsymbol{\xi}, \mathbf{p})$ has the singularity in $\boldsymbol{\xi}$. □

Lemma 5.1. *A point $\boldsymbol{\xi} \in B$ belongs to D if and only if $l^p_\xi \in$ Range$(\widetilde{G})$.*

Proof. For any $\varphi \in \mathring{H}^{-1/2}(\partial D)$ the problem (5.2.16) has a unique solution $w \in W$. Vice-versa, every function $w \in W$ is a solution to problem (5.2.16), with a well-defined normal derivative $\varphi \in \mathring{H}^{-1/2}(\partial D)$. The thesis follows as a consequence of Theorem 5.3. □

Remark 5.3. *The space $\mathring{H}^{1/2}(\partial D)$ is a Banach reflexive space, with dual containing $\mathring{H}^{-1/2}(\partial D)$.*

Theorem 5.4. *The operator $T : \mathring{H}^{1/2}(\partial D) \to \mathring{H}^{-1/2}(\partial D)$ is self-adjoint and coercive over $\mathring{H}^{1/2}(\partial D)$.*

Theorem 5.5. *Let $\widetilde{\Lambda} : \mathring{L}^2(\partial B) \to \mathring{L}^2(\partial B)$ and $\widetilde{G} : \mathring{H}^{-1/2}(\partial D) \to \mathring{L}^2(\partial B)$ and $T : \mathring{H}^{1/2}(\partial D) \to \mathring{H}^{-1/2}(\partial D)$. In addition, let $\widetilde{G}$ be injective with dense range and let T be self-adjoint and coercive on Range$(\widetilde{G}^*)$. Then Range$(\widetilde{G})$ = Range$(\widetilde{\Lambda}^{1/2})$ and*

$$y \in \text{Range}(\widetilde{G}) \quad \Longleftrightarrow \quad \sum_{j=1}^{\infty} \frac{|(y, y_j)_Y|^2}{\lambda_j} < \infty, \tag{5.2.20}$$

where $\{\lambda_j, y_j \ : \ j \in J\}$ denotes a spectral system of the self-adjoint and compact operator $\widetilde{\Lambda} = \widetilde{G} T \widetilde{G}^*$.

5.2.4 Description of the algorithm for the deterministic setting

Now we present our operative criterion to implement the range test. It is based on the spectral decomposition of the operator $|\widetilde{\Lambda}|^{1/2}$ and on the use of an appropriate regularization. We focus mainly on *Tikhonov Regularization* (TR), although also a straightforward application of the *Picard Criterion* (5.2.27) can give good results when the background is homogeneous (see Remark 5.4 below). To simplify the notation we set $G = |\widetilde{\Lambda}|^{1/2}$. To check whether a given l_ξ^k belongs to Range(G) we have to solve the problem

$$Gg_\xi^k = l_\xi^k, \tag{5.2.21}$$

which demands for regularization. A consequence of Theorem 5.1, which holds in the continuous setting, is that

$$\boldsymbol{\xi} \in D \quad \Longleftrightarrow \quad \Big(\|g_\xi^k\|_{L^2(\partial B)}\Big)^{-1} = 0.$$

The TR of (5.2.21) reads

$$\Big(\alpha + G^*G\Big)g_\xi^k = G^* l_\xi^k. \tag{5.2.22}$$

We choose the *Fourier basis* to discretize the operator G. Since we discretize l_ξ^k and g_ξ^k in (5.2.21) over the same orthogonal basis, then the discretization of G is a square matrix.

In the following we denote by σ_i the singular values of G and by u_i and v_i the corresponding left and right singular vectors. Both the sets $\{u_i\}_{i=1}^\infty$ and $\{v_i\}_{i=1}^\infty$ form an orthonormal basis of $L^2(\partial B)$. The regularized solution of problem (5.2.22) is

$$g_\xi^k = \sum_{i=1}^{\infty} \frac{\sigma_i}{\alpha + \sigma_i^2}(l_\xi^k, u_i)v_i. \tag{5.2.23}$$

We choose the regularization parameter α using the *Morozov principle*, *i.e.* imposing

$$\|Gg_\xi^k - l_\xi^k\|_{L^2(\partial B)} = \eta, \tag{5.2.24}$$

with the parameter $\eta > 0$ to be in some way related to the accuracy of the measurements. We pick

$$\eta = \gamma\sigma_1, \tag{5.2.25}$$

with σ_1 being the largest singular value of G, and γ a given threshold. The term l_ξ^k depends on $\boldsymbol{\xi}$, so to ensure that (5.2.24) is uniformly satisfied we normalize l_ξ^k as $\hat{l}_\xi^k = l_\xi^k / \|l_\xi^k\|_{L^2(\partial B)}$. The optimal value of α is computed plugging (5.2.23) into the nonlinear equation (5.2.24), and solving with respect to α. To show the inversion results, we display the isolines of the indicator function

$$C(\boldsymbol{\xi}) = \left(\log\Big(v_1(\boldsymbol{\xi}) + v_2(\boldsymbol{\xi})\Big)\right)^{-1}, \tag{5.2.26}$$

where $v_k(\boldsymbol{\xi}) := \|g_\xi^k\|_{L^2(\partial B)}$. We can summarize the algorithm by the following:

Remark 5.4 (Picard criterion)**.** *Another criterion to test the range condition is analyzed in [GH07]. It employs the indicator function:*

$$I_k(\boldsymbol{\xi}) = \frac{1}{\|l_\xi^k\|^2_{L^2(\partial B)}} \sum_{i=1}^{\infty} \frac{\Big|(l_\xi^k, u_i)_{L^2(\partial B)}\Big|^2}{|\sigma_i|^2}, \tag{5.2.27}$$

and it reads $\boldsymbol{\xi} \in D \quad \Longleftrightarrow \quad I_k(\boldsymbol{\xi}) < \infty$. Notice that the normalization of l_ξ^k is already embedded in (5.2.27)*. We refer to this criterion as the* Picard Criterion *(PC). It does not require to*

Algorithm 1 The FM in inhomogeneous deterministic background

Sample the region of B to be probed with a set of points $\mathcal{P} = \{\boldsymbol{\xi}_j\}_{j=1}^{P}$
for $\boldsymbol{\xi}$ in the set $\mathcal{P}$ **do**
  solve problem (5.2.5) to find its solution $N(\cdot, \boldsymbol{\xi})$,
  compute l_ξ^k from $N(\cdot, \boldsymbol{\xi})$ using (5.2.7),
  use $l_{\boldsymbol{\xi}}^k$ to compute the indicator $C(\boldsymbol{\xi})$ in (5.2.26),
  plot $\boldsymbol{\xi} \mapsto C(\boldsymbol{\xi})$.
end for

solve problem (5.2.21) *to find its regularized solution. However, as discussed in [HM11], the regularization technique TR is more accurate than PC when the background is not homogeneous, above all in presence of many inclusions. Hence, in this paper we present only the results obtained using the TR technique. See also [HM11] for a comparison between TR and PC for several numerical examples.*

We consider backgrounds that are piecewise constant in r regions $\{R_i\}_{i=1}^{r}$ that are a partition of the computational domain $B = \bigcup_{i=1}^{r} R_i$, see *e.g.* Fig. 5.3. Denote by $\mathbf{m}$ the r-dimensional vector of positive numbers that specifies the coefficient value in each region R_i of the domain B. The background coefficient σ_B is defined as

$$\sigma_B(\mathbf{x}) = \sum_{i=1}^{r} m_i \, \mathbb{I}_{R_i}(\mathbf{x}), \qquad \mathbf{x} \in B, \tag{5.2.28}$$

where $\mathbb{I}_X(\mathbf{x})$ denotes the indicator function of the subdomain $X \subseteq B$. If $\boldsymbol{m}$ is the vector $\boldsymbol{m} = [1, 1, \ldots, 1]^\top$ then we obtain a homogeneous deterministic coefficient. Since σ_B is a diffusion coefficient, it has to be positive: therefore we assume that $m_i > 0$, $\forall\, i$. More general deterministic inhomogeneous coefficients σ_B, nonlinearly depending on the spatial coordinates $\mathbf{x}$, have been analyzed in [HM11]. The effect of artificial noise has been analyzed in [HM11] as well.

5.3 The problem of EIT in a random background

A natural issue that arises in the mathematical modeling of physical phenomena concerns the uncertainty in the input data. There are mainly two interpretations on how the presence of uncertainty in the model could be characterized: the epistemic and the aleatory interpretations. We introduce them directly in the context of EIT and of the applications that we have in mind.

In the EIT framework, the most natural choice to incorporate in the model the uncertainty affecting the background media is to consider the background coefficient σ_B as a random variable. In the epistemic uncertainty, we can imagine that the coefficient that models the background material is deterministic but unknown, because of the lack of knowledge or because of inaccurate measurement instruments. Therefore we can model it as a random variable, although there are also other ways to cope with epistemic uncertainty. In the aleatory uncertainty, the coefficient is really random, according to a prescribed probability density. We could think of its realizations coming from a stationary or from a nonstationary stochastic process.

In general, the treatment of uncertainty through random variables distributed according to given probability densities is known as the probabilistic approach. For any positive integer q, we introduce the q-dimensional parameter space $\Gamma \subseteq \mathbb{R}^q$, the probability density $\rho : \Gamma \to \mathbb{R}$ and the Hilbert space $L^2(\Gamma, d\rho)$ equipped with the inner product $\langle u, v \rangle = \sqrt{\int_\Gamma u(\mathbf{y})\, v(\mathbf{y})\, d\rho(\mathbf{y})}$. Denote by $\mathbf{Y} \sim \rho$ a multidimensional random variable, and by $\mathbf{y} \in \Gamma$ any of its realizations. We denote by $L^2(\Gamma, L^\infty(B))$ the Bochner space of functions that take values into Γ and are $L^\infty(B)$-valued. Then we define the random diffusion coefficient $\sigma_B \in L^2(\Gamma, L^\infty(B))$ as

$$\sigma_B(\cdot, \mathbf{y}) \in L^\infty(B), \quad \forall\, \mathbf{y} \in \Gamma. \tag{5.3.1}$$

Moreover, we assume a smooth dependence of σ_B on $\mathbf{y}$ and that σ_B satisfies the following Uniform Ellipticity Assumption.

Assumption 5.2. $\forall\, \mathbf{y} \in \Gamma$ *the map* $\mathbf{y} \mapsto \sigma_B(\cdot, \mathbf{y})$ *is assumed to be smooth (analytical).*

Assumption 5.3 (Uniform Ellipticity Assumption)**.** *There exist* $c_1, c_2 \in \mathbb{R}^+$ *such that*

$$0 < c_1 \leq \sigma_B(\mathbf{x}, \mathbf{y}) \leq c_2 < +\infty, \quad \forall\, \mathbf{x} \in B, \quad \forall\, \mathbf{y} \in \Gamma.$$

We will consider two parametrizations of the random coefficient σ_B introduced in (5.3.1). Again, the spatial background is piecewise constant in r regions $\{R_i\}_{i=1}^r$ that are a partition of the computational domain $B = \bigcup_{i=1}^r R_i$. The randomness concerns the value of the coefficient in each region R_i. We associate each component of the q-dimensional random variable $\mathbf{Y}$ with a region, hence $q = r$.

The first parametrization is suitable for uncertainty analysis, and is an extension of (5.2.28) to treat the case of random background:

$$\sigma_B(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{r} \Big(m_i + d_i\ y_i\Big)\, \mathbb{I}_{R_i}(\mathbf{x}), \quad \mathbf{x} \in B, \quad \mathbf{y} \in \Gamma = [-1, 1]^r. \tag{5.3.2}$$

The vectors $\boldsymbol{m}$ and $\boldsymbol{d}$ quantify the mean and dispersion of the coefficient in each region, and $m_1 = 1$, $d_1 = 0$. The range of variation in each region is $[m_i - d_i, m_i + d_i]$, and we have to impose that $m_i - d_i > 0$ for all $i = 1, \ldots, r$, to satisfy Assumption 5.3. The second parametrization is suitable for large variations of the background coefficient σ_B, and we parametrize it with the exponential model

$$\sigma_B(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{r} 10^{y_i}\, \mathbb{I}_{R_i}(\mathbf{x}), \quad \mathbf{x} \in B, \quad \mathbf{y} \in \Gamma = [-1, 1]^r, \tag{5.3.3}$$

so that the coefficient σ_B jumps by up to two orders of magnitude in each one of the r regions. Assumption 5.3 imposes some limitations on the type of density ρ that can be associated with the parameter $\mathbf{y}$ in the parametrizations (5.3.2), (5.3.3) and (5.6.1). In this paper we focus only on the choice $\rho = \mathcal{U}(\Gamma)$, and consequently the parameter $\mathbf{y}$ is associated with a uniform density over the parameter space Γ.

Because of the randomness in σ_B, the diffusion coefficient (5.2.1) of problem (5.2.2) becomes random as well. Under Assumption 5.3 the solution $u = u(\mathbf{x}, \mathbf{y})$ to problem (5.2.2) belongs to the space $\mathring{H}^1(B) \otimes L^2(\Gamma, d\rho)$ and depends analytically on $\mathbf{y}$ (see [BNT10, CDS11] for the proof). In the same way the Neumann-to-Dirichlet maps $\Lambda : \mathring{L}^2(\partial B) \otimes L^2(\Gamma, d\rho) \to \mathring{L}^2(\partial B) \otimes L^2(\Gamma, d\rho)$ and $\Lambda_0 : \mathring{L}^2(\partial B) \otimes L^2(\Gamma, d\rho) \to \mathring{L}^2(\partial B) \otimes L^2(\Gamma, d\rho)$ become random themselves. Also the Green function $N : (\mathbf{x}, \boldsymbol{\xi}, \mathbf{y}) \in B \times B \times \Gamma \mapsto N(\mathbf{x}, \boldsymbol{\xi}, \mathbf{y}) : B \times B \times \Gamma \to \mathring{L}^2(B) \otimes L^2(\Gamma, d\rho)$ depends on the random variable $\mathbf{Y}$ because it solves problem (5.2.5) with the random coefficient (5.3.1). For the same reason, the quantities (5.2.7) depend now on any realization $\mathbf{y} \in \Gamma$ of $\mathbf{Y}$,

$$l_\xi^p = l_\xi^p(\mathbf{y}), \qquad \forall\, \mathbf{y} \in \Gamma. \tag{5.3.4}$$

We define a measurement of the random operator Λ (or Λ_0) to be an observation of Λ (or Λ_0) for a single realization of the random variable $\mathbf{Y}$. In practice the measurements of Λ come from experiments, while the measurements of Λ_0 could be simulated numerically. Denote by $M \geq 1$ the number of measurements of Λ. We will always assume that it is not possible to recover the value of the realization $\mathbf{y}$ that generated the measurement of Λ, because this would reduce the inverse problem in random background to an inverse problem with deterministic background. Moreover, we need also the following assumption:

Assumption 5.4. *We assume to be able to evaluate the operator* Λ *for any realization of the random variable* $\mathbf{Y}$*. To evaluate the operator* Λ *means that we are able to collect all the voltages corresponding to the current patterns in the Fourier basis, up to a certain accuracy.*

This assumption is completely fine in the epistemic uncertainty or in the stationary aleatory uncertainty, but could not be the case in the nonstationary aleatory uncertainty.

Now we discuss the crucial point of how to extend to the case of uncertain background the operator $\widetilde{\Lambda}$ defined in (5.2.4) for inhomogeneous but deterministic background. The main question is: how do we collect the measurements of the random operators Λ and Λ_0? Denote by $\{\boldsymbol{y}_m\}_m$ and $\{\widehat{\boldsymbol{y}}_t\}_t$ the sets of realizations drawn from the random variables $\{\boldsymbol{Y}_m\}_m \overset{\text{iid}}{\sim} \rho$ and $\{\widehat{\boldsymbol{Y}}_t\}_t \overset{\text{iid}}{\sim} \rho$ that generated the corresponding measurements of the random operators Λ and Λ_0, respectively. Even if the values of the realizations are unknown, at least two different situations can be identified, depending on how the measurements are collected:

- **arbitrary measurements**: the measurements of Λ and Λ_0 are collected separately, *i.e.* the sets $\{\boldsymbol{y}_m\}_m$ and $\{\widehat{\boldsymbol{y}}_t\}_t$ have nothing to do with each other. In the following we need only to assume that the realizations in each set are independent only among those belonging to the same set.

- **paired measurements**: the measurements of Λ and Λ_0 are collected in pair, *i.e.* the mth measurement of Λ and Λ_0 is collected with the same (but unknown) realization $\boldsymbol{y}_m = \widehat{\boldsymbol{y}}_m$ of the random variable $\boldsymbol{Y}$.

The case of arbitrary measurements is the general one, and it arises in both epistemic and aleatory uncertainties. The case of paired measurements arises in epistemic uncertainty. For example, in crack testing: several mechanical objects are built together with the same internal but unknown properties. The measurements of Λ are collected using the mechanical object to test, and the measurements of Λ_0 are collected using a reference object, where the absence of cracks has been checked by means of other more expensive techniques.

When the measurements are paired, we can define the operator $\widetilde{\Lambda}$ in the case of random background as in (5.2.4),

$$\widetilde{\Lambda}(\boldsymbol{y}) = \Lambda(\boldsymbol{y}) - \Lambda_0(\boldsymbol{y}), \quad \boldsymbol{y} \in \Gamma, \tag{5.3.5}$$

where now the operator is parametrized by the realization $\boldsymbol{y} \in \Gamma$ of the random variable $\boldsymbol{Y}$. Therefore

$$\widetilde{\Lambda} : \mathring{L}^2(\partial B) \otimes L^2(\Gamma, d\rho) \to \mathring{L}^2(\partial B) \otimes L^2(\Gamma, d\rho),$$

and $\widetilde{\Lambda}(\boldsymbol{y}) : \mathring{L}^2(\partial B) \to \mathring{L}^2(\partial B)$ is strongly compact for any value $\boldsymbol{y} \in \Gamma$. At this point, the only difficulty that remains is that the value of $\boldsymbol{y}$ that generated the measurement is unknown. We will see in Section 5.3.2 how to cope with this issue.

When the measurements are arbitrary, the operator $\widetilde{\Lambda}$ depends on $\boldsymbol{y}$ and $\widehat{\boldsymbol{y}}$,

$$\widetilde{\Lambda}(\boldsymbol{y}, \widehat{\boldsymbol{y}}) = \Lambda(\boldsymbol{y}) - \Lambda_0(\widehat{\boldsymbol{y}}), \quad \boldsymbol{y}, \widehat{\boldsymbol{y}} \in \Gamma, \tag{5.3.6}$$

and the strong compactness property of the operator $\widetilde{\Lambda}(\boldsymbol{y}, \widehat{\boldsymbol{y}}) : \mathring{L}^2(\partial B) \to \mathring{L}^2(\partial B)$ is lost in general. However, we know that the singular values of a strongly compact operator decay exponentially fast. So we can look at the singular values of $\widetilde{\Lambda}$ as defined in (5.3.6), and choose the value of $\widehat{\boldsymbol{y}}$ that yields the fastest decay of the spectrum of $\widetilde{\Lambda}$. The singular values of several strongly compact operators are computed in [HM11].

We introduce the quantity K defined as

$$K(\boldsymbol{y}, \widehat{\boldsymbol{y}}) = \Big|\text{supp}_{\boldsymbol{x} \in B}\Big\{|\sigma(\boldsymbol{x}, \boldsymbol{y}) - \sigma_B(\boldsymbol{x}, \widehat{\boldsymbol{y}})| \neq 0\Big\}\Big|, \quad \boldsymbol{y}, \widehat{\boldsymbol{y}} \in \Gamma, \tag{5.3.7}$$

that gives the size of the support of the contrast between σ and σ_B, for any realization $\boldsymbol{y}$ and $\widehat{\boldsymbol{y}}$ in Γ. Of course

$$|D| = K(\boldsymbol{y}, \boldsymbol{y}) \leq K(\boldsymbol{y}, \widehat{\boldsymbol{y}}) \leq |B|, \quad \forall\, (\boldsymbol{y}, \widehat{\boldsymbol{y}}) \in \Gamma \times \Gamma.$$

In general, the singular values of $\widetilde{\Lambda}$ increase with respect to the value of K. The next remark proves this statement in a particular situation where explicit calculations are made possible.

Remark 5.5. *Suppose that B is a circular domain with unitary radius, and D is a concentric circular inclusion with radius $\rho < 1$. The diffusion coefficient is assumed to be homogeneous both in the background and in the inclusion. We define $\mu = \sigma_B/\sigma_D$ and observe that K defined in* (5.3.7) *satisfies $K = |supp_{x\in B}\{\sigma - \sigma_B \neq 0\}| = |D|$. Of course ρ^2 is linearly proportional to K. The eigenvalues of the Dirichlet-to-Neumann map defined by problem* (5.2.2) *satisfy (see also [HK10, Section 2] and [SMI00, Lemma 4. 1])*

$$|\lambda_i^{DN}| = i\left|\frac{\mu+1-(\mu-1)\rho^{2i}}{\mu+1+(\mu-1)\rho^{2i}}\right|, \quad i = 1, 2, \dots$$

and therefore the eigenvalues of the dual Neumann-to-Dirichlet map decay as

$$|\lambda_i^{ND}| = \frac{1}{i}\left|\frac{\mu+1+(\mu-1)\rho^{2i}}{\mu+1-(\mu-1)\rho^{2i}}\right|, \quad i = 1, 2, \dots$$

From this we can derive the decay of the eigenvalues of the operator $\widetilde{\Lambda}$:

$$|\lambda_i(\widetilde{\Lambda})| = \frac{1}{i}\left|\frac{\mu+1+(\mu-1)\rho^{2i}}{\mu+1-(\mu-1)\rho^{2i}} - 1\right| = \frac{1}{i}\left|\frac{2}{\rho^{-2i}(2(\mu-1)^{-1}+1)-1}\right|, \quad i = 1, 2, \dots \tag{5.3.8}$$

and clearly the smallest $\rho < 1$, the fastest the decay, as shown in Fig. 5.2. See also the numerical sections where we tested inclusions with radius $\rho = 0.1$ and $\rho = 0.0275$. From (5.3.8)*, if we suppose $\mu > 1$ we derive*

$$|\lambda_i^{ND}| \sim \frac{K^i}{i}\frac{\mu-1}{\mu+1}, \quad i = 1, 2, \dots$$

The dependency of the eigenvalues on μ is very small, and it becomes unperceivable already from the third eigenvalue λ_3 on (see Fig. 5.2). Looking at the first eigenvalue we see a slight dependence of λ_1 on μ: the higher μ (that is equivalent to say the higher the contrast $\sigma_B - \sigma_D$) the slower the decay. The case $\rho = 0.1$ and $\mu = 10000$ has been tested in the numerical sections, since we have set $\sigma_D = 10^{-4}$ and $\sigma_B \equiv 1$.

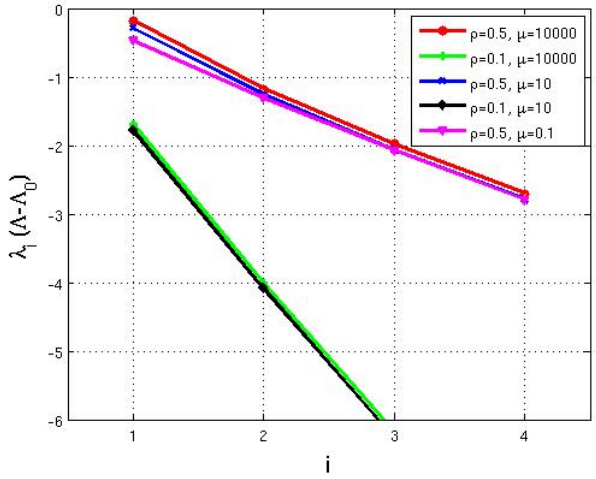


Figure 5.2: The decay of the eigenvalues $\lambda_i(\widetilde{\Lambda})$ in (5.3.8) for different choices of ρ and μ.

Remark 5.6. *The same argument outlined in Remark 5.5 is useful to investigate also the case of piecewise constant background with random values. Suppose that the circular domain B contains a concentric circular region R_2, and $R_1 = B \setminus R_2$. Consider $\sigma_B = 1$ in R_1 and $\sigma_B = y$ in R_2, where y is the realization of the univariate random variable Y observed when getting the measurement of Λ. When we get the measurement of Λ_0 we observe another realization $\widehat{y}$ of*

the same random variable Y, so that $\sigma_B = 1$ in R_1 but $\sigma_B = \widehat{y}$ in R_2 and the region R_2 is generating a contrast $|y - \widehat{y}|$ in the diffusion coefficient of the operator $\widetilde{\Lambda}$. Therefore the region R_2 is behaving as an inclusion in the operator $\widetilde{\Lambda}$. The real inclusion lies in the center of the domain B, but is obfuscated by the presence of the contrast region R_2. Due to the presence of the inclusion, the region R_2 is slightly inhomogeneous in its center, but as explained in Remark 5.5 the value of the contrast is practically negligible compared to ρ.

Remark 5.5 and Remark 5.6 justify analytically that the slower the decay of the eigenvalues, the larger the $K(\boldsymbol{y}, \widehat{\boldsymbol{y}})$ with respect to ρ. Formally $K(\boldsymbol{y}, \widehat{\boldsymbol{y}})$ can attain the value $|D|$, but when the random coefficient is associated with a continuous probability density a contrast is always present with probability one. The same occurs also in the applications: there is always a contrast somewhere in the domain between the background coefficient of Λ and Λ_0, because of modeling, measurement and numerical errors. As a consequence, the quantity K and the value of the contrast $|\sigma(\boldsymbol{x}, \boldsymbol{y}) - \sigma_B(\boldsymbol{x}, \widehat{\boldsymbol{y}})|$ have both influence on the decay of the eigenvalues. Since the background coefficients σ and σ_B are continuous w.r.t. the random variables (because of Assumption (5.2)), the magnitude of their contrast $|\sigma(\boldsymbol{x}, \boldsymbol{y}) - \sigma_B(\boldsymbol{x}, \widehat{\boldsymbol{y}})|$ satisfies

$$\sup_{\boldsymbol{x} \in B \setminus D} |\sigma(\boldsymbol{x}, \boldsymbol{y}) - \sigma_B(\boldsymbol{x}, \widehat{\boldsymbol{y}})| \leq C \|\widehat{\boldsymbol{y}} - \boldsymbol{y}\|_\infty, \quad \forall\, (\boldsymbol{y}, \widehat{\boldsymbol{y}}) \in \Gamma \times \Gamma.$$

In conclusion, we proved that in a circular domain with a circular concentric inclusion and piecewise constant background the decay of the eigenvalues of $\widetilde{\Lambda}$ is governed by $\|\widehat{\boldsymbol{y}} - \boldsymbol{y}\|_\infty$. Reasonably the same conclusion can be extended to more general geometries and partitions of the domain into regions with arbitrary shape.

Consequently, when $M = 1$ and $\Lambda(\mathbf{y})$ is the only measurement generated by the (unknown) realization $\mathbf{y}$, we propose to solve the inverse problem

$$\widehat{\boldsymbol{y}} = \underset{\boldsymbol{y}^* \in \Gamma}{\operatorname{argmin}} \, \|\widetilde{\Lambda}(\boldsymbol{y}, \boldsymbol{y}^*)\|_k, \tag{5.3.9}$$

to obtain $\widehat{\boldsymbol{y}}$ that satisfies

$$\|\widehat{\boldsymbol{y}} - \boldsymbol{y}\|_\infty \leq \varepsilon. \tag{5.3.10}$$

The norm $\|\cdot\|_k$ is defined as $\|\cdot\|_k = \sum_{i=1}^{k} \sigma_i(\cdot)$, and is known as *Ky Fan* k-norm. For any continuous linear operator in Hilbert spaces, the Ky Fan 1-norm corresponds to the usual operator norm equal to the largest singular value of the operator. The value of $\varepsilon > 0$ has to be sufficiently small, depending on the size of the inclusion $|D|$ and on the jumps of the background coefficient. At the discretization level, one can use also the *Frobenius norm*, that takes into account all the singular values including the smallest ones.

The problem (5.3.9) is an ill-posed inverse problem over the parameter space Γ, and the dimension of Γ can be moderately large. If more than one measurement of the operator Λ are available, *i.e.* $M > 1$, it is possible to exploit this further information solving

$$\widehat{\boldsymbol{y}} = \underset{\boldsymbol{y}^* \in \Gamma}{\operatorname{argmin}} \, \min_{1 \leq m \leq M} \|\widetilde{\Lambda}(\boldsymbol{y}_m, \boldsymbol{y}^*)\|_k. \tag{5.3.11}$$

The objective function in (5.3.11) contains more minimum points displaced in Γ, and the optimization algorithm can more easily detect one of them which satisfies (5.3.10), with $\mathbf{y}$ being the one among the M realizations that generated the M measurements. We will discuss in Section 5.6 how to efficiently solve problems (5.3.9) and (5.3.11).

5.3.1 The Factorization Method in the case of arbitrary measurements

In the case of random background with arbitrary measurements, we propose two versions of the FM.

