

# Numerical Approximation of Flows in Random Porous Media

THÈSE N° 6860 (2016)

PRÉSENTÉE LE 12 FÉVRIER 2016  
À LA FACULTÉ DES SCIENCES DE BASE  
MATHICSE - GESTION  
PROGRAMME DOCTORAL EN MATHÉMATIQUES

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Francesco TESEI

acceptée sur proposition du jury:

Prof. V. Panaretos, président du jury  
Prof. F. Nobile, directeur de thèse  
Prof. O. Ernst, rapporteur  
Prof. R. Scheichl, rapporteur  
Prof. A. Abdulle, rapporteur



ÉCOLE POLYTECHNIQUE  
FÉDÉRALE DE LAUSANNE

Suisse  
2015



# Abstract

The objective of this thesis is to develop efficient numerical schemes to successfully tackle problems arising from the study of groundwater flows in a porous saturated medium; we deal therefore with partial differential equations (PDE) having random coefficients and we are interested in computing statistics related to specific quantities of interest (QoI), e.g. a linear functional of the solution of the PDE or the solution itself. We mainly consider the approximation of the pressure in the medium through a stochastic Darcy problem with random lognormally distributed permeability relying on Matérn-type covariance functions to take into account a wide range of possible smoothness of the permeability field. Once the problem has been reformulated in terms of a countable number of random variables, we analyze sparse grid polynomial approximations of the QoI. We propose different strategies to exploit the anisotropy of the QoI with respect to the different random entries; to this end we consider “a priori” and “a posteriori” strategies to drive the exploration of the multi-index set that defines the sparse grid, associating a profit to each multi-index either by using explicit theoretical estimates or by actually solving the PDE and computing on the fly the corresponding sparse grid interpolant. We show on several numerical examples the effectiveness of this strategy in treating the case of smooth permeability fields. In order to cover also the case of rough input permeabilities we consider, instead, Multi Level Monte Carlo techniques based on the use of a suitable control variate. Such a control variate is obtained from the solution of an auxiliary Darcy problem with a regularized input permeability which leads to pressure distributions that are smoother and less oscillatory than the original ones, but still highly correlated with them. We use then a sparse grid approximation to compute effectively the mean of the control variate and provide explicit bounds for the corresponding estimator as well as a complexity result.

We also consider groundwater transport problems and focus, in particular, on arrival times properly defined starting from particle trajectories driven by the stochastic Darcy velocity and subject to molecular diffusion taking place at porous level. In this case, by using suitable PDEs whose solution can be linked to specific expectations (with respect to all Brownian motions) thanks to the famous Feynman-Kac formula, we compute statistics of such arrival times, e.g. their expected value or the probability of exiting the physical domain in a given time horizon. We discuss several scenarios and readapt the methodologies previously developed involving adaptive sparse grid stochastic collocation and Monte Carlo type schemes to this case.

Key words: Stochastic Darcy Problem, Lognormal random fields, Multi Level Monte Carlo, Control variate, Stochastic Collocation, Adaptive Sparse Grid, Matérn covariance, Groundwater problems, Well catchments, Capture zones





# Contents

<b>Abstract</b>	<b>i</b>
<b>List of figures</b>	<b>v</b>
<b>List of tables</b>	<b>ix</b>
<b>Introduction</b>	<b>1</b>
Motivations . . . . .	1
Forward UQ Problem . . . . .	2
Outline . . . . .	3
<b>1 Thesis Overview</b>	<b>5</b>
1.1 Problem setting . . . . .	5
1.2 Adaptive sparse grids . . . . .	8
1.3 Monte Carlo and Multi Level Monte Carlo methods with Control Variate . . . . .	10
1.4 Contaminant Transport in Groundwater Flows . . . . .	12
<b>2 The Darcy Problem</b>	<b>19</b>
2.1 Notation . . . . .	19
2.2 Problem setting . . . . .	20
2.3 Finite Element Approximation . . . . .	26
2.4 The Truncated Darcy Problem . . . . .	27
2.5 Sampling schemes . . . . .	28
2.5.1 Karhunen-Loève expansion . . . . .	28
2.5.2 Fourier expansion . . . . .	29
<b>3 Sparse Grid polynomial approximation for random PDEs in high dimensional probability spaces</b>	<b>35</b>
3.1 Introduction . . . . .	35
3.2 Preliminaries: a brief overview on the problem . . . . .	37
3.3 Sparse grid approximations in finite dimensions . . . . .	39
3.4 Possible criteria to select the multi-index set $\mathbf{I}$ . . . . .	40
3.4.1 An “a priori” strategy: quasi-optimal sparse grids . . . . .	42
3.4.2 An “a posteriori” strategy: adaptive sparse grids . . . . .	45
3.5 A dimension-adaptive sparse grid algorithm . . . . .	48
3.6 Numerical Results . . . . .	49
3.6.1 Reference case: $\nu = 2.5$ , $\sigma = 1$ , $L_c = 0.5$ . . . . .	51
3.6.2 Permeabilities with high variability: $\sigma^2 = 4$ . . . . .	56
3.6.3 Conditioned Darcy problem . . . . .	58

## Contents

---

3.7	Conclusions . . . . .	58
<b>4</b>	<b>Monte Carlo-type schemes with Control Variate</b>	<b>61</b>
4.1	Introduction . . . . .	61
4.2	Monte Carlo method with control variate (MCCV) . . . . .	63
4.3	Error analysis of the MCCV method . . . . .	65
4.4	Multi Level Monte Carlo method with control variate (MLMCCV) . . . . .	70
4.5	Error analysis of the MLMCCV method . . . . .	72
4.6	The Algorithm . . . . .	76
4.6.1	Optimization Problem . . . . .	78
4.7	Complexity of the Multi Level Monte Carlo with Control Variate method . . . . .	79
4.8	Numerical Results . . . . .	82
4.8.1	A level-dependent approach . . . . .	89
4.9	Conclusions . . . . .	92
	<b>Appendices</b>	<b>93</b>
4.A	Proof of Lemma 4.3.1 . . . . .	93
4.B	Optimal rates in Theorem 4.3 . . . . .	99
<b>5</b>	<b>Extension to Transport Problem</b>	<b>101</b>
5.1	Introduction . . . . .	101
5.2	Problem setting . . . . .	102
5.2.1	Pure Dirichlet case: $\Gamma_D = \partial D$ . . . . .	105
5.2.2	Mixed case: $ \Gamma_N  > 0$ . . . . .	109
5.3	Case 1: an undisturbed flow from left to right . . . . .	109
5.3.1	Spatial Discretization . . . . .	111
5.3.2	Approximation of the mean FPT through MLMCCV . . . . .	112
5.3.3	Delineation of capture zones for small temporal horizon by stochastic collocation on mean FPT . . . . .	114
5.3.4	Numerical results . . . . .	116
5.4	Case 2: a flow induced by the presence of an extracting well . . . . .	119
5.4.1	A MLMC approach to approximate expected arrival times . . . . .	121
5.4.2	Delineation of capture zones for small temporal horizon . . . . .	123
5.4.3	Delineation of capture zones for infinite temporal horizon: a streamline approach	124
5.4.4	Numerical results . . . . .	127
5.5	Conclusions . . . . .	129
<b>6</b>	<b>Conclusions and future work</b>	<b>131</b>

# List of Figures

1.1	Case $\nu = 2.5$ . Adaptive sparse grid errors computed with respect to a reference set $\mathbf{I}^*$ as $ \mathcal{Q}_{W_{\mathbf{I}}^m}[Q] - \mathcal{Q}_{W_{\mathbf{I}^*}^m}[Q] $ for different sets $\mathbf{I}$ adaptively constructed. Work measured as number of linear solves. QoI considered: flux on the outlet. $D = (0, 1)^2$ , $\Gamma_D = \{x_1 = 0\} \cup \{x_1 = 1\}$ , $\Gamma_N = \{x_2 = 0\} \cup \{x_2 = 1\}$ , $g(\mathbf{x}) = 1 - x_1$ . . . . .	9
1.2	Error vs Computational cost for $\nu = 0.5$ (left) and $\nu = 2.5$ (right). Error = sparse grid error + statistical error. Computational cost evaluated as estimated CPU time. . . . .	13
1.3	Case $\nu = 2.5$ : on the left the sparse grid approximation of $\mathbb{E}^\omega[\bar{\tau}_{\mathbf{x}}^{\sigma'}(\mathbf{y})]$ ; on the right the sparse grid error committed on such quantity in the $L^2$ spatial norm versus the work evaluated as number of solves needed to build such approximation. $\sigma' = 0.01$ , $L_C = 0.5$ , $p_0 = 1$ . . . . .	15
1.4	Variance of $Q_\ell$ and $Y_\ell(\mathbf{x}, \omega) = Q_\ell(\mathbf{x}, \omega) - Q_{\ell-1}(\mathbf{x}, \omega)$ on each level, for $\nu = 0.5$ and different values of $\epsilon$ . $\sigma' = 0.01$ , $L_C = 0.5$ , $p_0 = 1$ . . . . .	16
1.5	Case $\nu = 2.5$ : MLMC approximation of mean (left) and variance (right) of $Q(\mathbf{x}, \omega) = \mathbb{E}^{\omega'}[\mathbf{1}_{\{\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T\}}]$ . $\sigma' = 0$ , $L_C = 0.5$ , $p_0 = 3$ , $T = 0.3$ , $L = 5$ . . . . .	16
1.6	Case $\nu = 2.5$ : MLMC approximation of mean (left) and variance (right) of $Q(\mathbf{x}, \omega) = \mathbb{E}^{\omega'}[\mathbf{1}_{\{\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T\}}]$ . $\sigma' = 0.01$ , $L_C = 0.5$ , $p_0 = 1$ , $q = 0.5$ , $r = 0.01$ , $T = 0.5$ . . . . .	17
2.1	Decay of $\lambda_n$ . Left: radial covariance function as in (2.5). Right: tensor covariance function as in (2.6). Top: $\nu = 2.5$ . Bottom: $\nu = 0.5$ . $d = 2$ . . . . .	30
3.1	Case $\nu = 2.5$ radial covariance, adaptive sparse grids error. . . . .	54
3.2	Case $\nu = 2.5$ tensor covariance, adaptive sparse grids error. . . . .	54
3.3	Case $\nu = 2.5$ , quasi-optimal sparse grids error obtained with tensor (left) and radial (right) covariance function. . . . .	54
3.4	Case $\nu = 2.5$ , radial covariance function. Interpolation error with different weight functions for the Quasi Optimal (OPT) and Adaptive ( profit indicators Deltaint (AD-D) and Weighted Linf / new points (AD-WLNP) ) cases. $work = W_{\mathbf{I},m}$ , $N^* = 60$ , $\text{card}(\mathcal{R}) = 10000$ . Dashed lines represent the slopes -0.5 (black) and -1 (orange). . . . .	55
3.5	Case $\nu = 2.5$ , tensor covariance function. Interpolation error with different weight functions for the Quasi Optimal (OPT) and Adaptive ( profit indicators Deltaint (AD-D) and Weighted Linf / new points (AD-WLNP) ) cases. $work = W_{\mathbf{I},m}$ , $N^* = 60$ , $\text{card}(\mathcal{R}) = 10000$ . Dashed lines represent the slopes -0.5 (black) and -1 (orange). . . . .	55
3.6	Tensor covariance: sparse grids error for different regularities of the input random field. . . . .	56
3.7	Case $\nu = 2.5$ , $\sigma = 2$ , adaptive sparse grids error. . . . .	57
3.8	Case $\nu = 2.5$ , $\sigma = 2$ , adaptive sparse grids error. . . . .	57
3.9	Case $\nu = 2.5$ , $\sigma = 2$ , quasi-optimal sparse grids error obtained with tensor (left) and radial (right) covariance function. . . . .	57
3.10	Conditioned case: $\nu = 2.5$ radial covariance, adaptive sparse grids error. . . . .	59

3.11	Conditioned case: $\nu = 2.5$ tensor covariance, adaptive sparse grids error. . . . .	59
3.12	Conditioned case: $\nu = 2.5$ , $\sigma = 2$ , quasi-optimal sparse grids error obtained with tensor (left) and radial (right) covariance function. . . . .	59
4.1	Three different regularizations of the same realization of $a$ . $\nu = 0.5$ , $L_c = 0.5$ , $\sigma = 1$ . . .	64
4.2	Variance of $Q_\ell^1(\omega) = \int_D \kappa(x) p_\ell(x, \omega) dx$ and $Y_\ell^1(\omega) = Q_\ell^1(\omega) - Q_{\ell-1}^1(\omega)$ on each level, for $\nu = 0.5$ and different values of $\epsilon$ . . . . .	84
4.3	Variance of $Q_\ell^1(\omega) = \int_D \kappa(x) p_\ell(x, \omega) dx$ and $Y_\ell^1(\omega) = Q_\ell^1(\omega) - Q_{\ell-1}^1(\omega)$ on each level, for $\nu = 2.5$ and different values of $\epsilon$ . . . . .	84
4.4	SC error computed as $\left( \mathbb{E}[(Q^1)_{h_L}^\epsilon] - \mathcal{Q}_W[(Q^1)_{h_L}^\epsilon] \right)^2$ versus computational cost, computed as number of sparse grid knots $W$ for $\nu = 0.5$ (left) and $\nu = 2.5$ (right). . . . .	85
4.5	Discretization error computed as $\mathbb{E}[Q_{h_\ell}^1 - Q_{h_L}^1]^2$ for $\nu = 0.5$ (left) and $\nu = 2.5$ (right). . .	85
4.6	Error vs Computational cost for $\nu = 0.5$ (left) and $\nu = 2.5$ (right). Error = Sparse Grid error + Statistical error, computed as root mean square error. Computational cost estimated according to (4.18) and (4.19), with $C_\ell = h_\ell^{-d\rho}$ and $\rho = 1.05$ , fitted by using a Matlab backslash command. . . . .	87
4.7	Variance of $Q_\ell^2(\omega) = -\int_D a(x, \omega) \frac{\partial p_\ell}{\partial x_1}(x, \omega) dx$ and of $Y_\ell^2 = Q_\ell^2 - Q_{\ell-1}^2$ , for $\nu = 0.5$ and different $\epsilon$ . . . . .	88
4.8	Variance of $Q_\ell^2(\omega) = -\int_D a(x, \omega) \frac{\partial p_\ell}{\partial x_1}(x, \omega) dx$ and of $Y_\ell^2 = Q_\ell^2 - Q_{\ell-1}^2$ , for $\nu = 2.5$ and different $\epsilon$ . . . . .	88
4.9	SC error computed as $\left( \mathbb{E}[(Q^2)_{h_L}^\epsilon] - \mathcal{Q}_W[(Q^2)_{h_L}^\epsilon] \right)^2$ versus computational cost, computed as number of sparse grid knots $W$ for $\nu = 0.5$ (left) and $\nu = 2.5$ (right). . . . .	89
4.10	Discretization error computed as $\mathbb{E}[Q_{h_\ell}^1 - Q_{h_L}^1]^2$ for $\nu = 0.5$ (left) and $\nu = 2.5$ (right). . .	89
4.11	Error vs Computational cost for $\nu = 0.5$ (left) and $\nu = 2.5$ (right). Error = Sparse Grid error + Statistical error, computed as root mean square error. Computational cost estimated according to (4.18) and (4.19), with $C_\ell = h_\ell^{-d\rho}$ and $\rho = 1.05$ , fitted by using a Matlab backslash command. . . . .	90
4.12	Variance of $Y_\ell^{LD} = Q_\ell^{LD} - Q_{\ell-1}^{LD}$ , with $Q_\ell^{LD}(\omega) = -\int_D a_{N_\ell}(x, \omega) \frac{\partial p_\ell^{LD}}{\partial x_1}(x, \omega) dx$ for $\nu = 0.5$ and different $\epsilon$ . . . . .	91
4.13	Variance of $Y_\ell^{LD} = Q_\ell^{LD} - Q_{\ell-1}^{LD}$ , with $Q_\ell^{LD}(\omega) = -\int_D a_{N_\ell}(x, \omega) \frac{\partial p_\ell^{LD}}{\partial x_1}(x, \omega) dx$ for $\nu = 2.5$ and different $\epsilon$ . . . . .	91
4.14	Error vs Computational cost for $\nu = 0.5$ (left) and $\nu = 2.5$ (right). Error = Sparse Grid error + Statistical error. Computational cost estimated according to (4.18) and (4.19). .	92
5.1	Two realizations of the mean FPT through the outlet in the vanishing diffusion case for $\nu = 0.5$ (left) and $\nu = 2.5$ (right). Observe that the random event $\mathbf{y}$ used to generate the realizations is the same. . . . .	113
5.2	$\mathbb{P}^{\omega'}(\tau_{\mathbf{x}, t}^{\sigma'}(\omega) \leq T   \omega)$ for different values of $t$ . $\sigma = 1$ , $p_0 = 2$ , $\sigma' = 0$ , $\nu = 0.5$ , $L_c = 0.5$ . . . .	115
5.3	Case $\nu = 2.5$ : on the left the sparse grid approximation of $\mathbb{E}^\omega[Q(\mathbf{x}, \cdot)]$ ; on the right the sparse grid error committed on such quantity in the $L^2$ spatial norm versus the work evaluated as number of solves needed to build such approximation. $\sigma' = 0.01$ , $L_c = 0.5$ , $p_0 = 1$ . . . . .	117
5.4	Variance of $Q_\ell$ and $Y_\ell(\mathbf{x}, \omega) = Q_\ell(\mathbf{x}, \omega) - Q_{\ell-1}(\mathbf{x}, \omega)$ on each level, for $\nu = 0.5$ and different values of $\epsilon$ . $\sigma' = 0.01$ , $L_c = 0.5$ , $p_0 = 1$ . . . . .	117
5.5	Variance of $Q_\ell$ and $Y_\ell(\mathbf{x}, \omega) = Q_\ell(\mathbf{x}, \omega) - Q_{\ell-1}(\mathbf{x}, \omega)$ on each level, for $\nu = 2.5$ and different values of $\epsilon$ . $\sigma' = 0.01$ , $L_c = 0.5$ , $p_0 = 1$ . . . . .	118

5.6	Variance of $Q_\ell$ and $Y_\ell(\mathbf{x}, \omega) = Q_\ell(\mathbf{x}, \omega) - Q_{\ell-1}(\mathbf{x}, \omega)$ on each level, for $\nu = 0.5$ and different values of $\epsilon$ . $\sigma' = 0$ , $L_C = 0.5$ , $p_0 = 1$ . . . . .	118
5.7	Variance of $Q_\ell$ and $Y_\ell(\mathbf{x}, \omega) = Q_\ell(\mathbf{x}, \omega) - Q_{\ell-1}(\mathbf{x}, \omega)$ on each level, for $\nu = 2.5$ and different values of $\epsilon$ . $\sigma' = 0$ , $L_C = 0.5$ , $p_0 = 1$ . . . . .	118
5.8	Case $\nu = 2.5$ : MLMC approximation of mean (left) and variance (right) of $Q(\mathbf{x}, \omega) = \mathbb{E}^{\omega'}[\mathbf{1}_{\{\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T\}}]$ . $\sigma' = 0$ , $L_C = 0.5$ , $p_0 = 3$ , $T = 0.3$ , $L = 5$ . . . . .	119
5.1	Realizations of $\mathbb{E}^{\omega'}[\min\{\tau_{\mathbf{x},t}^{\sigma'}(\mathbf{y}, \omega'), T\} \mathbf{y}]$ for $\nu = 2.5$ in the vanishing diffusion case (left) and for a relatively large diffusion (right). Observe that the random event $\mathbf{y}$ used to generate the Darcy velocity is the same. . . . .	123
5.2	Realization of $\mathbb{P}^{\omega'}(\tau_{\mathbf{x},t}^{\sigma'}(\mathbf{y}) \leq T \mathbf{y})$ for $\nu = 2.5$ in the vanishing diffusion case (left) and for a relatively large diffusion (right). Observe that the random event $\mathbf{y}$ used to generate the realizations is the same. . . . .	124
5.3	Realizations of $\Psi(\mathbf{x}, \omega)$ (left) and $\Psi_R(\mathbf{x}, \omega)$ (right) corresponding to the random event $\omega_i$ . . . . .	125
5.4	Case $\nu = 2.5$ : MLMC approximation of mean (left) and variance (right) of $Q(\mathbf{x}, \omega) = \mathbb{E}^{\omega'}[\mathbf{1}_{\{\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T\}}]$ . $\sigma' = 0.01$ , $L_C = 0.5$ , $p_0 = 1$ , $q = 0.5$ , $r = 0.01$ , $T = 0.5$ . . . . .	127
5.5	MLMC variance reduction (continuous line) versus MC variance (dashed line) in the case of level dependent diffusion (in blue, label LD) and fixed diffusion (in red) for $\nu = 0.5$ (right) and $\nu = 2.5$ (left). The variances have been computed with respect the $L^2$ spatial norm. $Q(\mathbf{x}, \omega) = \mathbb{E}^{\omega'}[\mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) \leq T\}}]$ . $\sigma' = 0.01$ , $L_C = 0.5$ , $p_0 = 1$ , $q = 0.5$ , $r = 0.01$ . . . . .	128
5.6	Case $\nu = 2.5$ : sparse grid approximation of the stream function $\Psi_R(\mathbf{x}, \omega)$ . $L_C = 0.5$ , $p_0 = 1$ . . . . .	129
5.7	Case $\nu = 2.5$ : MC approximation of mean (left) and variance (right) of $Q(\mathbf{x}, \omega) = \mathbf{1}_{\{\tau_{\mathbf{x},0}(\omega) \leq \infty\}}$ . $L_C = 0.5$ , $p_0 = 2$ . . . . .	129





# List of Tables

3.1 Common choices of univariate collocation points for sparse grids. . . . . 41







# Introduction

In this thesis we study and analyze from both, a theoretical and a numerical point of view, the so called groundwater flow problem. The problem consists in characterizing a single phase flow taking place in a saturated porous medium with random properties. The first step to make this problem suitable for a proper analysis is to choose an appropriate underlying mathematical model able to well represent the phenomenon; with such mathematical model at our disposal then, the process that should ideally lead us towards an accurate numerical approximation of a desired quantity of interest, requires to properly estimate all the parameters / quantities involved in the model and, finally, to perform a discretization of the problem and apply numerical methods to compute its approximate solution. Of course such process involves several steps and, a priori, in each of those steps there is a potential source of errors.

The problem of studying a single phase flow taking place in a saturated porous medium has a great importance in hydrology: in such context it is pretty common to face a lack of knowledge of the properties of the porous medium, typically an aquifer. This makes an accurate prediction of specific quantities of interest unfeasible and suggests to consider the problem in a more general framework, namely the *Uncertainty Quantification* (UQ) framework.

In the following sections we go through the motivations that inspired this work, we introduce the problem in such UQ framework and we detail the structure of the dissertation.