The first version of the FM that we present is an optimized variant (Algorithm 2) where the background is reconstructed by means of an optimization approach. This optimization is

mandatory whenever only one measurement or few measurements of the operator Λ are available. The additional operations w.r.t. Algorithm (1) concern the approximation of the solution $\widehat{\boldsymbol{y}}$ to the optimization problem (5.3.9) when only one measurement of Λ is available, or (5.3.11) when more than one measurement are available. The second version is named pure variant of the

Algorithm 2 The optimized FM in uncertain background (arbitrary measurements)

Sample the region of B to be probed with a set of points $\mathcal{P} = \{\boldsymbol{\xi}_j\}_{j=1}^{P}$,
Choose the initial guess $\widehat{\boldsymbol{y}}^{(i)} \in \Gamma$, $i = 0$,
repeat
  Compute $\widehat{\boldsymbol{y}}^{(i+1)}$ performing one or several steps in the optimization procedure for solving problem (5.3.9) with initial guess $\widehat{\boldsymbol{y}}^{(i)}$.
  for $\boldsymbol{\xi}$ in the set $\mathcal{P}$ **do**
    solve problem (5.2.5) with $\sigma_B = \sigma_B(\cdot, \widehat{\boldsymbol{y}}^{(i+1)})$ to find its solution $N(\cdot, \boldsymbol{\xi}, \widehat{\boldsymbol{y}}^{(i+1)})$,
    compute $l^k_{\boldsymbol{\xi}}(\widehat{\boldsymbol{y}}^{(i+1)})$ from $N(\cdot, \boldsymbol{\xi}, \widehat{\boldsymbol{y}}^{(i+1)})$ using (5.2.7),
    use $l^k_{\boldsymbol{\xi}}(\widehat{\boldsymbol{y}}^{(i+1)})$ to compute the indicator $C(\boldsymbol{\xi})$ in (5.2.26),
    plot $\boldsymbol{\xi} \mapsto C(\boldsymbol{\xi})$,
  end for
  $i \leftarrow i + 1$
until $\Big($the background is NOT reconstructed in $C(\boldsymbol{\xi})$ $\Big)$ OR $\Big($ i $\geq$ maxit$\Big)$.

FM (Algorithm 3), because the final imaging reconstruction is a weighted linear combination of intermediate imaging reconstructions, each one obtained without any attempt to reconstruct the background. The approximation of the solution to problem (5.3.11) satisfying the convergence criterion (5.3.10) within a tolerance ε arbitrarily small is found by means of a sampling approach, rather than by optimization. This variant becomes attractive only when many measurements are available.

To set up an easy notation, we consider only the parametrization (5.3.2) of the coefficient σ_B with the vector $\mathbf{d}$. The same idea can be applied with any parametrization of σ_B. Define the l^∞ norm of an q-dimensional vector v, weighted w.r.t. the q-dimensional vector $\mathbf{d}$ as $\|\mathbf{v}\|_{\infty,\mathbf{d}} = \max\{d_1\, v_1, \ldots, d_q\, v_q\}$. Given the values of q, M, ε and the vector $\mathbf{d}$, then the number T of realizations in the sampling procedure is chosen such that

$$Pr\left(\min_{\substack{1\leq m\leq M\\ 1\leq t\leq T}} \|\boldsymbol{y}_m - \widehat{\boldsymbol{y}}_t\|_{\infty,\mathbf{d}} \leq \varepsilon\right) \geq 0.99. \tag{5.3.12}$$

Notice that ε in (5.3.12) plays the same role as in (5.3.10). If the dispersion vector $\mathbf{d}$ has components all equal to $d > 0$, so that the dispersion is the same for all the components of the random variable $\mathbf{Y}$, then (5.3.12) can be rewritten with the usual l^∞ norm as

$$Pr\left(\min_{\substack{1\leq m\leq M\\ 1\leq t\leq T}} \|\boldsymbol{y}_m - \widehat{\boldsymbol{y}}_t\|_{\infty} \leq \varepsilon\, d^{-1}\right) \geq 0.99. \tag{5.3.13}$$

In the pure FM the value of T is chosen such that (5.3.12) is satisfied, and this allows us to obtain an approximation of one of the realizations that generated the measurements, within a tolerance ε and with 99% probability. In this way the optimization of problem (5.3.11) can be avoided. Notice that, given M, q, d and ε, condition (5.3.13) states how to choose T such that the pure FM detects the inclusions with at least 99% probability. In Section 5.7.2 the calibration of ε is discussed and the value of T such that condition (5.3.13) holds is calculated for some values of M, q, d, ε and $\rho = \mathcal{U}(\Gamma)$.

Looking at Algorithms 2 and 3, it is clear that the most computationally expensive part of these two variants of the FM consists in the repetitive evaluation of Λ_0, for many realizations of the random variable $\mathbf{Y}$. In Algorithm 2 this operation is hidden inside the optimization step,

Algorithm 3 The pure FM in uncertain background (arbitrary measurements)

Collect M measurements $\Lambda(\boldsymbol{y}_1), \dots, \Lambda(\boldsymbol{y}_M)$,
Sample the region of B to be probed with a set of points $\mathcal{P} = \{\boldsymbol{\xi}_j\}_{j=1}^{P}$,
Sample T realizations $\widehat{\boldsymbol{y}}_1, \dots, \widehat{\boldsymbol{y}}_T$ from the random variables $\widehat{\mathbf{Y}}_1, \dots, \widehat{\mathbf{Y}}_T \overset{\text{iid}}{\sim} \rho$,
$i = 1$,
repeat
  Compute $F_{mi} = \|\Lambda(\boldsymbol{y}_m) - \Lambda_0(\widehat{\boldsymbol{y}}_i)\|_k$, $\forall\, m$ and $m_i = \text{argmin}_{m=1,\dots,M} F_{mi}$,
  for $\boldsymbol{\xi}$ in the set $\mathcal{P}$ **do**
    solve problem (5.2.5) with $\sigma_B = \sigma_B(\cdot, \widehat{\boldsymbol{y}}_i)$ to find its solution $N(\cdot, \boldsymbol{\xi}, \widehat{\boldsymbol{y}}_i)$,
    compute $l^k_{\boldsymbol{\xi}}(\widehat{\boldsymbol{y}}_i)$ from $N(\cdot, \boldsymbol{\xi}, \widehat{\boldsymbol{y}}_i)$ using (5.2.7),
    use $l^k_{\boldsymbol{\xi}}(\widehat{\boldsymbol{y}}_i)$ and the SVD decomposition of $\widetilde{\Lambda}(\boldsymbol{y_{m_i}}, \widehat{\boldsymbol{y}}_i)$ to compute $\|g^k_{\boldsymbol{\xi}}(\boldsymbol{y_{m_i}}, \widehat{\boldsymbol{y}}_i)\|_{L^2(\partial B)}$,
    use $\|g^k_{\boldsymbol{\xi}}(\boldsymbol{y_{m_i}}, \widehat{\boldsymbol{y}}_i)\|_{L^2(\partial B)}$ to compute the indicator $C_i(\boldsymbol{\xi})$ in (5.2.26),
    update $C(\boldsymbol{\xi}) \leftarrow C(\boldsymbol{\xi}) + C_i(\boldsymbol{\xi}) \exp\Big(-\frac{(\min_{1\le j\le i} F_{m_i j})}{F_{m_i i}}\Big)$,
    plot $\boldsymbol{\xi} \mapsto C(\boldsymbol{\xi})$,
  end for
  $i \leftarrow i + 1$
until $\Big($the background is NOT reconstructed in $C(\boldsymbol{\xi})\ \Big)$ OR $\Big(\ \text{i} \ge \text{T}\Big)$.

where pointwise evaluations of Λ_0 are needed. The number of evaluations of Λ_0 to achieve an accurate approximation of the solution to problems (5.3.9) and (5.3.11) can be reduced by means of polynomial approximation of problem (5.2.2), see Chapter 6.

5.3.2 The Factorization Method in the case of paired measurements

When a pair of measurements of Λ and Λ_0 is available, *i.e.* the measurements are generated from the same realization $\boldsymbol{y}$ of the random variable $\boldsymbol{Y}$, then we can build straightforwardly the operator $\widetilde{\Lambda}$ from (5.3.5). Since the value of the realization $\boldsymbol{y}$ that generated the measurements is unknown, we have to work with the sampling procedure to span the parameter space Γ looking for a good approximation of $\boldsymbol{y}$.

Let $\{\mathbf{s}_t\}_{t=1}^{T}$ be T realizations drawn from the random variables $\{\boldsymbol{Y}_t\}_{t=1}^{T} \overset{\text{iid}}{\sim} \rho$ that are used to explore the parameter space Γ. We propose the following indicator function, that substitutes the one defined in (5.2.26):

$$\widehat{C}(\boldsymbol{\xi}) = \sum_{t=1}^{T} \frac{1}{\log_{10}\Big(v_1(\boldsymbol{\xi}, \mathbf{s}_t) + v_2(\boldsymbol{\xi}, \mathbf{s}_t)\Big)}, \tag{5.3.14}$$

In (5.3.14) the functions $v_k : (\boldsymbol{\xi}, \mathbf{y}) \in (B \times \Gamma) \mapsto v_k(\boldsymbol{\xi}, \mathbf{y}) : B \times \Gamma \to \mathbb{R}$ are defined as $v_k(\boldsymbol{\xi}, \mathbf{y}) := \|g^k_{\xi}(\mathbf{y})\|_{L^2(\partial B)}$ and $g^k_{\xi}(\mathbf{y})$ is again given by (5.2.23) but with l^p_{ξ} that depends on $\mathbf{y}$, as in (5.3.4). The indicator (5.3.14) is based on the sample mean estimator, which is very sensitive to the outliers that correspond to the points where it diverges due to the presence of the inclusion. This variant of the FM that exploits the indicator function (5.3.14) is described in Algorithm 4. The number T of realizations to be evaluated with the FM has to be tuned according to the dimension q of the parameter space Γ.

5.4 Numerical tests featuring deterministic piecewise constant backgrounds

In this section we present several numerical tests to illustrate the capabilities of the FM in homogeneous and inhomogeneous but deterministic backgrounds. To show the results, we display the isolines of the indicator function $C(\boldsymbol{\xi})$ given by (5.2.26). A crucial issue is the tuning of

Algorithm 4 The FM in uncertain background (paired measurements)

Sample the region of B to be probed with a set of points $\mathcal{P} = \{\boldsymbol{\xi}_j\}_{j=1}^{P}$,
Sample T realizations $\mathbf{s}_1, \dots, \mathbf{s}_T$ from the random variables $\mathbf{S}_1, \dots, \mathbf{S}_T \overset{\text{iid}}{\sim} \rho$,
for $\boldsymbol{\xi}$ in the set $\mathcal{P}$ **do**
  for $t = 1, \dots T$ **do**
    solve problem (5.2.5) with $\sigma_B = \sigma_B(\cdot, \mathbf{s}_t)$ to find its solution $N(\cdot, \boldsymbol{\xi}, \mathbf{s}_t)$,
    compute $l_{\xi}^{k}(\mathbf{s}_t)$ from $N(\cdot, \boldsymbol{\xi}, \mathbf{s}_t)$ using (5.2.7),
    update the indicator $\widehat{C}(\boldsymbol{\xi})$ in (5.3.14) with $l_{\boldsymbol{\xi}}^{k}(\mathbf{s}_t)$,
  end for
  plot $\boldsymbol{\xi} \mapsto \widehat{C}(\boldsymbol{\xi})$.
end for

the scale and the choice of the isovalue that represents the inclusion. In practice, this requires additional information on the value of the diffusion coefficient σ_D in the inclusions. In all the numerical tests we plot the indicator $C(\boldsymbol{\xi})$ in the domain B choosing an uniform coloring scale, such that the black color is associated with the zero value and the white color is associated with $\max_{\boldsymbol{\xi} \in B} C(\boldsymbol{\xi})$. Throughout the paper the space $\mathring{L}^2(\partial B)$ is discretized using the Fourier basis with 128 terms. P2 finite elements over a 25000 nodes mesh have been used to numerically solve problem (5.2.2) with the diffusion coefficient (5.2.3) and problem (5.2.15).

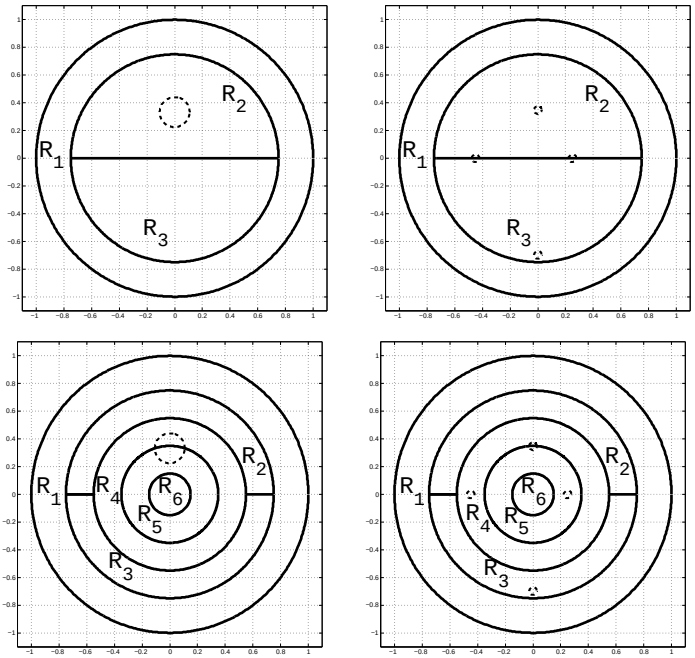


Figure 5.3: Geometries of the regions in the background and displacement of the inclusions (in dashed line). Top-left: configuration EN. Top-right: configuration ES. Bottom-left: configuration UN. Bottom-right: configuration US.

The domain B is always the unitary ball centered in $(0, 0)$. The sampling points consist of a 50×50 uniform grid over the square $[-1, 1] \times [-1, 1]$. In the FM we use only the points $\boldsymbol{\xi} \in B$ that fall at a distance larger than 0.05 from the boundary ∂B. In the test cases with small inclusions the resolution is increased to 100×100. The dashed purple line always marks the exact geometry of the inclusion(s). The value of the coefficient that identifies the inclusion is always set to $\sigma_D = 0.001$, although any value that satisfies Assumption 5.1 is allowed.

We focus on two configurations of the background diffusion coefficient σ_B:

- (E) the union of two semicircular regions, which are generated by splitting a circle with a radius equal to 0.75 and centered in the origin in its upper and lower part w.r.t. the

horizontal axis,

- (U) the union of five annular regions with radii 0.15, 0.35 0.55 and 0.75, where the last but one outer annular region is further split into two regions to break the radial symmetry of this configuration.

The inclusion(s) can be configured as

- (N) a circular inclusion with radius 0.1 centered in $(0, 0.33)$,
- (S) four small circular inclusions with radius 0.0275, centered in the points $(0.25, 0)$, $(0, 0.35)$, $(-0.45, 0)$, $(0, -0.7)$ that are progressively closer to the boundary ∂B.

Combining each background configuration with each inclusion configuration we obtain the four configurations EN, ES, UN, US depicted in Fig. 5.3.

5.4.1 The homogeneous case

We begin to apply the FM to the EIT problem with an homogeneous deterministic coefficient $\sigma_B \equiv 1$. Many numerical results in the literature show the potentiality of the method. We focus on two test cases presented in Table 5.1, that are named as hN and hS.

Example label	inclusion configuration	background configuration	parametrization of σ_B	value of $\mathbf{m}$	Figure with reconstruction
hN	N	homogeneous	(5.2.28)	$m_i = 1,\ i = 1, \dots, r$	Fig. 5.4(left)
hS	S	homogeneous	(5.2.28)	$m_i = 1,\ i = 1, \dots, r$	Fig. 5.4(right)

Table 5.1: Description of the numerical test cases hN and hS with homogeneous deterministic background.

Fig. 5.4 shows the results obtained with the inclusion configurations N and S, displaying the coloring of the indicator $C(\boldsymbol{\xi})$ in the domain B. In both cases, the reconstructions are very accurate and the presence and location of all the inclusions are clearly detected.

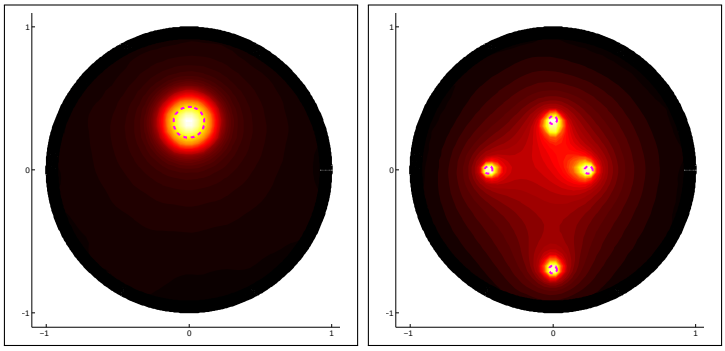


Figure 5.4: Coloring of $C(\boldsymbol{\xi})$. Left: test case hN, $\gamma = 10^{-2}$ in (5.2.25). Right: test case hS, $\gamma = 5 \times 10^{-3}$ in (5.2.25).

5.4.2 The inhomogeneous case

We now proceed to present the results concerning the inhomogeneous deterministic coefficient σ_B. In this case the FM is described in Algorithm 1. As will be explained in the next section, we pick the background coefficient equal to one in the region R_1 that is accessible from the boundary.

Example label	inclusion configuration	background configuration	parametrization of σ_B	value of $\mathbf{m}$	Figure with reconstruction
i1EN	N	E	(5.2.28)	$\mathbf{m} = [1, 5, 0.5]$	Fig. 5.5(left)
i1ES	S	E	(5.2.28)	$\mathbf{m} = [1, 5, 0.5]$	Fig. 5.5(right)
i2EN	N	E	(5.2.28)	$\mathbf{m} = [1, 10, 0.1]$	Fig. 5.6(left)
i2ES	S	E	(5.2.28)	$\mathbf{m} = [1, 0.1, 10]$	Fig. 5.6(right)
i1UN	N	U	(5.2.28)	$\mathbf{m} = [1, 5, 1, 5, 0.5, 0.5]$	Fig. 5.7(left)
i1US	S	U	(5.2.28)	$\mathbf{m} = [1, 5, 1, 0.5, 1, 0.5]$	Fig. 5.7(right)
i2UN	N	U	(5.2.28)	$\mathbf{m} = [1, 10, 1, 10, 0.1, 0.1]$	Fig. 5.8(left)
i2US	S	U	(5.2.28)	$\mathbf{m} = [1, 10, 1, 0.1, 1, 0.1]$	Fig. 5.8(right)

Table 5.2: Description of the numerical test cases i1EN, i1ES, i2EN, i2ES, i1UN, i1US, i2UN and i2US with inhomogeneous deterministic background.

We consider the test cases i1EN, i1ES, i2EN, i2ES, i1UN, i1US, i2UN and i2US presented in Table 5.2.

In Fig. 5.5 we present the results obtained when applying the FM in the test cases i1EN and i1ES, where σ_B is parametrized as in (5.2.28) and $\boldsymbol{m}$ generates one order of magnitude jumps between different regions. In Fig. 5.6 we present the results obtained in the test cases i2EN and i2ES, where σ_B is parametrized as in (5.2.28) and $\boldsymbol{m}$ generates two orders of magnitude jumps.

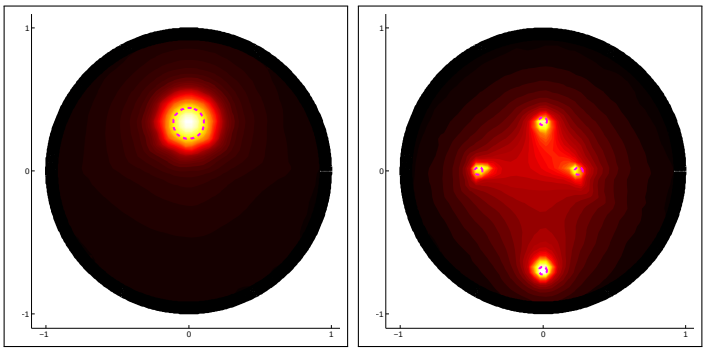


Figure 5.5: Coloring of $C(\boldsymbol{\xi})$. Left: test case i1EN, $\gamma = 10^{-2}$ in (5.2.25). Right: test case i1ES, $\gamma = 7 \cdot 10^{-3}$ in (5.2.25).

Next, in Fig. 5.7 we present the results obtained in the test cases i1UN and i1US, where σ_B is parametrized as in (5.2.28) and $\boldsymbol{m}$ generates one order jumps between different regions. Fig. 5.8 shows the results in the test cases i2UN and i2US, with two orders jumps.

The quality of the reconstructions with jumps of one order of magnitude in Fig. 5.5 and Fig. 5.7 is the same as the quality in Fig. 5.4 with the homogeneous background.

When the order of magnitude of the jumps increases to two, the effect of the jumps in the background coefficient becomes visible in the reconstruction, as in Fig. 5.8–left. Nonetheless the inclusion is still accurately detected. In Fig. 5.6–right we observe that the inclusions on the interface where the coefficient jumps are still detected, albeit with a fainter intensity of the indicator $C(\boldsymbol{\xi})$. The same occurs in Fig. 5.8–right with the most inner inclusion. All the four small inclusions are detected, but the method would hardly inspect the presence of additional inclusions inside their convex envelope. See also [HM11, Figs. 8, 21, 26], where several tests with ten small inclusions are presented.

We can conclude that the FM is able to accurately detect the presence and location of all the inclusions, when the jumps in the background are up to two orders of magnitude.

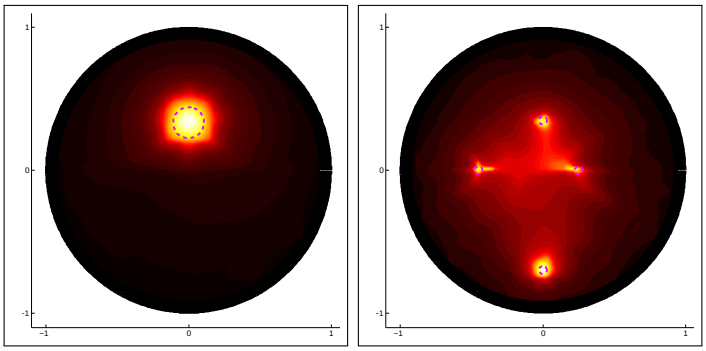


Figure 5.6: Coloring of $C(\boldsymbol{\xi})$. Left: test case i2EN, $\gamma = 10^{-2}$ in (5.2.25). Right: test case i2ES, $\gamma = 10^{-2}$ in (5.2.25).

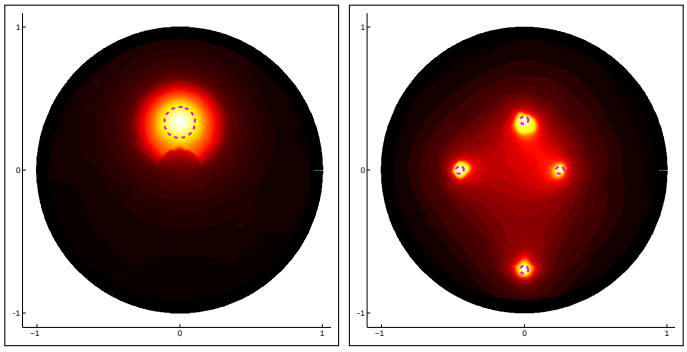


Figure 5.7: Coloring of $C(\boldsymbol{\xi})$. Left: test case i1UN, $\gamma = 5 \cdot 10^{-2}$ in (5.2.25). Right: test case i1US, $\gamma = 5 \cdot 10^{-3}$ in (5.2.25).

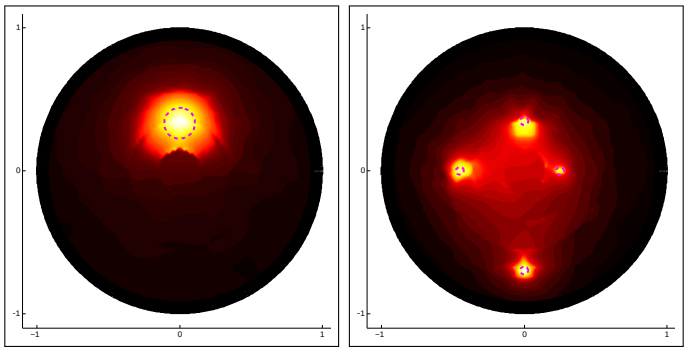


Figure 5.8: Coloring of $C(\boldsymbol{\xi})$. Left: test case i2UN, $\gamma = 5 \cdot 10^{-2}$ in (5.2.25). Right: test case i2US, $\gamma = 10^{-2}$ in (5.2.25).

5.5 Numerical tests featuring deterministic inhomogeneous background with nonlinear spatial dependence

In this section the performance of the numerical scheme proposed are shown, when the background coefficient is deterministic and nonlinearly dependent on the spatial coordinates. Two regularization techniques are compared, *i.e.* Tikhonov with the Morozov principle and the Picard Criterion. The presence of artificial noise is also investigated.

The Picard Criterion described in Remark 5.4 can be implemented to approximate the function (5.2.27) by $\widetilde{I}_k(\boldsymbol{\xi}) \approx I_k(\boldsymbol{\xi})$ with

$$\widetilde{I}_k(\boldsymbol{\xi}) = \sum_{i=1}^{m} \frac{(l_\xi^k, u_i)^2_{L^2(\partial B)}}{|\sigma_i|} \Bigg/ \sum_{i=1}^{m} (l_\xi^k, u_i)^2_{L^2(\partial B)}, \tag{5.5.1}$$

retaining only the m largest singular values above the expected measurement error. We refer to this criterion as the *Picard Criterion* (PC). It does not require to solve problem (5.2.21) to find its regularized solution.

Again, the results are shown by displaying the isolines of the indicator function (5.2.26), being $v_k(\boldsymbol{\xi}) = \|g_\xi^k\|_{L^2(\partial B)}$ in the case of TR, or $v_k(\boldsymbol{\xi}) = \widetilde{I}_k(\boldsymbol{\xi})$ in the case of PC. We plot the isolines of $C(\boldsymbol{\xi})$ in the range

$$\Big[C^f_{\min}, C_{\max}\Big] := \Big[f \cdot \min\{C(\boldsymbol{\xi}) \,:\, \boldsymbol{\xi} \in D\}, \max\{C(\boldsymbol{\xi}) \,:\, \boldsymbol{\xi} \in D\}\Big],$$

using the parameter $0 < f < 1$. The step between the isolines is kept fixed, and this provides information also on the gradient of the indicator function.

5.5.1 The homogeneous case

The tests featuring an homogeneous background $\sigma_B = 1$ and $\sigma_D = 2$ are named with uppercase letters:

- *test case A:* circular inclusion with radius 0.3 centered in $(0.3, 0.1)$;
- *test case B:* ten small circular inclusions with radius 0.025;
- *test case C:* two circular inclusions with radius 0.2 centered in $(-0.35, -0.35)$ and $(0.35, 0.35)$;
- *test case D:* ellipsoidal inclusion centered in $(0.3, 0.1)$, with semiaxes 0.1 and 0.3.

Fig. 5.9 displays the singular values of $\widetilde{\Lambda}^{1/2}$ in the aforementioned test cases. Then, Figs. 5.10, 5.11, 5.12 and 5.13 show the isolines of the indicator function (5.2.26) employing TR or PC. The test case A is classic and does not show any significative difference between TR and PC. If the value of the parameter γ is too low then the isolines exhibits oscillations, because the instability due to ill-posedness shows up. In the test case C the TR is less sensitive to the mutual disturbance between the two inclusions. Also in the test case D, the TR allows to recover the elongated shape, while the PC tends to reconstruct a circle. In the test case B both TR and PC can locate the outer inclusions, although the TR is more accurate and provides also some information for the internal inclusions.

In the homogeneous case, it is possible to clearly detect the inclusion directly looking at the values of the optimal regularization parameter $\alpha = \alpha(\boldsymbol{\xi})$.

5.5.2 The inhomogeneous case

We proceed to test our numerical scheme (5.2.15) in some cases featuring an inhomogeneous background. The roman numbers denote the following tests:

- *test case I:* piecewise constant conductivity defined as in Fig. 5.15-left. The radius of the concentric circle is 0.7. The inclusion is a circle with radius 0.1 centered in $(0.35, 0.35)$ and $\sigma_D = 10^{-3}$.

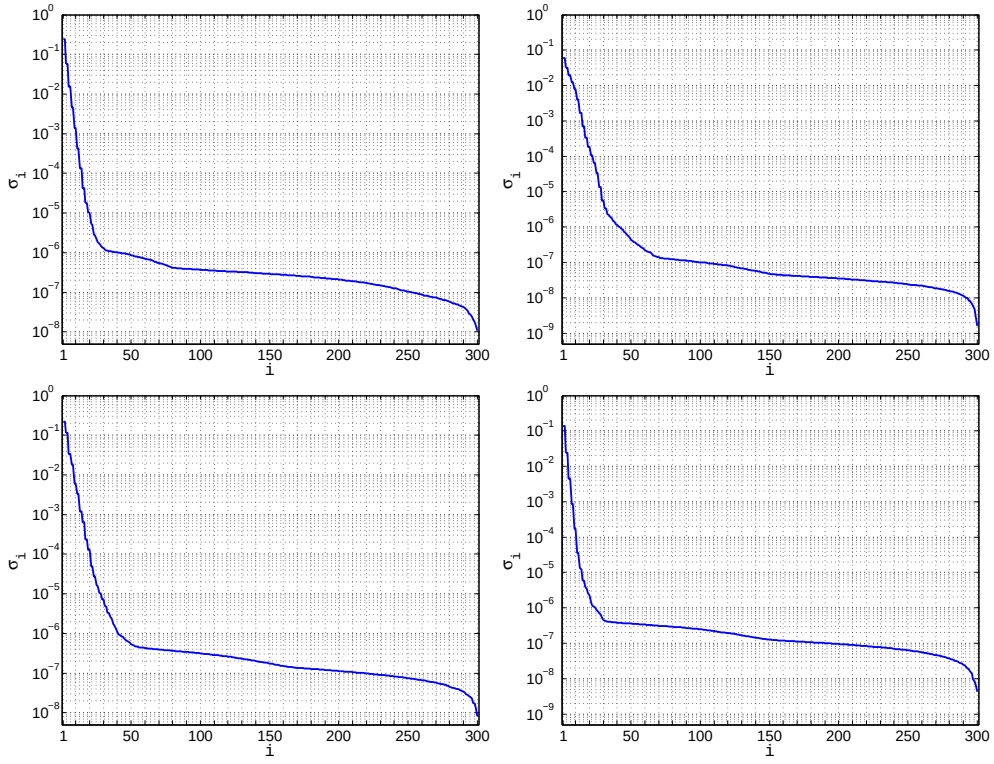


Figure 5.9: Singular values σ_i of $\widetilde{\Lambda}^{1/2}$. Test case A (top-left). Test case B (top-right). Test case C (bottom-left). Test case D (bottom-right).

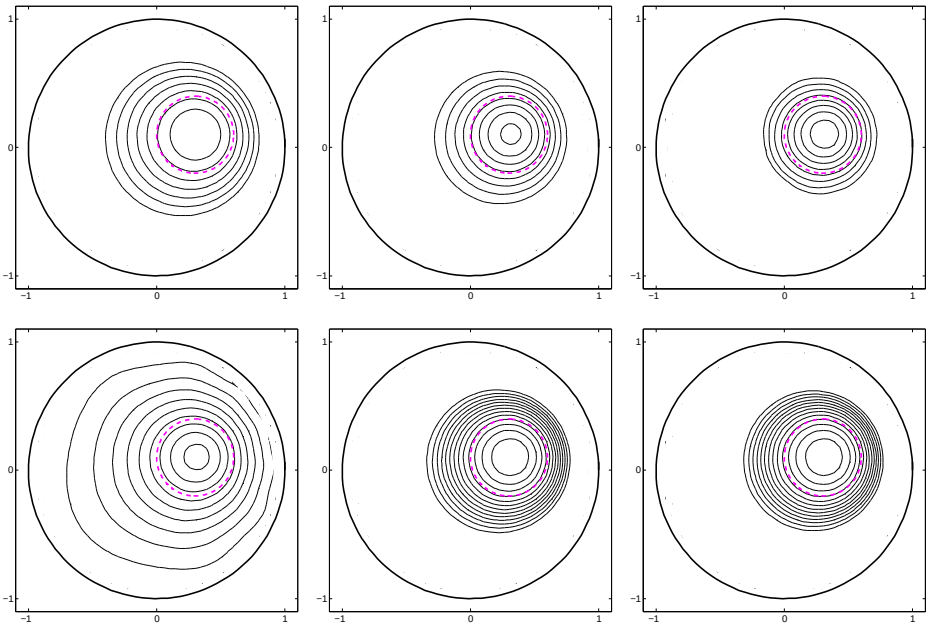


Figure 5.10: Test case A. Isolines of $C(\boldsymbol{\xi})$ in $[C_{\min}^{0.5}, C_{\max}]$. Top: Tikhonov regularization: $\gamma = 5 \times 10^{-1}$ (left), $\gamma = 5 \times 10^{-2}$ (center), $\gamma = 5 \times 10^{-3}$ (right). Bottom: Picard criterion, $m = 5$ (left), $m = 25$ (center), $m = 50$ (right).

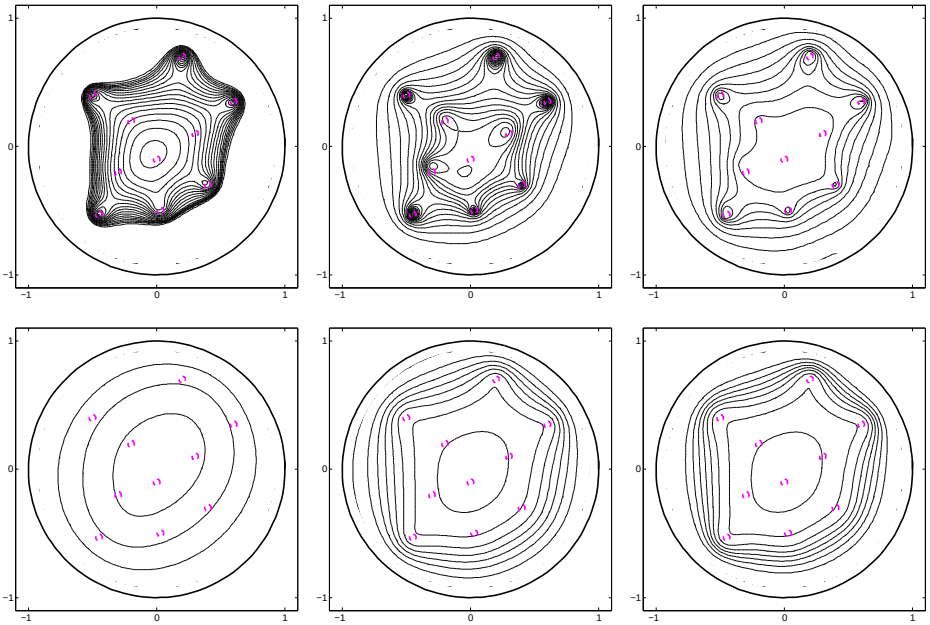


Figure 5.11: Test case B. Isolines of $C(\boldsymbol{\xi})$ in $[C_{\min}^{0.5}, C_{\max}]$. Top: Tikhonov regularization, $\gamma = 10^{-2}$ (left), $\gamma = 10^{-3}$ (center), $\gamma = 10^{-4}$ (right). Bottom: Picard criterion, $m = 10$ (left), $m = 25$ (center), $m = 50$ (right).

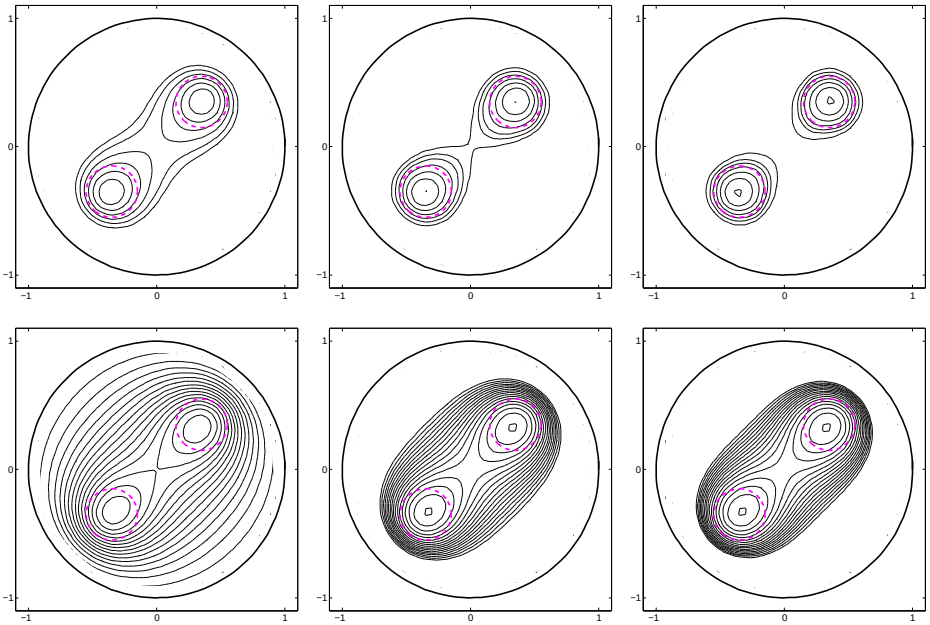


Figure 5.12: Test case C. Isolines of $C(\boldsymbol{\xi})$ in $[C_{\min}^{0.6}, C_{\max}]$. Top: Tikhonov regularization with $\gamma = 5 \times 10^{-3}$ (left), $\gamma = 10^{-3}$ (center), $\gamma = 5 \times 10^{-4}$ (right). Bottom: Picard criterion. $m = 10$ (left), $m = 25$ (center), $m = 50$ (right).

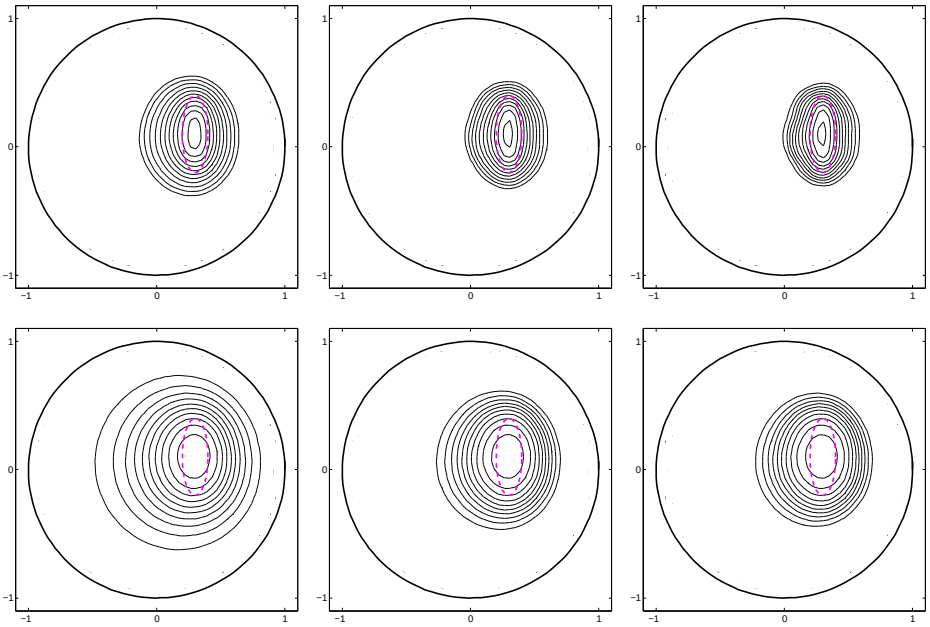


Figure 5.13: Test case D. Isolines of $C(\boldsymbol{\xi})$ in $[C_{\min}^{0.5}, C_{\max}]$. Top: Tikhonov regularization with $\gamma = 5 \times 10^{-2}$ (left), $\gamma = 10^{-2}$ (center), $\gamma = 5 \times 10^{-3}$ (right). Bottom: Picard criterion. $m = 10$ (left), $m = 25$ (center), $m = 50$ (right).

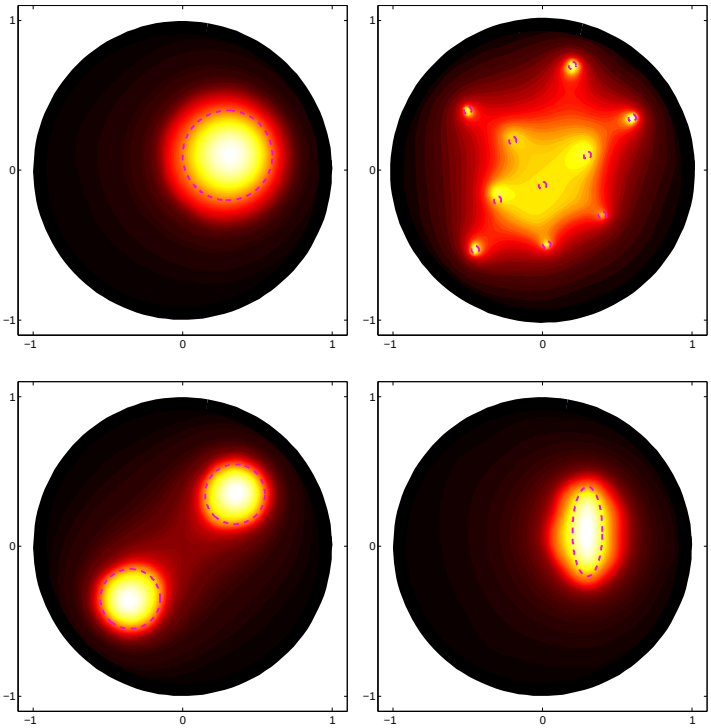


Figure 5.14: Tikhonov regularization. Test case A, $\gamma = 5 \times 10^{-3}$ (top-left). Test case B, $\gamma = 10^{-3}$ (top-right). Test case C, $\gamma = 5 \times 10^{-4}$ (bottom-left). Test case D, $\gamma = 5 \times 10^{-3}$ (bottom-right).

- *test case II:* piecewise constant conductivity defined as in Fig. 5.15-right. The inclusion falls across the interface where the value of σ_B jumps by one order of magnitude. The radius of the concentric circles are 0.35 and 0.75. The inclusion is a circle with radius 0.1 centered in $(0, 0.33)$ and $\sigma_D = 10^{-3}$.

- *test case IIIa:* $\sigma_B(\boldsymbol{x}) = 1 + \beta\Big(\sin(5x_1) + \cos(5x_2)\Big)$, $\beta = 0.15$. The inclusion is a circle with radius 0.15 centered in $(-0.3, 0.3)$ and $\sigma_D = 2$.

- *test case IIIb:* same as test case IIIa but with $\beta = 0.25$.

- *test case IVa:* ten small circular inclusions with radius (0.025), $\sigma_B(\boldsymbol{x}) = 1 + \beta\Big(\sin(5x_1) + \cos(5x_2)\Big)$, $\beta = 0.25$ and $\sigma_D = 2$.

- *test case IVb:* same as test case IVa but with $\beta = 0.3$.

- *test case V:* one small circular inclusion with radius 0.025 centered in $(-0.3, 0.3)$, $\sigma_B(\boldsymbol{x}) = 5 \times (x_1 + x_2) + 11$ and $\sigma_D = 10^{-3}$.

- *test case VI:* $\sigma_B(\boldsymbol{x}) = 1 + (x_1^2 + x_2^2)$, $\sigma_D = 10^{-3}$. The inclusion is a circle with radius 0.15 centered in $(-0.3, 0.3)$.

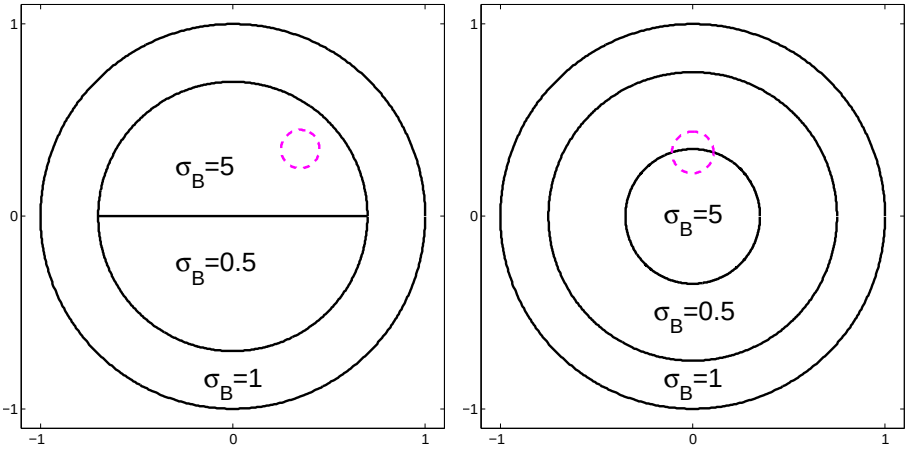


Figure 5.15: Values of σ_B in the test case I (left) and in the test case II (right).

Again, we display the singular values of $\widetilde{\Lambda}^{1/2}$ for all the test cases in Figs. 5.16 and 5.21. Despite the differences among the geometries and values of σ_B in the test cases, the singular values behave likewise. The most important difference is the order of magnitude of the largest singular values, since this affects the choice of the regularization parameters.

The numerical scheme (5.2.15) performs very well in the test cases I,II with a piecewise constant diffusion coefficient. In both cases σ_B jumps by one order of magnitude. In the figures 5.17,5.18 we reconstruct the inclusion using the Tikhonov regularization and the Picard criterion.

In the test case IIIa the coefficient σ_B features a $\pm 30\%$ variation, and Fig. 5.19 shows that both TR and PC are capable of accurately detecting the presence and location of the inclusion.

The test case IIIb features the same geometries as the test case IIIa, but now σ_B has a $\pm 50\%$ variation. Still both TR and PC are able to detect the location of the inclusion (Fig. 5.20), but the recovered shape begins to suffer a distortion, because of the strong nonlinearity in the coefficient. Analogously, in the test cases IVa and IVb the variations in σ_B are $\pm 50\%$ and $\pm 60\%$, respectively. As in the homogeneous case, the outer inclusions are well detected by both TR and PC, but only TR provides also some information in the internal region (Figs. 5.22 and 5.23).

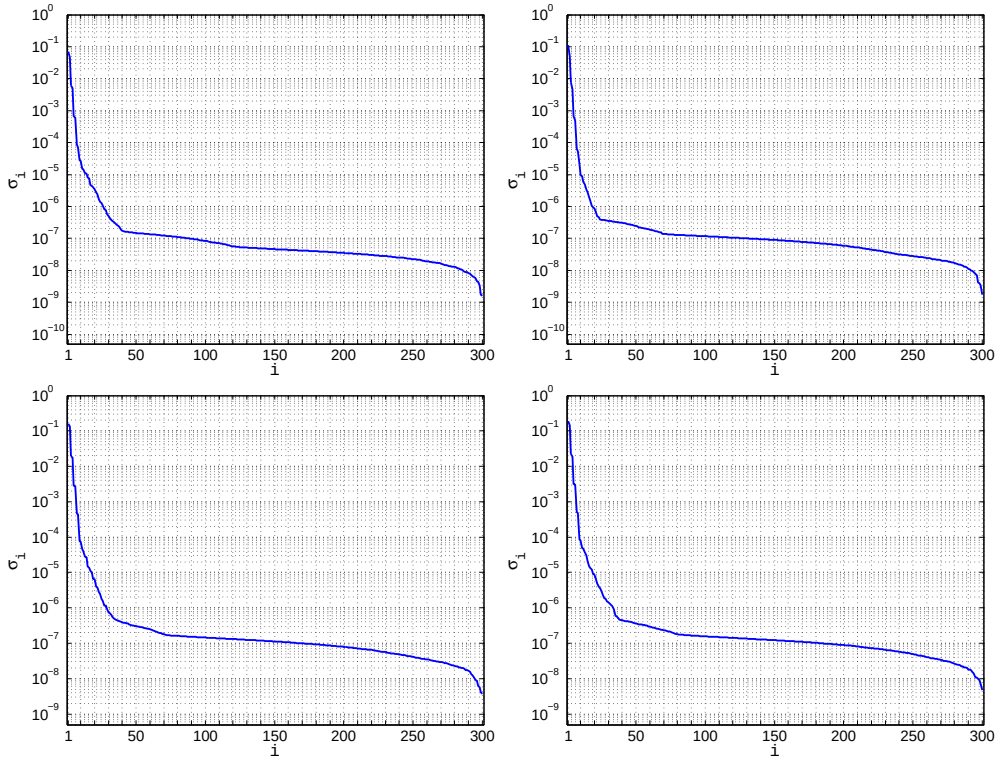


Figure 5.16: Singular values σ_i of $\widetilde{\Lambda}^{1/2}$. Test case I (top-left). Test case II (top-right). Test case IIIa . Test case IIIb (bottom-right).

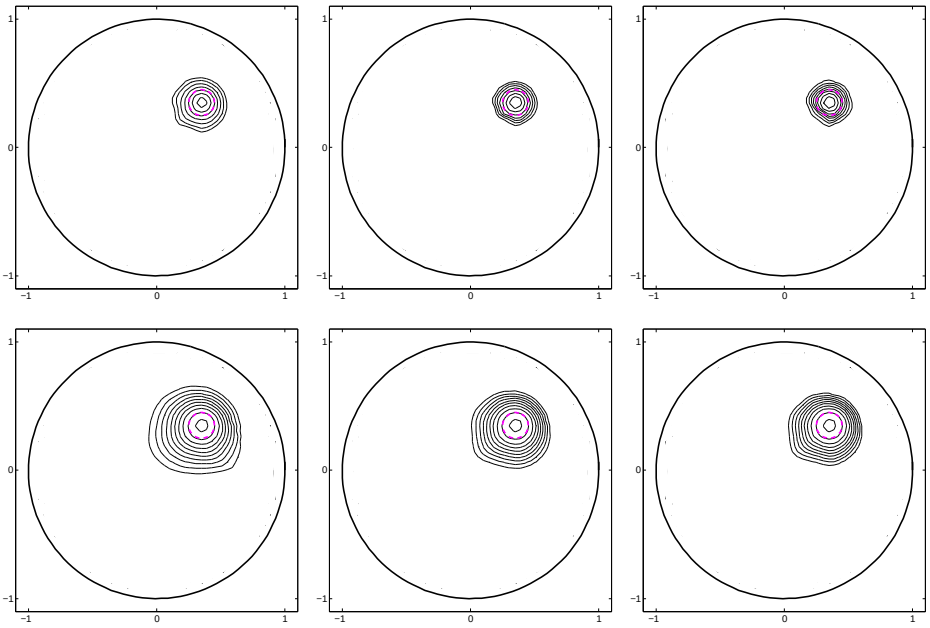


Figure 5.17: Test case I. Isolines of $C(\boldsymbol{\xi})$ in $[C_{\min}^{0.6}, C_{\max}]$. Tikhonov regularization: $\gamma = 5 \times 10^{-2}$ (left), $\gamma = 10^{-2}$ (center), $\gamma = 5 \times 10^{-3}$ (right). Picard criterion, $m = 10$ (left), $m = 25$ (center), $m = 50$ (right).

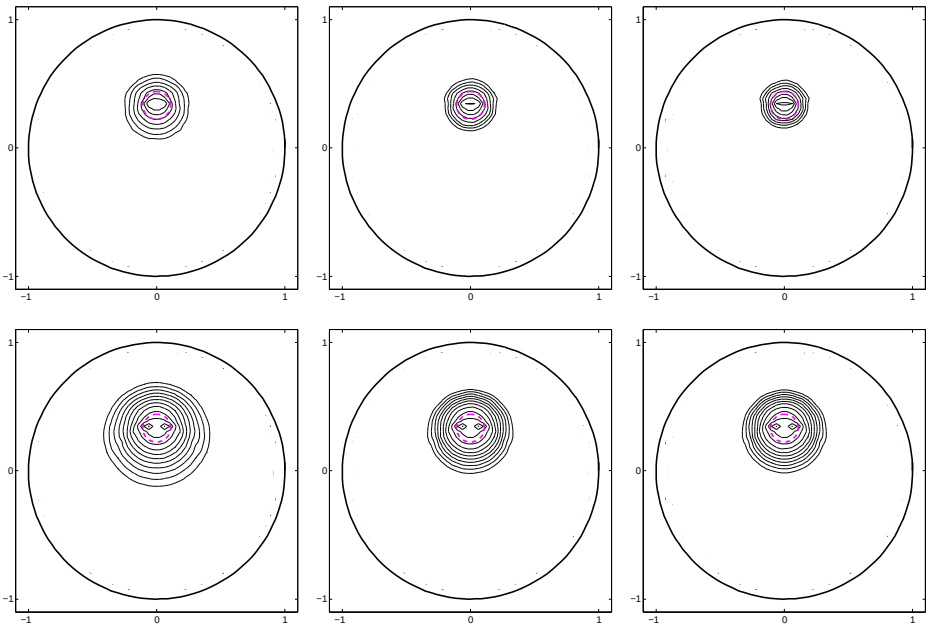


Figure 5.18: Test case II. Isolines of $C(\boldsymbol{\xi})$ in $[C_{\min}^{0.6}, C_{\max}]$. Tikhonov regularization, $\gamma = 5 \times 10^{-2}$ (left), $\gamma = 10^{-2}$ (center), $\gamma = 5 \times 10{-}3$ (right). Picard criterion, $m = 10$ (left), $m = 25$ (center), $m = 50$ (right).

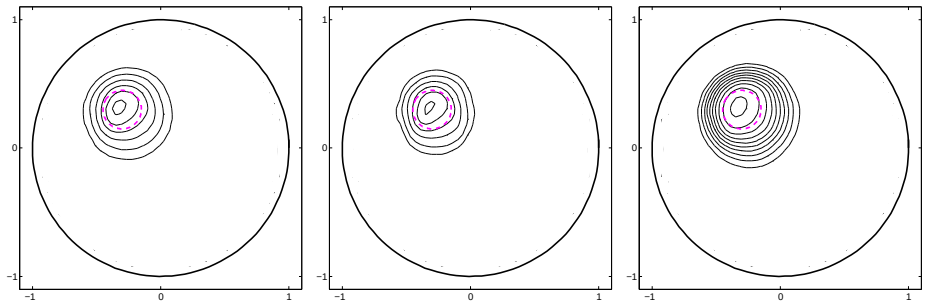


Figure 5.19: Test case IIIa. Isolines of $C(\boldsymbol{\xi})$ in $[C_{\min}^{0.5}, C_{\max}]$. Left: Tikhonov regularization, $\gamma = 10^{-1}$. Center: Tikhonov regularization, $\gamma = 5 \times 10^{-2}$. Right: Picard criterion, $m = 25$.

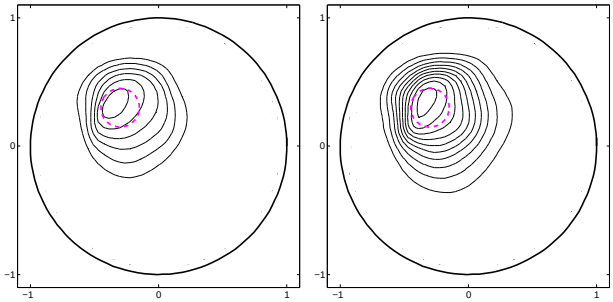


Figure 5.20: Test case IIIb. Isolines of $C(\boldsymbol{\xi})$ in $[C_{\min}^{0.5}, C_{\max}]$. Left: Tikhonov regularization, $\gamma = 5 \times 10^{-1}$. Right: Picard criterion, $m = 10$.

The test case V treats a linear coefficient σ_B that ranges between 4 and 18 approximately, while the test case VI treats a coefficient σ_B with radial dependence on the spatial coordinates. As before (in Figs 5.24 and 5.25), TR and PC can accurately detect the presence and location of the inclusion.

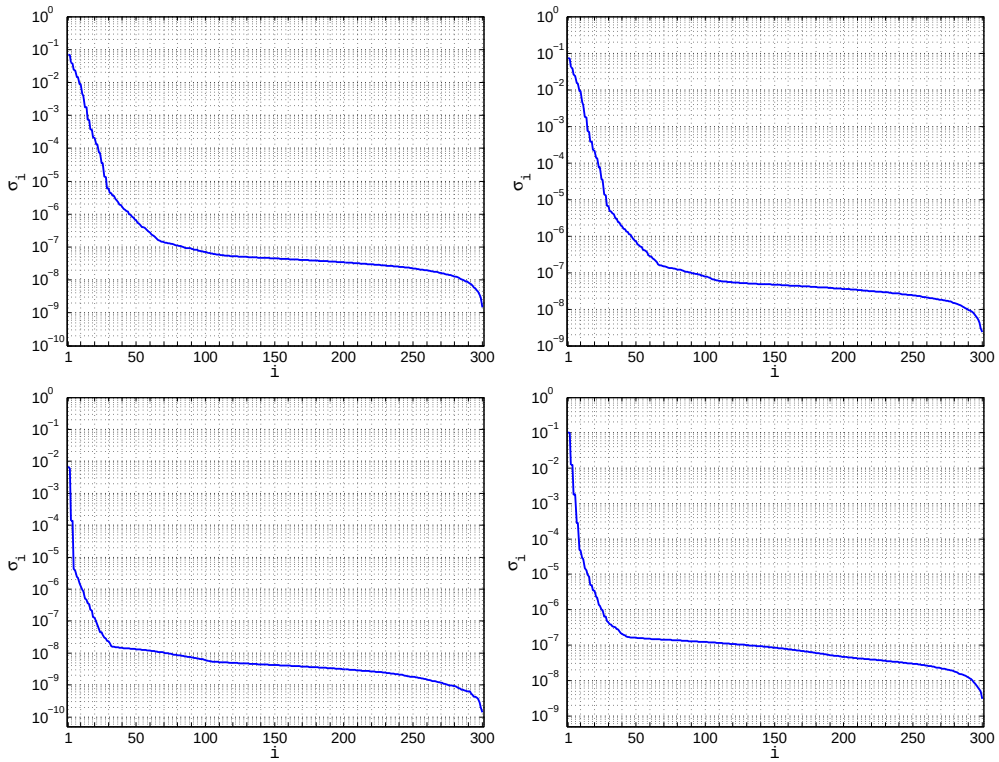


Figure 5.21: Singular values σ_i of $\widetilde{\Lambda}^{1/2}$. Test case IVa (top-left). Test case IVb (top-right). Test case V (bottom-left). Test case VI (bottom-right).

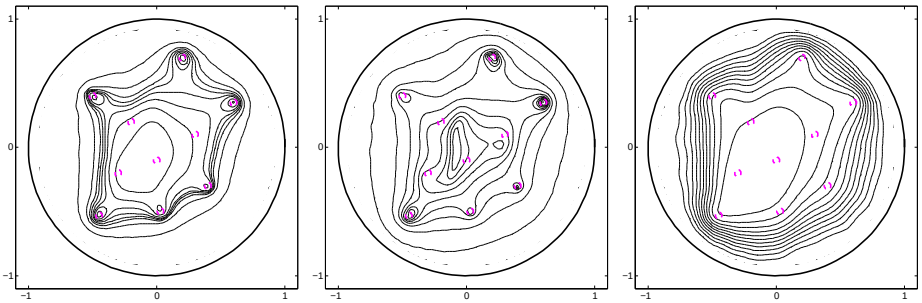


Figure 5.22: Test case IVa. Isolines of $C(\boldsymbol{\xi})$. Left: Tikhonov regularization, $\gamma = 5 \times 10^{-3}$. Center: Tikhonov regularization, $\gamma = 10^{-3}$. Right: Picard criterion, $m = 25$.

5.5.3 Measurements contaminated by artificial noise

Now we investigate the sensitivity of the reconstructions to noisy perturbations. To this aim we perturb the operator G with the random matrix U, $U_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{U}([-1,1])$, scaled to the spectral norm of G:

$$G_\mu = G + \mu \|G\|_2 \frac{U}{\|U\|_2}. \tag{5.5.2}$$

The parameter μ represents the magnitude of the noisy perturbation, and accordingly we will tune γ in (5.2.25) or m in (5.5.1). In the presence of noise the isolines obtained with the TR are

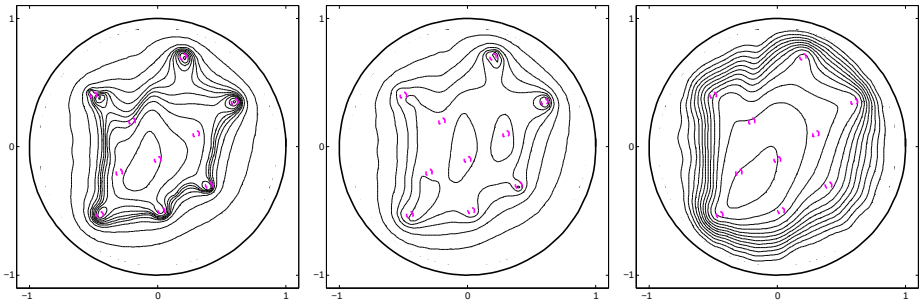


Figure 5.23: Test case IVb. Isolines of $C(\boldsymbol{\xi})$. Left: Tikhonov regularization, $\gamma = 5 \times 10^{-3}$. Center: Tikhonov regularization, $\gamma = 10^{-3}$. Right: Picard criterion, $m = 25$.