## Motivations

Our goal is to develop different tools to deal with the uncertainty in the properties of the subsurface. As previously anticipated, we have to face a lack of knowledge possibly due to different reasons, such as:

- lack of available measurements to accurately reconstruct some specific properties of the medium: permeability, porosity etc. can not be measured point-wise because of budget constraints, so, typically, we know their values only in a limited number of locations;
- errors coming from measurements: even if we were able to measure point-wise such properties we would have to deal with errors due to instruments' precision.

In the thesis we will always assume to have reliable mathematical models, i.e. not affected themselves by any uncertainty and potentially able to exactly describe the reality. At the same time we will assume to be able to characterize the uncertainty related to such specific properties in a probabilistic framework; this means that all the uncertain inputs we will deal with in this thesis will

be described in terms of suitable random variables or random fields. Throughout this work we will assume to know the law of such random input quantities.

Starting from these assumptions, we want to estimate how the uncertainty related to the input parameters propagates through the model and how it directly affects the desired quantity of interest (QoI), or system response (pressure or velocity).

## Forward UQ Problem

The general framework previously introduced has been widely adopted during the last few years. Since in such situation we assume to know the uncertainty related to the inputs of our system and we want to quantify the impact of such uncertainty on the outputs, we refer to this problem as a “forward uncertainty quantification problem”, in contrast to the so called “inverse uncertainty quantification problem”, in which we aim at recovering the properties of the medium by matching the values of the outputs of our system with the ones obtained with experimental measurements.

In this thesis we deal only with the former class of problems; the mathematical model used to represent the system will be always a suitable PDE, whose coefficients, boundary conditions or external forcing terms will be modeled as random quantities. In particular, we will always assume to be able to parametrize the input uncertainty with respect to a countable sequence of *i.i.d.* random variables  $y_n$ , with  $n \in \mathbb{N}$ .

The goal of solving such forward UQ problem is then to compute the law or some statistics of a QoI  $Q$  related to the output of the system that will depend on  $\mathbf{y} = \{y_n\}_{n \in \mathbb{N}}$ . Depending on the smoothness of the map  $\mathbf{y} \rightarrow Q(\mathbf{y})$  different strategies can be used in order to numerically solve the problem. In any case, to make the problem suitable for a numerical approximation, we need to consider only a finite number of random variables. The typical situation we will address is that of an input spatially distributed random field that can suitably be expanded in series involving random coefficients  $\{y_1, y_2, \dots\}$ . Since the series is convergent (in a proper norm) this naturally introduces an ordering and ranking of the random variables  $\{y_n\}_{n \in \mathbb{N}}$  whose influence on the QoI will become less and less important as  $n$  increases. This motivates truncating the series and retaining only a finite number  $N$  of the (most important) random variables. In some cases we will need to keep a large number  $N$  of random variables in order to accurately estimate the output of the system while in others it will suffice to consider only a moderate number  $N$ . This fact, in turn, significantly affects the choice of possible numerical methods that could be used to solve the problem.

Monte Carlo sampling [68] is the easiest approach to solve such forward UQ problems; it just requires generating  $M$  independent samples  $\{\mathbf{y}_i\}_{i=1}^M$ , solving the corresponding  $M$  PDEs and averaging over all the samples to obtain an estimation of the desired QoI  $Q$ . The method is very popular since it has a straightforward implementation and it is quite robust with respect to the size  $N$  of the vector  $\mathbf{y}$ . On the other hand it has the drawback of presenting poor performances in terms of convergence rate, since the associated error decays as  $\sigma / \sqrt{M}$ , being  $\sigma$  the standard deviation of  $Q(\mathbf{y})$ .

Many methods have been proposed to improve the convergence properties of MC; they can be essentially grouped in two categories: *sampling* and *deterministic* schemes. The first group includes mostly *MC*-type schemes, i.e. methods that generate different outputs every time they are launched due to their “random” (or “pseudo-random” nature), while in the second group we find mainly methods that use suitable grids built on specific classes of points to discretize the sample space

$\mathbb{R}^N$  and that reconstruct the solution by using projection or interpolation techniques and compute quantities of interest thanks to suitable quadrature formulas.

## Outline

The outline of the thesis is the following:

- **Chapter 1:** it represents the Thesis overview: after a very brief introduction of the Darcy problem the next subsections briefly report the main results obtained in the thesis that will be then fully detailed in the following chapters.
- **Chapter 2:** in this brief Chapter we introduce more in detail the stochastic Darcy Problem and the related theoretical results mainly concerning well posedness of the problem, regularity of the solution, finite element approximation and introducing the truncated Darcy problem, i.e.  $\mathbf{y} = \{y_n\}_{n=1}^N$ . In the last section of the chapter we present the sampling techniques that will be constantly used along the work.
- **Chapter 3:** here we deal with advanced sparse grid techniques supporting quadrature and interpolation on unbounded sets; in particular, we detail different strategies that allow to explore the anisotropy of the mapping  $\mathbf{y} \rightarrow Q(\mathbf{y})$ ,  $\mathbf{y} \in \mathbb{R}^N$ , and propose several indicators able to drive the exploration of the probabilistic space; then, in a second moment, we extend such routine to adaptively select the dimension  $N$  by relying on a certain ordering of the random variables. Numerical tests will be provided in order to show the effectiveness of the proposed methods.
- **Chapter 4:** here we use the methodology developed in the previous chapter and build an auxiliary smoothed Darcy problem that is used as a control variate to accelerate the computation of the original QoI. The Darcy problem is contextualized in a classical MC and in a more advanced Multi Level Monte Carlo (MLMC) framework; in both cases we use the control variate in order to reduce the variance of the corresponding estimator and the overall computational cost needed to achieve a desired tolerance. Numerical experiments and a comparison with MC and MLMC methods are reported.
- **Chapter 5:** in this chapter we develop some numerical tools to deal with groundwater transport problems. We consider suitable arrival times related to particle trajectories driven by the Darcy velocity and subject to molecular diffusion. We extend then the methodology previously developed to this class of problems and deal with problems such as the delineation of the so called *capture zones*, i.e. the zones starting from which a particle reaches a prescribed location of the aquifer, for instance an extracting pumping well, with a desired probability in a finite or infinite time horizon.

Most of the material presented in **Chapter 3** is based on the paper: F. Nobile, L. Tamellini, F. Tesei and R. Tempone, *An adaptive sparse grid algorithm for elliptic PDEs with lognormal diffusion coefficient*, and it is the last of a series of papers of my coauthors and other collaborators focusing on sparse grid approximations; based on these past works, in this paper my contribution has been mainly to develop the adaptive strategy for fixed  $N$  and its dimension adaptive counterpart to achieve an enhanced version of the code capable to deal also with high dimensional problems. The paper has

been accepted for publication as a proceeding of the workshop *Sparse Grid ans Application 2014*, held in Stuttgart; it is available as MATHISCE report 4/2015.

**Chapter 4** is based on the paper: F. Nobile and F. Tesei, *A Multi Level Monte Carlo method with control variate for elliptic PDEs with log-normal coefficients*. This work represents my main contribution in this Thesis: here we consider a particular MC-type estimator based on the introduction of a suitable control variate and provide a rigorous analysis of the associated statistical error as well as numerical results to show the effectiveness of this approach. The paper has been published on the Journal *Stochastic Partial Differential Equations: Analysis and Computations*.

The material in **Chapter 5** is still unpublished and it will be hopefully finalized in one or two publications in the next months. This is a joint work with F. Nobile and S. Krumscheid in which we deal with transport problems and propose basically two new approaches, one related to the use of the Feynman-Kac formula and one that uses a suitable streamline formulation, to efficiently compute specific quantities of interest which are particularly relevant in the hydrology field.

# 1 Thesis Overview

In this chapter, after having briefly introduced the stochastic Darcy problem, we put in evidence the main ideas and results of the Thesis. Each section refers to a specific subsequent Chapter. In Section 1.1 we introduce the Darcy problem and the main concepts and definitions that are needed to understand the next sections. Sections 1.2 and 1.3 present different methodologies able to efficiently solve the Darcy problem; in the former among deterministic methods we consider an adaptive sparse grid scheme to deal with the input uncertainty while in the latter we move to sampling schemes and consider a Multi Level Monte Carlo approach that is based on the use of a suitable control variate. In Section 1.4 we consider the problem of computing statistics related to the arrival time of particles to specific boundaries of the domain; we address here different scenarios, arising from hydrology applications, and focus on modeling and methodological issues.

## 1.1 Problem setting

In this section we provide all the necessary information to understand what is stated in the next sections of this Chapter. Therefore we will anticipate some of the results that then will be fully and rigorously addressed in the next Chapter 2.

Let us consider the mathematical problem given by the following elliptic stochastic Partial Differential Equation (PDE)

$$\begin{cases} -\operatorname{div}(a(\mathbf{x}, \omega) \nabla p(\mathbf{x}, \omega)) = f(\mathbf{x}) & \mathbf{x} \in D, \\ p(\mathbf{x}, \omega) = g(\mathbf{x}) & \mathbf{x} \in \Gamma_D, \\ \nabla p(\mathbf{x}, \omega) \cdot \mathbf{n} = 0 & \mathbf{x} \in \Gamma_N, \end{cases} \quad (1.1)$$

where  $D \in \mathbb{R}^N$  is a bounded open domain,  $\Gamma_D$  and  $\Gamma_N$  represent the Dirichlet and Neumann boundaries and  $g$  the corresponding Dirichlet datum. We are interested in studying equation (1.1) since it is widely used to describe the Darcy flow taking place in saturated porous media in which  $p$  represents the pressure and  $a$  plays the role of the permeability of the medium. In typical situations the input diffusion coefficient  $a$  (permeability) is not entirely known; hence, to take into account in the model this lack of knowledge, the permeability is modeled as a random field indexed with respect to the spatial variable  $\mathbf{x}$  and a random event  $\omega$ : this turns (1.1) into a random PDE (PDE with random input data). In particular  $\forall \mathbf{x}_i \in D$  the quantity  $a(\mathbf{x}_i, \omega)$  represents a positive real-valued random variable defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . By denoting with  $H^1(D)$  the space of

square integrable functions with square integrable partial derivatives on  $D$ , and with  $H_{\Gamma_D}^1(D)$  the space of  $H^1(D)$ -functions that vanish on the Dirichlet boundary  $\Gamma_D$ , the variational formulation associated with problem (1.1) is: find  $p \in H^1(D)$  such that  $p = g$  on  $\Gamma_D$  and

$$b_\omega(p(\cdot, \omega), v(\cdot)) = L(v(\cdot)), \quad \forall v \in H_{\Gamma_D}^1(D),$$

where the bilinear form  $b_\omega$  (parametrized by  $\omega$ ) and the linear functional  $L$  are defined as:

$$b_\omega(u, v) = \int_D a(\mathbf{x}, \omega) \nabla u(\mathbf{x}, \omega) \nabla v(\mathbf{x}) d\mathbf{x}, \quad L(v) = \int_D f(\mathbf{x}) v(\mathbf{x}) d\mathbf{x}.$$

Assuming to have sufficient regularity on the forcing term  $f$  and on the domain  $D$ , for any realization of the random field, i.e. by fixing a particular event  $\omega_i$ , and provided that the diffusion coefficient  $a$  is regular enough, it is possible to directly apply Lax-Milgram theorem to straightforwardly deduce existence and uniqueness of the solution  $p$  in the space  $H^1(D)$ . However, since in hydrology applications the permeability is often modeled as a lognormally distributed random field, we are not able to obtain a uniform coercivity of the bilinear form  $b_\omega$  with respect to  $\omega$ , and hence, in order to extend classical results on elliptic problems to this uncertain scenario, we need to ensure some  $L_{\mathbb{P}}^q(\Omega)$ -integrability of the continuity and coercivity constants related to the bilinear form  $b_\omega$ , where  $L_{\mathbb{P}}^q(\Omega)$  denotes the space of functions which are  $q$ -integrable with respect to the probability measure  $\mathbb{P}$ , i.e.

$$L_{\mathbb{P}}^q(\Omega) = \left\{ f : \Omega \rightarrow \mathbb{R} \text{ s.t. } \int_{\Omega} |f(\omega)|^q \mathbb{P}(d\omega) < \infty \right\}.$$

We can therefore say that equation (1.1) holds in an almost sure sense (a.s.) and that it admits a unique solution in the Bochner space  $L_{\mathbb{P}}^q(\Omega, H^1(D))$ .

For any positive  $\alpha$  we write  $\alpha = k + s$  with  $k \in \mathbb{N}$  and  $s \in (0, 1]$ , and we denote with  $C^\alpha(\bar{D})$  the space of Hölder continuous functions having continuous derivatives of order smaller or equal  $k$  and  $k$ -th order derivatives Hölder continuous with parameter  $s$ , i.e. the space of functions such that the following (spatial) norm is bounded:

$$\|v\|_{C^\alpha(\bar{D})} = \|v\|_{C^k(\bar{D})} + |v|_{C^\alpha(\bar{D})} = \sum_{j=1}^k \max_{|i|_1=j} \|D^i v\|_{C^0(\bar{D})} + \max_{|i|_1=k} \sup_{x, y \in \bar{D}} \frac{|D^i v(x) - D^i v(y)|}{|x - y|^s}.$$

Then, by further assuming  $a$  to belong to the Banach-space valued Bochner space  $L_{\mathbb{P}}^q(\Omega, C^\alpha(\bar{D}))$ , it is possible to obtain regularity results that bound Sobolev norms of the solution by Hölder norms of the diffusion coefficient  $a$ , i.e. an estimate of the type (see subsequent Lemma 2.2.1):

$$\|p(\cdot, \omega)\|_{H^{1+\beta}(D)} \leq \frac{1}{\alpha - \beta} C_1(\omega, \alpha) \|f\|_{H^{\alpha-1}(D)}, \quad \forall 0 < \beta < \alpha \quad \text{a.s. in } \Omega,$$

where  $C(\omega, \alpha)$  depends on the  $C^\alpha(\bar{D})$  norm of  $a$ .

In this thesis we will always consider a Matérn type covariance function [32, 72] for the log permeability  $\gamma(\mathbf{x}, \omega) = \log(a(\mathbf{x}, \omega))$ ; in particular we will consider either a stationary isotropic covariance function defined as

$$\text{cov}_\gamma(\mathbf{x}, \mathbf{x}') = \widetilde{\text{cov}}_\gamma(|\mathbf{x} - \mathbf{x}'|) = \frac{\sigma^2}{\Gamma(\nu) 2^{\nu-1}} \left( \sqrt{2\nu} \frac{|\mathbf{x} - \mathbf{x}'|}{L_c} \right)^\nu K_\nu \left( \sqrt{2\nu} \frac{|\mathbf{x} - \mathbf{x}'|}{L_c} \right), \quad \nu \geq 0.5,$$

or a tensor Matérn covariance function:

$$\text{cov}_\gamma(\mathbf{x}, \mathbf{x}') = \sigma^2 \prod_{i=1}^d \frac{\left( \sqrt{2\nu} \frac{|x_i - x'_i|}{L_c} \right)^\nu K_\nu \left( \sqrt{2\nu} \frac{|x_i - x'_i|}{L_c} \right)}{\Gamma(\nu) 2^{\nu-1}}, \quad \nu \geq 0.5.$$

The isotropic Matérn is the model more commonly used in applications while the tensor covariance model, although easier to implement from a numerical point of view, is often considered not suited to model physical processes because of the dependence of such a model on the choice of the axes; nevertheless a tensor covariance function can be used when a particular axis dependence is a priori known, as for instance in the modeling of stratified media. As will be fully detailed in Chapter 2, the regularity of these covariance functions (and then of the corresponding realizations of the log-permeability  $\gamma$ ) covers a wide range of values; in particular the parameter  $\nu$  governs the smoothness of both, the covariance function  $\text{cov}_\gamma$  and the realizations  $\gamma(\mathbf{x}, \omega_i)$ . We have the two extreme cases when  $\nu = 0.5$  and when  $\nu = \infty$ ; in the former we obtain a Lipschitz covariance function which generates realizations that are Hölder continuous with parameter  $\alpha < 0.5$  while in the latter we have an analytic covariance function leading to infinitely differentiable realizations of  $\gamma$ .

To numerically approximate the elliptic problem (1.1) in space we use linear finite elements. The choice of low order finite elements is motivated by the fact that we want to address also the case of low regularity permeability fields,  $a \in C^\alpha(\bar{D})$  with  $\alpha < \nu < 1$ . By using the previous regularity estimates, and by denoting with  $p_h$  the finite element solution of problem (1.1), a result (see subsequent Lemma 2.3.1) stating the rate of convergence of the finite element solution  $p_h$  to the true solution  $p$  can be derived:

$$\|p(\cdot, \omega) - p_h(\cdot, \omega)\|_{H^1(D)} \leq \frac{1}{\alpha - \beta} C_2(\omega, \alpha) \|f\|_{H^{\alpha-1}(D)} h^\beta, \quad \forall 0 \leq \beta < \alpha \quad \text{a.s. in } \Omega.$$

In this thesis, we will be often interested in computing specific quantities of interest  $Q$ , in our case linear functionals, related to the solution of the PDE,  $p$ . It is then possible to generalize the previous convergence estimate also to linear functionals:

$$|Q(p - p_h)(\omega)| \leq \frac{C_3(\omega, \alpha)}{(\alpha - \beta)^2} \|f\|_{H^{\alpha-1}(D)} \|Q\|_{H^{\alpha-1}(D)} h^{2\beta}, \quad \forall 0 \leq \beta < \alpha \quad \text{a.s. in } \Omega.$$

Such a bound basically tells us that, by considering linear functionals instead of the solution itself, we double the convergence rate thanks to the Galerkin orthogonality. Observe that all the constants appearing in the previous bounds are  $L^q_{\mathbb{P}}$ -integrable.

The second step concerns the approximation in the stochastic variable  $\omega$ . To do this, we first parametrize the input randomness with respect to a countable number of (possibly independent) random variables  $y_n$ ,  $n \in \mathbb{N}$ . In Section 2.4 of the next Chapter we will give further details on this; for the moment it suffices to say that, from a practical point of view, we will truncate this countable expansion by keeping only the first  $N$  most important random variables and we collect them in the vector  $\mathbf{y}$ . Then, we replace the exact quantity of interest  $Q = Q(\omega)$  with its finite dimensional counterpart  $Q = Q(\mathbf{y})$ .

## 1.2 Adaptive sparse grids

Given an input random vector  $\mathbf{y} \in \Gamma \subset \mathbb{R}^N$  in our system, we are interested in evaluating the corresponding QoI  $Q(\mathbf{y})$  starting from the solution of the truncated Darcy Problem. The goal in many situations is then to accurately estimate the expected value  $\mathbb{E}[Q]$  or some higher moments. By denoting with  $\rho$  the joint probability distribution of the Gaussian standard random vector  $\mathbf{y}$  we can write the expectation as

$$\mathbb{E}[Q] = \int_{\Gamma} Q(\mathbf{y}) \rho(\mathbf{y}) d\mathbf{y}.$$

The idea of sparse grid stochastic collocation methods is to derive a polynomial approximation of the map  $\mathbf{y} \rightarrow Q(\mathbf{y})$ , by solving the truncated Darcy problem in a suitable set of collocation points and by reconstructing the overall solution as a sum over the collocation points of the product between the corresponding solution of the Darcy problem and polynomial functions in the variable  $\mathbf{y}$ .

The general construction of sparse grid approximations will be detailed in chapter 3. Here we just sketch the method and highlight the main contributions obtained in this thesis.

By following the classical sparse grid approach [3] we define the sparse grid interpolant of  $Q(\mathbf{y})$  as

$$\mathcal{S}_{\mathbf{I}}^m[Q](\mathbf{y}) = \sum_{\mathbf{i} \in \mathbf{I}} \Delta^{m(\mathbf{i})}[Q](\mathbf{y}) \quad (1.2)$$

where  $\mathbf{I}$  is a suitable multi-index set, and  $\Delta^{m(\mathbf{i})}$  is the *hierarchical surplus* operator defined as

$$\Delta^{m(\mathbf{i})} = \bigotimes_{n=1}^N \Delta^{m(i_n)} = \bigotimes_{n=1}^N \left( \mathcal{U}_n^{m(i_n)} - \mathcal{U}_n^{m(i_n-1)} \right),$$

with  $\mathcal{U}_n^{m(i_n)}$  a one-dimensional interpolant operator in the variable  $y_n$  over  $m(i_n)$  points. With the sparse grid interpolant at our disposal, it is then possible to approximate the mean of our QoI through a sparse grid quadrature formula

$$\int_{\mathbb{R}^N} Q(\mathbf{y}) \rho(\mathbf{y}) d\mathbf{y} \approx \int_{\mathbb{R}^N} \mathcal{S}_{\mathbf{I}}^m[Q] \rho(\mathbf{y}) d\mathbf{y} = \sum_{j=1}^{W_{\mathbf{I}}^m} Q(\mathbf{y}_j) \beta_j = \mathcal{Q}_{W_{\mathbf{I}}^m}[Q],$$

where  $W_{\mathbf{I}}^m$  is the cardinality of the sparse grid,  $\mathbf{y}_i$  are suitably chosen collocation points and  $\beta_j$  the corresponding weights.

Once the choice of the collocation points has been made, the only quantity left to be defined in (1.2) in order to compute the sparse grid interpolant and quadrature of  $Q$ , is the multi-index set  $\mathbf{I}$  upon which to perform the summation.

At this stage, we aim at adaptively choosing the best multi-index set  $\mathbf{I}$ , based on a suitable definition of profits of the multi-indexes  $\mathbf{i} \in \mathbb{N}_+^N$ . In particular, for any multi-index  $\mathbf{i}$ , we define its profit  $P(\mathbf{i})$  as the ratio

$$P(\mathbf{i}) = \frac{\Delta E(\mathbf{i})}{\Delta W(\mathbf{i})}$$

where  $\Delta E(\mathbf{i})$  represents the error that we would commit by omitting the multi-index from the sparse grid interpolant, and it is related to the size of the hierarchical surplus  $\Delta^{m(\mathbf{i})}[Q](\mathbf{y})$ , while  $\Delta W(\mathbf{i})$  represents the additional work, in terms of linear solves of the PDE, that would require to add  $\mathbf{i}$  to the multi-index set  $\mathbf{I}$ . We will present two possible strategies to build the multi-index set  $\mathbf{I}$  with



“a priori” and “a posteriori” procedures. In the first case the construction of  $\mathbf{I}$  is based on a priori estimates or good ansatz for the estimation of  $\Delta E(\mathbf{i})$ ; having such computable estimates of  $\Delta E(\mathbf{i})$  at our disposal the set  $\mathbf{I}$  is built by picking the multi-indices  $\mathbf{i}$  having profit larger than a prescribed (small) threshold  $\epsilon$ , namely:  $\mathbf{I} = \{\mathbf{i} \in \mathbb{N}^N \text{ s.t. } \frac{\Delta E(\mathbf{i})}{\Delta W(\mathbf{i})} > \epsilon\}$ . In the second case we adaptively explore the multi-index space by looking at the neighbors of the current set  $\mathbf{I}$  and actually solving the PDE in a certain number of additional collocation points to estimate  $\Delta E(\mathbf{i})$ . We mainly focus on the second approach, by performing the sparse grid approximation for different choices of collocation points and different estimates of  $\Delta E(\mathbf{i})$ .