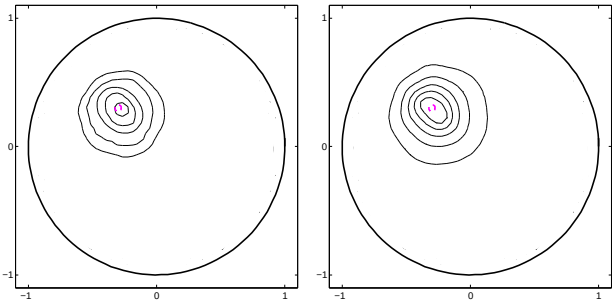


Figure 5.24: Test case V. Isolines of $C(\boldsymbol{\xi})$ in $[C_{\min}^{0.7}, C_{\max}]$. Left: Tikhonov regularization, $\gamma = 5$. Right: Picard criterion, $m = 25$.

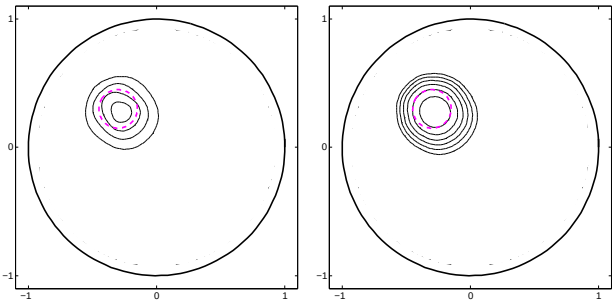


Figure 5.25: Test case VI. Isolines of $C(\boldsymbol{\xi})$ in $[C_{\min}^{0.7}, C_{\max}]$. Left: Tikhonov regularization, $\gamma = 10^{-1}$. Right: Picard criterion, $m = 25$.

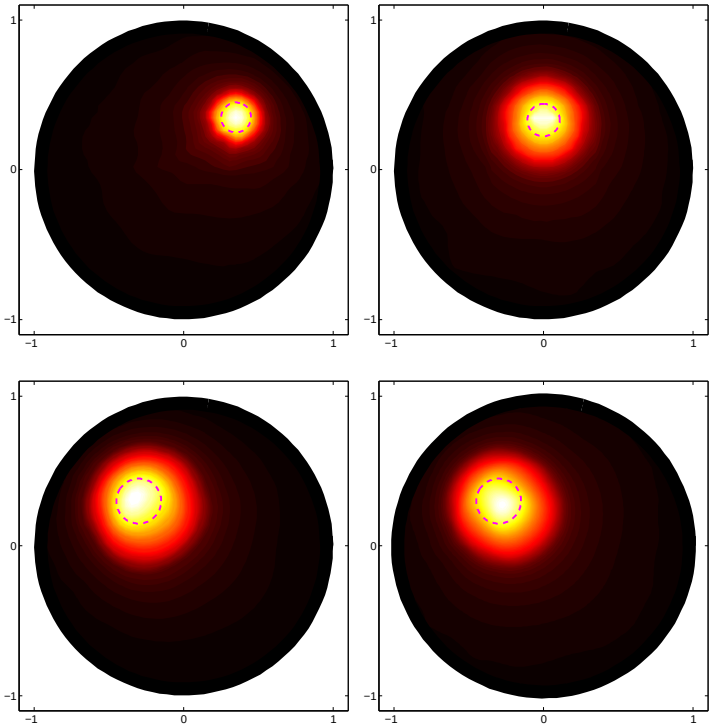


Figure 5.26: Tikhonov regularization. Test case I, $\gamma = 10^{-2}$ (top-left). Test case II, $\gamma = 5 \times 10^{-2}$ (top-right). Test case IIIa, $\gamma = 10^{-1}$ (bottom-left). Test case VI, $\gamma = 10^{-2}$ (bottom-right).

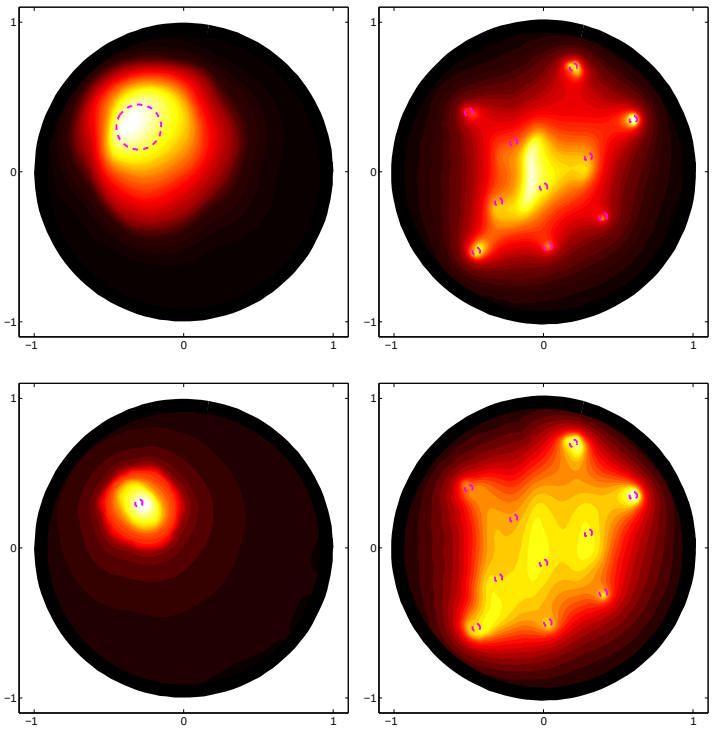


Figure 5.27: Tikhonov regularization. Test case IIIb, $\gamma = 5 \times 10^{-1}$ (top-left). Test case IVa, $\gamma = 10^{-3}$ (top-right). Test case V, $\gamma = 5$ (bottom-left). Test case IVb, $\gamma = 10^{-3}$ (bottom-right).

identical to those obtained without noise, as long as a sufficient number of the largest singular values are not obfuscated. The effect of noise is an attenuation of the values in the indicator function (5.2.26). This holds also for a large amount of noise, up to 10%.

Figs. 5.28, 5.29, 5.30, 5.31 and 5.32, show the results obtained for the test cases A,C,II,IIIa,IVa after adding 0.1% or 1% of noise. This corresponds to pick $\mu = 10^{-3}$ or $\mu = 10^{-2}$, respectively, in (5.5.2).

The indicator function (5.2.26) using PC is attenuated a lot faster, preventing the recovery of the inclusion also with small amounts of noise. Therefore the TR should be preferred to the PC in case of noisy data.

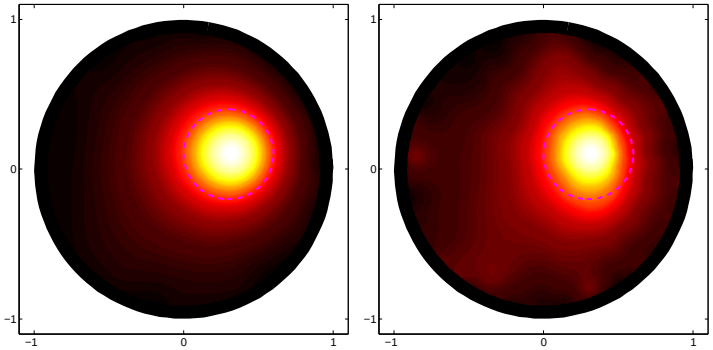


Figure 5.28: Test case A. Tikhonov regularization, $\gamma = 5 \times 10^{-3}$. Left: 0.1% noise. Right: 1% noise.

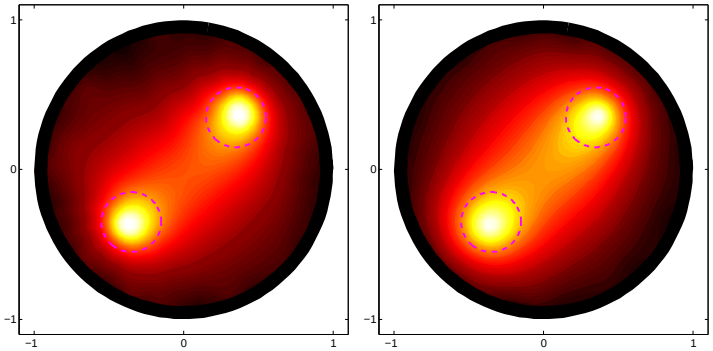


Figure 5.29: Test case C. Tikhonov regularization, $\gamma = 5 \times 10^{-4}$. Left: 0.1% noise. Right: 1% noise.

5.6 Numerical tests with one measurement in a random background

In this section we present some numerical tests with random backgrounds, in the arbitrary measurement and paired measurement cases. We are interested in large variations of the background coefficient σ_B, and therefore we parametrize it with the exponential model (5.3.3) so that the coefficient jumps by up to two orders of magnitude in each one of the r regions, or

$$\sigma_B(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{r} 0.5 \times 10^{y_i} \mathbb{I}_{R_i}(\mathbf{x}), \quad \mathbf{x} \in B, \quad \mathbf{y} \in \Gamma = [-0.5, 0.5]^r, \tag{5.6.1}$$

to have up to one order jumps.

Usually in the problem of EIT the boundary is accessible. This would allow to directly recover the value of the realization of the random variable Y_1 that generated the voltage and current

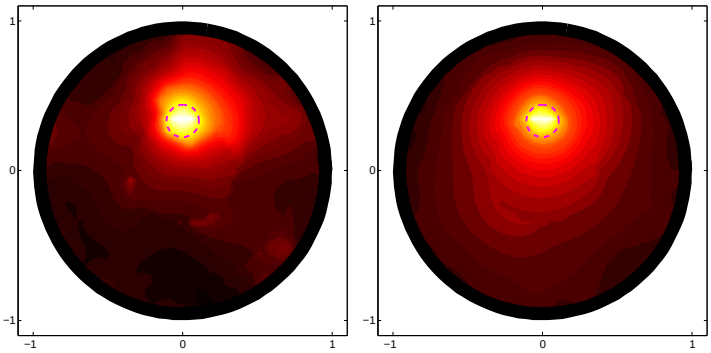


Figure 5.30: Test case II. Tikhonov regularization, $\gamma = 5 \times 10^{-2}$. Left: 0.1% noise. Right: 1% noise.

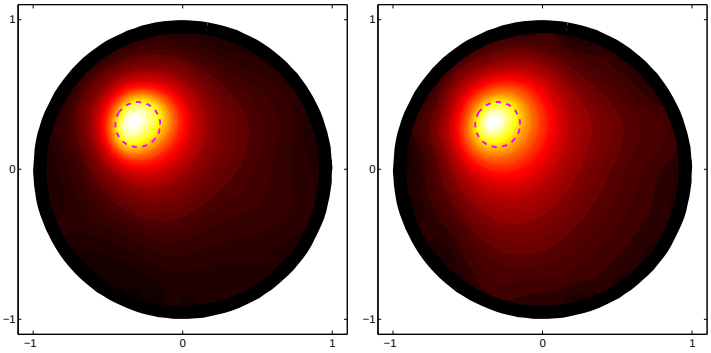


Figure 5.31: Test case IIIa. Tikhonov regularization, $\gamma = 10^{-1}$. Left: 0.1% noise. Right: 1% noise.

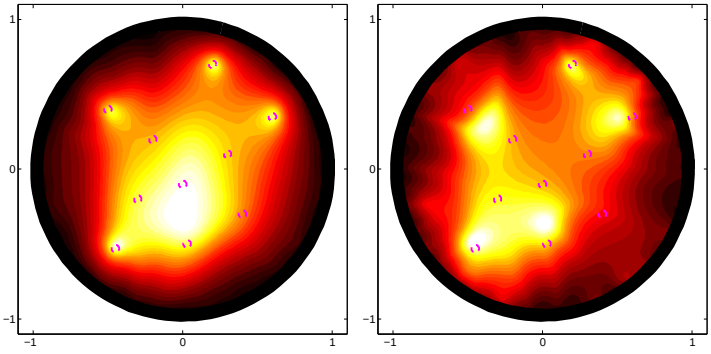


Figure 5.32: Test case IVa. Tikhonov regularization, $\gamma = 10^{-3}$. Left: 0.1% noise. Right: 1% noise.

measurement, corresponding to the observation of $\widetilde{\Lambda}$. As a consequence, the value of σ_B in R_1 could be retrieved, and the starting problem could be reduced to a problem with deterministic coefficient in R_1. For this reason, we will choose σ_B equal to 1 in R_1 and random only in $B \setminus R_1$. In this way, the parameter space Γ in the parametrizations (5.3.2), (5.3.3), (5.6.1) has to be dimensionally reduced by one, so that its dimension in practice becomes $q = r - 1$.

We proceed now to present several numerical tests performed with the optimized FM (Algorithm 2) in the case of one arbitrary measurement, and with the FM of Algorithm 4 in the case of one paired measurement.

5.6.1 The optimized Factorization Method with one arbitrary measurement

To show the capabilities of the proposed approach we set up three test cases aEN2, aES2 and aUN1, with an increasing difficulty in the optimization operations. These test cases feature a random background, and are presented in Table 5.3. The test case aEN2 is a standard configuration, with a not too small inclusion and a two-dimensional parameter space Γ. In aEN2 the measurement of $\Lambda(\boldsymbol{y})$ is generated by the realization $\boldsymbol{y} = (8.3567, 0.3558)$. The test case aES2 is more challenging than test case aEN2, because the inclusions have a smaller diameter and their presence is more easily obfuscated by the variations of the random background. The measurement of $\Lambda(\boldsymbol{y})$ in this test case is generated by the realization $\boldsymbol{y} = (4.5, 0.75)$. Lastly, in test case aUN1 we treat an ill-posed optimization problem over a five-dimensional parameter space, and the limitations of the optimization approach are highlighted. The measurement of $\Lambda(\boldsymbol{y})$ is generated by the realization $\boldsymbol{y} = (4.4950, 1.3450, 0.9450, 3.3450, 0.6240)$.

Example label	inclusion configuration	background configuration	parametrization of σ_B	Range(σ_B)	Figure with reconstruction
aEN2	N	E	(5.3.3)	$[0.1, 10]$	Fig. 5.33(top-left)
aES2	S	E	(5.3.3)	$[0.1, 10]$	Fig. 5.34(left)
aUN1	N	U	(5.6.1)	$[0.5, 5]$	Fig. 5.38(top-left)

Table 5.3: Description of the numerical test cases aEN2, aES2 and aUN1 with uncertain background in the case of arbitrary measurements.

From a computational standpoint, the optimization of the Ki Fan k-norm $\|\cdot\|_k$ can be equivalently replaced by the optimization of the real-valued operator $q_k(\cdot) = \sum_{i=1}^{k} \log \sigma_i(\cdot)$, which has the same minimum points as $\|\cdot\|_k$. In Figs. 5.33, 5.34 and 5.35 we display the function q_k evaluated over the parameter space Γ in the test cases aEN2 and aES2, for some values of k. As explained in Section 5.3, the solution to problem (5.3.9) that corresponds to the global minimum point is the realization $\boldsymbol{y}$ that generated the measurement of $\Lambda(\boldsymbol{y})$. The higher the value of k, and the steeper the objective function becomes, because more singular values are taken into account. Moreover, when k increases the objective function shows a product structure that suggests the use of Alternating Minimization, *i.e.* optimizing a given component of the variable while freezing the remaining components. When Γ is two-dimensional, we present the numerical results only using a global optimization approach, although we observed that the use of Alternating Minimization allows to reduce the overall computational cost.

When $\Gamma = [0.5, 5]^5$, the global optimization approach converges to the correct solution in the subspace of the parameter space associated with the three outer regions, but then gets stuck into local minima when exploring the two most inner regions. Alternating Minimization converges even more easily, but still gets stuck into local minima whenever the initial data of the two most inner regions is not sufficiently close to the exact solution. One possibility is to initialize with different initial data the optimization procedure, but since the dimension of the parameter space is moderately large this turns out to be very costly. Therefore, we think that

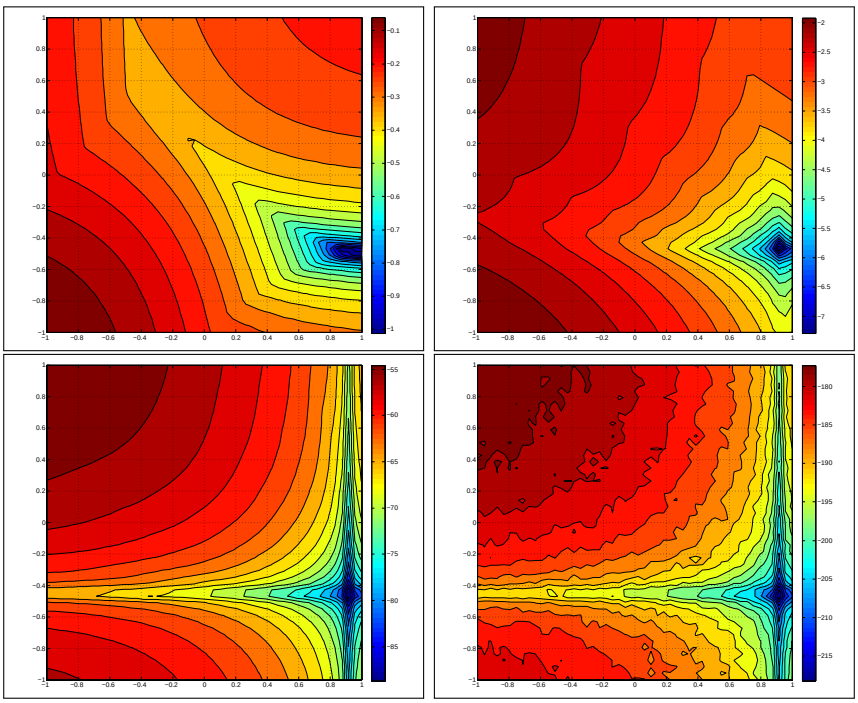


Figure 5.33: Test case aEN2. Function q_k over the parameter space Γ in log-log scale. Top-left: $k = 1$. Top-right: $k = 5$. Bottom-left: $k = 32$. Bottom-right: $k = 64$. The realization $\boldsymbol{y} = (8.3567, 0.3558)$ corresponds to the global minimum point $(0.9220, -0.4488)$ in log-log scale.

the optimization procedure could still be effectively employed but on problem (5.3.11) with more than one measurement of Λ and resizing the range of variation of the random variables associated with the most inner regions. For example, when Y_5, Y_6 deviate only 25% from their mean values the Alternating Minimization always easily finds one of the minimum points of problem (5.3.11). Some tests with many measurements are presented in Section 5.7.

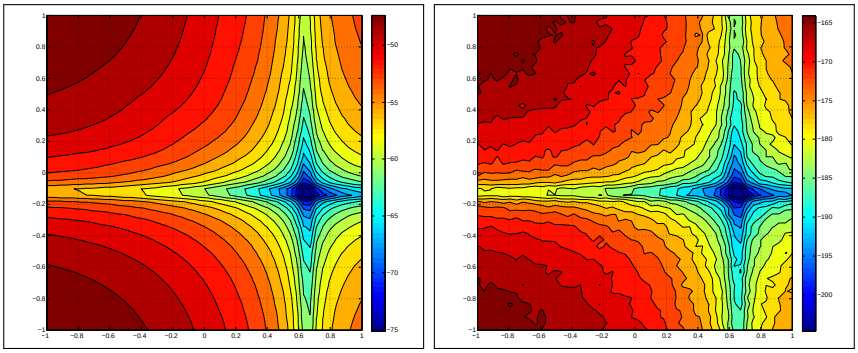


Figure 5.34: Test case aES2. Function q_k over the parameter space Γ in log-log scale. Left: $k = 32$. Right: $k = 64$. The realization $\boldsymbol{y} = (4.5, 0.75)$ corresponds to the global minimum point $(0.6532, -0.1249)$ in log-log scale.

To solve problem (5.3.9) we choose a derivative-free approach, since $\partial_{\boldsymbol{y}^*} \|\widetilde{\Lambda}(\boldsymbol{y}, \boldsymbol{y}^*)\|_k$ is not available. We employ the well-known Melder-Mead method [LRWW98] to perform a constrained optimization of the objective function q_k over the whole parameter space Γ.

In test case aEN2 the method converged in less than 100 iterations to the solution $\widehat{\boldsymbol{y}}$ of problem (5.3.9) within an accuracy $\varepsilon = 10^{-4}$ in (5.3.10), for several different choices of the initial point in the parameter space. The number of evaluations of Λ_0 has always been less than three times the number of iterations.

In test case aES2 the method converged in less than 100 iterations to the solution $\widehat{\boldsymbol{y}}$ of problem (5.3.9) with an accuracy $\varepsilon = 10^{-3}$, for several different choices of the initial point in the

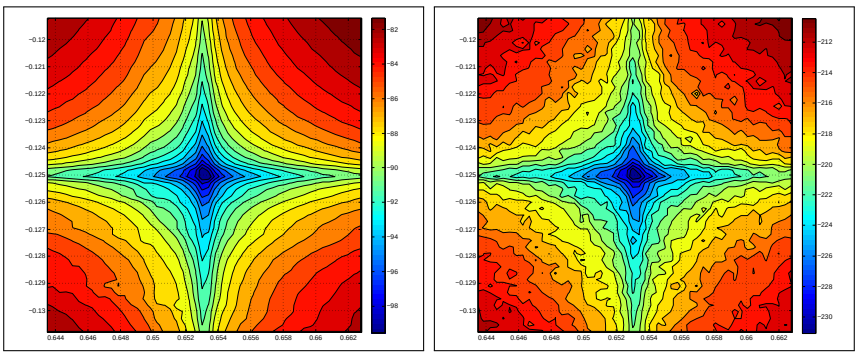


Figure 5.35: Test case aES2. Same as Fig. 5.34 but zoomed in a neighborhood of the global minimum point $(0.6532, -0.1249)$. Function q_k over the parameter space Γ in log-log scale. Left: $k = 32$. Right: $k = 64$.

parameter space.

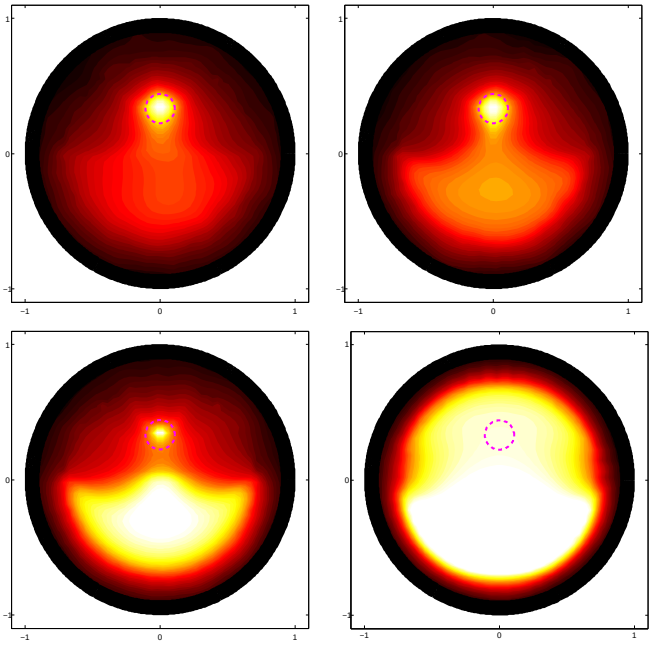


Figure 5.36: Coloring of $C(\boldsymbol{\xi})$, test case aEN2, $\gamma = 10^{-2}$ in (5.2.25). Top-left: using the solution $\widehat{\boldsymbol{y}}$ to problem (5.3.9) with $\varepsilon = 0.00014$. Top-right: using the solution $\widehat{\boldsymbol{y}}$ to problem (5.3.9) with $\varepsilon = 0.001$. Bottom-left: using the solution $\widehat{\boldsymbol{y}}$ to problem (5.3.9) with $\varepsilon = 0.01$. Bottom-right: using the point $\widehat{\boldsymbol{y}} = (5.05, 5.05)$.

Once we obtain an approximation $\widehat{\boldsymbol{y}}$ of the solution to problem (5.3.9), we can apply the FM with the realization $\widehat{\boldsymbol{y}}$. In Fig. 5.36 we show the reconstructions obtained in the test case aEN2, with different values of ε. We solve problem (5.3.9) with the Ky Fan 5-norm with initial point $(0.1, 10)$, and after 76 iterations and 223 evaluations of Λ_0 the method converges to the solution $\widehat{\boldsymbol{y}} = (8.356653, 0.355660)$ such that the convergence criterion (5.3.10) is satisfied with $\varepsilon = 0.00014$. The corresponding reconstruction obtained is depicted in Fig. 5.36-top-left. The other Figs. 5.36-top-right and 5.36-bottom-left show the effect of using a less accurate realization $\widehat{\boldsymbol{y}}$ for different values of ε. Fig. 5.36-bottom-right is obtained using $\widehat{\boldsymbol{y}} = (5.05, 5.05)$, that is the expected value of the random variable $\boldsymbol{Y}$ over Γ, and yields a reconstruction where the inclusion is completely obfuscated by the contrast of the background.

As justified in Section 5.3, when applying the FM with a realization $\widehat{\boldsymbol{y}}$ that is not accurate enough with respect to the criterion (5.3.10), the background begins to be detected in the reconstruction. This fact can be used to check a posteriori that the solution to problem (5.3.9) is accurate enough: if the region R_i of the background is sharply detected, then the ith component

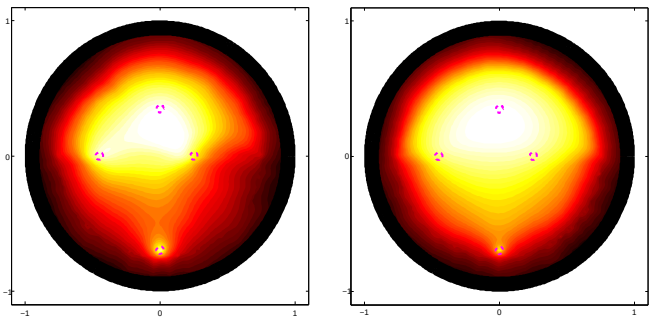


Figure 5.37: Coloring of $C(\boldsymbol{\xi})$, test case aES2, $\gamma = 10^{-2}$ in (5.2.25). Left: using the solution $\widehat{\boldsymbol{y}}$ to problem (5.3.9) with $\varepsilon = 10^{-3}$. Right: using the solution $\widehat{\boldsymbol{y}}$ to problem (5.3.9) with $\varepsilon = 10^{-2}$.

of the solution $\widehat{\boldsymbol{y}}$ to problem (5.3.9) is too inaccurate, and its accuracy needs to be improved. To achieve an improvement one can either utilize Alternating Minimization, or increase k or choose other types of optimization methods, *e.g.* Simulated Annealing or Global Pattern Search.

In Fig 5.37-left we show the reconstruction obtained in test case aES2 using the solution $\widehat{\boldsymbol{y}}$ to problem (5.3.9) that satisfies (5.3.10) with $\varepsilon = 10^{-3}$. This solution was computed solving problem (5.3.9) with the Ky Fan 32-norm and choosing the initial point $(0.1, 10)$. The optimization required 131 iterations and 312 evaluations of Λ_0.

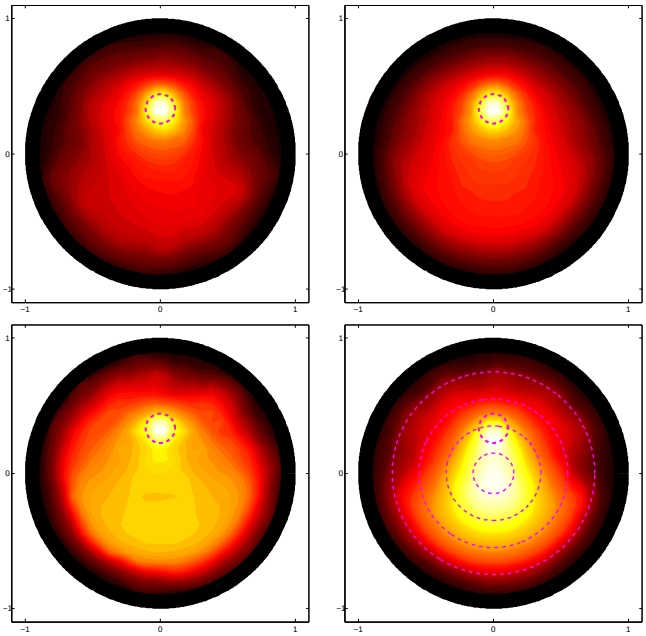


Figure 5.38: Coloring of $C(\boldsymbol{\xi})$, test case aUN1, $\gamma = 10^{-2}$ in (5.2.25). Top-left: using the solution $\widehat{\boldsymbol{y}}$ to problem (5.3.9) with $\varepsilon = 0.0005$. Top-right: using the solution $\widehat{\boldsymbol{y}}$ to problem (5.3.9) with $\varepsilon = 0.001$. Bottom-left: using the solution $\widehat{\boldsymbol{y}}$ to problem (5.3.9) with $\varepsilon = 0.01$. Bottom-right: using the solution $\widehat{\boldsymbol{y}}$ to problem (5.3.9) with $\varepsilon = 0.01$ in R_2, R_3, R_4 and $\varepsilon = 0.2$ in R_5, R_6.