In order to keep as small as possible the number of random variables activated during the sparse grid construction we expand the random field in a Karhunen Loève basis. Moreover, since the  $KL$  expansion somehow introduces a “weak ordering” of the random variables  $y_1, y_2, \dots, y_n, \dots$ , i.e. there exist  $N_b \geq 1$  (buffer) s.t.  $y_{n+N_b}$  is guaranteed to be less important than  $y_n$ , we implemented the algorithm in an  $N$ -adaptive fashion, by gradually adding random variables whenever one of the “buffered” variables is activated. Note that this means that we no longer need to truncate “a priori” the  $KL$  expansion.

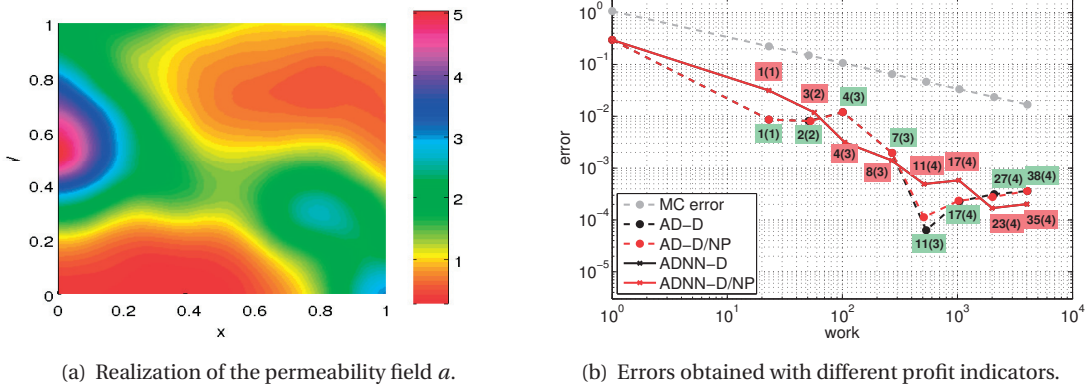


Figure 1.1 – Case  $\nu = 2.5$ . Adaptive sparse grid errors computed with respect to a reference set  $\mathbf{I}^*$  as  $|\mathcal{Q}_{W_{\mathbf{I}^*}^m}[Q] - \mathcal{Q}_{W_{\mathbf{I}^*}^m}[Q]|$  for different sets  $\mathbf{I}$  adaptively constructed. Work measured as number of linear solves. QoI considered: flux on the outlet.  $D = (0, 1)^2$ ,  $\Gamma_D = \{x_1 = 0\} \cup \{x_1 = 1\}$ ,  $\Gamma_N = \{x_2 = 0\} \cup \{x_2 = 1\}$ ,  $g(\mathbf{x}) = 1 - x_1$ .

This strategy gives satisfying results when applied to a smooth input random field. In Figure 1.1 we consider a random field  $\gamma$  with tensor covariance function (2.6) and parameters  $\sigma = 1$ ,  $L_c = 0.5$  and show the convergence of the adaptive scheme when using nested Kronrod Patterson Gauss nodes and non nested (suffix “NN” in Figure 1.1) Hermite nodes. The labels with the factor “NP” indicate that the adaptive algorithm is based on estimated profits  $P(\mathbf{i})$ , whereas those without “NP” refer to algorithms based on estimated errors  $\Delta E(\mathbf{i})$  only. Finally the numbers next to each point on the error curve give information about the shape of the multi-index sets  $\mathbf{I}(w)$  generated by the adaptive algorithm: the first number (outside the brackets) indicates the number of active directions, while the second number (inside the brackets) denotes the maximum number of directions that have been activated at the same time.

In Chapter 3 we detail the dimension adaptive algorithm developed and present several other test cases, comparing “a priori” and “a posteriori” sparse grid approximations on unbounded sets, including different choices of covariance function (tensor / radial Matérn), collocation points

(nested / non nested) and profits indicators. Moreover we perform some tests to understand how the variance  $\sigma^2$  of the random field  $\gamma$  influences the performance of the method. Finally we study also the case in which we have a number  $K$  of permeability measurements, in  $K$  locations  $\mathbf{x}_1, \dots, \mathbf{x}_K$  over the domain  $D$ , and we include them into the model by considering a conditioned log-permeability field  $\gamma_{cond}(\mathbf{x}, \omega) = \gamma(\mathbf{x}, \omega) | \gamma(\mathbf{x}_1, \omega) = \gamma_1, \dots, \gamma(\mathbf{x}_K, \omega) = \gamma_K$ . In this case we are interested in quantifying the effects of the measurements on the performance of the adaptive sparse grid approximations.

### 1.3 Monte Carlo and Multi Level Monte Carlo methods with Control Variate

The convergence properties of the methodology so far introduced are satisfactory when considering sufficiently smooth input random fields, while they deteriorate when dealing with rough input permeabilities. Since in the latter case a moderately small number of random variables is not sufficient anymore to guarantee that the truncation error is negligible, we need necessarily to deal with high dimensional problems in the stochastic variable  $\mathbf{y}$ . Also advanced sparse grid techniques such as the one mentioned in Section 1.2, suffer when dealing with problems characterized by low regularity and intrinsic high dimensionality; it seems natural then to switch to Monte Carlo type methods to solve problem (1.1). The MC estimator of  $Q$  is given by  $\hat{Q}_{h,M}^{MC} = \frac{1}{M} \sum_{i=1}^M Q_h(\mathbf{y}_i)$ , being  $\mathbf{y}_i$ ,  $i=1, \dots, M$ , identically distributed draws from  $\rho(\mathbf{y})$ , and its mean square error is

$$e(\hat{Q}_{h,M}^{MC})^2 := \mathbb{E}[(\hat{Q}_{h,M}^{MC} - \mathbb{E}[Q])^2] = \frac{\text{Var}(Q_h)}{M} + (\mathbb{E}[Q_h - Q])^2,$$

where the first term represents the statistical error coming from the sampling and the second term the bias due to the finite element discretization. Observe that here we are assuming to include a number  $N$  of random variables  $y_n$ ,  $n = 1, \dots, N$ , large enough to consider negligible the truncation error. Unfortunately this strategy, although robust with respect to  $N$ , presents a slow convergence rate with respect to the sample size  $M$ .

In Chapter 4 we propose and analyze a method that combines the advantages of a sparse grid approximation for smooth fields with the robustness of MC samplers to treat effectively the case of rough random fields (hence high dimensional problems). More specifically, since we know that sparse grid schemes are well suited to solve the problem when the input permeability field is smooth, we build an auxiliary Darcy problem having a smoothed permeability  $a^\epsilon = e^{\gamma^\epsilon}$  obtained through convolution of the original field  $\gamma$  with a Gaussian kernel  $\phi_\epsilon$ , namely:

$$\gamma^\epsilon(\mathbf{x}, \omega) = \gamma(\mathbf{x}, \omega) * \phi_\epsilon(\mathbf{x}) = \gamma(\mathbf{x}, \omega) * \frac{e^{-\frac{\|\mathbf{x}\|^2}{2\epsilon^2}}}{(2\pi\epsilon^2)^{\frac{d}{2}}}.$$

Then, since for small values of  $\epsilon$  we expect the solutions of the original and of the auxiliary Darcy problem to be highly correlated, we use the smoothed QoI  $Q^\epsilon$ , obtained starting from the solution  $p^\epsilon$  of the auxiliary Darcy problem, as control variate in a MC sampling strategy. More precisely we write the mean of our QoI as

$$\mathbb{E}[Q(\mathbf{y})] = \mathbb{E}[Q(\mathbf{y}) - Q^\epsilon(\mathbf{y})] + \mathbb{E}[Q^\epsilon(\mathbf{y})]$$

and we estimate the first expectation via MC sampling and the second one with a sparse grid scheme which is reasonable since  $Q^\epsilon(\mathbf{y})$  is smooth with respect to the stochastic parameters. The resulting

MC estimator with control variate (MCCV) reads

$$\hat{Q}_{h,M,W}^{MCCV} = \frac{1}{M} \sum_{i=1}^M (Q_h(\mathbf{y}_i) - Q_h^\epsilon(\mathbf{y}_i)) + \mathcal{Q}_W[Q_h^\epsilon], \quad (1.3)$$

where  $W$  represents the number of PDE solves needed to build the sparse grid approximation and again  $\mathbf{y}_i$  are i.i.d. draws from a standard multivariate Gaussian distribution, and  $Q_h(\mathbf{y}_i)$  the corresponding realization of the QoI approximated by finite elements. The corresponding mean square error can be bounded by

$$e(\hat{Q}_{h,M,W}^{MCCV})^2 := \mathbb{E}[(\hat{Q}_{h,M,W}^{MCCV} - \mathbb{E}[Q])^2] \leq \frac{\text{Var}(Q_h - Q_h^\epsilon)}{M} + 2(\mathcal{Q}_W[Q_h^\epsilon] - \mathbb{E}[Q_h^\epsilon])^2 + 2(\mathbb{E}[Q_h - Q])^2.$$

With respect to the standard MC case, we have an additional error term given by the sparse grid error: the idea is then to choose  $\epsilon$  sufficiently large to have good convergence properties of the sparse grid approximation, and at the same time sufficiently small to get a variance reduction sufficiently large to compensate for the introduction of the new error term.

In Theorem 4.1 we show how the term  $\text{Var}(Q_h - Q_h^\epsilon)$  scales with respect to  $\epsilon$  and the regularity of the original random permeability field  $a$ : in particular, given a permeability obtained from a Matérn covariance of parameter  $\nu$ , we prove the following (conservative) bound

$$\text{Var}(Q_h - Q_h^\epsilon) \leq \mathbb{E}[(Q_h - Q_h^\epsilon)^2] \lesssim \epsilon^{2\min(2,\alpha)}, \quad \forall \alpha < \nu.$$

At the same time, for sufficiently large values of  $\epsilon$  and for input permeability with limited regularity, we numerically observe an algebraic decay of the sparse grid error, with convergence rate depending on  $\epsilon$  and degenerating for  $\epsilon \rightarrow 0$ , namely  $|\mathcal{Q}_W[Q_h^\epsilon] - \mathbb{E}[Q_h^\epsilon]| \lesssim W^{-\eta(\epsilon)}$  for some  $\eta(\epsilon) > 0$ .

By properly choosing the parameter  $\epsilon$ , the sample size  $M$  and the number of sparse grid knots  $W$ , this strategy is shown to significantly reduce the computational cost needed to achieve a prescribed tolerance, in terms of mean square error, compared to the standard MC method. In particular, since the sparse grid converges with a better rate than the MC sampler, asymptotically the ratio  $W/M$  tends to zero, when considering smaller and smaller tolerances, and the computational gain is  $\mathcal{O}(\epsilon^{2\min(2,\alpha)}) \forall \alpha < \nu$ .

The same idea has been extended to a Multi Level framework; in this case, by introducing a sequence of increasingly fine triangulations  $\mathcal{T}_{h_\ell}$ ,  $\ell = 0, \dots, L$ , having mesh sizes  $h_0 > h_1 > \dots > h_L > 0$ , and exploiting the linearity of the expectation operator, we write the expectation of  $Q$  on the finest level as:

$$\mathbb{E}[Q_L] = \underbrace{\mathbb{E}[Q_0] + \sum_{\ell=1}^L \mathbb{E}[Q_\ell - Q_{\ell-1}]}_{(i)} = \mathbb{E}[Q_0 - Q_0^\epsilon] + \underbrace{\sum_{\ell=1}^L \mathbb{E}[Q_\ell - Q_{\ell-1} - (Q_\ell^\epsilon - Q_{\ell-1}^\epsilon)]}_{(ii)} + \mathbb{E}[Q_L^\epsilon].$$

By independently estimating the terms coming from different levels in (i) with MC estimators we obtain the so called Multi Level Monte Carlo estimator (MLMC)

$$\hat{Q}_{\{h_\ell\}, \{M_\ell\}}^{MLMC} = \sum_{\ell=0}^L \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} (Q_{h_\ell}(\mathbf{y}_{\ell,i}) - Q_{h_{\ell-1}}(\mathbf{y}_{\ell,i})), \quad \text{with } Q_{h_{-1}}(\mathbf{y}_{\ell,i}) = 0,$$

while by independently estimating the terms coming from different levels in (ii) with MC estimators

and the last expectation in (ii) with sparse grid quadrature we obtain the MLMC estimator with control variate (MLMCCV)

$$\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV} = \sum_{\ell=0}^L \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} \left( Q_{h_\ell}(\mathbf{y}_{\ell,i}) - Q_{h_{\ell-1}}(\mathbf{y}_{\ell,i}) - (Q_{h_\ell}^\epsilon(\mathbf{y}_{\ell,i}) - Q_{h_{\ell-1}}^\epsilon(\mathbf{y}_{\ell,i})) \right) + \mathcal{Q}_W[Q_{h_L}^\epsilon], \quad (1.4)$$

again with  $Q_{h_{-1}}(\mathbf{y}_{\ell,i}) = Q_{h_{-1}}^\epsilon(\mathbf{y}_{\ell,i}) = 0$ .

The mean square error related to the MLMC estimator is

$$e(\hat{Q}_{\{h_\ell\}, \{M_\ell\}}^{MLMC})^2 := \mathbb{E}[(\hat{Q}_{\{h_\ell\}, \{M_\ell\}}^{MLMC} - \mathbb{E}[Q])^2] = \sum_{\ell=0}^L \frac{\text{Var}(Q_{h_\ell} - Q_{h_{\ell-1}})}{M_\ell} + (\mathbb{E}[Q_h - Q])^2.$$

The idea of the method is to take advantage of the computations done on the coarsest levels which are, computationally cheap, and take only few samples on the finest levels, where the associated variance  $\text{Var}(Q_{h_\ell} - Q_{h_{\ell-1}})$  is expected to be small. By doing so it is possible to achieve the same overall tolerances, in terms of mean square error, of a standard MC scheme, with a much smaller computational cost. Our control variate approach produces an estimator whose mean square error can be bounded as

$$\begin{aligned} e(\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV})^2 &:= \mathbb{E}[(\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV} - \mathbb{E}[Q])^2] \leq \sum_{\ell=0}^L \frac{\text{Var}(Q_{h_\ell} - Q_{h_{\ell-1}} - (Q_{h_\ell}^\epsilon - Q_{h_{\ell-1}}^\epsilon))}{M_\ell} \\ &\quad + 2 \left( \mathcal{Q}_W[Q_{h_L}^\epsilon - \mathbb{E}[Q_{h_L}]] \right)^2 + 2 (\mathbb{E}[Q_{h_L} - Q])^2. \end{aligned}$$

Again, in order to compare the performances of the MLMCCV approach with the MLMC one, we need to compare the statistical errors: it has been proven that the variance of the difference between QoI computed in two successive levels, when using linear finite elements for the spatial discretization, scales as  $\text{Var}(Q_{h_\ell} - Q_{h_{\ell-1}}) \lesssim h_\ell^{2\alpha}$  for any  $\alpha < \min(\nu, 1)$ . On the other hand, in Theorem 4.3, we show that the variance of the double difference appearing in the MLMCCV estimator mean square error scales as

$$\text{Var}(Q_{h_\ell} - Q_{h_{\ell-1}}) \lesssim h_\ell^\alpha \inf_{\substack{\beta < \min(\nu, 1) \\ 0 < \eta + \beta \leq \nu}} \frac{h_\ell^\beta \epsilon^{\nu - \beta - \eta}}{(\nu - \beta)^2 \sqrt{\eta}} h_\ell^{2\alpha} \text{ for any } \alpha < \min(\nu, 1).$$

Numerical tests show that this strategy improves the performance of the MLMC method when applied to problem with both, rough and smooth permeability, see Figure 1.2. The computational cost appearing in Figure 1.2, has been obtained after having optimized over the sample sizes  $M_\ell$  for  $\ell = 0, \dots, L$  and the number of sparse grid knots  $W$  to achieve optimal complexity.

In Chapter 4 we go through a rigorous analysis of the statistical error related to the estimators (1.3) and (1.4) using the control variate approach and we present again several numerical results obtained with different scenarios.

## 1.4 Contaminant Transport in Groundwater Flows

After having deeply investigated the flow problem governed by the Darcy law in different frameworks, in Chapter 5 we aim at extending the methodology introduced so far to study contaminant transport phenomena associated with groundwater flows. Starting from the Darcy problem, we want to model

## 1.4. Contaminant Transport in Groundwater Flows

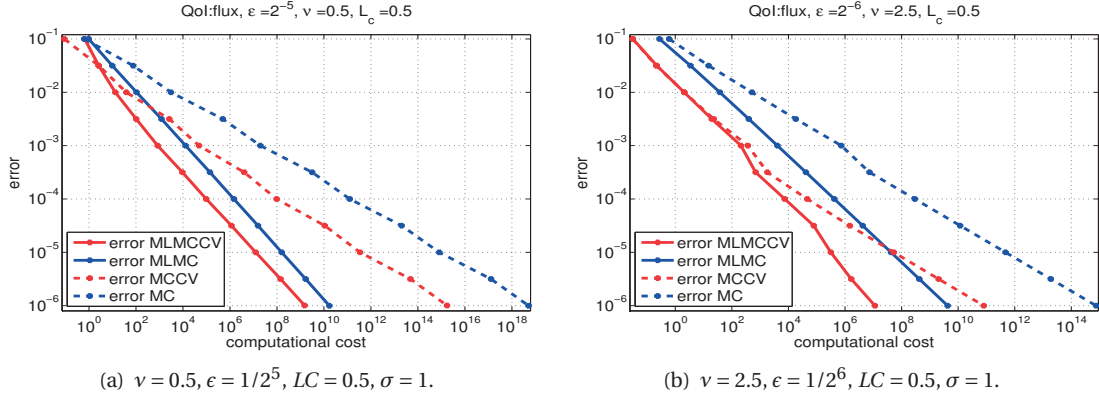


Figure 1.2 – Error vs Computational cost for  $\nu = 0.5$  (left) and  $\nu = 2.5$  (right). Error = sparse grid error + statistical error. Computational cost evaluated as estimated CPU time.

how contaminant particles are transported by the Darcy velocity in two different scenarios:

- in the case of an undisturbed flow induced by a pressure difference between two opposite boundaries;
- in the case of a flow induced by the presence of a pumping well in a prescribed location of our physical domain  $D$ .

In both cases, we model the motion of a single particle starting from the point  $\mathbf{x} \in D$  at time  $t$  with the following stochastic differential equation (SDE)

$$dX_{\mathbf{x},t}(s, \omega) = -a(X_{\mathbf{x},t}(s, \omega), \omega) \nabla p(X_{\mathbf{x},t}(s, \omega), \omega) ds + \sqrt{2}\sigma' dW(s, \omega'), \quad X_{\mathbf{x},t}(t, \omega) = \mathbf{x},$$

where  $W(t, \omega')$  is a Brownian Motion and  $\sigma'$  is a constant coefficient modeling the molecular diffusion into the medium. Notice that each single trajectory depends on two sources of randomness  $\omega = (\omega, \omega')$ , namely the randomness of the Darcy velocity and the Brownian motion here assumed to be independent. The goal of such analysis is to estimate arrival times to specific locations of the domain and the probability of these arrival times being smaller than a certain (finite or infinite) time horizon  $T$ .

Since the Darcy velocity is given only in our physical bounded domain  $D$ , we have to enforce suitable reflecting and absorbing conditions on the boundary  $\partial D$  to properly model the arrival time we are interested in. As we will detail in Chapter 5, we model our arrival time as the first passage time

$$\tau_{\mathbf{x},t}^{\sigma'}(\omega) = \inf\{s > t : X_{\mathbf{x},t}^{\sigma'}(s, \omega) \notin \bar{D}\}, \quad (1.5)$$

which represents the first time at which a particle starting from  $\mathbf{x}$  at time  $t$  exits the domain  $D$ . With this definition at our disposal, one of the goals will be the delineation of the so called capture zones, i.e. the zone starting from which a particle will reach a selected boundary of the domain with a certain probability in a prescribed time horizon  $T$ , that is:

$$\Sigma_{T,\alpha} = \{\mathbf{x} : \mathbb{P}^{\omega}(\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T) \geq \alpha\}, \quad 0 < \alpha < 1.$$

After discussing existence and uniqueness issues related to the solution of the previous stochastic differential equation, in Chapter 5 we aim at representing our quantity of interest by using the Feynman-Kac representation formula to link the arrival time, defined through the SDE, to the solution of a suitable time dependent PDE. More specifically, by denoting with  $\mathbf{u}(\mathbf{x}, \omega)$  the Darcy velocity field, we look at the backward PDE

$$\begin{cases} \frac{\partial \theta_T^{\sigma'}(\mathbf{x}, t, \omega)}{\partial t} + \mathbf{u}(\mathbf{x}, \omega) \cdot \nabla \theta_T^{\sigma'}(\mathbf{x}, t, \omega) + (\sigma')^2 \Delta \theta_T^{\sigma'}(\mathbf{x}, t, \omega) = 0 & \text{for } \mathbf{x} \in D, t \in [0, T), \\ \theta_T^{\sigma'}(\mathbf{x}, t, \omega) = g(\mathbf{x}, t) & \text{for } \mathbf{x} \in \Gamma_D, t \in [0, T), \\ \partial_{\mathbf{n}} \theta_T^{\sigma'}(\mathbf{x}, t, \omega) = 0 & \text{for } \mathbf{x} \in \Gamma_N, t \in [0, T), \\ \theta_T^{\sigma'}(\mathbf{x}, T, \omega) = h(\mathbf{x}) & \text{for } \mathbf{x} \in D. \end{cases} \quad \text{a.s. in } \Omega. \quad (1.6)$$

Under suitable conditions the solution  $\theta_T^{\sigma'}(\mathbf{x}, t, \omega)$  can be represented with a probabilistic formula, namely

$$\theta_T^{\sigma'}(\mathbf{x}, t, \omega) = \mathbb{E}^{\omega'} [g(X_{\mathbf{x},t}^{\sigma'}(\tau_{\mathbf{x},t}^{\sigma'}(\omega)), \tau_{\mathbf{x},t}^{\sigma'}(\omega)) \mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) \leq T\}} + h(X_{\mathbf{x},t}^{\sigma'}(T)) \mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) > T\}} | \omega],$$

which, upon properly selecting the boundary and the final data  $g$  and  $h$ , can represent our desired quantity of interest.

The total expectation of the QoI w.r.t.  $\omega = (\omega, \omega')$  is then computed using the conditional expectation formula as  $\mathbb{E}^{\omega, \omega'}[Q] = \mathbb{E}^{\omega}[\mathbb{E}^{\omega'}[Q|\omega]]$ . The Feynman-Kac formula will allow us to compute the internal conditional expectation by solving a parabolic PDE with dominant transport (since we will always consider small values of diffusion  $\sigma'$ ); then we compute the second expectation by applying one of the methods proposed in chapters 3 and 4, i.e. the sparse grid approximation as well as the MLMC and MLMCCV approaches. The probabilistic capture zones can then be delineated by considering  $g$  and  $h$  in such a way to obtain  $Q(\mathbf{x}, \omega) = \mathbf{1}_{\{\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T\}}$  and by looking at the level sets of the function  $\mathbb{E}^{\omega}[Q(\mathbf{x}, \cdot)]$ .