Finally in Fig. 5.38 we present the results obtained in test case aUN1 for different values of ε. In Fig. 5.38-top-left we show the reconstruction obtained when solving accurately problem (5.3.9) with $\varepsilon = 5 \times 10^{-4}$, in Fig. 5.38-bottom-right we display the reconstruction obtained when the optimization of problem (5.3.9) is accurate with $\varepsilon = 10^{-2}$ only in the three most outer regions, but converged only up to $\varepsilon = 0.2$ in the subspace of Γ associated with the two most inner regions. As a result, the two most inner regions are detected as an inclusion, and the real inclusion is obfuscated.

In general, we observed that the global optimization over the whole five-dimensional parameter space is always successful, when applied to test cases featuring small variations of the coefficient in the most inner regions, *e.g.* up to 25%. On the other hand, when the starting point is not

sufficiently close to the exact solution and the variations of the coefficient are of one order of magnitude in the two most inner regions, the global optimization over the five-dimensional parameter space got stuck into local minima. Clearly, this is a limitation of the proposed approach, and the availability of more than one measurement to facilitate the optimization seems required to overcome it.

5.6.2 The Factorization Method with one paired measurement

In the case of paired measurements, the inclusion is always easily detected in the configurations EN and UN , also when the background coefficient jumps by two orders of magnitude. Therefore we will not present these results, but focus only on the more challenging configurations ES and US, again with two orders of magnitude jumps in the background coefficient. We consider the test cases pES and pUS presented in Table 5.4, featuring uncertain background in the case of paired measurements. The results obtained are presented in Fig. 5.39 for the test case pES, and in Fig. 5.40 for the test case pUS.

Example label	inclusion configuration	background configuration	parametrization of σ_B	Range(σ_B)	Figure with reconstruction
pES	S	E	(5.3.3)	$[0.1, 10]$	Fig. 5.39
pUS	S	U	(5.3.3)	$[0.1, 10]$	Fig. 5.40

Table 5.4: Description of the numerical test cases pES and pUS with uncertain background in the case of paired measurements.

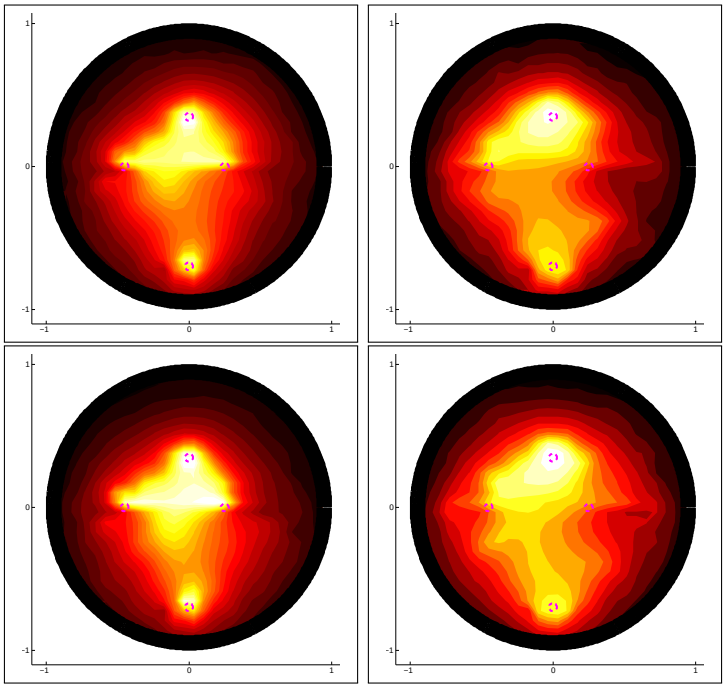


Figure 5.39: Coloring of $\widehat{C}(\boldsymbol{\xi})$ with $T = 10$, test case pES, $\gamma = 5 \times 10^{-2}$ in (5.2.25). Top-left: $\sigma_B = (1, 0.5, 5)$. Top-right: $\sigma_B = (1, 5, 0.5)$. Bottom-left: $\sigma_B = (1, 0.1, 10)$. Bottom-right: $\sigma_B = (1, 10, 0.1)$.

We use the indicator $\widehat{C}(\boldsymbol{\xi})$ in (5.3.14) with the number T of realizations chosen according to the number of regions r. We choose $T = 10$ when $r = 3$ and $T = 20$ when $r = 6$. Then we apply the FM when observing only one measurement of the random operator $\widetilde{\Lambda}$, generated from

the unknown realization $\boldsymbol{y}$. The results are presented when "common" realizations are observed, *e.g.* $\sigma_B = (1, 5, 0.5)$ or $\sigma_B = (1, 0.26, 4.17, 0.89, 7.05, 0.16)$, and when "extremal" realizations are observed as well. We distinguish between "common" and "extremal" realizations in the following sense: an extremal realization lies on the boundary of the support of the random variable, *e.g.* $\sigma_B = (1, 10, 0.1)$ and $\sigma_B = (1, 10, 0.1, 0.1, 10, 10)$. If a realization is not extremal, than it is a "common" realization. Extremal realizations provide a worst-case benchmark for the FM.

With all the realizations observed, the FM detects at least three inclusions over four. The inclusion closest to the center of the domain is the hardest to detect. The reconstructions rely on a single realization of the random operator $\widetilde{\Lambda}$, and this justifies their slight dependence on the particular realization observed. In spite of this, at least three inclusions are always sharply detected, but the most inner is occasionally undetected.

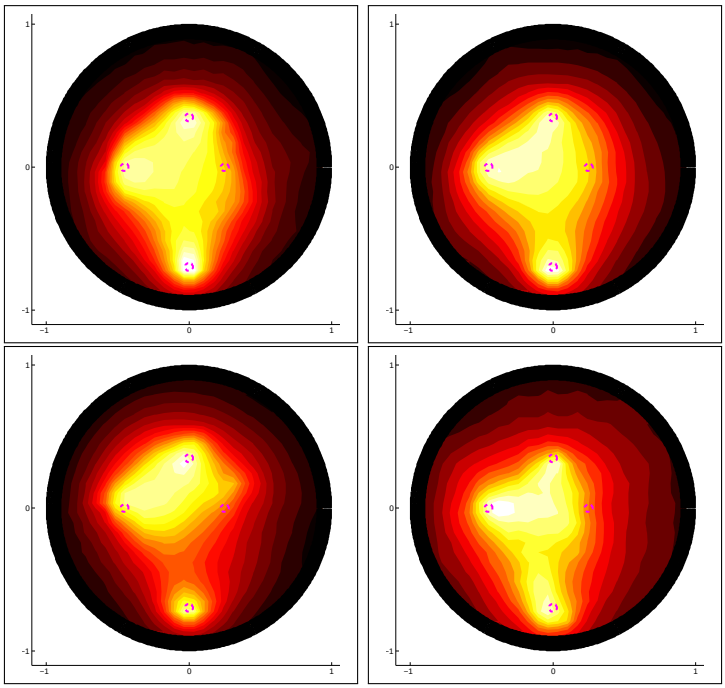


Figure 5.40: Coloring of $\widehat{C}(\boldsymbol{\xi})$ with $T = 20$, test case pUS, $\gamma = 5 \times 10^{-2}$ in (5.2.25). Top-left: $\sigma_B = (1, 0.26, 4.17, 0.89, 7.05, 0.16)$. Top-right: $\sigma_B = (1, 10, 0.1, 0.1, 10, 10)$. Bottom-left: $\sigma_B = (1, 10, 1, 0.1, 1, 0.1)$. Bottom-right: $\sigma_B = (1, 0.1, 0.1, 10, 0.1, 10)$.

5.7 Numerical tests with many measurements in a random background

5.7.1 The optimized Factorization Method with many arbitrary measurements

As described previously, the optimization in problem (5.3.9) can be facilitated, optimizing problem (5.3.11) where additional measurements of $\Lambda(\boldsymbol{Y})$ are included. Fig. 5.41 shows the evaluation over Γ of the objective function of problem (5.3.11) with $\|\cdot\|_k$ replaced by $q_k(\cdot)$, in the test case aEN2 and when four additional measurements are added to the previous measurement of Λ.

5.7.2 The pure Factorization Method with many arbitrary measurements

In this section we discuss some details about the application of the pure FM. The critical issue concerns the choice of ε. In Figs. 5.36-top-right, 5.37-left, 5.38-top-right several accurate reconstructions are obtained with $\varepsilon = 10^{-3}$ and large variations in the coefficient up to two orders of magnitude. Based on this observation, and on many other numerical experiments that we do

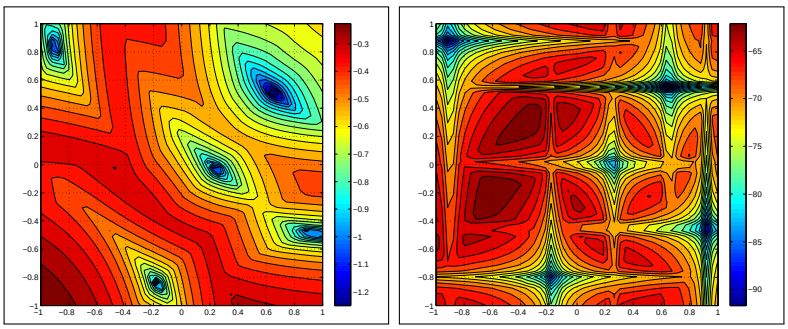


Figure 5.41: Test case aEN2 with 5 measurements of $\Lambda(\boldsymbol{Y})$. Evaluation over Γ in log-log scale of the objective function of problem (5.3.11), with $\|\cdot\|_k$ replaced by q_k. The realizations $\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3, \boldsymbol{y}_4, \boldsymbol{y}_5$ correspond to the five minimum points. Left: $k = 1$. Right: $k = 32$.

not report, an accuracy $\varepsilon = 10^{-3}$ is always enough to clearly detect inclusions of normal size, *e.g.* with radius 0.1, when there are up to $r = 4$ regions, d is not larger than 0.125 and the vector $\mathbf{m}$ does not produce jumps in the coefficient larger than one order of magnitude. When interested in smaller inclusions, with more regions $r > 4$, or d and $\mathbf{m}$ that produce larger variations in σ_B, then also the value of ε has to be decreased further.

Table 5.5 reports the number T of realizations needed to satisfy (5.3.13), for some values of M, q, d, $\varepsilon = 10^{-3}$, σ_B parametrized by (5.3.2) and d up to 0.125, yielding 25% variations in the coefficient. Notice that this variation is much smaller than those considered in Figs. 5.36, 5.37, 5.38.

In Table 5.6 we report the value of T such that a condition analogous to (5.3.13) holds, but with the exponential parametrizations (5.3.3) and (5.6.1). The ranges of variation achieved with these parametrizations are the same as those in Figs. 5.36, 5.37, 5.38.

	$d = 0.025$	$d = 0.05$	$d = 0.1$	$d = 0.125$
$M = 3$, $q = 2$ $(r = 3)$	1300	4500	15800	-
$M = 10$, $q = 2$ $(r = 3)$	400	1200	4500	7200
$M = 10$, $q = 3$ $(r = 4)$	9000	60000	-	-
$M = 25$, $q = 2$ $(r = 3)$	200	600	1900	2800
$M = 50$, $q = 2$ $(r = 3)$	100	300	1000	1600
$M = 50$, $q = 3$ $(r = 4)$	2000	14000	63000	133000

Table 5.5: Uncertainty analysis. The value of T such that $Pr(\min_{mt} \|\boldsymbol{y}_m - \widehat{\boldsymbol{y}}_t\|_\infty \leq \varepsilon\, d^{-1}) \geq 0.99$, given $\varepsilon = 10^{-3}$ and some values of q, M, d. Parametrization of σ_B by (5.3.2), where $\boldsymbol{y}_m, \widehat{\boldsymbol{y}}_t \in \Gamma = [-1, 1]^q$ and $\mathbf{Y} \sim \mathcal{U}(\Gamma)$. The dash advises for a larger value of M.

	$[0.5, 5]$	$[0.1, 10]$
$M = 100$, $q = 1$ $(r = 2)$	40	70
$M = 100$, $q = 2$ $(r = 2)$	10800	36400
$M = 1000$, $q = 1$ $(r = 2)$	10	30
$M = 1000$, $q = 2$ $(r = 3)$	1300	4500

Table 5.6: Large variations. The value of T such that $Pr(\min_{mt} \|\boldsymbol{y}_m - \widehat{\boldsymbol{y}}_t\|_\infty \leq \varepsilon) \geq 0.99$, given $\varepsilon = 10^{-3}$ and some values of q, M. Left column: one order of magnitude variations with the parametrization (5.6.1) of σ_B, where $\boldsymbol{y}_m, \widehat{\boldsymbol{y}}_t \in \Gamma = [0.5, 5]^q$ and $\mathbf{Y} \sim \mathcal{U}(\Gamma)$. Right column: two orders of magnitude variations with the parametrization (5.3.3) of σ_B, where $\boldsymbol{y}_m, \widehat{\boldsymbol{y}}_t \in \Gamma = [0.1, 10]^q$ and $\mathbf{Y} \sim \mathcal{U}(\Gamma)$.

We remark that the value of T corresponds to the number of evaluations of the operator Λ_0 in the parameter space Γ. These evaluations can be efficiently performed by means of polynomial approximation techniques, *e.g.* the random discrete L^2 projection presented in Chapters 1, 2, 4 and 6.

5.8 Conclusions

In this paper we provided a numerical scheme to solve the Neumann dipole-like boundary value problem in inhomogeneous deterministic background. Then we employed this scheme to apply the FM to the Continuous Model for EIT. We focused on two types of piecewise constant background diffusion coefficient: inhomogeneous deterministic and uncertain. With both types of background the FM is able to detect the presence and location of one or many inclusions, and to supply rough information on their geometry. The quality of the reconstructions obtained in the inhomogeneous deterministic case is comparable to the quality of those in the homogeneous case, with jumps in the background up to two orders of magnitude.

When the background is uncertain we proposed three variants of the FM, based on the distinction between arbitrary and paired measurements of the random operator Λ. In the paired measurement case the FM always detects a normal or small inclusion, using only one paired measurement of the operators Λ and Λ_0. In the presence of many inclusions those harder to detect could remain undetected. In the case of arbitrary measurements, we devised an optimization scheme to recover the realization of the background diffusion coefficient that generated a given measurement of the operator Λ. With this approach the FM can be applied also to situations where only one measurement of Λ is available. The quality of the reconstruction depends on the accuracy ε to which the optimization can be performed. We focused mainly on large variations of the background coefficient up to two orders of magnitude, and showed that the optimization can be performed up to a tolerance ε such that the FM is successful when the parameter space is low-dimensional. When the dimension of the parameter space increases the optimization becomes challenging, in particular with large variations of the background coefficient. Additional measurements of the operator Λ can be easily incorporated in the objective function to facilitate the optimization procedure, but in moderately large-dimensional parameter spaces, the applicability of the proposed optimization approach remains confined to uncertainty analysis. Lastly we proposed a pure variant of the FM, that does not require the optimization step and is suitable for uncertainty analysis when many measurements of Λ are available.

Chapter 6

Acceleration of the Factorization Method by means of polynomial approximation

The present chapter links Chapter 5 with the first part of the thesis. In Chapter 5 the problem of Electrical Impedance Tomography in uncertain background is presented as an identification problem, where the aim is to determine the presence and location of unknown inclusions. The case of piecewise constant background affected by uncertainty is addressed, and the diffusion coefficient is modeled by means of random variables. As in Chapter 5, we confine to random variables $\mathbf{Y}$ distributed according to the uniform density $\rho = \mathcal{U}(\Gamma)$. For example, the background diffusion coefficient σ_B has been parametrized by the exponential model

$$\sigma_B(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{r} 10^{y_i} \mathbb{I}_{R_i}(\mathbf{x}), \quad \mathbf{x} \in B, \quad \mathbf{y} \in \Gamma = [-1, 1]^r, \tag{6.0.1}$$

so that the coefficient jumps up to two orders of magnitude in each one of the r regions that form a partition of the computational domain $B = \bigcup_{i=1}^{r} R_i$, see *e.g.* Fig. 5.3. Two different situations are distinguished, depending on how the measurements of the random operators Λ and Λ_0 are collected.

In the case of arbitrary measurements, two variants of the Factorization Method are presented: the optimized Factorization Method (Algorithm 2) and the pure Factorization Method (Algorithm 3). These variants are both based on the repetitive evaluation of the Neumann-to-Dirichlet map

$$\mathbf{y} \mapsto \Lambda_0(\mathbf{y}), \qquad \mathbf{y} \in \Gamma, \tag{6.0.2}$$

which demands for a repetitive calculation of the solution to the PDE model (5.2.2) with the random diffusion coefficient σ_B introduced in (5.3.1). We recall here problem (5.2.2) and the definition of the map Λ_0. Given $\eta \in \mathring{H}^{-\frac{1}{2}}(\partial B)$, consider the $\mathbf{y}$-parametric Neumann boundary-value problem,

$$\text{find } u \in \mathring{H}^1(B) \text{ s.t. } \begin{cases} \nabla \cdot (\sigma_B(\mathbf{x}, \mathbf{y}) \nabla u(\mathbf{x})) = 0, & \mathbf{x} \in B, \quad \mathbf{y} \in \Gamma, \\ \sigma(\mathbf{x}, \mathbf{y}) \nabla u(\mathbf{x}) \cdot \boldsymbol{\nu}(\mathbf{x}) = \eta(\mathbf{x}), & \mathbf{x} \text{ on } \partial B, \quad \mathbf{y} \in \Gamma, \end{cases} \tag{6.0.3}$$

where $\boldsymbol{\nu}$ denotes the outward normal to ∂B. Problem (5.2.2) defines the Neumann-to-Dirichlet map $\Lambda_0 : \eta \mapsto u|_{\partial B}$, in the same way as in (5.2.3).

The case of paired measurements differs in how the measurements of Λ and Λ_0 are collected. In this case, the variant of the Factorization Method proposed (Algorithm 4) strongly relies on the sampling procedure, and the Monte Carlo-type estimator (5.3.14) converges with the well-known low convergence rate with respect to the number T of random samples. For a given sample point $\mathbf{y}$, the evaluation of the estimator (5.3.14) demands for the calculation of the

solution φ_N^k to problem (5.2.15), with the random diffusion coefficient σ_B introduced in (5.3.1). This corresponds to evaluate the following map:

$$\mathbf{y} \mapsto \varphi_N^k(\mathbf{y}), \qquad \mathbf{y} \in \Gamma, \tag{6.0.4}$$

for any realization $\mathbf{y} \in \Gamma$. We remark that, although a singular point source is present in problem (5.2.5), under Assumption 5.1 the forcing term in problem (5.2.15) is a smooth function, as explained at the end of Section 5.2.3. Therefore, polynomial approximation is the right tool to approximate the solution φ_N^k to problem (5.2.15).

The outline of this chapter is the following. In Section 6.1 we recall the results developed in Chapter 1 and Chapter 2 concerning the polynomial approximation of random functions and of the solution to elliptic PDE models with stochastic coefficients. Then the same polynomial approximation techniques are applied to the framework of EIT introduced in Chapter 5 and, more specifically, to accelerate the parametric evaluation of the Neumann-to-Dirichlet map (6.0.2) that is required to apply the Factorization Method in the case of arbitrary measurements. In Section 6.2 we consider the polynomial approximation of the solution to the Neumann problem (6.0.3). Then in Section 6.3 we consider the polynomial approximation of the Neumann-to-Dirichlet map Λ_0 defined in (6.0.2). Section 6.4 presents a low computational cost approach based on the polynomial approximation of the Ky-Fan k-norm. Section 6.5 presents some details about the polynomial approximation of the map (6.0.4), in the case of paired measurements.

6.1 Convergence estimates for the least squares approximation

Denoting by u the solution to model (5.2.2) or (5.2.15), both with the random diffusion coefficient σ_B introduced in (5.3.1), then the following results on the polynomial approximation of u hold. Note that the models (5.2.2) and (5.2.15) fall in the class of "inclusion type" PDE models with stochastic diffusion coefficient that has been introduced in Section 2.4.2, and that the Total Degree polynomial space has been shown in Chapter 2 to be the right polynomial space to use in this case.

In this section we use the same notation for the projections P_Λ and P_Λ^M that have been introduced in (2.4.7), as well as the truncated projection $\widetilde{P}_\Lambda^M$ by the truncation operator defined in (2.4.5), when needed.

Consider a Hilbert space V over the physical domain B. As in Section 2.4.1, we introduce the tensor space $\mathcal{V} = L^2(\Gamma, d\rho) \otimes V$, equipped with the inner product (2.4.3), and the space $\mathcal{W} = L^\infty(\Gamma, V)$ of bounded V-valued functions. Throughout this chapter, we will set $V = H^1(B)$. Moreover, we will use the constants c_1, c_2, c_3, c_4 defined in (2.4.17), (2.4.10), (2.4.11), (2.4.18).

Consider the parametrization (2.4.2) of the diffusion coefficient $a(x, y)$ of problem (2.4.1). In the general case of an infinite expansion of the coefficient with overlapping basis functions ψ_j, an algebraic convergence rate of the best N-term polynomial approximation has been proven.

Theorem 6.1. *If the coefficient $a(x, y)$ of problem (2.4.1) satisfies Assumption 2.2, and if the sequence $(\|\psi_j\|_{L^\infty(D)})_{j\geq 0}$ belongs to $\ell^p(\mathbb{N})$ for some $p < 1$ then the best N-term approximation $P_\Lambda u$ of u converges with the algebraic rate*

$$\|u - P_\Lambda u\|_{\mathcal{W}} \leq CN^{-s}, \quad s = \frac{1}{p} - 1. \tag{6.1.1}$$

Proof. See [CDS11, Theorem 1.2]. □

In the parametrizations (6.0.1) (or (5.3.2)) of the background diffusion coefficient σ_B of problems (5.2.2) or (5.2.15), the support of the inclusions do not overlap, and this is the situation that has been specifically addressed in our analysis in Chapter 2. In this case exponential convergence has been proven.

Lemma 2.6. *In the isotropic case,* i.e. $g_n = g$ *for all* $n = 1, \dots, d$*, the following estimate on the error of the* L^2 *projection* P_Λ *on the quasi-optimal TD sets holds*

$$\|u - P_\Lambda u\|_{\mathcal{V}}^2 \leq c_3 \exp\left\{-c_2 N^{\frac{1}{d}}\right\}. \tag{6.1.2}$$

Theorem 2.6. *In the aforementioned PDE model class, the convergence rate of the least squares approximation with optimal choice of the polynomial space satisfies*

$$\mathbb{E}\Big[\|u - \widetilde{P}_\Lambda^M u\|_{\mathcal{V}}^2\Big] \leq (c_4 + c_1)\, \exp\left\{-\left(c_2^{2d}\, \beta\, M\right)^{\frac{1}{2d+1}}\right\}. \tag{6.1.3}$$

Denote by $\widetilde{C}_{\text{inv}}^{\rho}$ the constant of the inverse inequalities between $L^\infty(V)$ and $L^2_\rho \otimes V$ introduced in (B.2.2) in Appendix B.2,

$$\widetilde{C}_{\text{inv}}^{\rho}(\Lambda) = \sup_{v \in \mathbb{P}_\Lambda \otimes V} \frac{\|v\|_{L^\infty(V)}}{\|v\|_{L^2_\rho \otimes V}} < +\infty. \tag{6.1.4}$$

Notice that $\widetilde{C}_{\text{inv}}^{\rho}$ is bounded when ρ is the uniform density, *e.g.* $\widetilde{C}_{\text{inv}}^{\rho}$ can be bounded by $\widetilde{C}_{\text{inv}}^{\rho}(\text{TD}) \leq \#\text{TD}$ in the case of the Total Degree space. In the sequel we will extend the estimate (6.1.3) in norm $\mathcal{V} = L^2_\rho \otimes V$ to the norm $\mathcal{W} = L^\infty(\Gamma, V)$, since the final aim is to obtain a pointwise estimate of the error when approximating the parametric map $\Lambda_0 = \Lambda_0(\mathbf{y})$ for any $\mathbf{y} \in \Gamma$.

Moreover, we introduce the same random variable C^ω as in (1.2.15), but adapted to the space $\mathcal{V}$ and using the discrete norm defined in (2.4.8):

$$C^\omega(M, \Lambda) := \sup_{v \in \{\mathbb{P}_\Lambda \setminus \{v \equiv 0\}\} \otimes V} \frac{\|v\|_{\mathcal{V}}^2}{\left(\|v\|_M^V\right)^2}. \tag{6.1.5}$$

Lemma 6.1. *With* $C^\omega(M, \Lambda)$ *defined as in* (6.1.5)*, it holds that*

$$\|u - P_\Lambda^M u\|_{\mathcal{W}} \leq \left(1 + \widetilde{C}_{inv}^{\rho} \sqrt{C^\omega(M, \Lambda)}\right) \inf_{v \in \mathbb{P}_\Lambda \otimes V} \|u - v\|_{\mathcal{W}}. \tag{6.1.6}$$

Proof. For any $v \in \mathbb{P}_\Lambda \otimes V$ it holds

$$\begin{aligned}
\|u - P_\Lambda^M u\|_{\mathcal{W}} \leq & \|u - v\|_{\mathcal{W}} + \|v - P_\Lambda^M u\|_{\mathcal{W}} \\
\leq & \|u - v\|_{\mathcal{W}} + \widetilde{C}_{\text{inv}}^{\rho} \|v - P_\Lambda^M u\|_{\mathcal{V}} \\
\leq & \|u - v\|_{\mathcal{W}} + \widetilde{C}_{\text{inv}}^{\rho} \sqrt{C^\omega(M, \Lambda)}\, \|v - P_\Lambda^M u\|_M^V \\
\leq & \left(1 + \widetilde{C}_{\text{inv}}^{\rho} \sqrt{C^\omega(M, \Lambda)}\right) \|u - v\|_{\mathcal{W}}.
\end{aligned}$$

Since $v \in \mathbb{P}_\Lambda \otimes V$ is arbitrary, we deduce (6.1.6). □

We now state the final result that will be needed in the following sections.

Corollary 1 (of Theorem 2.3)**.** *For any* $\gamma > 0$*, if* M *is such that* $K(\mathbb{P}_\Lambda)$ *satisfies* (2.2.6)*, then for any* $u \in \mathcal{W}$ *it holds*

$$Pr\left(\|u - P_\Lambda^M u\|_{\mathcal{W}} \leq \left(1 + \sqrt{2}\, \widetilde{C}_{inv}^{\rho}\right) \inf_{v \in \mathbb{P}_\Lambda \otimes V} \|u - v\|_{\mathcal{W}}\right) \geq Pr\left(\Omega_+^M\right) \geq 1 - 2M^{-\gamma}. \tag{6.1.7}$$

Proof. Same as in the proof of Theorem 2.3 but with the $L^2_\rho(\Gamma)$ norm replaced by the $\mathcal{W}$ norm, and with the constant of the inverse inequality $\widetilde{C}_{\text{inv}}^{\rho}$ between L^∞ and L^2_ρ. The switch from the $\mathcal{W}$ norm to the $\mathcal{V}$ norm is by inverse inequality, as in the proof of Lemma 6.1. □

Theorem 6.2. *For any $\gamma > 0$, if M is such that $K(\mathbb{P}_\Lambda)$ satisfies (2.2.6), then for any $u \in \mathcal{W}, \|u\|_{\mathcal{W}} \leq \tau$, it holds*

$$\mathbb{E}\left[\|u - \widetilde{P}_\Lambda^M u\|_{\mathcal{W}}^2\right] \leq \left(1 + \sqrt{2}\,\widetilde{C}_{inv}^{\rho}\right)^2 \inf_{v \in \mathbb{P}_\Lambda \otimes V} \|u - v\|_{\mathcal{W}}^2 + (\tau + \|u\|_{\mathcal{W}})^2\, 2M^{-\gamma}. \tag{6.1.8}$$

Proof. First we note that

$$\|u - \widetilde{P}_\Lambda^M u\|_{\mathcal{W}} \leq \|u - P_\Lambda^M u\|_{\mathcal{W}}, \quad \forall u \in \mathcal{W} : \|u\|_{\mathcal{W}} \leq \tau,$$

where $\widetilde{P}_\Lambda^M$ is the same truncation operator defined in (2.4.5). Thanks to (6.1.7) we obtain

$$\begin{aligned}
\mathbb{E}\left[\|u - \widetilde{P}_\Lambda^M u\|_{\mathcal{W}}^2\right] \leq & \mathbb{E}\Big[\|u - \widetilde{P}_\Lambda^M u\|_{\mathcal{W}}^2 \mathbb{I}_{\Omega_+^M}\Big] + \mathbb{E}\Big[\|u - \widetilde{P}_\Lambda^M u\|_{\mathcal{W}}^2 \mathbb{I}_{\Omega_-^M}\Big] \\
\leq & \Big(1 + \sqrt{2}\,\widetilde{C}_{\text{inv}}^{\rho}\Big)^2 \inf_{v \in \mathbb{P}_\Lambda \otimes V} \|u - v\|_{\mathcal{W}}^2 \Pr(\Omega_+^M) + (\tau + \|u\|_{\mathcal{W}})^2 \Pr(\Omega_-^M) \\
\leq & \Big(1 + \sqrt{2}\,\widetilde{C}_{\text{inv}}^{\rho}\Big)^2 \inf_{v \in \mathbb{P}_\Lambda \otimes V} \|u - v\|_{\mathcal{W}}^2 + (\tau + \|u\|_{\mathcal{W}})^2\, 2M^{-\gamma}
\end{aligned}$$

and the proof is completed. □

6.2 Polynomial approximation of the solution to the Neumann problem

In this section we consider the polynomial approximation of the solution to the Neumann problem (6.0.3). We introduce the Hilbert space $L_\rho^2(\Gamma) \otimes \mathring{L}^2(\partial B)$ endowed with the norm

$$\|u\|_{L_\rho^2(\Gamma)\otimes\mathring{L}^2(\partial B)}^2 = \int_\Gamma \|u(\mathbf{y})\|_{\mathring{L}^2(\partial B)}^2\, d\mathbf{y}. \tag{6.2.1}$$

The discrete norm to approximate (6.2.1) is given by

$$\|u\|_{M,\mathring{L}^2(\partial B)}^2 = \frac{1}{M} \sum_{m=1}^{M} \|u(\mathbf{y}_m)\|_{\mathring{L}^2(\partial B)}^2. \tag{6.2.2}$$

At the discretization level the space $\mathring{L}^2(\partial B)$ is discretized with the Fourier basis $\mathbb{F}_K(\partial B) = \{\eta_k\}_{k=1}^K$ with zero mean, and $L_\rho^2(\Gamma)$ is discretized with the usual polynomial space $\mathbb{P}_\Lambda$. An L^2 projection on the tensor space $\mathbb{P}_\Lambda(\Gamma) \otimes \mathbb{F}_K(\partial B)$ would lead to the following problem

$$\Pi^{\mathbb{P}_\Lambda \otimes \mathbb{F}_K} u = \operatorname*{argmin}_{v \in \mathbb{P}_\Lambda(\Gamma)\otimes\mathbb{F}_K(\partial B)} \|u - v\|_{L_\rho^2(\Gamma)\otimes\mathring{L}^2(\partial B)}, \tag{6.2.3}$$

such that the solution $u \in L_\rho^2(\Gamma)\otimes\mathring{L}^2(\partial B)$ to problem (5.2.2) with the random diffusion coefficient (5.2.1) is approximated in the subspace $\mathbb{P}_\Lambda(\Gamma) \otimes \mathbb{F}_K(\partial B) \subseteq L_\rho^2(\Gamma) \otimes \mathring{L}^2(\partial B)$. In this way

$$\Pi^{\mathbb{P}_\Lambda \otimes \mathbb{F}_K} u = \sum_{j=1}^{\#\Lambda} \sum_{k=1}^{K} u_{jk}\, \eta_k(\mathbf{x})\, l_j(\mathbf{y}), \quad \mathbf{x} \text{ on } \partial B, \quad \mathbf{y} \in \Gamma. \tag{6.2.4}$$

In this expression the basis functions of the spaces $L_\rho^2(\Gamma)$ and $\mathring{L}^2(\partial B)$ are coupled through the coefficients u_{jk}. For any $u \in L_\rho^2(\Gamma) \otimes \mathring{L}^2(\partial B)$ we can write

$$\begin{aligned}
\|u - v\|_{L_\rho^2(\Gamma)\otimes\mathring{L}^2(\partial B)}^2 &= \left\| \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} u_{jk}\eta_k(\mathbf{x}) l_j(\mathbf{y}) - \sum_{j=1}^{\#\Lambda}\sum_{k=1}^{\infty} v_{jk}\, \eta_k(\mathbf{x})\, l_j(\mathbf{y}) \right\|_{L_\rho^2(\Gamma)\otimes\mathring{L}^2(\partial B)}^2 \\
&= \left\| \sum_{k=1}^{\infty} \widehat{u}_k(\mathbf{y})\eta_k(\mathbf{x}) - \sum_{k=1}^{\infty} \widehat{v}_k(\mathbf{y})\, \eta_k(\mathbf{x}) \right\|_{L_\rho^2(\Gamma)\otimes\mathring{L}^2(\partial B)}^2 \\
&= \sum_{k=1}^{K} \|\widehat{u}_k - \widehat{v}_k\|_{L_\rho^2(\Gamma)}^2 + \sum_{k=K+1}^{\infty} \|\widehat{u}_k\|_{L_\rho^2(\Gamma)}^2 .
\end{aligned} \tag{6.2.5}$$

The terms $\|\eta_k(\mathbf{x})\|^2_{\mathring{L}^2(\partial B)}$ are all equal to one, since the basis is orthonormal. Notice that

$$\begin{aligned}\min_{v\in\mathbb{P}_\Lambda(\Gamma)\otimes\mathbb{F}_K(\partial B)} \|u-v\|^2_{L^2_\rho(\Gamma)} &= \min_{(\widehat{v}_1,\dots,\widehat{v}_K)\in(\mathbb{P}_\Lambda(\Gamma))^K} \sum_{k=1}^{K} \|\widehat{u}_k-\widehat{v}_k\|^2_{L^2_\rho(\Gamma)} + \sum_{k=K+1}^{\infty} \|\widehat{u}_k\|^2_{L^2_\rho(\Gamma)} \\ &= \sum_{k=1}^{K} \min_{\widehat{v}_k\in\mathbb{P}_\Lambda(\Gamma)} \|\widehat{u}_k-\widehat{v}_k\|^2_{L^2_\rho(\Gamma)} + \sum_{k=K+1}^{\infty} \|\widehat{u}_k\|^2_{L^2_\rho(\Gamma)}\,. \end{aligned} \tag{6.2.6}$$

Each one of the terms $\|\widehat{u}_k-\widehat{v}_k\|^2_{L^2_\rho(\Gamma)}$ on the right side of (6.2.6) does not depend on the basis functions $\{\eta_k\}_k$.

In the application of the Factorization Method to EIT presented in Chapter 5, the solution u to problem (5.2.2) with the random diffusion coefficient σ_B given in (5.3.1) can only be evaluated for some snapshots, corresponding to the realizations $\mathbf{y}_1,\dots,\mathbf{y}_M$ of the random variable $\mathbf{Y}$. Moreover, the M snapshots $u_m = u(\cdot,\mathbf{y}_m)$ of the solution are already projected on the basis $\mathbb{F}_K$,

$$u(\mathbf{x},\mathbf{y}_m) = \sum_{k=1}^{K} \widehat{u}_k(\mathbf{y}_m)\eta_k(\mathbf{x}), \quad \mathbf{x} \text{ on } \partial B, \quad m=1,\dots M, \tag{6.2.7}$$

where

$$\widehat{u}_k(\mathbf{y}) = \sum_{j=1}^{\infty} u_{jk}\, l_j(\mathbf{y}), \quad \mathbf{y}\in\Gamma,$$

and a further projection on $\mathbb{F}_K(\partial B)$ is useless. Therefore, the approach by the global projection (6.2.3) is not suitable for the application to EIT, where a first discretization in space is needed to collect the M measurements. For this reason, we will develop an approach based on the random discrete L^2 projection

$$\Pi^M_{\Lambda,K} u = \operatorname*{argmin}_{v\in\mathbb{P}_\Lambda\otimes\mathbb{F}_K(\partial B)} \|u-v\|_{M,\mathring{L}^2(\partial B)}, \tag{6.2.8}$$

using the discrete norm (6.2.2). The general element $v\in\mathbb{P}_\Lambda(\Gamma)\otimes\mathbb{F}_K(\partial B)$ is of the same form as (6.2.4),

$$v(\mathbf{x},\mathbf{y}) = \sum_{k=1}^{K} \widehat{v}_k(\mathbf{y})\,\eta_k(\mathbf{x}), \quad \mathbf{x} \text{ on } \partial B, \quad \mathbf{y}\in\Gamma,$$

where again

$$\widehat{v}_k(\mathbf{y}) = \sum_{j=1}^{\#\Lambda} v_{jk}\, l_j(\mathbf{y}), \quad \mathbf{y}\in\Gamma.$$

We can rewrite the projection (6.2.3) using (6.2.6) as

$$\Pi^{\mathbb{P}_\Lambda\otimes\mathbb{F}_K} u = \sum_{k=1}^{K} \widehat{w}_k\eta_k,$$

where the coefficients $\widehat{w}_k$ are given by the projection

$$\widehat{w}_k = P^M_\Lambda \widehat{u}_k = \operatorname*{argmin}_{\widehat{v}_k\in\mathbb{P}_\Lambda} \|\widehat{u}_k-\widehat{v}_k\|_{L^2_\rho(\Gamma)}.$$

Therefore

$$\Pi^{\mathbb{P}_\Lambda\otimes\mathbb{F}_K} u = \sum_{k=1}^{K} P_\Lambda\widehat{u}_k\eta_k.$$

The error committed by the projection (6.2.3) is given by

$$\|u - \Pi^{\mathbb{P}_\Lambda \otimes \mathbb{F}_K} u\|_{L^2_\rho(\Gamma)\otimes \mathring{L}^2(\partial B)} = \sum_{j=\#\Lambda+1}^{\infty} \sum_{k=1}^{K} u_{jk}^2 + \sum_{k=K+1}^{\infty} \|\widehat{u}_k\|^2.$$

Now we define the following discrete projections,

$$P_\Lambda^M \widehat{u}_k = \underset{\widehat{v}_k \in \mathbb{P}_\Lambda(\Gamma)}{\operatorname{argmin}} \|\widehat{u}_k - \widehat{v}_k\|_M = \underset{\widehat{v}_k \in \mathbb{P}_\Lambda(\Gamma)}{\operatorname{argmin}} \frac{1}{M} \sum_{m=1}^{M} \Big(\widehat{u}_k(\mathbf{y}_m) - \widehat{v}_k(\mathbf{y}_m)\Big)^2. \tag{6.2.9}$$

From (6.2.6), it is evident that we can approximate each term in the summation on the right with the corresponding discrete projection $P_\Lambda^M \widehat{u}_k$ given by (6.2.9). Finally, repeating the steps (6.2.5) with the norm (6.2.2), the global discrete projector $\Pi_{\Lambda,K}^M$ introduced in (6.2.8) can be written as

$$\Pi_{\Lambda,K}^M u = \sum_{k=1}^{K} P_\Lambda^M \widehat{u}_k \eta_k. \tag{6.2.10}$$

6.3 Polynomial approximation of the parametric Neumann-to-Dirichlet map

In this section we consider the polynomial approximation of the Neumann-to-Dirichlet map Λ_0 defined in (6.0.2), exploiting the results obtained in the previous section.

Following [NT09, BNTT12], we give here an approximation result for problem (6.0.3) when the boundary datum $\eta(\mathbf{x})$ is an element of the Fourier basis denoted by η_k. It has been shown in [BNTT12] that the solution to (6.0.3) with a coefficient σ_B of the form (6.0.1) featuring nonoverlapping random inclusions is analytic in a polydisk w.r.t. the random variables. Denote by $PAR \subset \mathbb{C}^r$ the Polydisk Analyticity Region centered in the origin of Γ of problem (6.0.3), where r indicates the dimension of the parameter set Γ as in Chapter 5.

Lemma 6.2. *Let U^k be the solution to problem* (6.0.3) *with Neumann boundary condition $\eta_k \in \mathbb{F}_K$ and $u^k = U^k|_{\partial B}$. Then*

$$\inf_{v \in \mathbb{P}_\Lambda \otimes V} \|u^k - v\|_{L^\infty(\Gamma, \mathring{L}^2(\partial B))} \le C k^{-\frac{1}{2}} e^{-\alpha \mathrm{w}}, \quad \forall\ k = 1, \dots, K, \tag{6.3.1}$$

with $C, \alpha > 0$ independent of k and w.

Proof. First note that, for any Neumann boundary condition η, the solution $U(z) : PAR \to H^1(B)$ to problem (6.0.3) is analytic in the PAR, [NT09, BNTT12], as well as its trace on the boundary $u^k(z) : PAR \to \mathring{L}^2(\partial B)$. Moreover, when $\eta_k \in \mathbb{F}_K$ is the kth element of the Fourier basis, then it holds:

$$\|u^k(z)\|_{\mathring{L}^2(\partial B)} \le C_T \|U^k(z)\|_{H^1(B)} \le C_T C_D \|\eta_k\|_{H^{-1/2}(\partial B)} \le C_T C_D k^{-\frac{1}{2}}, \quad \forall\, z \in PAR, \tag{6.3.2}$$

since $\|\eta_k\|_{H^{-1/2}} = k^{-\frac{1}{2}}$, see [Ada75]. In (6.3.2) C_T is the constant of the trace inequality, and C_D is the stability constant of the solution w.r.t. the input data, that holds uniformly on the analiticity region PAR. Furthermore, denote by $T_{\mathrm{w}} u(z)$ the Taylor polynomial obtained expanding u in the origin of the parameter set up to degree w. Then, using the result in [BNTT12, Lemma 10] it holds

$$\|u^k(z) - T_{\mathrm{w}} u^k(z)\|_{\mathring{L}^2(\partial B)} \le \widehat{C} \sup_{z \in PAR} \|u^k(z)\|_{\mathring{L}^2(\partial B)} e^{-\alpha \mathrm{w}}, \quad \forall\, z \in PAR. \tag{6.3.3}$$

See [BNTT12, Lemma 10] for the value of α and the expression of the proportionality constant $\widehat{C}$. The thesis (6.3.1) follows from (6.3.3) and (6.3.2) with $C = \widehat{C} C_T C_D$. □

Finally we can define the following operatorial random discrete projection Π_O that acts on the operator Λ_0 for each entry $u^k = \Lambda_0\eta_k(x)$ as:

$$\Pi_O\Lambda_0\eta_k = T_{\tau(k)}\Pi^M_{\Lambda,K}u^k = T_{\tau(k)}\left(\sum_{i=1}^{K} P^M_\Lambda \widehat{u}^k_i\eta_i\right), \quad \forall\, k = 1,\dots,K, \tag{6.3.4}$$

where $T_{\tau(k)}$ is the same truncation operator

$$T_\tau(v) = \begin{cases} v, & \text{if} \quad \|v\|_{\mathring{L}^2(\partial B)} \le \tau, \\ \dfrac{\tau(k)}{\|v\|_{\mathring{L}^2(\partial B)}}\, v, & \text{if} \quad \|v\|_{\mathring{L}^2(\partial B)} > \tau, \end{cases} \tag{6.3.5}$$

introduced in (2.4.6) in Chapter 2.

We prove a convergence theorem for the discretized operator Λ_0 over the Fourier basis $\mathbb{F}_K$. Denote by $L^\infty(\Gamma, F)$ the space of operators parametrized by $\mathbf{y} \in \Gamma$ with a finite Frobenius norm for all $\mathbf{y} \in \Gamma$.

Theorem 6.3. *For any $\gamma > 0$, if M is such that $K(\mathbb{P}_\Lambda)$ satisfies* (2.2.6)*, then the following convergence estimate holds:*

$$\mathbb{E}\left[\|\Lambda_0 - \Pi_O\Lambda_0\|^2_{L^\infty(\Gamma,F)}\right] \le \left(C_1 e^{-2\alpha \mathrm{w}} + C_2 M^{-\gamma}\right)(1 + \log K), \tag{6.3.6}$$

with $C_1 = \left(1 + \sqrt{2}\,\widetilde{C}^\rho_{inv}\right)^2 \widehat{C}^2 C_T^2 C_D^2$, $C_2 = 8C_T^2C_D^2$, $\alpha > 0$ and choosing the truncation (6.3.5) *with thresholds $\tau(k) = \|u^k\|_{L^\infty(\Gamma,\mathring{L}^2(\partial B))}$ for all $k = 1,\dots,K$ in the definition* (6.3.4) *of Π_O.*

Proof. Introducing the space $\mathcal{Z} = L^\infty(\Gamma, \mathring{L}^2(\partial B))$, we can apply the thesis (6.1.8) of Theorem 6.2, but with the space $\mathcal{Z}$ instead of $\mathcal{W}$. Moreover, from (6.3.2) we have for all $k = 1,\dots,K$ that $(\tau(k) + \|u^k\|_{\mathcal{Z}}) \le 2C_TC_Dk^{-\frac{1}{2}}$, choosing $\tau(k) = \|u^k\|_{\mathcal{Z}}$. We obtain

$$\begin{aligned}
\mathbb{E}\left[\|\Lambda_0 - \Pi_O\Lambda_0\|^2_{L^\infty(\Gamma,F)}\right] &= \mathbb{E}\left[\left(\sum_{k=1}^{K} \|u^k - T_{\tau(k)}\Pi^M_{\Lambda,K}u^k\|^2_{\mathcal{Z}}\right)\right] \\
&= \sum_{k=1}^{K} \mathbb{E}\left[\|u^k - T_{\tau(k)}\Pi^M_{\Lambda,K}u^k\|^2_{\mathcal{Z}}\right] \\
&\le \sum_{k=1}^{K} \left(1 + \sqrt{2}\,\widetilde{C}^\rho_{\mathrm{inv}}\right)^2 \inf_{v\in\mathbb{P}_\Lambda\otimes\mathring{L}^2(\partial B)} \|u^k - v\|^2_{\mathcal{Z}} + (\tau(k) + \|u^k\|_{\mathcal{Z}})^2\, 2M^{-\gamma} \\
&\le \sum_{k=1}^{K} \left(1 + \sqrt{2}\,\widetilde{C}^\rho_{\mathrm{inv}}\right)^2 \widehat{C}^2 C_T^2 C_D^2 k^{-1} e^{-2\alpha \mathrm{w}} + 8C_T^2C_D^2k^{-1}M^{-\gamma} \\
&\le \left(C_1 e^{-2\alpha \mathrm{w}} + C_2 M^{-\gamma}\right)(1 + \log K),
\end{aligned}$$

with $C_1 = \left(1 + \sqrt{2}\,\widetilde{C}^\rho_{\mathrm{inv}}\right)^2 \widehat{C}^2 C_T^2 C_D^2$ and $C_2 = 8C_T^2C_D^2$. □

Theorem 6.3 shows that the polynomial approximation of the discretized operator Λ_0 is convergent. The non-optimal factor $\log K$ in (6.3.6) does not allow the extension of the proof to the continuous operator Λ_0, replacing the Frobenius norm by the more general Hilbert-Schmidt norm. However, given that in practical applications of EIT the number of frequencies used is relatively small, this factor gives only a mild multiplicative constant to the error estimate in w and M. As shown in Chapter 2, the number of sampling points M and the polynomial space Λ with degree w have to be related according to condition (2.2.6), to achieve stability and optimality of the random discrete projection.

Remark 6.1. *The choice of the truncation operators* (6.3.5) *does not preserve the symmetry of* $\Pi_O\Lambda_0$*, as any row of the matrix is truncated differently. However, if the number of sampling points is sufficiently high,* e.g. *such that condition* (2.2.6) *is satisfied, then the truncation is not effective with high probability, and the symmetry of* $\Pi_O\Lambda_0$ *is almost preserved. Moreover, if the polynomial approximation* $\Pi_O\Lambda_0$ *is accurate enough to reproduce the operator* Λ_0, i.e. *if* M *and* w *are sufficiently large, then the lack of symmetry does not affect the accuracy in the evaluation of the Ky-Fan norm that is needed to apply the Factorization Method. In practice, the lost of symmetry of the measurements operator always occurs, due to the presence of noise contaminating the measurements. Nonetheless, see* e.g. *the results with noise presented in Section 5.5.3, within a reasonable range of perturbations the successful application of the Factorization Method is still possible.*

In the end, we obtained a reduced model $\Lambda_0^M(\mathbf{y}) = \Pi_O\Lambda_0(\mathbf{y})$ of the random operator Λ_0 and proved in Theorem 6.3 that its accuracy is arbitrarily high, depending on the choice of the parameters M and w. Then we can replace $\|\Lambda(\mathbf{y}^*) - \Lambda_0(\mathbf{y})\|_F$ by $\|\Lambda(\mathbf{y}^*) - \Lambda_0^M(\mathbf{y})\|_F$ in the optimization of problem (5.3.9), since

$$\left\|\Lambda(\mathbf{y}^*) - \Lambda_0^M(\mathbf{y})\right\|_F \leq \|\Lambda(\mathbf{y}^*) - \Lambda_0(\mathbf{y})\|_F + \left\|\Lambda_0(\mathbf{y}) - \Lambda_0^M(\mathbf{y})\right\|_F, \quad \forall\, \mathbf{y} \in \Gamma, \quad \forall\, \mathbf{y}^* \in \Gamma,$$

with the last term on the right converging to zero. The convergence of the Frobenius norm ensures the convergence of all the Ky-Fan k-norms with $k \leq K$.

6.4 Polynomial approximation of the Ky-Fan k-norm

The approach outlined in the previous section requires to approximate the solution to problem (6.0.3) by the random discrete projection. This requires to solve many (up to K) uncoupled least squares problems embedded in the projection (6.2.10), depending on the number of Fourier coefficients to retain and thus on the desired final accuracy level.

An alternative approach to deal with the optimization of problem (5.3.9) is to accomplish a polynomial approximation of the the Ky-Fan k-norm itself, *i.e.* to build a polynomial approximation of the following real-valued scalar target function

$$\phi_k(\mathbf{y}) = \|\Lambda(\mathbf{y}^*) - \Lambda_0(\mathbf{y})\|_k, \quad \mathbf{y} \in \Gamma, \tag{6.4.1}$$

with the simpler random discrete projection P_Λ^M introduced in (2.2.2). The reduced model $P_\Lambda^M\phi_k$ can then be used in the optimization procedure. This approach allows to perform the optimization procedure with a surrogate but smoother function that adequately resembles the original target function in problem (5.3.9). The evaluations of the surrogate model are so cheap that even the pointwise evaluations of the target function ϕ_k over a very fine tensor product grid on Γ can be employed to look for the point of global minimum.

On the other hand this approach is not rigorously justified as the one outlined in the previous sections, because in general (6.4.1) is only continuous w.r.t. $\mathbf{y}$, and a further smoothness has still to be investigated. Nonetheless, several numerical tests have shown that the smoothing effect of the polynomial interpolation preserves the global minimum point of the original target function, and at the same time allows to use any optimization algorithm based on derivatives. Therefore, this approach is a useful tool to identify with a good accuracy the realization $\mathbf{y}^*$ that generated the measurement $\Lambda(\mathbf{y}^*)$. Fig. 6.1 shows the evaluation over the parameter set Γ of the reduced model $P_\Lambda^M\phi_k$ for some values of k and w, always choosing $M = 225$ sampling points. The red cross corresponds to the point $\mathbf{y}^* \in \Gamma$. The polynomial set Λ is chosen as the Total Degree space with level w. A 3000 nodes mesh with P1 finite elements has been chosen to discretize the domain B of problem (6.0.3). In the results displayed in Fig. 6.1 the global minimum point $\mathbf{y}^*$ that generated the measurement of Λ has been recovered up to a tolerance equal to 10^{-2}. Depending on the choice of the discretization parameters M, w, k and the finite element gridsize an arbitrarily accurate approximation of $\mathbf{y}^*$ can be recovered.

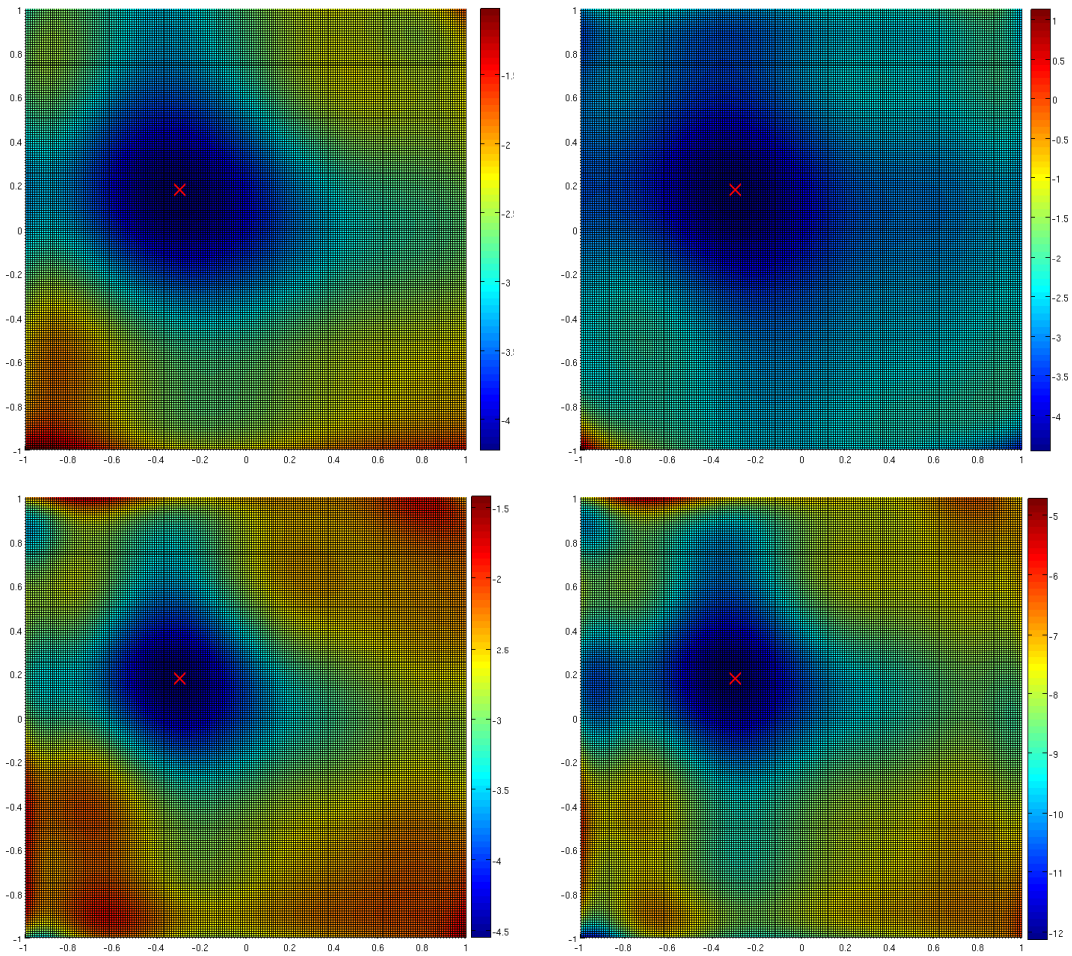


Figure 6.1: The function $\mathbf{y} \mapsto P_\Lambda^M \phi_k(\mathbf{y})$ displayed over the parameter set Γ. The red cross denotes the realization $\mathbf{y}^*$. Top-left: $k = 5$, w $= 6$, $M = 225$. Top-right: $k = 5$, w $= 7$, $M = 225$. Bottom-left: $k = 5$, w $= 8$, $M = 225$. Bottom-right: $k = 10$, w $= 8$, $M = 225$.

6.5 The case of paired measurements

In the case of paired measurements, the variant of the Factorization Method proposed in Algorithm 4 requires a repetitive calculation of the solution to problem (5.2.15) for different realizations of the random variable which parametrizes the stochastic diffusion coefficient. The solutions are then combined in the indicator function (5.3.14) to provide a final indicator, defined over the physical domain, that is able to detect the presence and location of the inclusions.

The methodology presented in Section 6.2 is directly applicable to compute the polynomial approximation via random discrete projection of the solution to the Neumann problem (5.2.15), in the same way as for the solution to (6.0.3).

However, the numerical tests shown in Section 5.6 point out that in the case of paired measurements a limited number of realizations is enough to detect the presence and location of the inclusions with a reasonable accuracy, also when the parameter space features a moderately large dimension, *e.g.* in Fig. 5.40 with a five-dimensional parameter space. Therefore, the acceleration via polynomial approximation seems to be less necessary in the case of paired measurements than in the case of arbitrary measurements.

6.6 Conclusions

A methodology for the polynomial approximation based on the random discrete projection of the Factorization Method for EIT has been proposed in Section 6.2 and Section 6.3. The theoretical analysis proves the convergence of the error committed by the random discrete projection, when approximating the spectrum of the operator Λ_0. In Theorem 6.3 an estimate in the Frobenius norm has been derived, and it ensures the convergence of all the Ky-Fan k-norms with $k \leq K$.

In the case of arbitrary measurements, the two variants of the Factorization Method proposed (Algorithm 2 and Algorithm 3) have been accelerated by computing the polynomial approximation $\Lambda_0^M = \Pi_O \Lambda_0$ of the operator Λ_0. Hence the optimization of problem (5.3.9) can be performed as in Section 5.6, but using the costless pointwise evaluations of the reduced model $\Lambda_0^M = \Lambda_0^M(\mathbf{y})$ for each value of the parameter $\mathbf{y} \in \Gamma$, instead of the evaluations of $\Lambda_0 = \Lambda_0(\mathbf{y})$ for which is required each time to numerically solve problem (6.0.3).

The approach based on polynomial acceleration proposed in Section 6.2 and Section 6.3, becomes useful when the parameter space Γ features a large dimension, to keep affordable the overall computational cost required to successfully apply the Factorization Method, *i.e.* the computational cost required to find an accurate solution, according to the convergence criterion (5.3.10), to the optimization problem (5.3.9).

In the case of paired measurements, the random discrete projection can be applied to approximate the map (6.0.4) instead of the map (6.0.2). This turns out in the approximation of the solution φ_N^k to problem (5.2.15) instead of the solution to problem (6.0.3). Notice that both problems have the same background diffusion coefficient σ_B, *e.g.* the one defined in (6.0.1), and that the elliptic problem (5.2.15) is suitable for polynomial approximation, since it falls in the same class of elliptic problems treated in Chapter 2. In the end, the same approach can be applied to accelerate via polynomial approximation the Factorization Method proposed in Algorithm 4, although the results presented in Section 5.3.2 show that this case is computationally less demanding than the case with arbitrary measurements.

The result proved in Theorem 6.3 ensures that the same reconstructions obtained in Section 5.6 could be obtained using the polynomial approximation technique proposed, under condition (2.2.6). This condition requires the number of sampling points to scale as the square of the dimension of the polynomial space. However, as observed in Chapter 1, Chapter 2 and Chapter 4, the condition (2.2.6) has been shown to be very conservative. In dimension higher than one a less strict condition, that requires only a linear proportionality between the number of sampling points and the dimension of the polynomial space, could be employed, making the random discrete projection attractive in high dimensions.

As discussed in Section 6.2 the Stochastic Galerkin method cannot be applied in this context.

The efficiency of the approach with the random discrete projection should be compared with the efficiency of other techniques such as Stochastic Collocation on Sparse Grids.

Conclusions and perspectives

In this thesis the polynomial approximation methodology based on the random discrete L^2 projection has been presented, to approximate aleatory functions depending on random variables distributed according to a given probability density. In the former part of the thesis, the stability and optimality of the random discrete projection have been analyzed. A key role condition between the number of sampling points and the dimension of the polynomial space has been pointed out. With the uniform density, the optimal convergence rate is achieved when this condition is satisfied, *i.e.* when the number of sampling points is proportional to the square of the dimension of the polynomial space. Moreover, the analysis proposed clarifies how the dimension of the parameter space and the smoothness of the target function affect the optimal convergence rate. Several types of probability densities other than the uniform have been investigated, *i.e.* the more general beta family and the Gaussian density.