### Case of an undisturbed flow

In this first test case we consider a Darcy velocity coming from the solution of problem (1.1) defined on the unit square  $(0, 1)^2$  with a zero external forcing term  $f$ , Dirichlet conditions on the left and right boundary, respectively equal to  $p_0 > 0$  and 0, and homogeneous Neumann conditions on the top and bottom boundaries. As usual the permeability is modeled as a lognormally distributed random field having a Matérn covariance structure. Such problem produces a velocity field inducing a flow of particles from left to right. We impose here reflecting conditions on all boundaries but  $\Gamma_{out}$ , where we absorb the particles and register the corresponding arrival time.

As we will see in this case, since the arrival times are bounded random variables, it is possible to consider a more convenient (from a computational point of view) elliptic PDE having as solution the expected arrival time (with respect to all Brownian motions)  $\bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega)$ ; by solving then the following PDE

$$\begin{cases} \mathbf{u}(\mathbf{x}, \omega) \cdot \nabla \bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega) + (\sigma')^2 \Delta \bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega) = -1 & \text{in } D \times [0, T), \\ \bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega) = 0 & \text{on } \{x_1 = 1\} \times [0, T), \\ \partial_{\mathbf{n}} \bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega) = 0 & \text{elsewhere} \end{cases}$$



## 1.4. Contaminant Transport in Groundwater Flows

we can identify the solution  $\theta_T^{\sigma'}(\mathbf{x}, t, \omega)$  with the quantity  $\bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega) = \mathbb{E}^{\omega'}[\tau_{\mathbf{x},0}^{\sigma'}(\omega)|\omega]$ .

We apply then a MLMCCV strategy in order to compute the expectation with respect to the randomness coming from the Darcy velocity. We define then the quantity of interest as

$$\bar{\tau}_{\mathbf{x}}^{\sigma',CV}(\mathbf{y}) = \bar{\tau}_{\mathbf{x}}^{\sigma'}(\mathbf{y}) - \bar{\tau}_{\mathbf{x}}^{\sigma',\epsilon}(\mathbf{y}) + \mathbb{E}^{\omega}[\bar{\tau}_{\mathbf{x}}^{\sigma',\epsilon}(\cdot)]; \quad (1.7)$$

where  $\bar{\tau}_{\mathbf{x}}^{\sigma',\epsilon}(\mathbf{y})$  represents the expected arrival time for particles transported by a smoothed Darcy velocity, obtained starting from the regularized problem with smoothed permeability  $a^\epsilon$  as input. Such arrival times are again the solution of the PDE with  $\mathbf{u}$  replaced with  $\mathbf{u}^\epsilon = -a^\epsilon \nabla p^\epsilon$ . By following the strategy adopted in Chapter 4, we define the MLMCCV estimator as

$$\hat{\tau}_{\{h_\ell\},\{M_\ell\},W}^{MLMCCV}(\mathbf{x}) = \sum_{\ell=0}^L \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} \left( \bar{\tau}_{\ell}^{\sigma',CV}(\mathbf{x}, \mathbf{y}_{i,\ell}) - \bar{\tau}_{\ell-1}^{\sigma',CV}(\mathbf{x}, \mathbf{y}_{i,\ell}) \right) + \mathcal{Q}_W[\bar{\tau}_L^{\sigma',\epsilon}(\mathbf{x}, \cdot)], \quad \bar{\tau}_{-1}^{\sigma',CV}(\mathbf{x}, \mathbf{y}_{i,\ell}) = 0. \quad (1.8)$$

We give a bound on the mean square error of the estimator and verify the effectiveness of such strategy by comparing the results to the ones obtained with a standard MLMC scheme. In Figure 1.3 we report the results obtained by applying a sparse grid scheme to compute the expectation of the arrival time when starting with a smooth permeability function. Similar results can be obtained

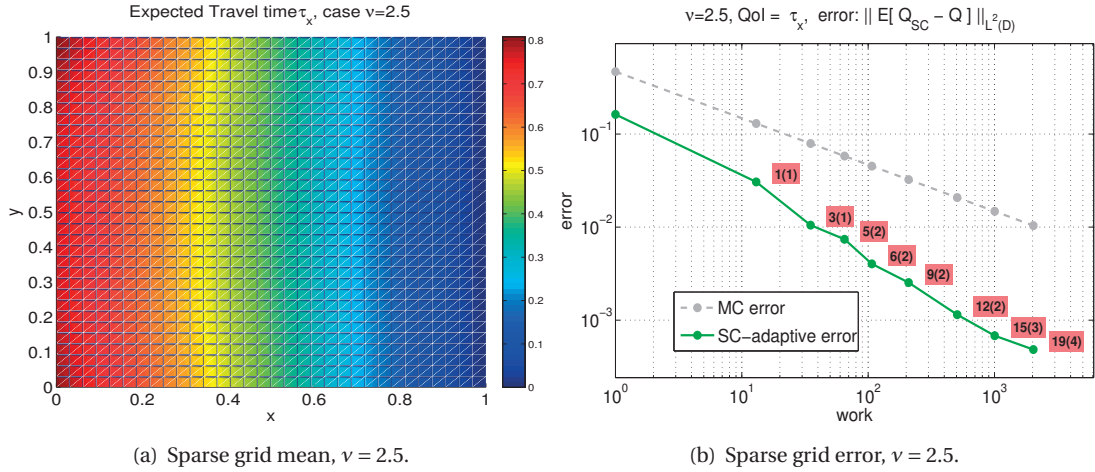


Figure 1.3 – Case  $\nu = 2.5$ : on the left the sparse grid approximation of  $\mathbb{E}^{\omega}[\bar{\tau}_{\mathbf{x}}^{\sigma'}(\mathbf{y})]$ ; on the right the sparse grid error committed on such quantity in the  $L^2$  spatial norm versus the work evaluated as number of solves needed to build such approximation.  $\sigma' = 0.01$ ,  $L_C = 0.5$ ,  $p_0 = 1$ .

when starting with a rough permeability and considering the smoothed arrival time  $\bar{\tau}_{\mathbf{x}}^{\sigma',\epsilon}(\mathbf{y})$ .

On the other hand, in Figure 1.4 we report the variance reduction obtained with the control variate multi level strategy in the case of a rough field and compare the results with a standard MLMC approach.

Then we look at the delineation of capture zones; in this case we propose two strategies: the first consists in applying a MLMC scheme on the solution of the parabolic problem previously introduced, equipped with boundary condition  $g = 1$  on the outlet and homogeneous Neumann condition elsewhere. This approach has generated the results presented in Figure 1.5. The second approach, by contrast, is based on the idea of reusing and properly adapting the results obtained when considering as quantity of interest the expected arrival time  $\bar{\tau}_{\mathbf{x}}^{\sigma'}(\mathbf{y})$ . In this case we use again the

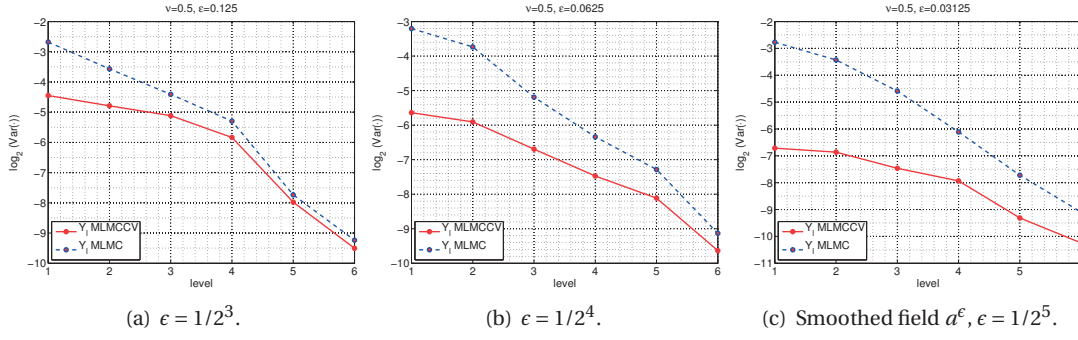


Figure 1.4 – Variance of  $Q_\ell$  and  $Y_\ell(\mathbf{x}, \omega) = Q_\ell(\mathbf{x}, \omega) - Q_{\ell-1}(\mathbf{x}, \omega)$  on each level, for  $\nu = 0.5$  and different values of  $\epsilon$ .  $\sigma' = 0.01$ ,  $L_C = 0.5$ ,  $p_0 = 1$ .

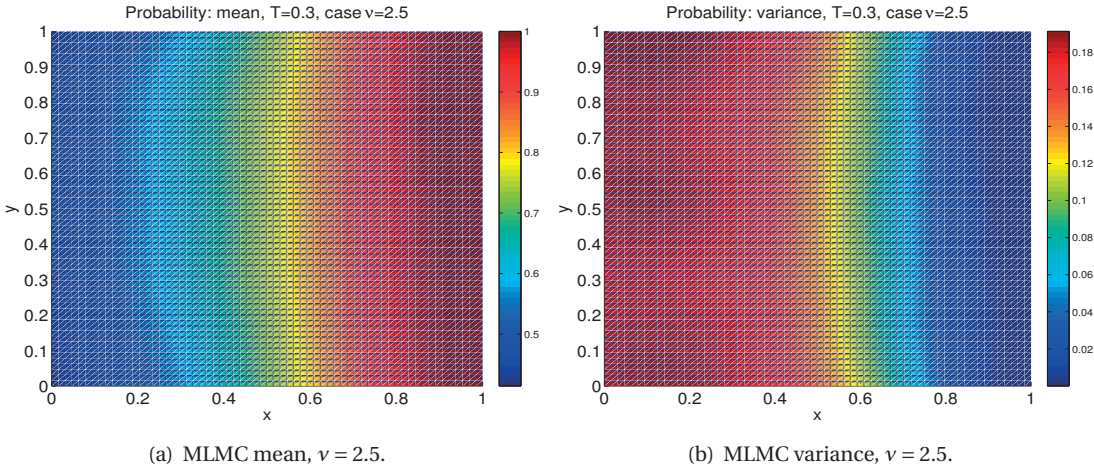


Figure 1.5 – Case  $\nu = 2.5$ : MLMC approximation of mean (left) and variance (right) of  $Q(\mathbf{x}, \omega) = \mathbb{E}^{\omega'}[\mathbf{1}_{\{\tau_{\mathbf{x}_0}^{\sigma'}(\omega) \leq T\}}]$ .  $\sigma' = 0$ ,  $L_C = 0.5$ ,  $p_0 = 3$ ,  $T = 0.3$ ,  $L = 5$ .

sparse grid approximation of the arrival time  $\bar{\tau}_{\mathbf{x}}^{\sigma'}(\mathbf{y})$ , i.e.  $\mathcal{S}_1^m[\bar{\tau}_{\mathbf{x}}^{\sigma'}](\mathbf{y})$ , and sample from the indicator function  $\mathbf{1}_{\{\bar{\tau}_{\mathbf{x}}^{\sigma'}(\mathbf{y}) \leq T\}}$  through sparse grid interpolation, namely  $\mathbf{1}_{\{\bar{\tau}_{\mathbf{x}}^{\sigma'}(\mathbf{y}_i) \leq T\}} \approx \mathbf{1}_{\{\mathcal{S}_1^m[\bar{\tau}_{\mathbf{x}}^{\sigma'}](\mathbf{y}_i) \leq T\}}$ ; hence, we can build a convenient MC estimator of the probability function involved in the definition of the capture zones by averaging over all these indicator functions; in this strategy the cost of the MC sampling is low since no PDE solve is involved and the MC statistical error can be made negligible by taking a very large sample size.

### Case of a flow induced by the presence of a pumping well

In this second test case we are interested in the computation of arrival times and capture zones when, superposed to an underlying flow modeled as in the previous case, we have also a well in the domain which extracts drinkable water with a fixed extraction rate. We model the presence of the well with a forcing term  $f$  given by a Dirac function located in the center of the well  $\mathbf{x}_0$  and look at



the following Darcy problem

$$\begin{cases} -\operatorname{div}(a(\mathbf{x}, \omega) \nabla p(\mathbf{x}, \omega)) = -q \delta_{\mathbf{x}_0}(\mathbf{x}), & \mathbf{x} \in D, \\ p(\mathbf{x}, \omega) = p_0(1 - x_1) & \mathbf{x} \in x_1 = \{0, 1\}, \\ a(\mathbf{x}, \omega) \partial_{\mathbf{n}} p(\mathbf{x}, \omega) = 0 & \mathbf{x} \in x_2 = \{0, 1\}, \end{cases}$$

which is well posed for any pressure  $a \in C^\alpha(\bar{D})$  (see Chapter 5). To properly define our arrival time we consider then a small region surrounding the well, for instance the ball  $B(\mathbf{x}_0, r)$ ,  $r \ll 1$ , and define the arrival time as

$$\tau_{\mathbf{x}, t}^{\sigma'}(\omega) = \inf(s \geq t : X_{\mathbf{x}, t}^\sigma(s) \in B(\mathbf{x}_0, r)),$$

where now we are imposing reflecting condition on all the external boundaries and absorbing conditions only on  $\partial B(\mathbf{x}, r)$ . Observe that by doing this, we are implicitly assuming that a particle reaches the location of the well as soon as it enters the ball  $B(\mathbf{x}, r)$ . Analogously to what we did in the previous case, we put homogeneous Neumann boundary conditions where we have reflection, Dirichlet conditions where we have absorption and select the final condition of the associated parabolic PDE to model the desired quantities, i.e. the expected arrival times or probabilities. In Chapter 5 we will see how, in this case, it is not possible to apply the MLMCCV scheme to compute the expectation with respect to  $\omega$ . This is intuitively due to the fact that now the solution of the parabolic PDE (1.6) presents very sharp moving fronts; this fact causes a lack of smoothness of the QoI with respect to the random variables  $\mathbf{y}$  and a poor performance of any sparse grid approximation. To compute the expectation of our QoI, we propose then a more classical MLMC scheme however with a level dependent diffusion coefficient  $\sigma'$  which improves the stability of the finite element approximation on coarse meshes as well as the variance reduction in the MLMC estimator. In Figure 1.6 we show the MLMC estimate of the probability  $\mathbb{P}^\omega(\tau_{\mathbf{x}, 0}^{\sigma'}(\omega) \leq T)$  for a finite temporal horizon  $T$ , and the corresponding MLMC estimate of the variance of the random quantity  $\mathbb{E}^{\omega'}[\tau_{\mathbf{x}, 0}^{\sigma'}(\cdot, \omega) \leq T | \omega]$ .

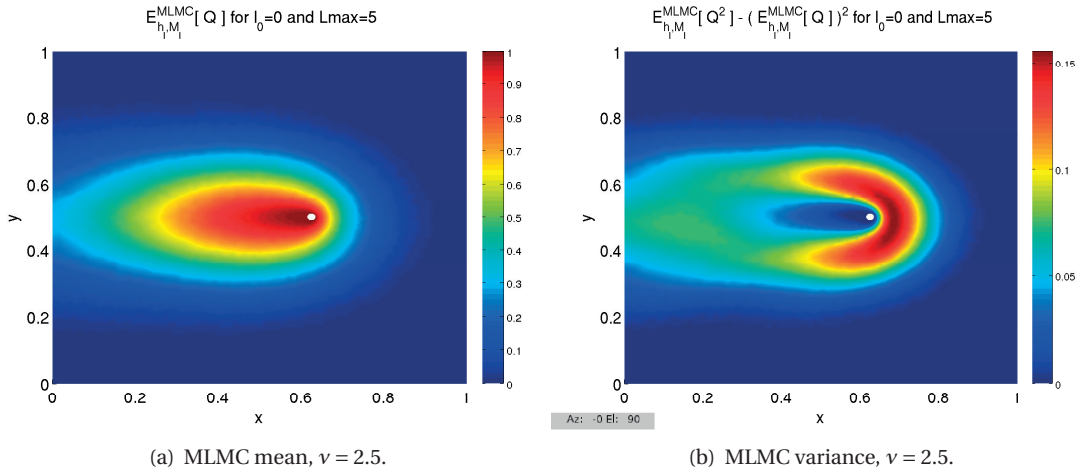


Figure 1.6 – Case  $\nu = 2.5$ : MLMC approximation of mean (left) and variance (right) of  $Q(\mathbf{x}, \omega) = \mathbb{E}^{\omega'}[\mathbf{1}_{\{\tau_{\mathbf{x}, 0}^{\sigma'}(\omega) \leq T\}}]$ .  $\sigma' = 0.01$ ,  $L_C = 0.5$ ,  $p_0 = 1$ ,  $q = 0.5$ ,  $r = 0.01$ ,  $T = 0.5$ .



## 2 The Darcy Problem

In this chapter we introduce more specifically the Darcy problem and recall the main results available in the literature that will be used in our analysis and in the development of the numerical methods discussed in the following Chapters. In particular, after having introduced some notation in Section 2.1, in Section 2.2 we introduce the continuous problem and the main related theoretical results; then in Section 2.3 we introduce the finite element approximation of the continuous problem again by recalling the related convergence result; in Section 2.4 we discuss the problem of truncating the infinite dimensional probability space for computational purposes while in Section 2.5 we detail specific choices of parametrization and truncation of the input randomness.

### 2.1 Notation

Given a bounded Lipschitz domain  $D \in \mathbb{R}^d$ , we introduce the following notation. For any  $k \in \mathbb{N}$  we denote with  $\mathcal{C}^k(\overline{D})$  the space of continuously  $k$  times differentiable functions with the usual norms. For any positive real  $\alpha$  we set  $\alpha = k + s$  with  $k \in \mathbb{N}$  and  $s \in (0, 1]$ . We recall as we denote with  $C^\alpha(\overline{D})$  the Hölder space for which the following norm is bounded

$$\|v\|_{C^\alpha(\overline{D})} = \|v\|_{\mathcal{C}^k(\overline{D})} + |v|_{C^\alpha(\overline{D})} = \sum_{j=1}^k \max_{|i|_1=j} \|D^i v\|_{\mathcal{C}^0(\overline{D})} + \max_{|i|_1=k} \sup_{x,y \in \overline{D}} \frac{|D^i v(x) - D^i v(y)|}{|x - y|^s},$$

where  $i$  is a multi-index of  $\mathbb{N}^d$  with  $|i|_1 = \sum_{k=1}^d i_k$ ,  $D^i v = \frac{\partial^{|i|_1}}{\partial x_1^{i_1} \dots \partial x_d^{i_d}} v$  and  $|\cdot|$  denotes the Euclidean norm in  $\mathbb{R}^d$ . Notice that with this definition, the space  $C^1$  denotes the space of Lipschitz continuous functions and not the usual space  $\mathcal{C}^1$  of continuously differentiable functions. We will also use the usual Sobolev spaces  $H^k(D)$ ,  $k \in \mathbb{N}$ , characterized by corresponding norm and seminorm

$$|v|_{H^k(D)}^2 = \int_D \sum_{|i|_1=k} |D^i v(x)|^2 dx, \quad \|v\|_{H^k(D)}^2 = \int_D \sum_{|i|_1 \leq k} |D^i v(x)|^2 dx,$$

as well as the fractional Sobolev spaces  $H^\alpha(D)$ ,  $\alpha \in \mathbb{R}$ , using the Sobolev-Slobodetskii seminorm  $|v|_{H^\alpha(D)}$ :

$$\|v\|_{H^\alpha(D)}^2 = \|v\|_{H^k(D)}^2 + |v|_{H^\alpha(D)}^2 = \|v\|_{H^k(D)}^2 + \sum_{|i|_1=k} \int_{D \times D} \frac{|D^i v(x) - D^i v(y)|^2}{|x - y|^{d+2s}} dx dy.$$

In the following, whenever possible, instead of the usual  $H^1(D)$  norm, we will use the equivalent  $H_0^1(D)$  norm, defined as

$$\|v\|_{H_0^1(D)} = \int_D |\nabla v|^2 dx.$$

Given a Banach space  $B$  and a complete probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , it is also useful to introduce the Bochner space  $L_{\mathbb{P}}^q(\Omega, B)$  as the space of strongly measurable functions  $v : \Omega \rightarrow B$  equipped with the norm  $\|v\|_{L_{\mathbb{P}}^q(\Omega, B)} = \mathbb{E}[\|v\|_B^q]^{\frac{1}{q}}$ .

Finally, to simplify the notation, sometimes we will use the symbol  $\lesssim$  to indicate a bound in which the hidden constant is just a positive real number that does not depend on any relevant parameter (such as mesh size  $h$ , regularity parameter  $\nu$  of a random field, etc.).