Afterwards the random discrete projection has been applied to approximate the solution to a class of PDE models with stochastic data. The aforementioned class contains elliptic PDEs and the linear elasticity model, both with stochastic coefficients described by means of random variables with a disjoint support in the physical domain. A convergence estimate has been proven, showing that the random discrete L^2 projection on suitably chosen sequences of polynomial spaces converges at least with sub-exponential convergence rate w.r.t. the number of sampling points. The same estimate established a relation between the convergence rate of the random discrete L^2 projection on a priori chosen polynomial spaces and the convergence rate of the Stochastic Galerkin method.

In addition, the random discrete projection has been applied to approximate Quantities of Interest, defined as integral functionals of the solution to PDE models with random coefficients and random domains. Several examples with the Darcy model have been presented, when the diffusion coefficient is parametrized by univariate and multivariate random variables. Furthermore, the linear elasticity model with random Young's modulus and the Navier–Stokes model with random viscosity and random geometry of the computational domain have been tested.

In the latter part of the thesis, the Factorization Method has been considered in the context of the *Continuous Model* in Electrical Impedance Tomography. To begin with, a numerical scheme to solve the dipole-like singular Neumann problem has been proposed, when the background diffusion coefficient is deterministic but spatially inhomogeneous. Then, this numerical scheme has been incorporated in the Factorization Method to solve the EIT inverse problem in inhomogeneous and uncertain background, with a specific focus on the case of piecewise constant background.

In the inhomogeneous deterministic case, the proposed approach yields reconstructions of the same accuracy level as those obtained in the homogeneous case, with the range of variation of the background coefficient up to two orders of magnitude. The method showed to be robust also when the observations of the measurement operator are contaminated by noise.

In presence of uncertainty in the background coefficient, three variants of the Factorization Method have been proposed, in both cases of arbitrary and paired measurements of the random observation operator. Also with uncertain background, the inversion method can detect the presence and location of several inclusions, again with the range of variation of the background coefficient up to two orders of magnitude. When uncertainty is present the reconstructions feature a lower accuracy than those obtained in the deterministic case. When many inclusions are present those harder to detect could remain undetected.

Finally the random discrete projection has been applied to accelerate the variants of the Factorization Method proposed in the case of uncertain background, and some convergence estimates of the proposed approach have been derived.

Perspectives

Natural continuations of the analysis of the random discrete projection will concern the application to PDE models with random field coefficients, useful *e.g.* in the modeling of flow in porous media, and a fair comparison of this methodology with the Stochastic Collocation method on Sparse Grids.

Concerning the Factorization Method, further developments could address more realistic models such as the *Complete Electrode Model*, the framework of Hybrid Electrical Impedance Tomography and inverse scattering in uncertain media.

Appendix A

Appendix on polynomial spaces

A.1 Anisotropic polynomial spaces

In Section 1.2.1 several isotropic polynomial spaces were introduced: the Tensor Product space (1.2.8), the Total Degree space (1.2.9), the Hyperbolic Cross space (1.2.10) and the Smolyak space (1.2.11). Anisotropic versions of these space can be derived as well. Let

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N) \tag{A.1.1}$$

be a vector of real positive weights, each one being greater or equal than one. Denote by

$$\alpha_{min} = \min_{n=1,\ldots,N} \alpha_n. \tag{A.1.2}$$

Then we can define the following anisotropic variants of the polynomial spaces:

Anisotropic Tensor Product space (ATP)

$$\Lambda_{\mathrm{w}} = \Big\{ \mathbf{p} \in \mathbb{N}^N : \max_{n=1,\ldots,N} \alpha_n p_n \leq \alpha_{min} \mathrm{w} \Big\}, \tag{A.1.3}$$

Anisotropic Total Degree space (ATD)

$$\Lambda_{\mathrm{w}} = \Big\{ \mathbf{p} \in \mathbb{N}^N : \sum_{n=1}^{N} \alpha_n p_n \leq \alpha_{min} \mathrm{w} \Big\}, \tag{A.1.4}$$

Anisotropic Hyperbolic Cross space (AHC)

$$\Lambda_{\mathrm{w}} = \Big\{ \mathbf{p} \in \mathbb{N}^N : \prod_{n=1}^{N} (p_n + 1)^{\frac{\alpha_n}{\alpha_{min}}} \leq \mathrm{w} + 1 \Big\}, \tag{A.1.5}$$

Anisotropic Smolyak polynomial space (ASM)

$$\Lambda_{\mathrm{w}} = \Big\{ \mathbf{p} \in \mathbb{N}^N : \sum_{n=1}^{N} \alpha_n f(p_n) \leq \alpha_{min} f(\mathrm{w}) \Big\}. \tag{A.1.6}$$

The function f in (A.1.6) is the same function as in (1.2.11). Our choice to avoid weights lower than one ensures that each anisotropic space is always included in the corresponding isotropic space. Of course, when taking $\boldsymbol{\alpha} = \mathbf{1}$ the usual isotropic spaces are recovered.

A.2 Dimension of the Hyperbolic Cross space

In this section we derive an upper bound for the dimension of the HC space, in any dimension N of the parameter space Γ. This bound is sharp when $N = 2$, but it becomes very conservative as the dimension N increases.

Proposition A.1. *An upper estimate of the dimension of the isotropic Hyperbolic Cross space* (1.2.10) *is*

$$\#HC(\mathrm{w}, N) \le (\mathrm{w}+1)\Big(1+\log_e(\mathrm{w}+1)\Big)^{N-1}. \tag{A.2.1}$$

Proof. Clearly $\#\mathrm{HC}(0, N) = 1$ for any N, because only the null multi-index belongs to the set Λ. Moreover, when $N = 2$ and for any $\mathrm{w} \ge 1$ it holds

$$\begin{aligned}
\#\mathrm{HC}(\mathrm{w}, 2) \le & \Big\lfloor \sum_{k=1}^{\mathrm{w}+1} \frac{\mathrm{w}+1}{k} \Big\rfloor \\
= & \Big\lfloor (\mathrm{w}+1) \sum_{k=1}^{\mathrm{w}+1} \frac{1}{k} \Big\rfloor && \text{[using (A.2.4)]} \\
\le & \Big\lfloor (\mathrm{w}+1)\Big(1+\log_e(\mathrm{w}+1)\Big) \Big\rfloor .
\end{aligned} \tag{A.2.2}$$

This estimate can be extended to larger dimensions by induction

$$\begin{aligned}
\#\mathrm{HC}(\mathrm{w}, N+1) = & \sum_{\prod_{n=1}^{N+1}(i_n+1) \le \mathrm{w}+1} 1 \\
= & \sum_{i_1=0}^{\mathrm{w}} \sum_{\prod_{n=2}^{N+1}(i_n+1) \le \left\lfloor \frac{\mathrm{w}+1}{i_1+1} \right\rfloor} 1 \\
= & \sum_{i_1=0}^{\mathrm{w}} \#\mathrm{HC}\Big(\Big\lfloor \frac{\mathrm{w}+1}{i_1+1} \Big\rfloor - 1, N \Big) && \text{[inductive step]} \\
= & \sum_{i_1=0}^{\mathrm{w}} \Big\lfloor \frac{\mathrm{w}+1}{i_1+1} \Big\rfloor \left(1 + \log\left(\Big\lfloor \frac{\mathrm{w}+1}{i_1+1} \Big\rfloor \right) \right)^{N-1} \\
\le & (\mathrm{w}+1)\Big(1+\log(\mathrm{w}+1)\Big)^{N-1} \sum_{i_1=0}^{\mathrm{w}} \frac{1}{i_1+1} && \text{[using (A.2.4)]} \\
= & (\mathrm{w}+1)\Big(1+\log(\mathrm{w}+1)\Big)^{N}.
\end{aligned} \tag{A.2.3}$$

□

Remark A.1. *The partial sum of the harmonic series can be sandwiched as*

$$\log_e(n+1) < \sum_{k=1}^{n} \frac{1}{k} < \log_e(e\,n)$$

because

$$\log_e(n+1) = \int_1^{n+1} \frac{1}{x}\,dx = \sum_{k=1}^{n} \int_k^{k+1} \frac{1}{x}\,dx < \sum_{k=1}^{n} \int_k^{k+1} \frac{1}{k}\,dx = \sum_{k=1}^{n} \frac{1}{k},$$

and

$$\sum_{k=1}^{n} \frac{1}{k} < 1 + \int_1^n \frac{1}{x}\,dx = 1 + \log_e x \Big|_1^n = 1 + \log_e n = \log_e(e\,n). \tag{A.2.4}$$

Remark A.2. *The estimate* (A.2.1) *is very sharp when* $N = 2$*, but its accuracy deteriorates as* N *increases. To quantify the overestimation of the real value of* $\#HC$ *we define*

$$E(\mathrm{w}, N) = \frac{\#HC(\mathrm{w}, N)_{estimated} - \#HC(\mathrm{w}, N)_{exact}}{\#HC(\mathrm{w}, N)_{exact}}. \tag{A.2.5}$$

In the case $N = 2$ *it holds*

$$\begin{aligned} E(10, 2) &< 0.28, \\ E(25, 2) &< 0.21, \\ E(50, 2) &< 0.19, \end{aligned}$$

and a geometrical argument shows that

$$\lim_{\mathrm{w}\to\infty} E(\mathrm{w}, 2) = 0.$$

In higher dimension

$$\begin{aligned} E(5, 5) &< 5.508, \\ E(10, 5) &< 9.399, \\ E(10, 10) &< 1142. \end{aligned}$$

Remark A.3. *It is possible to derive sharper estimates exploiting the recursive relation*

$$\#HC(\mathrm{w}, N) \le I(\mathrm{w}, N) + N\,\mathrm{w} + N\,\#HC(\mathrm{w}, N-1) + 1,$$

where the integral $I(\mathrm{w}, N)$ *is equal to*

$$I(\mathrm{w}, N) = \int_0^{\mathrm{w}} \int_0^{\frac{\mathrm{w}+1}{x_1+1}-1} \cdots \int_0^{\left(\frac{\mathrm{w}+1}{\prod_{n=1}^{N-1}(x_n+1)}-1\right)} dx_N \cdots dx_1.$$

For example, in the case $N = 3$*,*

$$I(\mathrm{w}, 3) = (\mathrm{w}+1)\Big(\log_e(\mathrm{w}+1)\Big)\left(\frac{\log_e(\mathrm{w}+1)}{2} - 1\right) + \mathrm{w},$$

and therefore

$$\#HC(\mathrm{w}, N) \le (\mathrm{w}+1)\Big(\log_e(\mathrm{w}+1)\Big)\left(\frac{\log_e(\mathrm{w}+1)}{2} - 1\right) + \mathrm{w} + 3\mathrm{w} + 1 + 3\,\#HC(\mathrm{w}, 2),$$

that at least does not deteriorates

$$\begin{aligned} E(10, 3) &\approx 3.586, \\ E(25, 3) &\approx 3.692, \\ E(50, 3) &\approx 3.870. \end{aligned}$$

A.3 Comparison condition number TP vs TD vs HC

The comparison of the condition number (1.4.9) obtained with the three spaces TP, TD, HC is shown in Fig. A.1, in the case $N = 4$ and with the number of sampling points chosen as $M = 5 \cdot \#\Lambda$. The TP space features the fastest growth of the condition number w.r.t. $\#\Lambda$. On the other side, the growth of the HC space is the slowest one. With any space, it is evident that the relation $M = 5 \cdot \#\Lambda$ yields a stable Least Squares approximation for the TD and HC space. Clearly, the TP space does not allow to further increase the maximum polynomial degree.

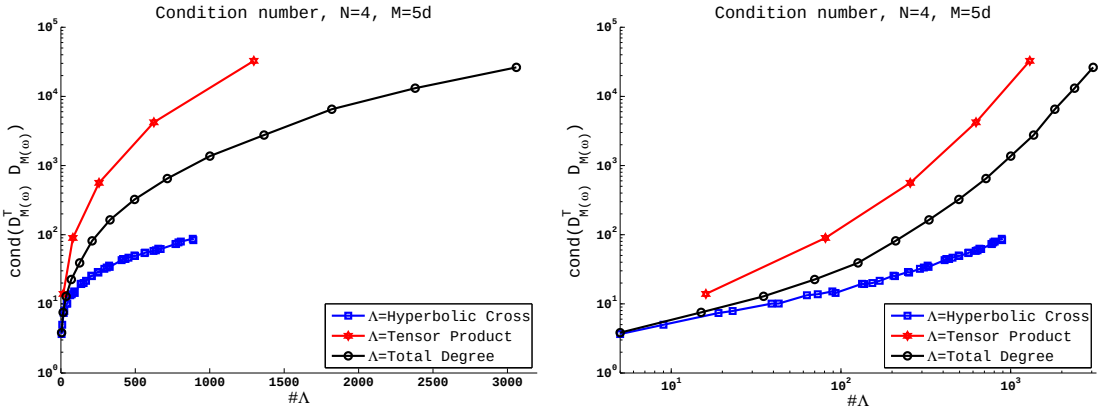


Figure A.1: Left: Condition number (1.4.9). TP, TD, HC spaces, $N = 4$, $M = 5 \cdot \#\Lambda$. Right: same as left but in log–log scale.

Table A.1 reports selected values from Fig. A.1 of the condition number and of $\#\Lambda$, for some values of the parameter w.

	cond(D^TD)			$\#\Lambda$		
w	TP	TD	HC	TP	TD	HC
1	$1.4 \cdot 10^1$	$3.8 \cdot 10^0$	3.7	16	5	5
2	$9.0 \cdot 10^1$	$7.5 \cdot 10^0$	5.0	81	15	9
3	$5.6 \cdot 10^2$	$1.3 \cdot 10^1$	7.4	256	35	19
4	$4.2 \cdot 10^3$	$2.2 \cdot 10^1$	7.9	625	70	23
5	$3.3 \cdot 10^4$	$3.9 \cdot 10^1$	10	1296	126	39
6		$8.1 \cdot 10^1$	10		210	43
8		$3.2 \cdot 10^2$	14		495	73
10		$1.4 \cdot 10^3$	14		1001	93
14		$2.6 \cdot 10^4$	22		3060	169
20			33			308
30			55			566
40			83			890

Table A.1: Selected values from Fig. A.1 of $\#\Lambda$ and cond(D^TD), corresponding to several values of w, for the spaces TP, TD, HC, $N = 4$, $M = 5 \cdot \#\Lambda$.

Appendix B

Multidimensional inverse inequalities with uniform distribution

B.1 Inequalities between $\|\cdot\|_{L^\infty(\Gamma)}$ and $\|\cdot\|_{L^2_\rho(\Gamma)}$

In this section, we present some inverse inequalities on polynomial spaces. The same notation introduced in Remark 1.1 is used. As in Chapter 1, N denotes the dimension of Γ.

One-dimensional inverse inequality.

From the inequality in Remark 1.1, by interpolation we deduce an analogous result in L^4

$$\|v\|_{L^4} \leq \frac{\sqrt{(\mathrm{w}+1)}}{2^{1/4}}\|v\|_{L^2},$$
$$\|v\|_{L^4_\rho} \leq \sqrt{(\mathrm{w}+1)}\|v\|_{L^2_\rho}.$$

Inverse inequality in multiD

Let $v \in \mathbb{P}_\Lambda([-1,1]^N)$. Again v can be expanded in Legendre series

$$v(\mathbf{y}) = \sum_{\mathbf{p}\in\Lambda} \alpha_{\mathbf{p}}\phi_{\mathbf{p}}(\mathbf{y}), \qquad \mathbf{y} = (y_1,\dots,y_N) \in [-1,1]^N,$$

where $\phi_{\mathbf{p}}(\mathbf{y}) = \prod_{n=1}^N \phi_{p_n}(y_n)$ and $\|\phi_{\mathbf{p}}\|_{L^\infty} = \prod_{n=1}^N \sqrt{p_n+\frac{1}{2}}$. Define $\widetilde{C}_{\mathrm{inv}}(\Lambda,N)$ as

$$\widetilde{C}_{\mathrm{inv}}(\Lambda,N) = \left(\sum_{\mathbf{p}\in\Lambda}\prod_{n=1}^N (p_n+\frac{1}{2})\right)^{\frac{1}{2}}. \tag{B.1.1}$$

Notice that the constant $\widetilde{C}_\Lambda$ defined in (1.2.17) satisfies also

$$\widetilde{C}_\Lambda \leq \widetilde{C}_{\mathrm{inv}}(\Lambda,N),$$

and the equality holds whenever the expressions of $\widetilde{C}_{\mathrm{inv}}(\Lambda,N)$ are optimal.

Then we can generalize the one-dimensional inequality given in Remark 1.1:

$$\begin{aligned}\|v\|_{L^\infty} \leq& \sum_{\mathbf{p}\in\Lambda}|\alpha_{\mathbf{p}}|\prod_{n=1}^N\sqrt{p_n+\frac{1}{2}}\\ \leq& \|v\|_{L^2}\left(\sum_{\mathbf{p}\in\Lambda}\prod_{n=1}^N(p_n+\frac{1}{2})\right)^{\frac{1}{2}}\\ =& \widetilde{C}_{\mathrm{inv}}(\Lambda,N)\|v\|_{L^2}.\end{aligned}$$

Notice that for $N = 1$ we recover the 1D estimate given in Remark 1.1.

Moreover, we can deduce an analogous result in L^4 by interpolation

$$\|v\|_{L^4} \leq \sqrt{\widetilde{C}_{\text{inv}}(\Lambda, N)}\|v\|_{L^2}.$$

TP estimate. In the case of the isotropic Tensor Product space (1.2.8):

$$\begin{aligned}\widetilde{C}_{\text{inv}}(\Lambda(\mathrm{w}), N)^2 &= \sum_{\mathbf{p}\in\Lambda(\mathrm{w})}\prod_{n=1}^{N}(p_n + \frac{1}{2}) \\ &= \prod_{n=1}^{N}\sum_{j=0}^{\mathrm{w}}(j + \frac{1}{2}) \\ &= \frac{(\mathrm{w}+1)^{2N}}{2^N} \\ &= \frac{\#TP^2}{2^N}.\end{aligned}$$

Therefore

$$\begin{aligned}&\|v\|_{L^\infty} \leq \frac{\#TP}{\sqrt{2}^N}\|v\|_{L^2} = \#TP\|v\|_{L^2_\rho}, \\ &\|v\|_{L^4} \leq \frac{\sqrt{\#TP}}{2^{N/4}}\|v\|_{L^2}, \\ &\|v\|_{L^4_\rho} \leq \sqrt{\#TP}\|v\|_{L^2_\rho}.\end{aligned}$$

TD estimate. In the case of the isotropic Total Degree space (1.2.9), we recall that

$$\#TD = \binom{\mathrm{w}+N}{N} = \prod_{n=1}^{N}\left(1 + \frac{\mathrm{w}}{n}\right),$$

and we have

$$\begin{aligned}\widetilde{C}_{\text{inv}}(\Lambda(\mathrm{w}), N)^2 &= \sum_{\mathbf{p}\in\Lambda(\mathrm{w})}\prod_{n=1}^{N}(p_n + \frac{1}{2}) \\ &\leq \left(\frac{\mathrm{w}}{N} + \frac{1}{2}\right)^N \#TD \\ &\leq \frac{1}{2^N}\left(1 + \frac{\mathrm{w}}{N/2}\right)^N \#TD. \qquad \text{(B.1.2)}\end{aligned}$$

A consequence of Proposition 2.1 and Lemma 2.2 is that $\widetilde{C}_{\text{inv}}(\Lambda(\mathrm{w}), N) \leq \#TD$. As explained in Remark 2.1, the relation in (B.1.2) gives a sharper upper bound of $\widetilde{C}_{\text{inv}}(\Lambda(\mathrm{w}), N)$ when $d > 5$. However, we write the inverse inequalities in the case of TD using the simpler bound $\widetilde{C}_{\text{inv}}(\Lambda(\mathrm{w}), N) \leq \#TD$ rather than (B.1.2):

$$\begin{aligned}&\|v\|_{L^\infty} \leq \frac{\#TD}{\sqrt{2}^N}\|v\|_{L^2} = \#TD\|v\|_{L^2_\rho}, \\ &\|v\|_{L^4} \leq \frac{\sqrt{\#TD}}{2^{N/4}}\|v\|_{L^2}, \\ &\|v\|_{L^4_\rho} \leq \sqrt{\#TD}\|v\|_{L^2_\rho}.\end{aligned}$$

HC estimate. In the case of the isotropic Hyperbolic Cross space (1.2.10) we have

$$\begin{aligned}\widetilde{C}_{\text{inv}}(\Lambda(\mathrm{w}),N)^2 &= \sum_{\mathbf{p}\in\Lambda(\mathrm{w})}\prod_{n=1}^{N}(p_n+\frac{1}{2})\\ &\leq \sum_{\mathbf{p}\in\Lambda(\mathrm{w})}\prod_{n=1}^{N}(p_n+1)\\ &\leq (\mathrm{w}+1)\#HC.\end{aligned}$$

Therefore

$$\begin{aligned}&\|v\|_{L^\infty}\leq\sqrt{(\mathrm{w}+1)\#\mathrm{HC}}\|v\|_{L^2}=\sqrt{2^N(\mathrm{w}+1)\#\mathrm{HC}}\|v\|_{L^2_\rho},\\ &\|v\|_{L^4}\leq((\mathrm{w}+1)\#\mathrm{HC})^{\frac{1}{4}}\|v\|_{L^2},\\ &\|v\|_{L^4_\rho}\leq\left(2^N(\mathrm{w}+1)\#\mathrm{HC}\right)^{\frac{1}{4}}\|v\|_{L^2_\rho}.\end{aligned}$$

B.2 Inequalities between $\|\cdot\|_{L^\infty(\Gamma,V)}$ and $\|\cdot\|_{L^2_\rho(\Gamma)\otimes V}$

Consider an Hilbert space V, *e.g.* $V=H^1(B)$ or $V=L^2(\partial B)$ with B bounded subset of $\mathbb{R}^n$, for any positive integer n. In this section we extend the multidimensional inequalities derived in the Section B.1 between $\|\cdot\|_{L^\infty(\Gamma)}$ and $\|\cdot\|_{L^2_\rho(\Gamma)}$ to the case of V-valued functions, *i.e.* between the spaces $\|\cdot\|_{L^\infty(\Gamma,V)}$ and $\|\cdot\|_{L^2_\rho(\Gamma)\otimes V}$.

Consider $v\in L^2_\rho(\Gamma)\otimes V$, and denote by $\{\psi_j\}_j$ an orthonormal basis of $L^2_\rho(\Gamma)$ and by $\{\varphi_j\}_j$ an orthonormal basis of V. Defining $v_j(\mathbf{x})=\sum_{i=1}^\infty\widehat{v}_{ij}\varphi_i(\mathbf{x})$, the expansion of v is

$$v=v(\mathbf{x},\mathbf{y})=\sum_{ij=1}^{\infty}\widehat{v}_{ij}\varphi_i(\mathbf{x})\psi_j(\mathbf{y})=\sum_{j=1}^{\infty}v_j(\mathbf{x})\psi_j(\mathbf{y}),$$

and of course

$$\|v\|^2_{L^2_\rho\otimes V}=\sum_{ij=1}^{\infty}\widehat{v}_{ij}^2=\sum_{j=1}^{\infty}\left(\sum_{i=1}^{\infty}\widehat{v}_{ij}^2\right)=\sum_{j=1}^{\infty}\|v_j\|_V^2,$$

since $\|v_j\|_V^2=\sum_{i=1}^\infty\widehat{v}_{ij}^2$. Now we can write

$$\begin{aligned}\|v\|_{L^\infty(\Gamma,V)}&=\left\|\sum_{j=1}^{\infty}v_j(\mathbf{x})\psi_j(\mathbf{y})\right\|_{L^\infty(V)}\\ &\leq\operatorname*{ess\,sup}_{\mathbf{y}\in\Gamma}\left\|\sum_{j=1}^{\infty}v_j(\mathbf{x})\psi_j(\mathbf{y})\right\|_V\\ &\leq\sup_{\mathbf{y}\in\Gamma}\sum_{j=1}^{\infty}\|v_j\|_V\,|\psi_j(\mathbf{y})|\\ &\leq\sum_{j=1}^{\infty}\|v_j\|_V\max_{\mathbf{y}\in\Gamma}|\psi_j(\mathbf{y})|\\ &\leq\left(\sum_{j=1}^{\infty}\|v_j\|_V^2\right)^{\frac{1}{2}}\left(\sum_{j=1}^{\infty}\max_{\mathbf{y}\in\Gamma}|\psi_j(\mathbf{y})|^2\right)^{\frac{1}{2}}\\ &=\|v\|_{L^2_\rho\otimes V}\left(\sum_{j=1}^{\infty}\max_{\mathbf{y}\in\Gamma}|\psi_j(\mathbf{y})|^2\right)^{\frac{1}{2}}.\end{aligned}\tag{B.2.1}$$

Finally we can define the constant of the inverse inequality $\widetilde{C}^{\rho}_{\text{inv}}(\Lambda)$ as

$$\widetilde{C}^{\rho}_{\text{inv}}(\Lambda) = \sup_{v \in \mathbb{P}_\Lambda \otimes V} \frac{\|v\|_{L^\infty(V)}}{\|v\|_{L^2_\rho \otimes V}} \leq \left(\sum_{j=1}^{\#\Lambda} \max_{\mathbf{y} \in \Gamma} |\psi_j(\mathbf{y})|^2 \right)^{\frac{1}{2}} < +\infty. \tag{B.2.2}$$

When the density is uniform, the $\{\psi_j\}_j$ are the Legendre polynomials, that are bounded on Γ. In this case, observe that $\widetilde{C}^{\rho}_{\text{inv}}(\Lambda)$ is finite, since $\forall\, v \in \mathbb{P}_\Lambda \otimes V$ the summation on the right side of (B.2.2) has a finite number of bounded terms.

Appendix C

On the probability distribution of the random variable ΔY

The aim of this section is to quantify through some numerical tests the probability distribution of the random variable ΔY, which realizations are defined in (1.3.10). The dimension of the parameter space is $N = 1$, and the density ρ is chosen to be uniform over $\Gamma = [-1, 1]$.

From Lemma 1.2 it is clear that the distribution of the random variable ΔY concentrates around M^{-1} when M increases. The Figs. C.1, C.2, C.3, C.4 and C.5 report the frequency distribution of ΔY and its Cumulative Distribution Function, estimated over 10^6 realizations of the random sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_M$. The values of M are $5, 10, 25, 100, 1000$. The frequency distributions are displayed by means of histograms with 1000 classes, or 10^5 classes in the case $M = 1000$.

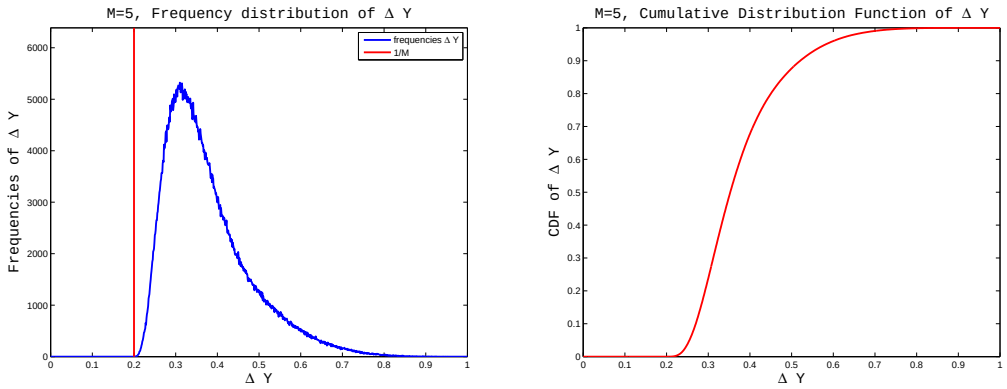


Figure C.1: Distribution of the random variable ΔY with $M = 5$. Left: frequency distribution of ΔY. Right: Cumulative Distribution Function of ΔY.

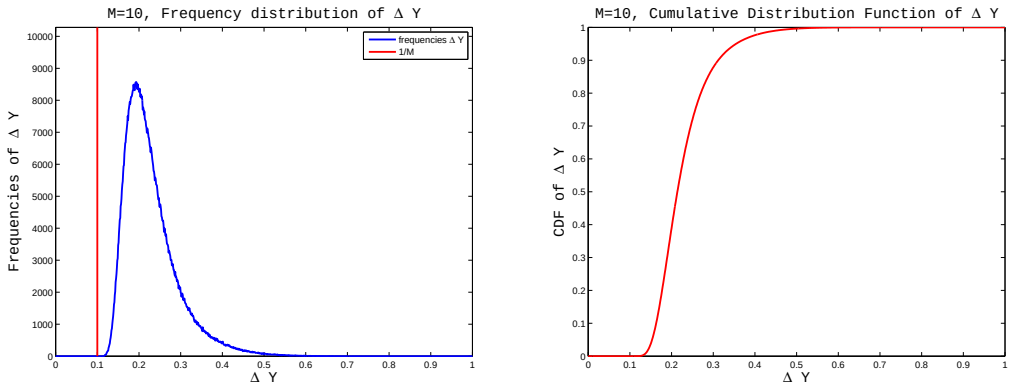


Figure C.2: Distribution of the random variable ΔY with $M = 10$. Left: frequency distribution of ΔY. Right: Cumulative Distribution Function of ΔY.