## 2.2 Problem setting

In this thesis from a mathematical point of view we are interested in developing efficient numerical methods to solve uncertainty quantification problems, and in particular partial differential equations with random input data. We will then test such methodology to solve the so called stochastic Darcy problem, which consists of an elliptic partial differential equation with random entries, that may be for instance the diffusion coefficient, the forcing term, the boundary data or also the domain itself. Such problem arises in hydrology and models the groundwater flow in a highly heterogeneous saturated porous medium. We will assume in what follows that the input stochasticity directly affects only the diffusion coefficient, while all the other quantities characterizing the elliptic PDE are considered to be perfectly known. The mathematical mixed formulation of the problem, described by the Darcy law relating the volumetric flux (in the following we will refer to this quantity as velocity) to the hydraulic head (in the following pressure), together with a mass balance, is given by the following equations:

$$\begin{cases} \mathbf{u} = -a\nabla p & \text{in } D, \\ \operatorname{div}(\mathbf{u}) = f & \text{in } D, \\ p = g_j & \text{on } \Gamma_j^D, j = 1, \dots, m_D, \\ \mathbf{u} \cdot \mathbf{n} = 0 & \text{on } \Gamma_j^N, j = 1, \dots, m_N, \end{cases} \quad (2.1)$$

where  $\Gamma_D = \cup_{j=1}^{m_D} \Gamma_j^D$  denotes the Dirichlet boundary,  $\Gamma_N = \cup_{j=1}^{m_N} \Gamma_j^N$  denotes the Neumann boundary and  $\bar{\Gamma}_D \cup \bar{\Gamma}_N = \partial D$ ,  $\bar{\Gamma}_D \cap \bar{\Gamma}_N = \emptyset$ . Here  $p$  is the pressure,  $\mathbf{u}$  the velocity field and  $a = \frac{\kappa}{\mu}$  represents the intrinsic permeability, i.e. the ratio between the hydraulic conductivity of the medium  $\kappa$  and the dynamic viscosity  $\mu$  of the fluid;  $f$  is an external source or sink term and  $D \in \mathbb{R}^d$  a bounded open domain. In the following, we assume the dynamic viscosity  $\mu$  to be constant and refer to  $a$  as a permeability. A key issue in the study of groundwater flows in heterogeneous media concerns the characterization of the subsurface proprieties. In many cases we have only a very limited knowledge of the input data of the problem and particularly the permeability field. In order to deal with this uncertainty and lack of knowledge the permeability is often modeled as a spatially correlated random field depending on a random event  $\omega$  of a suitable probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  [21]. Hence also the solution  $(p, \mathbf{u})$  of (1.1) will depend on  $\omega$  and the problem (1.1) is interpreted in

a probabilistic sense as

$$\begin{cases} \mathbf{u}(x, \omega) = -a(x, \omega) \nabla p(x, \omega) & \text{in } D, \\ \operatorname{div}(\mathbf{u}(x, \omega)) = f(x) & \text{in } D, \\ p(x, \omega) = g_j(x) & \text{on } \Gamma_j^D, j = 1, \dots, m_D, \\ \mathbf{u}(x, \omega) \cdot \mathbf{n} = 0 & \text{on } \Gamma_j^N, j = 1, \dots, m_N, \end{cases} \quad \text{a.s. in } \Omega \quad (2.2)$$

where *a.s.* means “almost surely” with respect to the probability measure  $\mathbb{P}$ . A widely used model for the permeability field  $a$  describes it as a lognormal random field [21, 32, 78], namely  $a(x, \omega) = e^{\gamma(x, \omega)}$  with  $\gamma(x, \omega)$  a Gaussian random field having mean  $\mathbb{E}[\gamma(x, \cdot)]$  and covariance function  $\operatorname{cov}_\gamma(x_1, x_2) = \mathbb{E}[\gamma(x_1, \cdot)\gamma(x_2, \cdot)] - \mathbb{E}[\gamma(x_1, \cdot)]\mathbb{E}[\gamma(x_2, \cdot)]$ . The choice of the covariance function is a delicate issue. It directly relates to the spatial smoothness of the random field realizations and strongly influences the choice of the numerical method to use. Equations (2.2) have been extensively studied during the last few years from both the theoretical and numerical point of view. They can be reformulated as an elliptic partial differential equation with random diffusion coefficient, given by the permeability. Denoting  $V_g = \{v \in H^1(D) : v = g \text{ on } \Gamma_D\}$ , the variational formulation associated with the Darcy problem is: find  $p \in V_g$  such that

$$b_\omega(p, v) = L(v) \quad \forall v \in V_0, \quad (2.3)$$

where the bilinear form  $b_\omega$  (parametrized by  $\omega$ ) and the linear functional  $L$  are defined as:

$$b_\omega(u, v) = \int_D a(x, \omega) \nabla u(x, \omega) \nabla v(x) dx, \quad L(v) = \int_D f(x) v(x) dx.$$

Well posedness results for the problem (2.3) can be found in [16, 43, 38] where it is shown that the solution  $p$  is unique in the space  $L_{\mathbb{P}}^q(\Omega, V_g)$ ,  $\forall q \in \mathbb{R}_+$ . Moreover, the following regularity result is shown in [16, 19]:

**Lemma 2.2.1.** *Let  $D$  be a  $\mathcal{C}^2$  bounded domain,  $f \in H^{\alpha-1}(D)$  for some  $0 < \alpha \leq 1$  and  $g = 0$  on  $\Gamma_D = \partial D$ . Let  $a(x, \omega)$  be the input random field of problem (2.3) a.s. continuous and denote  $a_{\max}(\omega) = \max_{x \in \overline{D}} a(x, \omega)$  and  $a_{\min}(\omega) = \min_{x \in \overline{D}} a(x, \omega)$ . If*

- $a_{\min}(\omega) \in L_{\mathbb{P}}^q(\Omega)$ ,  $\forall q \in \mathbb{R}_+$ ,
- $a(x, \omega) \in L_{\mathbb{P}}^q(\Omega, C^\alpha(\overline{D}))$   $\forall q \in \mathbb{R}_+$ ;

then for the problem (2.3) the following regularity result holds:

$$\|p(\cdot, \omega)\|_{H^{1+\beta}(D)} \lesssim \frac{1}{\alpha - \beta} C_{2.2.1}(\omega, \alpha) \|f\|_{H^{\alpha-1}(D)}, \quad \forall 0 < \beta < \alpha \quad \text{a.s. in } \Omega.$$

If the hypotheses hold also for  $\alpha > 1$ , then

$$\|p(\cdot, \omega)\|_{H^2(D)} \lesssim C_{2.2.1}(\omega, \alpha) \|f\|_{L^2(D)}, \quad \text{a.s. in } \Omega,$$

where

$$C_{2.2.1}(\omega, \alpha) = \begin{cases} \frac{a_{\max}(\omega) \|a(\cdot, \omega)\|_{C^\alpha(\bar{D})}}{a_{\min}^3(\omega)} & \text{if } \alpha \leq 1, \\ \frac{a_{\max}(\omega) \|a(\cdot, \omega)\|_{\mathcal{C}^1(\bar{D})}}{a_{\min}^3(\omega)} & \text{if } \alpha > 1. \end{cases} \quad (2.4)$$

Moreover the constant  $C_{2.2.1}(\omega, \alpha)$  is  $q$ -integrable for any  $q \in \mathbb{R}_+$ , i.e.  $C_{2.2.1}(\omega, \alpha) \in L^q_{\mathbb{P}}(\Omega) \forall q \in \mathbb{R}_+$ .

*Proof.* The proof of this result follows the one given in [19] by replacing [19, Lemma A.2] with the following Lemma, in order to make explicit the dependence of the bound with respect to the regularity  $\alpha$  of the random field  $a$ .

**Lemma 2.2.2.** *Let  $b \in C^\alpha(\bar{D})$  and let  $v$  be a function in  $H^\beta(D)$  for some  $0 < \beta < \alpha \leq 1$ . It holds*

$$\|bv\|_{H^\beta(D)} \lesssim \frac{1}{\sqrt{\eta}} \|b\|_{C^{\beta+\eta}(\bar{D})} \|v\|_{H^\beta(D)} \quad \forall 0 < \eta \leq \alpha - \beta.$$

*Proof.* By definition the  $H^\beta$  norm of the function  $bv$  is

$$\|bv\|_{H^\beta(D)}^2 = \|bv\|_{L^2(D)}^2 + |bv|_{H^\beta(D)}^2 = \|bv\|_{L^2(D)}^2 + \int_{D \times D} \frac{|b(x)v(x) - b(y)v(y)|^2}{|x - y|^{d+2\beta}} dx dy.$$

The first term can be easily bounded as  $\|bv\|_{L^2(D)}^2 \leq \|b\|_{\mathcal{C}^0(\bar{D})}^2 \|v\|_{L^2(D)}^2$ . For the second term we obtain

$$\begin{aligned} \int_{D \times D} \frac{|b(x)v(x) - b(y)v(y)|^2}{|x - y|^{d+2\beta}} dx dy &\leq 2\|b\|_{\mathcal{C}^0(\bar{D})}^2 |v|_{H^\beta(D)}^2 + 2 \int_{D \times D} \frac{|b(x) - b(y)|^2}{|x - y|^{2(\beta+\eta)}} \frac{v(y)^2}{|x - y|^{d-2\eta}} dx dy \\ &\leq 2\|b\|_{\mathcal{C}^0(\bar{D})}^2 |v|_{H^\beta(D)}^2 + 2\|b\|_{C^{\beta+\eta}(\bar{D})}^2 \int_{D \times D} \frac{v(y)^2}{|x - y|^{d-2\eta}} dx dy. \end{aligned}$$

If we extend  $v$  by 0 in  $\mathbb{R}^d \setminus D$ , and denote  $\tilde{v}$  this extension and  $\rho = \max_{x \in D} |x|$ , the integral appearing in the right hand side of the above inequality can be bounded as

$$\begin{aligned} \int_{D \times D} \frac{v(y)^2}{|x - y|^{d-2\eta}} dx dy &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{\tilde{v}(y)^2}{|x - y|^{d-2\eta}} \mathbb{1}_{\{|x-y| \leq 2\rho\}} dx dy \leq \left\| \tilde{v}^2(x) * \frac{\mathbb{1}_{\{|x| \leq 2\rho\}}}{|x|^{d-2\eta}} \right\|_{L^1(\mathbb{R}^d)} \\ &\leq \|\tilde{v}^2\|_{L^1(\mathbb{R}^d)} \left\| \frac{\mathbb{1}_{\{|x| \leq 2\rho\}}}{|x|^{d-2\eta}} \right\|_{L^1(\mathbb{R}^d)} \lesssim \frac{\rho^{2\eta}}{\eta} \|v\|_{L^2(D)}^2. \end{aligned}$$

By putting everything together we obtain

$$\|bv\|_{H^\beta(D)} \lesssim \frac{1}{\sqrt{\eta}} \|b\|_{C^{\beta+\eta}(\bar{D})} \|v\|_{H^\beta(D)},$$

which is the desired result. □

□

**Remark 2.2.1.** *This result may seem slightly different than the one presented in [19] but actually it is not; in fact in our work we use the  $C^\alpha$  norm instead of the usual  $\mathcal{C}^\alpha$  one: this makes possible to recover a bound for the  $H^2$  norm only when  $\alpha$  is strictly larger than one. Secondly here we explicitly*

write the dependence of the constant with respect to the degenerating part which is  $\mathcal{O}(\frac{1}{\alpha-\beta})$  when  $\beta \rightarrow \alpha$ .

**Remark 2.2.2.** Although here we refer to the case of a  $\mathcal{C}^2$  bounded domain, it should be possible to weaken this assumption to the case of a convex Lipschitz domain by following the results given in [47, Chapter 3]. Observe that the expressions of the constant  $C_{2.2.1}(\omega, \alpha)$  and of the degenerating term  $\mathcal{O}(\frac{1}{\alpha-\beta})$  appearing in the regularity estimate depends on the assumptions made on the domain, so they would change if we consider a convex Lipschitz domain instead of a  $\mathcal{C}^2$  bounded one. For instance, in [74], an explicit form of the constant  $C_{2.2.1}(\omega, \alpha)$  has been provided in the case of polygonal domains.

In what follows we focus on the case in which the log-permeability Gaussian random field  $\gamma$  is stationary and has a “Matérn-type” covariance function; in particular we will consider two cases:

- an isotropic covariance function belonging to the classical Matérn family [32]:

$$\text{cov}_\gamma(\mathbf{x}, \mathbf{x}') = \widetilde{\text{cov}}_\gamma(|\mathbf{x} - \mathbf{x}'|) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left( \sqrt{2\nu} \frac{|\mathbf{x} - \mathbf{x}'|}{L_c} \right)^\nu K_\nu \left( \sqrt{2\nu} \frac{|\mathbf{x} - \mathbf{x}'|}{L_c} \right), \quad \nu \geq 0.5, \quad (2.5)$$

- a tensor Matérn covariance function:

$$\text{cov}_\gamma(\mathbf{x}, \mathbf{x}') = \widetilde{\text{cov}}_\gamma(|\mathbf{x} - \mathbf{x}'|) = \sigma^2 \prod_{i=1}^d \frac{\left( \sqrt{2\nu} \frac{|x_i - x'_i|}{L_c} \right)^\nu K_\nu \left( \sqrt{2\nu} \frac{|x_i - x'_i|}{L_c} \right)}{\Gamma(\nu)2^{\nu-1}}, \quad \nu \geq 0.5, \quad (2.6)$$

where  $\sigma^2$  is the pointwise variance,  $L_c$  is a correlation length,  $\Gamma$  is the gamma function,  $K_\nu$  is the modified Bessel function of the second kind and  $\nu$  is a parameter that governs the regularity of the covariance function and, consequently, of the realizations of the random field. In particular the following holds:

- the covariance function is Hölder continuous, namely  $\text{cov}_\gamma \in C^{2\nu}(\bar{D} \times \bar{D})$  (see the following Lemma 2.2.3),
- the realizations of the random field are a.s. Hölder continuous,  $\gamma(\cdot, \omega) \in C^\alpha(\bar{D})$ ,  $\forall 0 < \alpha < \nu$  (see the following Lemma 2.2.4),
- the random variables  $a_{\min}(\omega)$ ,  $a_{\max}(\omega)$  and  $\|a(\cdot, \omega)\|_{C^\alpha(\bar{D})}$  with  $0 < \alpha < \nu$  are all  $L^q_{\mathbb{P}}(\Omega)$ -integrable  $\forall q \in \mathbb{R}_+$  (see e.g. [16, 19]).

**Lemma 2.2.3.** Let  $\text{cov}_\gamma$  be a covariance function belonging to the Matérn family defined in (2.5) on an open bounded convex domain  $D$ . Then, if  $\nu$  is not an integer,  $\text{cov}_\gamma \in C^{2\nu}(\bar{D} \times \bar{D})$ , otherwise  $\text{cov}_\gamma \in C^\alpha(\bar{D} \times \bar{D})$  for any  $\alpha < 2\nu$  if  $\nu \in \mathbb{N}_+$ .

*Proof.* By definition we have

$$\text{cov}_\gamma(x_1, x_2) = \widetilde{\text{cov}}_\gamma(|x_1 - x_2|) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left( \sqrt{2\nu} \frac{|x_1 - x_2|}{L_c} \right)^\nu K_\nu \left( \sqrt{2\nu} \frac{|x_1 - x_2|}{L_c} \right);$$

$K_\nu : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is given by  $K_\nu(\rho) = \frac{\pi}{2 \sin \pi \nu} (I_{-\nu}(\rho) - I_\nu(\rho))$  where  $I_\alpha(\rho) = \sum_{m=0}^{\infty} \frac{1}{m! \Gamma(m+\alpha+1)} \left( \frac{\rho}{2} \right)^{2m+\alpha}$ . This formula is valid when  $\nu$  is not an integer, i.e.  $\nu = n + s$  with  $n \in \mathbb{N}$  and  $s \in (0, 1)$ . Since  $\forall \epsilon > 0$  the function  $K_\nu \in \mathcal{C}^\infty[\epsilon, +\infty)$ , and consequently  $\widetilde{\text{cov}}_\gamma$  as well, in order to prove the result we focus on

the asymptotic behavior of the function  $\widetilde{\text{cov}}_\gamma(|x - y|)$  in a neighborhood of  $|x - y| = 0$ . By denoting  $\lambda_\nu = \frac{\sqrt{2\nu}}{2L_c}$  and by recalling that, for any  $x \in \mathbb{R} \setminus \mathbb{Z}$  it holds  $\Gamma(-x) = \frac{-\pi}{\sin \pi x \Gamma(x+1)}$ , it is possible to obtain

$$\begin{aligned} \widetilde{\text{cov}}_\gamma(|x_1 - x_2|) &= \frac{\sigma^2 \pi}{\Gamma(\nu) \sin \pi \nu} \left( \sum_{m=0}^{\infty} \frac{\lambda_\nu^{2m} |x_1 - x_2|^{2m}}{m! \Gamma(m - \nu + 1)} - \sum_{m=0}^{\infty} \frac{\lambda_\nu^{2(m+\nu)} |x_1 - x_2|^{2(m+\nu)}}{m! \Gamma(m + \nu + 1)} \right) \\ &= \sigma^2 \left( \sum_{m=0}^n \frac{(-1)^m \Gamma(\nu - m) \lambda_\nu^{2m}}{m! \Gamma(\nu)} |x_1 - x_2|^{2m} - \frac{\lambda_\nu^{2\nu} |x_1 - x_2|^{2\nu}}{\Gamma(\nu) \Gamma(\nu + 1) \sin(\pi \nu)} \right) \\ &\quad + \frac{\sigma^2}{\Gamma(\nu) \sin \pi \nu} \left( \sum_{m=1}^{\infty} \frac{\lambda_\nu^{2(m+n)} |x_1 - x_2|^{2(m+n)}}{(m+n)! \Gamma(m + n - \nu + 1)} - \frac{\lambda_\nu^{2(m+\nu)} |x_1 - x_2|^{2(m+\nu)}}{m! \Gamma(m + \nu + 1)} \right). \end{aligned}$$

Hence, the asymptotic behavior is

$$\widetilde{\text{cov}}_\gamma(|x_1 - x_2|) \sim \sigma^2 \left\{ \sum_{m=0}^n \frac{(-1)^m \Gamma(\nu - m) \lambda_\nu^{2m}}{m! \Gamma(\nu)} |x_1 - x_2|^{2m} - \frac{\lambda_\nu^{2\nu} |x_1 - x_2|^{2\nu}}{\Gamma(\nu) \Gamma(\nu + 1) \sin(\pi \nu)} \right\}. \quad (2.7)$$

Since the function  $f(z) = |z|^{2\nu} : \mathbb{R}^d \rightarrow \mathbb{R}$  belongs to the space  $C^{2\nu}(A)$  for any bounded set  $A \in \mathbb{R}^d$  we can conclude that  $\text{cov}_\gamma \in C^{2\nu}(\bar{D} \times \bar{D})$ .

When  $\nu = n \in \mathbb{N}_+$  the previous definition gives removable indeterminate values of the form  $\frac{0}{0}$ ; in this case the Bessel function  $K_\nu$  can be defined through the limit  $K_n(\rho) = \lim_{\nu \rightarrow n} K_\nu(\rho)$ . The covariance function becomes:

$$\begin{aligned} \widetilde{\text{cov}}_\gamma(|x_1 - x_2|) &= \lim_{\nu \rightarrow n} \frac{\sigma^2 \pi}{\Gamma(\nu) \sin \pi \nu} \left( \sum_{m=0}^{\infty} \frac{\lambda_\nu^{2m} |x_1 - x_2|^{2m}}{m! \Gamma(m - \nu + 1)} - \sum_{m=0}^{\infty} \frac{\lambda_\nu^{2(m+\nu)} |x_1 - x_2|^{2(m+\nu)}}{m! \Gamma(m + \nu + 1)} \right) \\ &= \sigma^2 \sum_{m=0}^{n-1} \frac{(-1)^m (n - m - 1)! \lambda_n^{2m}}{m! \Gamma(n)} |x_1 - x_2|^{2m} \\ &\quad + \lim_{\nu \rightarrow n} \frac{\sigma^2}{\Gamma(\nu) \sin \pi \nu} \left( \sum_{m=0}^{\infty} \frac{\lambda_\nu^{2(m+n)} |x_1 - x_2|^{2(m+n)}}{(m+n)! \Gamma(m + n - \nu + 1)} - \frac{\lambda_\nu^{2(m+\nu)} |x_1 - x_2|^{2(m+\nu)}}{m! \Gamma(m + \nu + 1)} \right) \\ &= \sigma^2 \sum_{m=0}^{n-1} \frac{(-1)^m (n - m - 1)! \lambda_n^{2m}}{m! \Gamma(n)} |x_1 - x_2|^{2m} \\ &\quad + \lim_{\nu \rightarrow n} \frac{\sigma^2}{\Gamma(\nu) \sin \pi \nu} \left( \sum_{m=0}^{\infty} \frac{\lambda_\nu^{2(m+n)} |x_1 - x_2|^{2(m+n)}}{(m+n)! \Gamma(m + n - \nu + 1) m! \Gamma(m + \nu + 1)} \left( m! (\Gamma(m + \nu + 1) - \Gamma(m + n + 1)) + \right. \right. \\ &\quad \left. \left. (m + n)! (\Gamma(m + 1) - \Gamma(m + n - \nu + 1)) + m! (m + n)! (1 - \lambda_\nu^{2(\nu-n)} |x_1 - x_2|^{2(\nu-n)}) \right) \right) \\ &= \sigma^2 \sum_{m=0}^{n-1} \frac{(-1)^m (n - m - 1)! \lambda_n^{2m}}{m! (n - 1)!} |x_1 - x_2|^{2m} \\ &\quad + \frac{(-1)^n \sigma^2}{(n - 1)!} \sum_{m=0}^{\infty} \lambda_n^{2(m+n)} \left( \frac{m! \Gamma'(m + n + 1) + (m + n)! \Gamma'(m + 1)}{((m + n)!)^2 (m!)^2} - \frac{2 \log(\lambda_n |x - y|)}{(m + n)! m!} \right) |x_1 - x_2|^{2(m+n)} \end{aligned}$$

Again we focus on the asymptotic behavior of the function  $\widetilde{\text{cov}}_\gamma(|x_1 - x_2|)$  in a neighborhood of  $|x_1 - x_2| = 0$ . We obtain

$$\widetilde{\text{cov}}_\gamma(|x_1 - x_2|) \sim \sigma^2 \left\{ \sum_{m=0}^{n-1} \frac{(-1)^m (n - m - 1)! \lambda_n^{2m}}{m! (n - 1)!} |x_1 - x_2|^{2m} - \frac{(-1)^n 2 \lambda_n^{2n}}{n! (n - 1)!} |x_1 - x_2|^{2n} \log(\lambda_n |x_1 - x_2|) \right\} \quad (2.8)$$

Since the function  $f(z) = |z|^{2n} \log(|z|) : \mathbb{R}^d \rightarrow \mathbb{R}$  belongs to the space  $C^\alpha(A)$  for any  $\alpha < 2\nu$  and for



any bounded set  $A \in \mathbb{R}^d$  we can conclude that  $\text{cov}_\gamma \in C^\alpha(\bar{D} \times \bar{D})$  for any  $\alpha < 2\nu$ .  $\square$

**Remark 2.2.3.** Let  $\widetilde{\text{cov}}_\gamma(|x - y|)$  be a covariance function belonging to the Matérn family defined in (2.5) and let  $\gamma(x, \omega)$  be a centered Gaussian random field defined on  $\bar{D}$ . Denote  $\nu = n + \alpha$  with  $n \in \mathbb{N}$  and  $\alpha \in (0, 1]$ . Then, for any multi-index  $i \in \mathbb{N}^d$  such that  $|i|_1 \leq n$ , it holds:

$$\mathbb{E}[D^i \gamma(x, \cdot) D^i \gamma(y, \cdot)] = \frac{\partial^{2|i|_1}}{\partial x_1^{i_1} \dots \partial x_d^{i_d} \partial y_1^{i_1} \dots \partial y_d^{i_d}} \widetilde{\text{cov}}_\gamma(|x - y|).$$

**Lemma 2.2.4.** Let  $\gamma(x, \omega)$  be a centered Gaussian random field with covariance function  $\text{cov}_\gamma$  as in Lemma 2.2.3. Then  $\gamma$  admits a version with trajectories a.s. in  $C^\alpha(\bar{D})$  for any  $0 < \alpha < \nu$  (see also [45]).