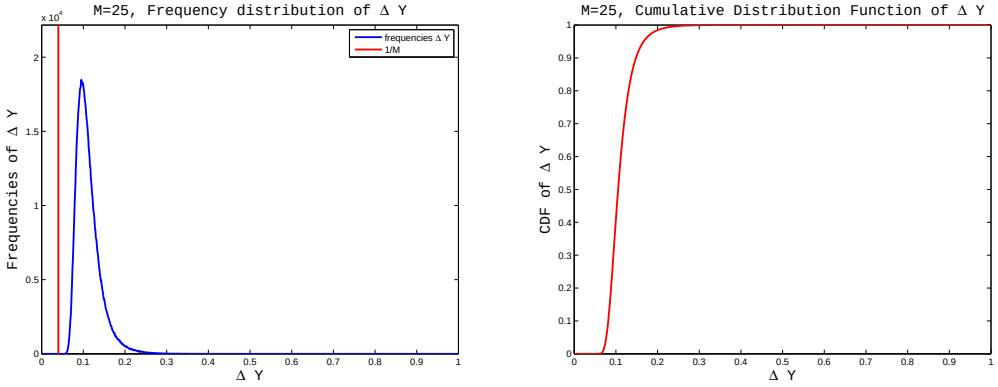


Figure C.3: Distribution of the random variable ΔY with $M = 25$. Left: frequency distribution of ΔY. Right: Cumulative Distribution Function of ΔY.

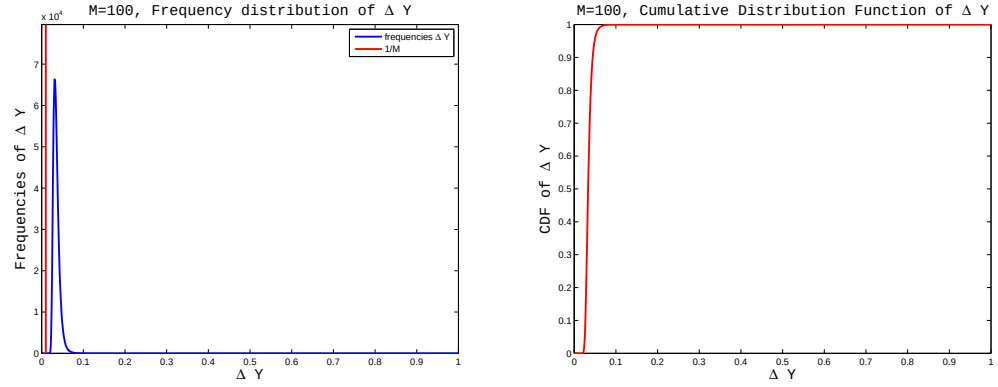


Figure C.4: Distribution of the random variable ΔY with $M = 100$. Left: frequency distribution of ΔY. Right: Cumulative Distribution Function of ΔY.

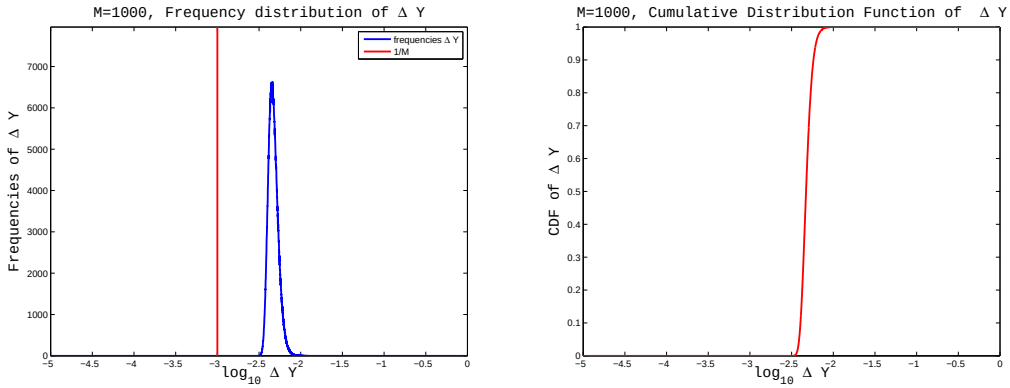


Figure C.5: Distribution of the random variable ΔY with $M = 1000$. Left: frequency distribution of ΔY. Right: Cumulative Distribution Function of ΔY.

List of Figures

List of Tables

List of Notations

$\mathbf{Y}$	random variable
N or d or r or q	dimension of the parameter space Γ
M	number of sampling points
ρ	probability density
L^2_ρ	L^2 functional space weighted w.r.t. the ρ density
$\mathcal{U}$	uniform probability density
$\mathcal{B}(a,b)$	beta probability density with parameters a and b
$\mathcal{N}$	Gaussian probability density
Γ	support of the random variable $\mathbf{Y}$
Λ	multi-index set
$\mathbb{P}_\Lambda$	polynomial space associated to the multi-index set Λ
$\mathbf{y}$	realization of the random variable $\mathbf{Y}$
ϕ	target function
σ_B	background diffusion coefficient
σ_D	inclusion diffusion coefficient
w	spectral accuracy
HC	the isotropic Hyperbolic Cross polynomial space
TD	the isotropic Total Degree polynomial space
TP	the isotropic Tensor Product polynomial space

Bibliography

[Ada75] R. A. Adams, *Sobolev Spaces*, Academic Press, 1975.

[AHOS07] M. Azzouz, M. Hanke, C. Oesterlein, and K. Schilcher, *The Factorization Method for Electrical Impedance Tomography Data from a New Planar Device*, International Journal of Biomedical Imaging (2007).

[AK07] H. Ammari and H. Kang, *Polarization and moment tensors*, Applied Mathematical Sciences, vol. 162, Springer, New York, 2007.

[Ale88] G. Alessandrini, *Stable determination of conductivity by boundary measurements*, Appl. Anal. **27** (1988), no. 1–3, 153–172.

[AO00] M. Ainsworth and J. T. Oden, *A posteriori error estimation in finite element analysis*, Pure and Applied Mathematics (New York), Wiley-Interscience, New York, 2000.

[AP06] K. Astala and L. Päivärinta, *Calderón's inverse conductivity problem in the plane*, Ann. of Math. (2) **163** (2006), no. 1, 265–299.

[Ask75] R. Askey, *Orthogonal polynomials and special functions*, Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1975.

[BBG+11] L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, L. Tenorio, B. v. B. Waanders, K. Willcox, and Y. Marzouk, *Large-Scale Inverse Problems and Quantification of Uncertainty*, Wiley, 2011.

[BBP+09] G. Bozza, M. Brignone, M. Pastorino, A. Randazzo, and M. Piana, *Imaging of unknown targets inside inhomogeneous backgrounds by means of qualitative inverse scattering*, Inverse Probl. Imaging **3** (2009), no. 2, 231–241.

[BCD+05] P. Binev, A. Cohen, W. Dahmen, R. DeVore, and V. Temlyakov, *Universal algorithms for learning theory. I. Piecewise constant functions*, J. Mach. Learn. Res. **6** (2005), 1297–1321.

[BCDD07] P. Binev, A. Cohen, W. Dahmen, and R. DeVore, *Universal algorithms for learning theory. II. Piecewise polynomial functions*, Constr. Approx. **26** (2007), no. 2, 127–152.

[BCK09] A. M. Bagirov, C. Clausen, and M. Kohler, *Estimation of a regression function by maxima of minima of linear functions*, IEEE Trans. Inform. Theory **55** (2009), no. 2, 833–845.

[BDG08] L. Borcea, V. Druskin, and V. F. Guevara, *Electrical impedance tomography with resistor networks*, Inverse Problems **24** (2008).

[BE09] J. Burkardt and M. S. Eldred, *Comparison of Non-Intrusive Polynomial Chaos and Stochastic Collocation Methods for Uncertainty Quantification*, Tech. report, Sandia National Laboratoires, 2009.

[BG04] H. J. Bungartz and M. Griebel, *Sparse grids*, Acta Numer. **13** (2004), 147–269.

[BHV03] M. Brühl, M. Hanke, and M. S. Vogelius, *A direct impedance tomography algorithm for locating small inhomogeneities*, Numer. Math. **93** (2003), no. 4, 635–654.

[BMNP04] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera, *An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations*, C. R. Math. Acad. Sci. Paris **339** (2004), no. 9, 667–672.

[BNR00] V. Barthelmann, E. Novak, and K. Ritter, *High dimensional polynomial interpolation on sparse grids*, Adv. Comput. Math. **12,** (2000), 273–288.

[BNT07] I. Babuška, F. Nobile, and R. Tempone, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM J. Numer. Anal. **45** (2007), no. 3, 1005–1034 (electronic).

[BNT10] ———, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM Rev. **52** (2010), no. 2, 317–355.

[BNTT11a] J. Bäck, F. Nobile, L. Tamellini, and R. Tempone, *Stochastic spectral Galerkin and collocation methods for PDEs with random coefficients: a numerical comparison*, Spectral and High Order Methods for Partial Differential Equations (J.S. Hesthaven and E.M. Ronquist, eds.), vol. 76, Springer, 2011, pp. 43–62.

[BNTT11b] J. Beck, F. Nobile, L. Tamellini, and R. Tempone, *On the optimal polynomial approximation of stochastic PDEs by Galerkin and Collocation methods*, no. 23-2011, To appear in Math. Mod. Methods. Appl. Sci. (M3AS).

[BNTT12] ———, *Convergence of quasi-optimal stochastic Galerkin methods for a class of PDEs with random coefficients*, MATHICSE Technical Report 24.2012, submitted (2012).

[Bor02] L. Borcea, *Electrical impedance tomography*, Inverse Problems **18** (2002), no. 6, R99–R136.

[BR03] W. Bangerth and R. Rannacher, *Adaptive finite element methods for differential equations*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2003.

[Brü01] M. Brühl, *Explicit characterization of inclusions in electrical impedance tomography*, SIAM J. Math. Anal. **32** (2001), no. 6, 1327–1341 (electronic).

[BS08] G. Blatman and B. Sudret, *Sparse Polynomial Chaos expansions and adaptive stochastic finite elements using regression approach*, C.R.Mechanique **336** (2008), 518–523.

[BS09] M. Bieri and C. Schwab, *Sparse high order FEM for elliptic sPDEs*, Comput. Methods Appl. Mech. Engrg. **198** (2009), no. 13-14, 1149–1170.

[BS11] G. Blatman and B. Sudret, *Adaptive sparse polynomial chaos expansion based on least angle regression*, Journal of Computational Physics **230** (2011), no. 6, 2345 – 2367.

[BT03] R. M. Brown and R. H. Torres, *Uniqueness in the inverse conductivity problem for conductivities with* $3/2$ *derivatives in* L^p, $p > 2n$, J. Fourier Anal. Appl. **9** (2003), no. 6, 563–574.

[BTWG08] T. Bui-Thanh, K. Willcox, and O. Ghattas, *Model reduction for large-scale systems with high-dimensional parametric input space*, SIAM J. Sci. Comput. **30** (2008), no. 6, 3270–3288.

[BTZ04] I. Babuška, R. Tempone, and G. E. Zouraris, *Galerkin finite element approximations of stochastic elliptic partial differential equations*, SIAM J. Numer. Anal. **42** (2004), no. 2, 800–825.

[Cat04] O. Catoni, *Statistical learning theory and stochastic optimization*, Lecture Notes in Mathematics, vol. 1851, Springer-Verlag, Berlin, 2004, Lecture notes from the 31st Summer School on Probability Theory held in Saint-Flour, July 8–25, 2001.

[CC06] F. Cakoni and D. Colton, *Qualitative methods in inverse scattering theory*, Springer-Verlag, 2006.

[CCDS13] A. Chkifa, A. Cohen, R. DeVore, and C. Schwab, *Sparse adaptive Taylor approximation algorithms for parametric and stochastic elliptic PDEs*, M2AN Math. Model. Numer. Anal. **47** (2013), 253–280.

[CCM+13] A. Chkifa, A. Cohen, G. Migliorati, F. Nobile, and R. Tempone, *Discrete least squares polynomial approximation with random evaluations; application to parametric and stochastic elliptic PDEs*, in preparation (2013).

[CDL13] A. Cohen, M. A. Davenport, and D. Leviatan, *On the Stability and Accuracy of Least Squares Approximations*, to appear in Foundations of Computational Mathematics (2013).

[CDS10] A. Cohen, R. DeVore, and C. Schwab, *Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs*, Found. Comput. Math. **10** (2010), no. 6, 615–646.

[CDS11] ———, *Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's*, Anal. Appl. (Singap.) **9** (2011), no. 1, 11–47.

[CFMV98] D. J. Cedio-Fengya, S. Moskow, and M. S. Vogelius, *Identification of conductivity imperfections of small diameter by boundary measurements. Continuous dependence and computational reconstruction*, Inverse Problems **14** (1998), no. 3, 553–595.

[Che02] H. F. Chen, *Stochastic approximation and its applications*, Kluwer Academic Publishers, 2002.

[CHP03] D. Colton, H. Haddar, and M. Piana, *The linear sampling method in inverse electromagnetic scattering theory*, Inverse Problems **19** (2003), no. 6, S105–S137.

[CIN99] M. Cheney, D. Isaacson, and J. Newell, *Electrical impedance tomography*, SIAM Rev. **41** (1999), no. 1, 85–101 (electronic).

[CK98] D. Colton and R. Kress, *Inverse acoustic and electromagnetic scattering theory*, second ed., Applied Mathematical Sciences, vol. 93, Springer-Verlag, Berlin, 1998.

[CLMM09] T. Crestaux, O. Le Maître, and J. M. Maritinez, *Polynomial Chaos Expansion for Sensitivity Analysis*, Reliability Engineering and System Safety (2009).

[CM47] R. H. Cameron and W. T. Martin, *The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals*, Ann. of Math. (2) **48** (1947), 385–392.

[CPP97] D. Colton, M. Piana, and R. Potthast, *A simple method using Morozov's discrepancy principle for solving inverse scattering problems*, Inverse Problems **13** (1997), no. 6, 1477–1493.

[CQ82] C. Canuto and A. Quarteroni, *Approximation results for orthogonal polynomials in Sobolev spaces*, Math. Comp. **38** (1982), no. 157, 67–86.

[DGRH07] A. Doostan, R. G. Ghanem, and J. Red-Horse, *Stochastic model reduction for chaos representations*, Comput. Methods Appl. Mech. Engrg. **196** (2007), no. 37-40, 3951–3966.

[DGS05] C. Desceliers, R. Ghanem, and C. Soize, *Polynomial chaos representation of a stochastic preconditioner*, Internat. J. Numer. Methods Engrg. **64** (2005), no. 5, 618–634.

[DN03] H. A. David and H. N. Nagaraja, *Order statistics*, Third ed., Wiley-Interscience, 2003.

[DNP+04] B. Debusschere, H. Najm, P. Pebray, O. Knio, R. Ghanem, and O. Le Maître, *Numerical Challenges in the Use of Polynomial Chaos Representations for Stochastic Processes*, SIAM Journal on Scientific Computing **26** (2004), 698–719.

[DO11] A. Doostan and H. Owhadi, *A non-adapted sparse approximation of PDEs with stoachastic inputs*, Journal of Computational Physics **230** (2011), 3015–3034.

[DP10] J. Dick and F. Pillichshammer, *Digital nets and sequences: Discrepancy Theory and Quasi–Monte Carlo Integration*, Cambridge Univ. Press, 2010.

[EEU07] M. Eiermann, O. G. Ernst, and E. Ullmann, *Computational aspects of the stochastic finite element method*, Comput. Vis. Sci. **10** (2007), no. 1, 3–15.

[EHN96] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, Mathematics and its Applications, vol. 375, Kluwer Academic Publishers Group, 1996.

[Eld09] M. S. Eldred, *Recent Advances in Non-Intrusive Polynomial Chaos and Stochastic Collocation Methods for Uncertainty Analysis and Design*, American Institute of Aeronautics and Astronautics, 2009.

[Eld11] _____, *Design under uncertainty employing stochastic expansion methods*, International Journal for Uncertainty Quantification **1** (2011), 119–146.

[EMSU12] O. G. Ernst, A. Mugler, H. Starkloff, and E. Ullmann, *On the convergence of generalized polynomial chaos expansions*, ESAIM Math. Model. Numer. Anal. **46** (2012), no. 2, 317–339.

[EWC08] M. S. Eldred, C. G. Webster, and P. G. Constantine, *Evaluation of Non-Intrusive Approaches for Wiener-Askey Generalized Polynomial Chaos*, American Institute of Aeronautics and Astronautics (2008).

[Fis96] G. S. Fishman, *Monte Carlo*, Springer Series in Operations Research, Springer-Verlag, New York, 1996.

[GG03] T. Gerstner and M. Griebel, *Dimension-adaptive tensor-product quadrature*, Computing **71** (2003), no. 1, 65–87.

[GG07] R. Ghanem and D. Ghosh, *Efficient characterization of the random eigenvalue problem in a polynomial chaos decomposition*, Internat. J. Numer. Methods Engrg. **72** (2007), no. 4, 486–504.

[GH07] B. Gebauer and N. Hyvönen, *Factorization method and irregular inclusions in electrical impedance tomography*, Inverse Problems **23** (2007), no. 5, 2159–2170.

[Gha99] R. Ghanem, *Ingredients for a general purpose stochastic finite elements implementation*, Comput. Methods Appl. Mech. Engrg. **168** (1999), no. 1-4, 19–34.

[Gil08] M. B. Giles, *Multilevel Monte Carlo path simulation*, Oper. Res. **56** (2008), no. 3, 607–617.

[GKKW02] L. Gyørfi, M. Kohler, A. Krzyżak, and H. Walk, *A Distribution-Free Theory of Nonparametric Regression*, first ed., Springer Series in Statistics, Springer-Verlag, Berlin, 2002.

[GKN+11] I. G. Graham, F. Y. Kuo, D. Nuyens, R. Scheichl, and I. H. Sloan, *Quasi-Monte Carlo methods for elliptic PDEs with random coefficients and applications*, J. Comput. Phys. **230** (2011), no. 10, 3668–3694.

[GKW+08] B. Ganis, H. Klie, M. F. Wheeler, T. Wildey, I. Yotov, and D. Zhang, *Stochastic collocation and mixed finite elements for flow in porous media*, Comput. Methods Appl. Mech. Engrg. **197** (2008), no. 43-44, 3547–3559.

[GMNP07] M. A. Grepl, Y. Maday, N. C. Nguyen, and A. T. Patera, *Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations*, M2AN Math. Model. Numer. Anal. **41** (2007), no. 3, 575–605.

[GR11] A. N. Gorban and D. Roose, *Coping with Complexity: Model Reduction and Data Analysis*, Springer, 2011.

[GS03] G.R. Ghanem and P.D. Spanos, *Stochastic Finite Elements: a spectral approach*, Dover Publications, 2003.

[GZ03] Z. Guo and Y. Zou, *A review of electrical impedance techniques for breast cancer detection*, Medical Engineering and Physics **25** (2003), no. 2, 79 – 90.

[GZ07a] B. Ganapathysubramanian and N. Zabaras, *Sparse grid collocation schemes for stochastic natural convection problems*, J. Comput. Phys. **225** (2007), no. 1, 652–685.

[GZ07b] ———, *Sparse grid collocation schemes for stochastic natural convection problems*, J. Comput. Phys. **225** (2007), no. 1, 652–685.

[HD03] J. C. Helton and F. J. Davis, *Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems*, Reliability Engineering and System Safety **81** (2003), 23–69.

[HHP07] N. Hyvönen, H. Hakula, and S. Pursiainen, *Numerical implementation of the factorization method within the complete electrode model of electrical impedance tomography*, Inverse Probl. Imaging **1** (2007), no. 2, 299–317.

[HK10] H. Haddar and R. Kress, *Conformal mapping and impedance tomography*, Inverse Problems **26** (2010), no. 7, 074002, 18.

[HM11] H. Haddar and G. Migliorati, *Numerical Analysis of the Factorization Method for Electrical Impedance Tomography in Inhomogeneous Medium*, Rapport de recherche 7801, INRIA, France, 2011.

[HM13] ———, *Numerical analysis of the factorization method for EIT with a piecewise constant uncertain background*, Inverse Problems **29** (2013), 065009 (29pp).

[HS08] M. Hanke and B. Schappel, *The Factorization Method for electrical impedance tomography in the half-space*, SIAM J. Appl. Math. **68** (2008), no. 4, 907–924.

[HS09] B. Harrach and J. K. Seo, *Detecting inclusions in electrical impedance tomography without reference measurements*, SIAM J. Appl. Math. **69** (2009), no. 6, 1662–1681.

[HSST12] H. Hoel, E. von Schwerin, A. Szepessy, and R. Tempone, *Adaptive multilevel Monte Carlo simulation*, Lecture Notes in Computational Science and Engineering vol.82, vol. 82, Numerical Analysis of Multiscale Computations, 2012, pp. 217–234.

[HWB10] S. Hosder, R. W. Walters, and M. Balch, *Point-Collocation Nonintrusive Polynomial Chaos Method for Stochastic Computational Fluid Dynamics*, AIAA Journal **48** (2010), 2721–2730.

[KG08] A. Kirsch and N. Grinberg, *The Factorization Method for inverse problems*, Oxford University Press, 2008.

[Kir96] A. Kirsch, *An introduction to the mathematical theory of inverse problems*, Applied Mathematical Sciences, vol. 120, Springer-Verlag, New York, 1996.

[KKSV00] J. P. Kaipio, V. Kolehmainen, E. Somersalo, and M. Vauhkonen, *Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography*, Inverse Problems **16** (2000), 1487–1522.

[Koh00] M. Kohler, *Inequalities for uniform deviations of averages from expectations with applications to nonparametric regression*, J. Statist. Plann. Inference **89** (2000), no. 1-2, 1–23.

[Koh03] ———, *Nonlinear orthogonal series estimates for random design regression*, J. Statist. Plann. Inference **115** (2003), no. 2, 491–520.

[KS05] J. Kaipio and E. Somersalo, *Statistical and computational inverse problems*, Applied Mathematical Sciences, vol. 160, Springer-Verlag, New York, 2005.

[LARH11] Y. Li, M. Anitescu, O. Roderick, and F. Hickernell, *Orthogonal bases for polynomial regression with derivative information in uncertainty quantification*, Int. J. Uncertain. Quantif. **1** (2011), no. 4, 297–320.

[Lem09] C. Lemieux, *Monte Carlo and Quasi-Monte Carlo Sampling*, Springer, 2009.

[LGH11] M. Liu, Z. Gao, and J. S. Hesthaven, *Adaptive sparse grid algorithms with applications to electromagnetic scattering under uncertainty*, Appl. Numer. Math. **61** (2011), no. 1, 24–37.

[LHH08] A. Lechleiter, N. Hyvönen, and H. Hakula, *The Factorization Method applied to the complete electrode model of impedance tomography*, SIAM J. Appl. Math. **68** (2008), no. 4, 1097–1121.

[LK04] D. Lucor and G. E. Karniadakis, *Adaptive generalized polynomial chaos for nonlinear random oscillators*, SIAM J. Sci. Comput. **26** (2004), 720–735.

[LM06] O. Le Maître, *Polynomial chaos expansion of a Lagrangian model for the flow around an airfoil*, Comptes Rendus de Mecanique **334** (2006), 693–698.

[LMK10] O. P. Le Maître and O. M. Knio, *Spectral Methods for Uncertainty Quantification*, Springer, 2010.

[LMKNG01] O. P. Le Maître, O. M. Knio, H. N. Najm, and R. G. Ghanem, *A stochastic projection method for fluid flow. I. Basic formulation*, J. Comput. Phys. **173** (2001), no. 2, 481–511.

[LMKNG04] ———, *Uncertainty propagation using Wiener-Haar expansions*, J. Comput. Phys. **197** (2004), no. 1, 28–57.

[LMNGK04] O. P. Le Maître, H. N. Najm, R. G. Ghanem, and O. M. Knio, *Multi-resolution analysis of Wiener-type uncertainty propagation schemes*, J. Comput. Phys. **197** (2004), no. 2, 502–531.

[LMRN+02] O. P. Le Maître, M. T. Reagan, H. N. Najm, R. G. Ghanem, and O. M. Knio, *A stochastic projection method for fluid flow. II. Random process*, J. Comput. Phys. **181** (2002), no. 1, 9–44.

[Loh96] W. L. Loh, *On Latin hypercube sampling*, Ann. Statist. **24** (1996), no. 5, 2058–2080.

[LRWW98] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, *Convergence properties of the Nelder-Mead simplex method in low dimensions*, SIAM Journal of Optimization **9** (1998), no. 1, 112 – 147.

[LU89] J. M. Lee and G. Uhlmann, *Determining anisotropic real-analytic conductivities by boundary measurements*, Comm. Pure Appl. Math. **42** (1989), no. 8, 1097–1112.

[MBC79] M. D. McKay, R. J. Beckman, and W. J. Conover, *A comparison of three methods for selecting values of input variables in the analysis of output from a computer code*, Technometrics **21** (1979), no. 2, 239–245.

[MNvST11] G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone, *Analysis of the Discrete L^2 Projection on Polynomial Spaces with Random Evaluations*, MOX-Report 46-2011, Department of Mathematics, Politecnico di Milano, Italy, 2011, submitted.

[MNvST13] ———, *Approximation of Quantities of Interest for stochastic PDEs by the random discrete L^2 projection on polynomial spaces*, SIAM Journal on Scientific Computing **35** (2013), A1440–A1460.

[MS12] J. L. Mueller and S. Siltanen, *Linear and Nonlinear Inverse Problems with Practical Applications*, SIAM Computational Science and Engineering, 2012.

[MT02] K. Moosegard and A. Tarantola, *Probabilistic Approach to Inverse Problems*, 2002.

[NT09] F. Nobile and R. Tempone, *Analysis and implementation issues for the numerical approximation of parabolic equations with random coefficients*, Internat. J. Numer. Methods Engrg. **80** (2009), no. 6-7, 979–1006.

[NTW08a] F. Nobile, R. Tempone, and C. G. Webster, *An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM J. Numer. Anal. **46** (2008), no. 5, 2411–2442.

[NTW08b] ———, *A sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM J. Numer. Anal. **46** (2008), no. 5, 2309–2345.

[RC04] C. P. Robert and G. Casella, *Monte Carlo statistical methods*, second ed., Springer Texts in Statistics, Springer-Verlag, New York, 2004.

[RHP08] G. Rozza, D. B. P. Huynh, and A. T. Patera, *Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations: application to transport and continuum mechanics*, Arch. Comput. Methods Eng. **15** (2008), no. 3, 229–275.

[Rud76] W. Rudin, *Principles of mathematical analysis*, Third ed., McGraw-Hill Book Co., New York, 1976.

[RW91] R. T. Rockafellar and R. J. B. Wets, *Scenarios and policy aggregation in optimization under uncertainty*, Math. Oper. Res. **16** (1991), no. 1, 119–147.

[RW98] ———, *Variational analysis*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 317, Springer-Verlag, Berlin, 1998.

[Sch10] O. Scherzer (ed.), *Handbook of Mathematical methods in Imaging*, Springer, 2010.

[SG04] C. Soize and R. Ghanem, *Physical systems with random uncertainties: chaos representations with arbitrary probability measure*, SIAM J. Sci. Comput. **26** (2004), no. 2, 395–410 (electronic).

[SG09] C. Soize and R. G. Ghanem, *Reduced chaos decomposition with random coefficients of vector-valued random variables and random fields*, Comput. Methods Appl. Mech. Engrg. **198** (2009), no. 21-26, 1926–1934.

[SG11] C. Schwab and C. J. Gittelson, *Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs*, Acta Numerica **20** (2011), 291–467.

[SJ94] I. H. Sloan and S. Joe, *Lattice methods for multiple integration*, Oxford Science Publications, The Clarendon Press Oxford University Press, New York, 1994.

[SMI00] S. Siltanen, J. Mueller, and D. Isaacson, *An implementation of the reconstruction algorithm of A Nachman for the 2D inverse conductivity problem*, Inverse Problems **16** (2000), 681–699.

[Smo63] S. Smolyak, *Quadrature and interpolation formulas for tensor products of certain classes of functions.*, Dokl. Akad. Nauk SSSR **4** (1963), 240–243.

[Soi10] C. Soize, *Identification of high-dimension polynomial chaos expansions with random coefficients for non-Gaussian tensor-valued random fields using partial and limited experimental data*, Comput. Methods Appl. Mech. Engrg. **199** (2010), no. 33-36, 2150–2164.

[Stu10] A. M. Stuart, *Inverse problems: a Bayesian perspective*, Acta Numer. **19** (2010), 451–559.

[Sud07] B. Sudret, *Global sensitivity analysis using polynomial chaos expansion*, Reliability Engineering and System Safety (2007).

[SVV+01] A. Seppänen, M. Vauhkonen, P. J. Vauhkonen, E. Somersalo, and J. P. Kaipio, *State estimation with fluid dynamical evolution models in process tomography—an application to impedance tomography*, Inverse Problems **17** (2001), no. 3, 467–483.

[SW98] I. H. Sloan and H. Woźniakowski, *When are quasi-Monte Carlo algorithms efficient for high-dimensional integrals?*, J. Complexity **14** (1998), no. 1, 1–33.

[Sze39] G. Szego, *Orthogonal polynomials*, Americal Mathematical Society, 1939.

[TA77] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems*, John Wiley & Sons Inc., 1977.

[Tar05] A. Tarantola, *Inverse problem theory and methods for model parameter estimation*, Society for Industrial and Applied Mathematics (SIAM), 2005.

[Vog02] C. R. Vogel, *Computational methods for inverse problems*, vol. 23, Society for Industrial and Applied Mathematics (SIAM), 2002.

[Wie38] N. Wiener, *The Homogeneous Chaos*, Amer. J. Math. **60** (1938), no. 4, 897–936.

[XH05] D. Xiu and J. S. Hesthaven, *High-order collocation methods for differential equations with random inputs*, SIAM J. Sci. Comput. **27** (2005), no. 3, 1118–1139 (electronic).

[Xiu09] D. Xiu, *Fast numerical methods for stochastic computations: a review*, Commun. Comput. Phys. **5** (2009), no. 2-4, 242–272.

[XK02a] D. Xiu and G. E. Karniadakis, *Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos*, Comput. Methods Appl. Mech. Engrg. **191** (2002), no. 43, 4927–4948.

[XK02b] ———, *The Wiener-Askey polynomial chaos for stochastic differential equations*, SIAM J. Sci. Comput. **24** (2002), no. 2, 619–644 (electronic).

[Zha02] D. Zhang, *Stochastic Methods for Flow in Porous Media: Coping with Uncertainties*, Academic Press, 2002.