*Proof.* Let us start with the case in which  $\nu$  is not an integer. Lemma 2.2.3 tells us that  $\widetilde{\text{cov}}_\gamma \in C^{2\nu}(A)$  for any bounded set  $A \in \mathbb{R}^d$ . Therefore, thanks to (2.8), by writing  $\nu = n + s$  with  $n \in \mathbb{N}$  and  $s \in (0, 1)$ , for any multi-index  $i \in \mathbb{N}^d$  such that  $|i|_1 = n$ , we obtain

$$\begin{aligned} \mathbb{E}[(D^i \gamma(x, \cdot) - D^i \gamma(y, \cdot))^2] &= \mathbb{E}[(D^i \gamma(x, \cdot))^2] + \mathbb{E}[(D^i \gamma(y, \cdot))^2] - 2\mathbb{E}[D^i \gamma(x, \cdot) D^i \gamma(y, \cdot)] = \\ &= 2 \left( \frac{\partial^{2|i|_1} \widetilde{\text{cov}}_\gamma}{\partial x_1^{i_1} \dots \partial x_d^{i_d} \partial y_1^{i_1} \dots \partial y_d^{i_d}}(0) - \frac{\partial^{2|i|_1} \widetilde{\text{cov}}_\gamma}{\partial x_1^{i_1} \dots \partial x_d^{i_d} \partial y_1^{i_1} \dots \partial y_d^{i_d}}(|x - y|) \right) \\ &\leq C(\nu) |x - y|^{2s}, \end{aligned}$$

where the last inequality comes from the fact that the coefficients appearing in the covariance function decay sufficiently fast. Since for any positive integer  $p$  it holds  $\mathbb{E}[(D^i \gamma(x, \cdot) - D^i \gamma(y, \cdot))^{2p}] \leq c_p \mathbb{E}[(D^i \gamma(x, \cdot) - D^i \gamma(y, \cdot))^2]^p$  with  $c_p = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^{2p} e^{-\frac{x^2}{2}} dx$  we have

$$\mathbb{E}[(D^i \gamma(x, \cdot) - D^i \gamma(y, \cdot))^{2p}] \leq c_p C(\nu)^p |x - y|^{2ps}.$$

Thanks to the Kolmogorov continuity theorem (see e.g. [29]) we can deduce that there exists a version of  $D^i \gamma$  which belongs to  $C^a(\bar{D})$  for any  $a < \frac{2ps-d}{2p}$ ; by taking the limit for  $p \rightarrow +\infty$  we can conclude that there exist a version of  $D^i \gamma$  which belongs to  $C^a(\bar{D})$  for any  $a$  strictly smaller than  $s$ . Consequently, since this reasoning can be repeated for every  $k < n$ ,  $k \in \mathbb{N}$ , by picking 1 instead of  $s$ , we deduce that there exist a version of  $\gamma$  which belongs to  $C^\alpha(\bar{D})$  for any  $\alpha$  strictly smaller than  $\nu$ . The proof in the case  $\nu \in \mathbb{N}$  is similar. By writing  $\nu = n + 1$  in this case, thanks to (2.8), for any  $\epsilon > 0$  and for any multi-index  $i \in \mathbb{N}^d$  such that  $|i|_1 = n$  we obtain

$$\mathbb{E}[(D^i \gamma(x, \cdot) - D^i \gamma(y, \cdot))^2] \leq \tilde{C}(\nu) \frac{\partial^{2|i|_1} |x - y|^{2n} \log(|x - y|)}{\partial x_1^{i_1} \dots \partial x_d^{i_d} \partial y_1^{i_1} \dots \partial y_d^{i_d}} \leq C_\epsilon(\nu) |x - y|^{2-\epsilon}.$$

Again, thanks to the Kolmogorov continuity theorem we can deduce that there exists a version of  $D^i \gamma$  which belongs to  $C^a(\bar{D})$  for any  $a < \frac{p(2-\epsilon)-d}{2p}$ ; thanks to the arbitrariness of  $\epsilon$  by taking the limit for  $p \rightarrow +\infty$  we can conclude that there exist a version of  $D^i \gamma$  which belongs to  $C^a(\bar{D})$  for any  $a$  strictly smaller than 1. Consequently we deduce that there exist a version of  $\gamma$  which belongs to  $C^\alpha(\bar{D})$  for any  $\alpha$  strictly smaller than  $\nu$ .  $\square$

Hence for  $\nu = 0.5$  the covariance function is only Lipschitz continuous and the field is Hölder continuous  $\gamma(\cdot, \omega) \in C^\alpha(\bar{D})$  with  $\alpha < 0.5$ . On the other hand, for  $\nu \rightarrow \infty$  the covariance function as well as the field are continuous with all their derivatives, namely  $\text{cov}_\gamma(\cdot) \in C^\infty(\bar{D} \times \bar{D})$  and  $\gamma(\cdot, \omega) \in$

$C^\infty(\bar{D})$  a.s. in  $\Omega$ . It is important to notice that every log-normal random field  $a$  that can be obtained starting from a Gaussian log-permeability  $\gamma$  having a covariance function belonging to the Matérn family satisfies the hypothesis of Lemma 2.2.1.

The goal of the analysis is to compute statistics of some quantities of interest given by a linear functional  $Q(p) \in \mathbb{R}$  related to the solution of (2.2).

## 2.3 Finite Element Approximation

In order to numerically solve problem (2.3) we consider a piecewise linear finite element approximation  $p_h$  of  $p$  on a regular triangulation  $\mathcal{T}_h$  of the domain. The approximate solution  $p_h \in V_{h,g}$  solves the problem

$$b_\omega(p_h, v) = L(v) \quad \forall v \in V_{h,0}, \quad (2.9)$$

where  $V_{h,g} = \{v_h \in \mathcal{C}^0(\bar{D}) : v_h|_K \in \mathbb{P}_1 \ \forall K \in \mathcal{T}_h \text{ and } v_h = I_h g \text{ on } \Gamma_D\}$ , and  $I_h g$  is a suitable interpolation of the Dirichlet boundary datum.

Concerning the finite element approximation error of the original problem (2.2) the following result, taken from [19], holds:

**Lemma 2.3.1.** *Let  $D$  be a  $\mathcal{C}^2$  convex bounded domain and let  $a(x, \omega)$  be a log-normal stationary random field with realizations a.s. in  $C^\alpha(\bar{D})$ ,  $\Gamma_D = \partial D$ ,  $g = 0$  and  $f \in H^{r-1}(D)$  with  $r = \min(\alpha, 1)$ . If  $\alpha \leq 1$ , by using linear finite elements for the spatial discretization, and assuming all integrals are computed exactly in (2.9), it holds:*

$$\|p(\cdot, \omega) - p_h(\cdot, \omega)\|_{H_0^1(D)} \lesssim \frac{1}{\alpha - \beta} C_{2.3.1}(\omega, \alpha) \|f\|_{H^{\alpha-1}(D)} h^\beta, \quad \forall 0 \leq \beta < \alpha \quad \text{a.s. in } \Omega.$$

where

$$C_{2.3.1}(\omega, \alpha) = \sqrt{\frac{a_{\max}(\omega)}{a_{\min}(\omega)}} C_{2.2.1}(\omega, \alpha).$$

If  $\alpha > 1$  it holds

$$\|p(\cdot, \omega) - p_h(\cdot, \omega)\|_{H_0^1(D)} \lesssim C_{2.3.1}(\omega, \alpha) \|f\|_{L^2(D)} h \quad \text{a.s. in } \Omega.$$

Moreover the random variable  $C_{2.3.1}(\omega, \alpha)$  is  $q$ -integrable for any  $q \in \mathbb{R}_+$ , i.e.  $C_{2.3.1}(\omega, \alpha) \in L_{\mathbb{P}}^q(\Omega) \ \forall q \in \mathbb{R}_+$ .

**Remark 2.3.1.** Since  $C_{2.3.1}(\omega, \alpha) \in L_{\mathbb{P}}^q(\Omega) \ \forall q \in \mathbb{R}_+$ , we deduce immediately from Lemma 2.3.1 in the case  $\alpha \leq 1$  the bound

$$\|p - p_h\|_{L_{\mathbb{P}}^q(\Omega; H_0^1(D))} \lesssim \frac{c_{2.3.1}(\alpha, q)}{\alpha - \beta} \|f\|_{H^{\alpha-1}(D)} h^\beta, \quad \forall 0 \leq \beta < \alpha,$$

where  $c_{2.3.1}(\alpha, q) = \|C_{2.3.1}(\cdot, \alpha)\|_{L_{\mathbb{P}}^q(\Omega)}$ , and in the case  $\alpha > 1$  the bound

$$\|p - p_h\|_{L_{\mathbb{P}}^q(\Omega; H_0^1(D))} \lesssim c_{2.3.1}(\alpha, q) \|f\|_{L^2(D)} h.$$

Starting from this result, it is straightforward to obtain a bound on a specific QoI represented by a linear functional. The following result holds:

**Lemma 2.3.2.** *Let  $a$ ,  $f$ ,  $D$  and  $r$  be as in Lemma 2.3.1 and let  $Q(\cdot)$  be a functional on  $H^{1-r}(D)$ , i.e.  $Q \in H^{r-1}(D)$ , representing our QoI. Then, by using linear finite elements for the spatial discretization,*

for  $\alpha \leq 1$  a.s. in  $\Omega$  it holds:

$$|Q(p - p_h)(\omega)| \lesssim \frac{C_{2.3.2}(\omega, \alpha)}{(\alpha - \beta)^2} \|f\|_{H^{\alpha-1}(D)} \|Q\|_{H^{\alpha-1}(D)} h^{2\beta}, \quad \forall 0 \leq \beta < \alpha \quad \text{a.s. in } \Omega.$$

where  $C_{2.3.2}(\omega, \alpha, \beta) = C_{2.3.1}^2(\omega, \alpha)$ . If the assumptions hold also for  $\alpha > 1$  then, it is also valid the bound

$$|Q(p - p_h)(\omega)| \lesssim C_{2.3.2}(\omega, \alpha) \|f\|_{L^2(D)} \|Q\|_{L^2(D)} h^2, \quad \text{a.s. in } \Omega.$$

**Remark 2.3.2.** Since  $C_{2.3.1}(\omega, \alpha) \in L_{\mathbb{P}}^q(\Omega) \quad \forall q \in \mathbb{R}_+$ , the same goes for  $C_{2.3.2}$ . Hence the bounds in Lemma 2.3.2 can be rewritten as

$$\|Q(p) - Q(p_h)\|_{L_{\mathbb{P}}^q(\Omega)} \lesssim \frac{c_{2.3.2}(\alpha, q)}{(\alpha - \beta)^2} \|f\|_{H^{\alpha-1}(D)} \|Q\|_{H^{\alpha-1}(D)} h^{2\beta}, \quad \forall 0 \leq \beta < \alpha,$$

where  $c_{2.3.2}(\alpha, q) = \|C_{2.3.2}(\cdot, \alpha)\|_{L_{\mathbb{P}}^q(\Omega)}$ , and in the case  $\alpha > 1$  the bound

$$\|Q(p) - Q(p_h)\|_{L_{\mathbb{P}}^q(\Omega)} \lesssim c_{2.3.2}(\alpha, q) \|f\|_{L^2(D)} \|Q\|_{L^2(D)} h^2.$$

## 2.4 The Truncated Darcy Problem

For the methods proposed in this work we need to expand the input random field in a countable number of independent identically distributed standard normal random variables  $y_n$ , namely

$$\gamma(\mathbf{x}, \omega) = \mu(\mathbf{x}) + \sum_{n=1}^{\infty} \sqrt{\lambda_n} y_n(\omega) b_n(\mathbf{x}), \quad (2.10)$$

where  $\lambda_n$  are coefficients whose decay depends on the smoothness of the covariance function and  $b_n$  are suitably normalized functions in  $D$ , which may be for instance Fourier modes or Karhunen-Loève modes, namely the eigenfunctions of the covariance operator  $T_\gamma : L^2(D) \rightarrow L^2(D)$  given by

$$T_\gamma v(\mathbf{x}) = \int_D v(\mathbf{x}') \text{cov}_\gamma(\mathbf{x}, \mathbf{x}') d\mathbf{x}'.$$

Then, in order to use numerical methods to solve the problem we will consider truncated versions of the random field:

$$\gamma_N(\mathbf{x}, \mathbf{y}(\omega)) = \mu(\mathbf{x}) + \sum_{n=1}^N \sqrt{\lambda_n} y_n(\omega) b_n(\mathbf{x}) \quad \text{where } \mathbf{y} = (y_1, \dots, y_N), \quad (2.11)$$

with  $N$  chosen so that the error due to the truncation (see [18, 45]) of the input random field is sufficiently small compared to the space discretization error induced by the finite element approximation. Observe that the truncated random field  $\gamma_N(x, \mathbf{y})$  can be seen as the original random field  $\gamma$  evaluated at  $\mathbf{y} = (y_1, \dots, y_N, 0, 0, \dots)$ ; in the following we will refer to these first  $N$  random variables as “active variables”. In the case of a random field with limited regularity, since the decay of the coefficients  $\lambda_n$  in (2.10) is slow, many terms will have to be included in (2.11) to have a truncation error sufficiently small. The truncated Darcy problem, in its continuous and discretized form, is then obtained starting respectively from (2.3) and (2.9) by replacing  $\gamma$  with  $\gamma_N$  and the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with  $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N), \varrho(\mathbf{y}))$ , being  $\mathcal{B}(\mathbb{R}^N)$  the  $\sigma$ -algebra of Borel subsets of  $\mathbb{R}^N$  and  $\varrho(\mathbf{y}) = \prod_{n=1}^N \varrho_n(y_n)$  the probability density of the standard Gaussian vector  $\mathbf{y}$ ;

consistently we denote  $p_N$  and  $p_{N,h}$  the solutions of the two corresponding variational problems

$$b_{\mathbf{y}}(p_N, v) = L(v) \quad \forall v \in V_0, \quad b_{\mathbf{y}}(p_{N,h}, v) = L(v) \quad \forall v \in V_{h,0}, \quad (2.12)$$

where the bilinear form  $b_{\mathbf{y}}$  is defined as  $b_{\mathbf{y}}(u, v) = \int_D a_N(x, \mathbf{y}) \nabla u(x, \mathbf{y}) \nabla v(x) dx$

By taking into account only the error coming from the truncation of the input permeability field  $a_N = e^{\gamma_N}$  and from the spatial discretization through finite elements we can bound the error related to the corresponding QoI as

$$\|Q - Q_{N,h}\| \leq \|Q - Q_N\| + \|Q_N - Q_{N,h}\|.$$

The first term represents the truncation error that we are committing by replacing the permeability  $a$  with its finite dimensional version  $a_N$  (see [17]); the second term represents the error that we are committing by solving with finite elements the truncated problem. In the following we will assume that the truncation error will be negligible compared to the finite element one and for this reason, when there is no ambiguity, the subscript  $N$  will be omitted.

## 2.5 Sampling schemes

The final goal of the analysis is to well approximate the QoI  $Q(\mathbf{y})$  with its finite dimensional version  $Q_{N,h}(\mathbf{y})$ , where  $\mathbf{y}$  represent a random Gaussian vector with 0 mean and identity covariance matrix, and  $h$  is the spatial discretization parameter. In this Thesis, depending on the specific case, we will parametrize the random field  $\gamma$  with a finite number of random variables according to equation (2.11), in which the function  $b_n$  are either trigonometric functions forming a Fourier basis or Karhunen-Loève basis functions. Whenever we will use methods, such as Monte Carlo (MC) sampling, which are basically not affected by the so called “curse of dimensionality” we will perform the sampling on a Fourier basis, by using the tools provided by the FFT algorithm which well behaves in terms of complexity; on the other hand, when we deal with numerical schemes whose convergence performance deteriorates with respect to  $N$ , and hence for numerical purposes we would like to keep as small as possible the number of variables that “significantly affect” the QoI  $Q(\mathbf{y})$ , we perform the sampling on a Karhunen-Loève basis, which is known to be the “best”  $N$ -term expansion, i.e. the one minimizing  $\|\gamma - \gamma_N\|_{L^2_{\mathbb{P}}(\Omega; L^2(D))}$ .

### 2.5.1 Karhunen-Loève expansion

In order to expand the random field  $\gamma$  on a KL basis we need to solve the continuous eigenvalue problem

$$\int_D v(\mathbf{x}') \text{cov}_{\gamma}(\mathbf{x}, \mathbf{x}') d\mathbf{x}' = \lambda v(\mathbf{x}). \quad (2.13)$$

From a practical point of view, for a given triangulation  $\mathcal{T}_h$  of  $D$  (not necessarily the same used to solve the Darcy problem with finite elements), this leads to the computation of an approximate KL-expansion which requires the numerical solution of the matrix eigenvalue problem

$$M \Sigma M^T \mathbf{b} = \lambda M \mathbf{b}. \quad (2.14)$$

For instance, if we consider piecewise linear F.E. and denote by  $\{x_i\}_{i=1}^{N_v}$  the vertices of the triangulation and  $\{\ell_i\}_{i=1}^{N_v}$  the corresponding Lagrangian  $\mathbb{P}^1$  basis functions, then  $M$  represents the mass matrix such that  $M_{i,j} = \int_D \ell_i(x) \ell_j(x) dx$  and  $\Sigma$  is the covariance matrix  $\Sigma_{i,j} = \text{cov}_\gamma(\mathbf{x}_i, \mathbf{x}_j)$  associated to the vertices of the triangulation.

The size  $N_v$  of the KL eigenvalue problem (2.14) can be very large and, since the matrix  $\Sigma$  is non sparse, its solution can pose numerical challenges. More efficient strategy have been proposed to compute KL expansions, for instance in [70] the authors achieved a complexity in terms of number of operations of  $O(Nh^{-d} \log(h)^{-d\beta})$ , for some positive  $\beta$ , being  $N$  the number of modes to be kept in the KL expansion; such approximation guarantees accuracy  $O(\exp(-\beta N^{1/d}))$ .

Moreover it is important to say that, by performing a discretization of the physical domain  $D$ , we are implicitly restricting ourselves to consider a finite number of pairs  $(\lambda_n, b_n(\mathbf{x}))$ . Since the exact eigenvalues form a decreasing sequence converging to 0, the solution of the matrix eigenproblem (2.14) will produce good estimates only of the pairs corresponding to the largest eigenvalues. This means that, if we want to keep a high number of random variables in our truncated field  $\gamma_N$  in order to obtain a negligible truncation error on the solution of the PDE  $p$ , we are forced to solve an eigenproblem having size  $N_v \gg N$ . The construction of the basis becomes much easier when considering a tensor covariance function as in (2.6): in this case it suffices to solve the eigenvalue problem in dimension  $D = 1$  and then tensorising to get the first  $N$  pairs  $(\lambda_n, b_n(\mathbf{x}))$  in higher dimension.

The truncation error related to the solution  $p$  of the Darcy problem has been shown to decay algebraically with respect to  $N$ , at least for finite differentiable field  $\gamma$  (see [18, 45, 17]); the convergence properties of such a truncation error strongly depend on the decay of the eigenvalues  $\lambda_n$ : the faster the decay the better the convergence. In Figure 2.1 we show the decay of the eigenvalues of a two dimensional covariance function in the Matérn family of either tensor type (2.6) or isotropic type (2.5) for different choices of correlation length  $L_c$ , that typically influences the pre-asymptotic behavior of the sequence  $\{\lambda_n\}_{n \in \mathbb{N}}$ , and  $\nu$ , that, on the contrary, affects the asymptotic behavior. Moreover, the eigenvalues coming from a tensor covariance function present a better rate than the ones coming from a radial one which follow the predicted asymptotic behavior  $\lambda_n \lesssim n^{-(1+2\nu/d)}$  (see [46]). It is then clear that, when dealing with rough permeability fields, it is necessary to include many more terms than when dealing with smooth ones to keep the truncation error at acceptable levels. This, of course, may represent a problem in case of radial covariance functions since, to achieve small values of tolerances, we might need many terms (order of  $10^4 - 10^5$ ) in the truncated field. In such cases we switch to Fourier expansions of the random field and use a standard FFT approach to sample from the field at a low computational cost.

### 2.5.2 Fourier expansion

We detail hereafter the 2D case. If the domain  $D$  is not a rectangle, in order to sample from a Fourier expansion, we consider the smallest rectangle  $[a, b] \times [c, d]$  that contains  $D$  and we write the expansion on this set. Without loss of generality, suppose we want to expand the random field  $\gamma$  on the unitary square  $D = (0, 1)^2$ . The idea is to expand the covariance function on the square  $(-L, L)^2$ , with  $L \geq 1$  properly chosen, namely:

$$\widetilde{\text{cov}}_\gamma(|\mathbf{x} - \mathbf{x}'|) = \sum_{\mathbf{n} \in \mathbb{N}^2} c_{\mathbf{n}}^2 \prod_{i=1}^2 \cos\left(\frac{\pi n_i}{L} (x_i - x'_i)\right). \quad (2.15)$$

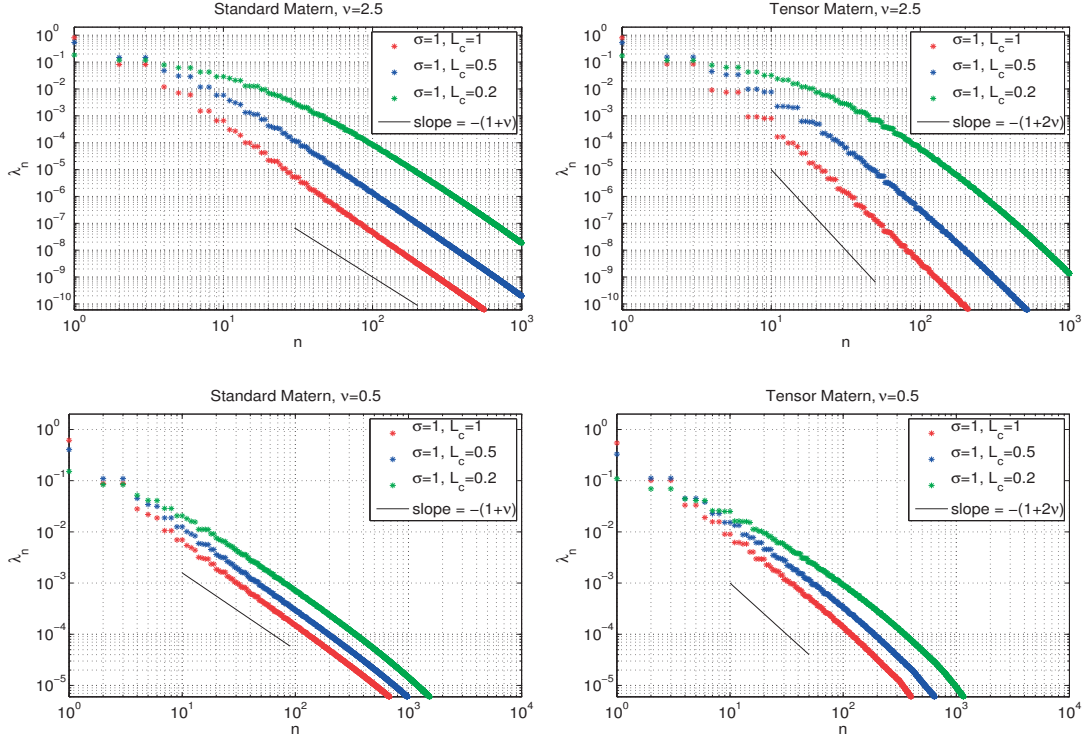


Figure 2.1 – Decay of  $\lambda_n$ . Left: radial covariance function as in (2.5). Right: tensor covariance function as in (2.6). Top:  $\nu = 2.5$ . Bottom:  $\nu = 0.5$ .  $d = 2$ .

By doing this we are periodically replicating the function  $\widetilde{\text{cov}}_\gamma(|\mathbf{x} - \mathbf{x}'|)$ ,  $\mathbf{x}, \mathbf{x}' \in (-L, L)^2$ , in  $\mathbb{R}^2$ : the resulting function will therefore present, in general, discontinuities of the first derivatives along the edges of the reference domain  $(-L, L)^2$ . We have to be careful in operating such a replication of the field in  $\mathbb{R}^2$  since we want to end up with a periodic covariance with Fourier coefficients presenting the same decay of the original ones (coming from the non periodic covariance function). More specifically, suppose we start from an exponential covariance function, which is not differentiable in the origin; in this case, since our original covariance has discontinuous derivatives in the origin, even if we introduce an additional discontinuity of the derivatives in correspondence of the boundary by replicating the field on  $\mathbb{R}^2$ , we would not worsen the decay of the Fourier coefficient of the original field; conversely, if we start from a smooth field, we have to choose  $L > L_c$  sufficiently large to mitigate the effect produced by these singularities in order to avoid any worsening of the Fourier decay of the replicated covariance with respect to the original one. Since the parameter that regulates the “actual support” of the covariance function, i.e. the set of points on which the covariance function takes values larger than a prescribed “small” threshold, is the correlation length  $L_c$ , we compute the reference domain size as  $L = \max(6L_c, 1)$ . This choice has been verified to generate accurate enough estimates of the Fourier coefficients, by checking that the results obtained would have not changed by considering a larger factor than 6. Of course the larger the domain size  $L$  is, the larger the number of modes to be kept will be in order to include in the series prescribed frequencies.

Starting from this Fourier representation of the covariance, the field  $\gamma(\mathbf{x}, \mathbf{y})$  can be written as

$$\gamma(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{n} \in \mathbb{N}^2} c_{\mathbf{n}} \sum_{\ell \in \{0,1\}^2} y_{\mathbf{n}}^{\ell}(\omega) \prod_{i=1}^2 \cos(\kappa_i x_i)^{\ell_i} \sin(\kappa_i x_i)^{1-\ell_i}, \quad (2.16)$$

where  $y_{\mathbf{n}}^{\ell}(\omega)$  are i.i.d. standard Gaussian random variables and  $\boldsymbol{\kappa} = \frac{\pi \mathbf{n}}{L}$ . It is immediate to notice as  $\mathbb{E}[\gamma(\cdot, \mathbf{x})] = 0$ ; at the same time, by recalling that the random variables  $y_{\mathbf{n}}^{\ell}(\omega)$  are mutually independent, we have,

$$\mathbb{E}[\gamma(\cdot, \mathbf{x})\gamma(\cdot, \mathbf{x}')] = \sum_{\mathbf{n} \in \mathbb{N}^2} c_{\mathbf{n}}^2 \sum_{\ell \in \{0,1\}^2} \prod_{i=1}^2 \cos(\kappa_i x_i)^{\ell_i} \sin(\kappa_i x_i)^{1-\ell_i} \cos(\kappa_i x'_i)^{\ell_i} \sin(\kappa_i x'_i)^{1-\ell_i} = \sum_{\mathbf{n} \in \mathbb{N}^2} c_{\mathbf{n}}^2 \cos(\kappa_i(x_i - x'_i)),$$

which is the Fourier series of the even extension over the square  $(-L, L)^2$  of the considered covariance function. We recall that, given an  $N$ -term vector with entries  $v_n \in \mathbb{C}$ , its discrete Fourier transform is a  $N$ -term vector with entries

$$\hat{v}_n = \sum_{j=0}^{N-1} v_j e^{-i \frac{2\pi}{N} n j}.$$

To be able to use the tools coming from the FFT we first truncate the Fourier series (2.15) by considering  $\mathbf{n} \in \{0, \dots, N\}^2$  for some  $N \in \mathbb{N}$  and then we rewrite it with respect to the complex exponential function  $e^{i \frac{\pi \mathbf{z}}{L} \cdot \mathbf{x}}$  with  $\mathbf{z} \in \mathbb{Z}^2$ ; we obtain

$$\widetilde{\text{cov}}_{\gamma}(|\mathbf{x}|) = \sum_{\mathbf{z} \in \{-N, \dots, N\}^2} w_{\mathbf{z}} c_{(|z_1|, |z_2|)}^2 e^{i \frac{\pi}{L} \mathbf{z} \cdot \mathbf{x}}, \quad (2.17)$$

where  $w_{\mathbf{z}} = 1/4$  if  $z_1 z_2 \neq 0$ ,  $w_{\mathbf{z}} = 1/2$  if  $z_1 z_2 = 0$  and one of the two is different than 0,  $w_{\mathbf{z}} = 1$  if  $z_1 = z_2 = 0$ . Then we discretize the point  $\mathbf{x}$  in the reference domain  $(-L, L)^2$  by putting  $2N+1$  uniformly distributed points in any direction, namely  $\mathbf{x}_{j,k} = (x_j, x_k)$ , with  $x_j = -L + \frac{2L}{2N} j$  for  $j = 0, \dots, 2N$  and  $x_k = -L + \frac{2L}{2N} k$  for  $k = 0, \dots, 2N$ . Notice that  $\widetilde{\text{cov}}_{\gamma}(|\mathbf{x}_{0,k}|) = \widetilde{\text{cov}}_{\gamma}(|\mathbf{x}_{2N,k}|)$  and  $\widetilde{\text{cov}}_{\gamma}(|\mathbf{x}_{j,0}|) = \widetilde{\text{cov}}_{\gamma}(|\mathbf{x}_{j,2N}|)$ . By approximating the integral involved in the definition of  $c_{\mathbf{n}}^2$  with trapezoidal rule we obtain

$$w_{\mathbf{z}} c_{(|z_1|, |z_2|)}^2 = \frac{1}{(2L)^2} \int_{(-L,L)^2} \widetilde{\text{cov}}_{\gamma}(|\mathbf{x}|) e^{-i \frac{\pi}{L} \mathbf{z} \cdot \mathbf{x}} d\mathbf{x} \approx \frac{(-1)^{z_1+z_2}}{(2N)^2} \sum_{j,k=0}^{2N-1} \widetilde{\text{cov}}_{\gamma}(|\mathbf{x}_{j,k}|) e^{-i \frac{2\pi}{2N} (z_1 j + z_2 k)} := \tilde{c}_{\mathbf{z}}^2. \quad (2.18)$$

The double sum represents the two-dimensional discrete Fourier transform (DFT) and can be efficiently computed via FFT, especially when  $N$  is chosen as a power of 2. Once we have computed the approximated Fourier coefficients, we can generate realizations of the random field using formula (2.16). Notice that, even if we have to generate  $M$  realizations, we need to compute the Fourier coefficients in (2.17) only once. We can then afford to compute many more terms in the DFT and then just keep only the coefficients such that  $\mathbf{z} \in \{-N, \dots, N\}^2$ . Now in order to use again the inverse DFT to generate our realization we rewrite the truncated expansion  $\gamma_N$  in terms of the complex exponentials

$$\begin{aligned} \gamma(\mathbf{x}) &= \sum_{\mathbf{n} \in \mathbb{N}^2} \frac{c_{\mathbf{n}}}{4} \left( e^{i\kappa_1 x_1} e^{i\kappa_2 x_2} \left( y_{\mathbf{n}}^{(1,1)} - i y_{\mathbf{n}}^{(1,0)} - i y_{\mathbf{n}}^{(0,1)} - y_{\mathbf{n}}^{(0,0)} \right) + e^{i\kappa_1 x_1} e^{-i\kappa_2 x_2} \left( y_{\mathbf{n}}^{(1,1)} + i y_{\mathbf{n}}^{(1,0)} - i y_{\mathbf{n}}^{(0,1)} + y_{\mathbf{n}}^{(0,0)} \right) \right. \\ &\quad \left. + e^{-i\kappa_1 x_1} e^{i\kappa_2 x_2} \left( y_{\mathbf{n}}^{(1,1)} - i y_{\mathbf{n}}^{(1,0)} + i y_{\mathbf{n}}^{(0,1)} + y_{\mathbf{n}}^{(0,0)} \right) + e^{-i\kappa_1 x_1} e^{-i\kappa_2 x_2} \left( y_{\mathbf{n}}^{(1,1)} + i y_{\mathbf{n}}^{(1,0)} + i y_{\mathbf{n}}^{(0,1)} - y_{\mathbf{n}}^{(0,0)} \right) \right) \\ &= \sum_{\mathbf{n} \in \mathbb{N}^2} \frac{c_{\mathbf{n}}}{4} \left( e^{i\kappa_1 x_1} e^{i\kappa_2 x_2} Y_{\mathbf{n}}^{(1,1)} + e^{i\kappa_1 x_1} e^{-i\kappa_2 x_2} Y_{\mathbf{n}}^{(1,0)} + e^{-i\kappa_1 x_1} e^{i\kappa_2 x_2} Y_{\mathbf{n}}^{(0,1)} + e^{-i\kappa_1 x_1} e^{-i\kappa_2 x_2} Y_{\mathbf{n}}^{(0,0)} \right), \end{aligned} \quad (2.19)$$



where  $Y_{\mathbf{n}}^\ell$ ,  $\ell \in \{0, 1\}^2$  are proper complex standard normal random variables mutually independent defined through the previous steps. Again we truncate (2.19) by considering only the multi-indexes  $\mathbf{n} \in \{0, N\}^2$  and rewrite the series as

$$\gamma(\mathbf{x})_N = \sum_{\mathbf{z} \in \{-N, \dots, N\}^2} w_{\mathbf{z}} c_{(|z_1|, |z_2|)} \tilde{Y}_{\mathbf{z}} e^{i \frac{\pi}{L} \mathbf{z} \cdot \mathbf{x}} = \sum_{\mathbf{z} \in \{-N, \dots, N\}^2} \sqrt{w_{\mathbf{z}}} \tilde{c}_{\mathbf{z}} \tilde{Y}_{\mathbf{z}} e^{i \frac{\pi}{L} \mathbf{z} \cdot \mathbf{x}}, \quad (2.20)$$

where  $\tilde{Y}_{\mathbf{z}} = Y_{(|z_1|, |z_2|)}^{(\mathbf{1}_{z_1}, \mathbf{1}_{z_2})}$  when  $z_1 z_2 \neq 0$ ,  $\tilde{Y}_{\mathbf{z}} = Y_{(|z_1|, 0)}^{(\mathbf{1}_{z_1}, 1)} + Y_{(|z_1|, 0)}^{(\mathbf{1}_{z_1}, 0)}$  when  $z_2 = 0$  and  $z_1 \neq 0$ ,  $\tilde{Y}_{\mathbf{z}} = Y_{(0, |z_2|)}^{(1, \mathbf{1}_{z_2})} + Y_{(0, |z_2|)}^{(0, \mathbf{1}_{z_2})}$  when  $z_1 = 0$  and  $z_2 \neq 0$ , and finally  $\tilde{Y}_{\mathbf{z}} = 4Y_0^{1,1}$  when  $\mathbf{z} = \mathbf{0}$ . By evaluating the realization in the points  $\mathbf{x}_{j,k}$  we obtain

$$\gamma(\mathbf{x}_{i,j})_N = (-1)^{i+j} \sum_{\mathbf{z} \in \{-N, N\}^2} \sqrt{w_{\mathbf{z}}} \tilde{c}_{\mathbf{z}} \tilde{Y}_{\mathbf{z}} e^{i \frac{2\pi}{2N} (z_1 j + z_2 k)}.$$

To use the inverse discrete Fourier transform we want to consider a double sum involving  $(2N)^2$  terms; we then separate the sum as follows:

$$\begin{aligned} \gamma(\mathbf{x}_{i,j}, \mathbf{y})_N = & (-1)^{i+j} \left( \sum_{\mathbf{z} \in \{-N+1, \dots, N\}^2} \sqrt{w_{\mathbf{z}}} \tilde{c}_{\mathbf{z}} \tilde{Y}_{\mathbf{z}} e^{i \frac{2\pi}{2N} (z_1 j + z_2 k)} + \sum_{\substack{z_2 = -N, \\ z_1 \in \{-N+1, \dots, N\}}} \sqrt{w_{\mathbf{z}}} \tilde{c}_{\mathbf{z}} \tilde{Y}_{\mathbf{z}} e^{i \frac{2\pi}{2N} (z_1 j + z_2 k)} \right. \\ & \left. + \sum_{\substack{z_1 = -N, \\ z_2 \in \{-N+1, \dots, N\}}} \sqrt{w_{\mathbf{z}}} \tilde{c}_{\mathbf{z}} \tilde{Y}_{\mathbf{z}} e^{i \frac{2\pi}{2N} (z_1 j + z_2 k)} + \tilde{c}_{(-N, -N)} \tilde{Y}_{(-N, -N)} e^{i \frac{2\pi}{2N} (-Nj - Nk)} \right). \quad (2.21) \end{aligned}$$

The double sum represent again the two-dimensional inverse DFT, while the two successive sums represent two one-dimensional inverse DFT; the overall cost to generate a discrete realization of the random field, evaluated in the tensor points  $\mathbf{x}_{j,k}$ ,  $j, k = 0, \dots, 2N$ , thanks to the efficiency of the inverse FFT algorithm is then  $\mathcal{O}(N^2 \log(N)^2)$ , while a naive implementation to compute the DFT would have had a computational cost scaling as  $\mathcal{O}(N^4)$ . Notice that, since this realization is defined on  $(-L, L)^2$ , for our purposes we will just keep the vertices that belong to  $(0, 1)^2$ .

Observe also that, by operating in this way, there is an intrinsic link between the number of spatial vertices and the number of Fourier coefficients (hence of random variables). If we want to consider a fine mesh, automatically we have to include in the expansion of the field many random variables; on the other hand, if we want to consider many random variables on a coarse mesh we will have to sample on a finer mesh (being the two possibly nested in order to avoid any interpolation error) and then pick up only the values corresponding to the vertices of the coarse mesh.

The strategy described above is very efficient since it uses FFT both to compute the Fourier coefficients and generate realizations of the random field on a uniform grid. It shares some analogies with the popular circulant embedding approach [31, 55]. In particular it presents the same computational cost, since it relies on the features of the FFT algorithm. However, this approximation does not ensure the positivity of the coefficients  $\tilde{c}_{\mathbf{z}}^2$ ; nevertheless, since we compute the coefficients  $\tilde{c}_{\mathbf{z}}^2$  only once by approximating the integral in (2.18), we can actually afford to run this calculation by using a value  $\tilde{N} \gg N$  and then keep only the coefficients such that  $\mathbf{z} \in \{-N, \dots, N\}^2$ ; by doing this we can reasonably assume to commit a negligible error in the computation of the Fourier coefficients. In what comes next we do not introduce any other source of error since we use the inverse DFT just to perform the summation appearing in the definition of the Fourier series.



The strategy to obtain a realization of the random field  $\gamma$  on a structured grid can be summarized as follows:

1. given a mesh size  $h_\ell = 2^{-\ell}$  and a covariance function  $\widetilde{\text{cov}}_\gamma$  set  $N = h_\ell^{-1}$  and compute the approximated Fourier coefficients according to (2.18);
2. generate the random variables  $y_{\mathbf{n}}^\ell$ ,  $\mathbf{n} \in \{0, \dots, N\}^2$ ,  $\ell \in \{0, 1\}^2$  and compute the corresponding variables  $Y_{\mathbf{n}}^\ell$  according to (2.19);
3. generate the realization by using FFT to evaluate the summation in (2.21).



# 3 Sparse Grid polynomial approximation for random PDEs in high dimensional probability spaces

## 3.1 Introduction

In this chapter we deal with the polynomial approximation in the random variables of the stochastic Darcy problem. We start the discussion without any specific assumption on the law of the finite random vector  $\mathbf{y} \in \Gamma \subset \mathbb{R}^N$  and on the permeability field  $a(\mathbf{x}, \mathbf{y})$ . Observe that, due to the *curse of dimensionality*, the problem can become particularly challenging when the dimension of the stochastic space  $N$  is large; for this reason such elliptic random PDE have been extensively studied during the last few years from both a theoretical and a numerical point of view. The goal of such analysis is trying to eliminate, or at least reduce as much as possible, the effects of the *curse of dimensionality*. It is clear that, in order to achieve such a goal, the polynomial approximation of the map  $\mathbf{y} \rightarrow p(\mathbf{y})$  has to exploit as much as possible the anisotropy of the solution with respect to the different random inputs  $y_n$ ,  $n = 1, \dots, N$ .

To achieve this goal, we consider the problem of building a sparse grid approximation of a multivariate function  $Q(\mathbf{y}) : \Gamma \rightarrow V$  with global polynomials, where  $\Gamma$  is an  $N$ -dimensional hypercube  $\Gamma = \Gamma_1 \times \Gamma_2 \times \dots \times \Gamma_N$  (with  $\Gamma_n \subseteq \mathbb{R}$ ,  $n = 1, \dots, N$ ), and  $V$  is a Hilbert space [15, 8, 71, 60, 4]. We point out that, although the methodology presented in this chapter will be entirely applied to the stochastic lognormal Darcy problem, it is possible to use it also for problems depending on uniform random variables by suitably changing the collocation points; further details will be provided in the following. We also assume the random variables  $y_1, \dots, y_N$  to be independent and identically distributed and each  $\Gamma_n$  to be endowed with a probability measure  $\varrho_n(y_n)dy_n$ , so that the joint probability density function of the vector  $\mathbf{y}$  is the product of the marginals, namely  $\varrho(\mathbf{y})d\mathbf{y} = \prod_{n=1}^N \varrho_n(y_n)dy_n$  on  $\Gamma$ .

This setting is common in many optimization and Uncertainty Quantification problems, where sparse grids have been increasingly used to perform tasks such as quadrature, interpolation and surrogate modeling, since they allow for trivial parallelization and maximal reuse of legacy codes, with little or no expertise required by the end-user. While very effective for moderate dimensions (say  $N \approx 10$ ), the basic sparse grid algorithms show a significant performance degradation when  $N$  increases (the so-called “curse of dimensionality” effect). The search for advanced sparse grid implementations, ideally immune to this effect, has thus become a very relevant research topic.

In particular we want to exploit the anisotropy of  $Q$  by assessing its variability with respect to each parameter  $y_n$  and enriching the sparse grid approximation accordingly. Two main classes of schemes can be individuated to this end: those exploiting the anisotropy of  $Q$  by using “a posteriori” strategies, i.e. at run-time, based on suitable indicators, and those using “a priori” strategies, based

### Chapter 3. Sparse Grid polynomial approximation for random PDEs in high dimensional probability spaces

---

on theoretical estimates or good ansatz and possibly aided by some preliminary computations (that is why sometimes people refer to the latter as “a-priori/a-posteriori” strategies).

A priori schemes might be somehow more tempting since, by having at our disposal computable formulas, we can build an anisotropic index-set without the need of exploring the multi-index space to assess its anisotropy structure. This basically means that they save the cost of the exploration; on the other hand a posteriori approaches require some extra exploration cost but are more flexible and robust since they do not rely on the sharpness of theoretical estimates or ansatz as in the “a priori” case. Examples of a-priori/a-posteriori algorithms can be found e.g. in [63, 11, 60], while the classical a-posteriori algorithm originally proposed in [40] has been further considered e.g. in [22, 69, 79].

A-posteriori sparse grid algorithms have always been used in the literature in combination with nested univariate quadrature rules, since this choice eases the computation of the anisotropy indicators, cf. [40]. In Uncertainty Quantification it is quite natural to choose univariate quadrature points according to the probability measures  $\rho_n(y_n)dy_n$ , see e.g. [4]: hence, one is left with the problem of computing good univariate nested quadrature rules for the probability measures at hand. While the case of the uniform measure has been thoroughly investigated and several choices of appropriate nested quadratures are available, like Leja, Gauss–Patterson or Clenshaw–Curtis points (see e.g. [61, 60] and references therein), non-uniform measures have been less explored. In the very relevant case of normal probability distribution a common choice is represented by Genz-Keister quadrature rules [39]; however, the cardinality of such quadrature rules increases very quickly when moving from one quadrature level to the following one, hence leading to heavy computational burden when tensorized in a high-dimensional setting. The very recent work [59] develops instead generalized Leja quadrature rules for arbitrary measures on unbounded intervals: the main advantage of such quadrature rules over the Genz-Keister points is that two consecutive quadrature rules differ by one point only, rendering the Leja points more suitable for sparse grid construction.

In this work we will approach the problem from a different perspective and propose a slight generalization of the classical a-posteriori adaptive algorithm that allows to use non-nested quadrature rules: this immediately permits to build adaptive sparse grids using Gaussian-type quadrature nodes, which are readily available for practically every common probability measure. We will also consider different profit indicators and compare the performances of the corresponding adaptive schemes.

We will then test our version of the adaptive algorithm on the stochastic Darcy problem [12, 35, 17, 43, 25] by using a truncated Karhunen–Loève expansion of the input log-permeability  $\gamma$ . We will consider the cases in which the covariance structure of the random field is described by a tensor and an isotropic Matérn covariance model [32], according to (2.6) and (2.5).

More specifically, we will mainly consider the case of a fairly smooth (twice differentiable) random permeability  $a$ . Starting from the value  $\nu = 2.5$  we analyze the convergence in the case of tensor and radial covariance, and with increasing values of the variance of the log-permeability  $\sigma^2$ , in order to show how the variability of the input random field affects the convergence properties of the algorithm.

In such cases we will compare the performance of the adaptive sparse grid procedure with the “a-priori” quasi-optimal sparse grid proposed in [12] for the same problem, that has been readapted

to cover also the case of non-nested nodes.

In the case  $\nu = 2.5$ , the lognormal random field can be very accurately described by including a moderate number of random variables in the Karhunen–Loève expansion. In this case a sparse grid approach to solve the Darcy problem is quite effective, at least for reasonably moderate variances of the input permeability. Note, however, that we will not fix a-priori the number of random variables to be considered, but rather propose a version of the adaptive algorithm that progressively adds dimensions to the search space, thus formally working with  $N = \infty$  random variables. In this sense, when we say that it is possible to accurately describe the random field by including a moderate number of random variables in the Karhunen–Loève expansion, we mean that the number of variables activated by the algorithm does not increase quickly when achieving smaller and smaller tolerances. However even such dimension adaptive sparse grids (as well as the quasi-optimal ones) may suffer from a deterioration in performance when the lognormal random field gets rougher. In particular, in the case  $\nu = 0.5$ , we show numerical tests indicating that even such advanced adaptive schemes might not present a substantial gain over a standard Monte Carlo method. In this case indeed the number of variables activated by the algorithm increases very quickly when considering decreasing tolerances.

The chapter is organized as follows. We start by introducing the general construction of sparse grids in Section 3.3. Then, in Section 3.4, we discuss in detail the construction of the quasi-optimal and adaptive sparse grids: in particular, we will set up a common framework for the two methods in the context of the solution of discrete optimization problems, and specify the details of the two algorithms in Subsections 3.4.1 and 3.4.2, respectively. In Section 3.5 we will introduce the proposed dimension adaptive algorithm in order to solve the Darcy problem. Then in Section 3.6 we present the numerical results, while Section 3.7 contains the conclusions of this chapter.

In what follows,  $\mathbb{N}$  will denote the set of integer numbers including 0, and  $\mathbb{N}_+$  that of integer numbers excluding 0. Given two vectors  $\mathbf{v}, \mathbf{w} \in \mathbb{N}^N$ ,  $|\mathbf{v}|_0, |\mathbf{v}|_1, |\mathbf{v}|_2$  denote respectively the number of non-zero entries of  $\mathbf{v}$ , the sum of their absolute values and the Euclidean norm of  $\mathbf{v}$ , and we write  $\mathbf{v} \leq \mathbf{w}$  if and only if  $v_j \leq w_j$  for every  $1 \leq j \leq N$ . Moreover,  $\mathbf{0}$  will denote the vector  $(0, 0, \dots, 0) \in \mathbb{N}^N$ ,  $\mathbf{1}$  the vector  $(1, 1, \dots, 1) \in \mathbb{N}^N$ , and  $\mathbf{e}_j$  the  $j$ -th canonical vector in  $\mathbb{R}^N$ , i.e. a vector whose components are all zero but the  $j$ -th, whose value is one. To close our introduction, we recall the definition of some function spaces that will be useful in the following. In particular, we will need the space of continuous functions with weighted maximum norm

$$C_\pi^0(\Gamma; V) = \left\{ f : \Gamma \rightarrow V \text{ s.t. } f \text{ is continuous and } \max_{\Gamma} \|f(\mathbf{y})\|_V \pi(\mathbf{y}) < \infty \right\},$$

where  $\pi = \prod_{n=1}^N \pi_n(y_n)$ ,  $\pi_n : \Gamma_n \rightarrow \mathbb{R}$ , is a positive and smooth function. The reasons for introducing two different weight functions  $\varrho$  and  $\pi$  will become clearer later on. Observe in particular that, since  $V$  and  $L_\varrho^2(\Gamma)$  are Hilbert spaces,  $L_\varrho^2(\Gamma; V)$  is isomorphic to the tensor product space  $V \otimes L_\varrho^2(\Gamma)$ , and is itself an Hilbert space.

### 3.2 Preliminaries: a brief overview on the problem

The problem of considering polynomial approximations of random elliptic PDEs has been widely investigated by several authors. Without any loss of generality, by denoting with  $\mathbb{P}_q(\Gamma_n)$  the set of univariate polynomials in  $y_n$  of degree at most  $q$ , the starting point is a series representation of  $Q(\mathbf{y})$

### Chapter 3. Sparse Grid polynomial approximation for random PDEs in high dimensional probability spaces

---

using multivariate polynomials:

$$Q(\mathbf{y}) = \sum_{\mathbf{q} \in \mathbb{N}^N} Q_{\mathbf{q}} P_{\mathbf{q}}(\mathbf{y}), \quad P_{\mathbf{q}}(\mathbf{y}) \in \bigotimes_{n=1}^N \mathbb{P}_{q_n}(\Gamma_n), \quad (3.1)$$

where  $P_{\mathbf{q}}(\mathbf{y})$  are general global polynomial functions that will depend on the specific kind of approximation. For instance they could be an orthonormal basis in  $L^2(\Gamma)$  upon which to perform a projection, Lagrange polynomials built over suitable collocation points to perform a sparse grid interpolation or monomials representing a Taylor expansion. All approximations of  $Q$ , namely  $Q_{\Lambda_W}(\mathbf{y})$ , are based on a suitable choice of a truncation set  $\Lambda_W$ , having cardinality  $W$ , upon which evaluating the summation (3.1).

Theoretical results on the convergence of polynomial approximation are available in the literature for some specific cases. The most studied situation is that of a diffusion coefficient depending affinely on the random parameters  $a(\mathbf{x}, \mathbf{y}) = a^*(\mathbf{x}) + \sum_{n=1}^N y_n \psi_n(\mathbf{x})$ , with  $y_n$  uniformly distributed and under the assumption of uniform ellipticity and boundedness. In [27] the authors perform a rigorous analysis in case of both a monomial expansion and a projection on Legendre polynomials. They showed that, in both cases, the approximations based on these polynomial spaces lead to algebraic convergence rates with respect to  $W$  independently of  $N$ , and then valid also when taking  $N \rightarrow \infty$ , namely

$$\|Q(\cdot) - Q_{\Lambda_W}(\cdot)\|_{L^\infty(\Gamma)} \leq C W^{-s}.$$

Here  $s$  is a parameter related to the decay of the modes  $\psi_n(\mathbf{x})$  modeling the diffusion coefficient; by assuming  $\{\|\psi_n(\cdot)\|_{L^\infty(D)}\}_{n \geq 1} \in \ell^p(\mathbb{N})$  for some  $p < 1$ , the result in [27] states that the achievable convergence rate is  $s = \frac{1}{p} - 1$ . Such a result is interesting since it directly links the  $p$ -summability of the sequence  $\{\|\psi_n(\cdot)\|_{L^\infty(D)}\}_{n \geq 1}$  with the convergence rate. This result has been extended in [24] to cover also the case in which the diffusion coefficient  $a(\mathbf{x}, \mathbf{y})$  depends nonlinearly on the random variables  $y_n$ ,  $n = 1, \dots, N$ . Unfortunately, all these results strongly rely on the assumption of uniform ellipticity of the random field  $a$ .

A straightforward extension of such results to the case of normal (unbounded) random variables is not obvious; in [35] the authors proved a convergence result for a standard Smolyak approximation [71] of the solution of the log-normal Darcy problem with truncated KL expansion; however, since the Smolyak construction does not exploit the anisotropy of the solution with respect to different random variables, it is not possible to extend this result also for  $N \rightarrow \infty$  since the predicted convergence rates would degenerate. This fact has therefore motivated many authors to consider different strategies to take into account the anisotropy of the solution in the construction of the sparse polynomial space [63, 23, 42].

We may conjecture that such a theoretical result could applies also to the Darcy problem with log-normal permeability: we can then link the regularity of the input random field with the summability of the sequence  $\{\|\psi_n(\cdot)\|_{L^\infty(D)}\}_{n \geq 1}$  in which now  $\psi_n$ ,  $n \in \mathbb{N}$  represent the KL modes; since such a sequence scales as  $n^{-\frac{d+2\nu}{2d}}$  (in the case of isotropic Matérn covariance function of the input permeability) we would obtain  $p$ -summability of the input permeability only for  $\frac{2d}{d+2\nu} < p < 1$ , which, for  $\nu$  sufficiently large, would imply a convergence rate equal to  $\frac{\nu-1}{2}$  in the two dimensional case with respect to the cardinality of the approximating space  $\Lambda_W$ , i.e.  $\|Q(\cdot) - Q_{\Lambda_W}(\cdot)\|_{L^\infty(\Gamma)} \leq C W^{-\frac{\nu-1}{2}}$ ; a good test for our adaptive scheme will be then to achieve a convergence rate as close as possible to the one predicted for an easier class of problems.

### 3.3 Sparse grid approximations in finite dimensions

As anticipated in the introduction, we aim at constructing a sparse grid approximation with global polynomials of a Hilbert space-valued function  $Q$ , defined on  $\Gamma \subseteq \mathbb{R}^N$  with associated probability measure  $\varrho(\mathbf{y})d\mathbf{y} = \prod_{n=1}^N \varrho_n(y_n)dy_n$ . More specifically, we will consider functions  $Q$  that are continuous with respect to  $\mathbf{y}$  and with finite variance, i.e. belonging to  $L^2_\varrho(\Gamma; V) \cap C_\pi^0(\Gamma; V)$  for some suitable weight function  $\pi$ : such weights can be often taken equal to  $\varrho$ , but not always as in fact we will consider in Section 3.6, see e.g. [4, 49]. Observe that approximating  $Q$  with global polynomials is a sound approach if  $Q$  is not just continuous, but actually a smooth function of  $\mathbf{y}$ , see [4, 60]. Sparse grids based on piecewise polynomial approximations, which are suitable for non-smooth or even discontinuous functions, have been developed e.g. in [52, 36].

The sparse grid polynomial approximation is built starting from a sequence of univariate Lagrangian interpolation operators  $\{\mathcal{U}_n^{m(i_n)}\}_{i_n \in \mathbb{N}}$  along each direction  $y_n$ ,  $n = 1, \dots, N$ , namely

$$\mathcal{U}_n^{m(i_n)} : C_{\pi_n}^0(\Gamma_n) \rightarrow \mathbb{P}_{m(i_n)-1}(\Gamma_n),$$

where  $m(i_n)$  denotes the number of collocation points used by the  $i_n$ -th interpolant. The function  $m : \mathbb{N} \rightarrow \mathbb{N}$  is called “level-to-nodes function” and is a strictly increasing function, with  $m(0) = 0$  and  $m(1) = 1$ ; consistently, we set  $\mathcal{U}_n^0[Q] = 0$ . Next, since the underlying hypercube  $\Gamma$  is the tensor product of each  $\Gamma_n$ , the sparse grid interpolant will be built as a linear combination of tensor products of univariate interpolants; therefore for any  $\mathbf{i} \in \mathbb{N}_+^N$  we define the tensor interpolant operator

$$\mathcal{T}_{\mathbf{i}}^{\mathbf{m}}[Q](\mathbf{y}) = \bigotimes_{n=1}^N \mathcal{U}_n^{m(i_n)}[Q](\mathbf{y}), \quad (3.2)$$

and the *hierarchical surplus* operator

$$\Delta^{\mathbf{m}(\mathbf{i})} = \bigotimes_{n=1}^N \left( \mathcal{U}_n^{m(i_n)} - \mathcal{U}_n^{m(i_n-1)} \right), \quad (3.3)$$

where we have denoted with  $\mathbf{m}(\mathbf{i})$  the vector  $[m(i_1), m(i_2), \dots, m(i_N)]$ . A sparse grid approximation is built as a sum of *hierarchical surplus* operators; more specifically, we consider for  $w \in \mathbb{N}$  a sequence of index sets  $\mathbf{I}(w) \subset \mathbb{N}_+^N$  such that  $\mathbf{I}(w) \subset \mathbf{I}(w+1)$ ,  $\mathbf{I}(0) = \{\mathbf{1}\}$  and  $\cup_{w \in \mathbb{N}} \mathbf{I}(w) = \mathbb{N}_+^N$ , and we define the sparse grid approximation of  $f(\mathbf{y})$  at level  $w \in \mathbb{N}$  as

$$\mathcal{S}_{\mathbf{I}(w)}^{\mathbf{m}} : L_\varrho^2(\Gamma; V) \cap C_\pi^0(\Gamma; V) \rightarrow L_\varrho^2(\Gamma; V), \quad \mathcal{S}_{\mathbf{I}(w)}^{\mathbf{m}}[Q](\mathbf{y}) = \sum_{\mathbf{i} \in \mathbf{I}(w)} \Delta^{\mathbf{m}(\mathbf{i})}[Q](\mathbf{y}). \quad (3.4)$$

To ensure good approximation properties for the sparse approximation, the sum (3.4) should be telescopic, cf. [40]: to this end we require that

$$\forall \mathbf{i} \in \mathbf{I}, \quad \mathbf{i} - \mathbf{e}_j \in \mathbf{I} \text{ for } 1 \leq j \leq N \text{ such that } i_j > 1.$$

A set  $\mathbf{I}$  satisfying the above property is said to be a *lower set* or a *downward closed set*, see e.g. [30]. Again we stress the fact that the choice of the set  $\mathbf{I}(w)$  strongly affects the effectiveness of such sparse grid schemes; to properly face this issue, Section 3.4 will be entirely devoted to the discussion of two possible strategies, namely the “a-posteriori” adaptive and the “a-priori” quasi-optimal procedures that have been mentioned in the introduction.

### Chapter 3. Sparse Grid polynomial approximation for random PDEs in high dimensional probability spaces

It is important to notice how, by properly rearranging the terms in the sum (3.4), the sparse grid interpolant can be rewritten as linear combination of tensor interpolant operators (3.2), see e.g. [76]. In particular, for downward closed multi-index sets  $\mathbf{I}(w)$ , we obtain

$$\mathcal{I}_{\mathbf{I}(w)}^{\mathbf{m}}[Q](\mathbf{y}) = \sum_{\mathbf{i} \in \mathbf{I}(w)} c_{\mathbf{i}} \mathcal{T}_{\mathbf{i}}^{\mathbf{m}}[Q](\mathbf{y}), \quad c_{\mathbf{i}} = \sum_{\substack{\mathbf{j} \in \{0,1\}^N \\ (\mathbf{i}+\mathbf{j}) \in \mathbf{I}(w)}} (-1)^{|\mathbf{j}|}. \quad (3.5)$$

Observe that many of the coefficients  $c_{\mathbf{i}}$  in (3.5) may be zero: in particular  $c_{\mathbf{i}}$  is zero whenever  $\mathbf{i} + \mathbf{j} \in \mathbf{I}(w) \ \forall \mathbf{j} \in \{0,1\}^N$ . The set of all collocation points needed by (3.5) is actually called a *sparse grid*, and we denote its cardinality by  $W_{\mathbf{I}(w)}^{\mathbf{m}}$ .

To make easier the dissertation in the following we introduce the operator  $\text{pts}(\mathcal{S})$  that returns the set of points associated to a tensor / sparse grid operator, and the operator  $\text{card}(\mathcal{S})$  that returns the cardinality of  $\text{pts}(\mathcal{S})$ :

$$\text{card}(\mathcal{T}_{\mathbf{i}}^{\mathbf{m}}) = \prod_{n=1}^N m(i_n), \quad \text{card}(\mathcal{I}_{\mathbf{I}(w)}^{\mathbf{m}}) = W_{\mathbf{I}(w)}^{\mathbf{m}}. \quad (3.6)$$

Finally, consider a sequence of univariate quadrature operators built over the same set of points of  $\{\mathcal{Q}_n^{m(i_n)}\}_{i_n \in \mathbb{N}}$ ; then, since it is straightforward to derive a tensor grid quadrature scheme for  $\mathcal{T}_{\mathbf{i}}^{\mathbf{m}}[Q]$ , we can use (3.5) to derive a sparse grid quadrature scheme for  $\mathcal{I}_{\mathbf{I}(w)}^{\mathbf{m}}[Q]$ ; by recombining the terms in (3.5) we get then

$$\int_{\Gamma} f(\mathbf{y}) \varrho(\mathbf{y}) d\mathbf{y} \approx \int_{\Gamma} \mathcal{I}_{\mathbf{I}(w)}^{\mathbf{m}}[Q] \varrho(\mathbf{y}) d\mathbf{y} = \sum_{\mathbf{i} \in \mathbf{I}(w)} c_{\mathbf{i}} \int_{\Gamma} \mathcal{T}_{\mathbf{i}}^{\mathbf{m}}[Q](\mathbf{y}) d\mathbf{y} = \sum_{j=1}^{W_{\mathbf{I}(w)}^{\mathbf{m}}} Q(\mathbf{y}_j) \beta_j = \mathcal{Q}_{\mathbf{I}(w)}^{\mathbf{m}}[Q], \quad (3.7)$$

for  $\mathbf{y}_j \in \text{pts}(\mathcal{I}_{\mathbf{I}(w)}^{\mathbf{m}})$  and suitable quadrature weights  $\beta_j \in \mathbb{R}$ .

Coming to the choice of the univariate collocation points used to build  $\mathcal{Q}_n^{m(i_n)}$ , as mentioned in the introduction they should be chosen according to the probability measure  $\varrho_n(y_n) dy_n$  on  $\Gamma_n$ . Although the use of nested points seems to be particularly indicated for the hierarchical construction (3.4), as the  $\Delta^{\mathbf{m}(\mathbf{i})}$  operator would entail evaluations only on the new points added going from the tensor grid  $\mathcal{T}_{\mathbf{i}-1}^{\mathbf{m}}$  to  $\mathcal{T}_{\mathbf{i}}^{\mathbf{m}}$ , at this point any choice of univariate collocation points is allowed (see Table 3.1), and in particular Gauss interpolation / quadrature points, associated to the underlying probability density functions  $\varrho_n(y_n)$ , have been widely used, cf. e.g. [5, 63, 34, 33]. Note however that non-nested interpolatory rules have not been used in the adaptive context, for reasons that will be clearer in a moment; one of our aims is then to extend the adaptive algorithm to non-nested quadrature rules.

#### 3.4 Possible criteria to select the multi-index set $\mathbf{I}$

As previously anticipated in this section we will detail two possible strategies, namely the *a priori* and the *a posteriori* ones, to properly choose the sequence of sets  $\mathbf{I}(w)$ . The QoI  $Q$  we want to approximate is defined starting from the solution  $p$  of the Darcy problem which is a Hilbert space-valued function. Since  $Q$  is a (spatial) linear functional  $Q(p)$ , in order to simplify the notation, we assume  $V = \mathbb{R}$ , i.e.  $Q$  is a real-valued  $N$ -variate function over  $\Gamma$ . However, all the considerations and algorithms extend directly to  $p$  as well.



### 3.4. Possible criteria to select the multi-index set $\mathbf{I}$

Table 3.1 – Common choices of univariate collocation points for sparse grids.

Collocation points			
	measure	nested	$\mathbf{m}(\mathbf{i})$
Gauss–Legendre	uniform	no	$i$
Clenshaw–Curtis	uniform	yes	$2^{i-1} + 1$
Gauss–Patterson	uniform	yes	$2^i - 1$
Leja	uniform	yes	$m(i) = i$ or $m(i) = 2i - 1$
Gauss–Hermite	gaussian	no	$i$
Genz–Keister	gaussian	yes	tabulated: $m(i) = 1, 3, 9, 19, 35$
generalized Leja	gaussian	yes	$i$

Our goal is to measure the sparse grid approximation error by using proper non-negative sublinear functionals<sup>1</sup>  $\mathcal{E}[\cdot]$ , e.g. a semi-norm on  $L^p_\rho(\Gamma)$  (three examples of such sublinear functionals will be provided in the following). To do this, we assume that formally the following equality holds:

$$Q = \mathcal{S}_{\mathbb{N}_+^N}^{\mathbf{m}}[Q] = \sum_{\mathbf{i} \in \mathbb{N}_+^N} \Delta^{\mathbf{m}(\mathbf{i})}[Q];$$

then, we can measure the error between the exact QoI  $Q$  and its sparse grid approximation  $\mathcal{S}_{\mathbf{I}(w)}^{\mathbf{m}}[Q]$  as

$$\mathcal{E}[Q - \mathcal{S}_{\mathbf{I}(w)}^{\mathbf{m}}[Q]] = \mathcal{E}\left[\sum_{\mathbf{i} \notin \mathbf{I}(w)} \Delta^{\mathbf{m}(\mathbf{i})}[Q]\right] \leq \sum_{\mathbf{i} \notin \mathbf{I}(w)} \mathcal{E}[\Delta^{\mathbf{m}(\mathbf{i})}[Q]]. \quad (3.8)$$

Since in general we are not able to exactly compute the contributions  $\mathcal{E}[\Delta^{\mathbf{m}(\mathbf{i})}[Q]]$ , we further define the *error contribution operator*  $\Delta E(\mathbf{i})$  as a computable (and hopefully tight) approximation of  $\mathcal{E}[\Delta^{\mathbf{m}(\mathbf{i})}[Q]]$ , namely  $\Delta E(\mathbf{i}) \approx \mathcal{E}[\Delta^{\mathbf{m}(\mathbf{i})}[Q]]$ . At the same time, we also define the *work contribution operator*  $\Delta W(\mathbf{i})$  as the number of additional evaluations of  $Q$  needed to include the hierarchical surplus operator  $\Delta^{\mathbf{m}(\mathbf{i})}[Q]$  to the sparse grid approximation. Observe that keeping track of these new evaluations is easy when using nested collocations points while it is actually a quite delicate issue when using non-nested points as we will see in the following.

Now, upon having assigned an error and a work contribution to each hierarchical surplus operator, the problem of finding optimal sets  $\mathbf{I}(w)$  can be rewritten as a “binary knapsack problem” [15, 56], namely:

$$\max \sum_{\mathbf{i} \in \mathbb{N}_+^N} \Delta E(\mathbf{i}) x_{\mathbf{i}} \quad \text{s.t.} \quad \sum_{\mathbf{i} \in \mathbb{N}_+^N} \Delta W(\mathbf{i}) x_{\mathbf{i}} \leq W_{\max}(w) \quad \text{and} \quad x_{\mathbf{i}} \in \{0, 1\},$$

where  $W_{\max}(w)$  is the maximum computational work allowed for the approximation level  $w$ . Note that we are not explicitly enforcing the resulting sets  $\mathbf{I}(w)$  to be downward closed (which will have to be verified a-posteriori).

While the binary knapsack problem is known to be computationally intractable (NP-hard), its *linear programming relaxation*, in which fractional values of  $x_{\mathbf{i}}$  are allowed, can be solved analytically by the so-called Dantzig algorithm [56]:

<sup>1</sup>A sublinear functional over a vector space  $X$  is a function  $\Theta : X \rightarrow \mathbb{R}$  such that

- $\Theta(\alpha x) = \alpha \Theta(x)$ ,  $\forall \alpha > 0$  and  $x \in X$ ;
- $\Theta(x + y) \leq \Theta(x) + \Theta(y)$ ,  $\forall x, y \in X$ .

### Chapter 3. Sparse Grid polynomial approximation for random PDEs in high dimensional probability spaces

1. Assign a “profit” to each multi-index  $\mathbf{i}$ ,

$$P(\mathbf{i}) = \frac{\Delta E(\mathbf{i})}{\Delta W(\mathbf{i})}; \quad (3.9)$$

2. sort multi-indices by decreasing profit;
3. set  $x_i = 1$ , i.e. add  $\mathbf{i}$  to  $\mathbf{I}(w)$ , until the constraint on the maximum work is fulfilled. In particular, whenever the multi-index  $\mathbf{i} = \mathbf{1} + \mathbf{e}_n$  enters the set  $\mathbf{I}(w)$  we say that the random variable  $y_n$  is *activated*.

Note that only the last multi-index included in the selection is possibly taken not entirely (i.e. with  $x_i < 1$ ), whereas all the previous ones are taken entirely (i.e. with  $x_i = 1$ ). This is due to the fact that, denoting with  $\mathbf{i}^*$  the last multi-index, we may have  $W_{\mathbf{I}(w) \setminus \mathbf{i}^*}^{\mathbf{m}} < W_{\max}(w) < W_{\mathbf{I}(w)}^{\mathbf{m}}$ . However, if this is the case, we assume that we could slightly adjust the computational budget, so that all  $x_i$  have integer values; observe that such integer solution is also the solution of the original binary knapsack problem with modified work constraint, i.e.  $W_{\max}(w) = W_{\mathbf{I}(w)}^{\mathbf{m}}$ .

Both the *a priori* quasi-optimal and the *a posteriori* adaptive sparse grids strategies fit in this general framework. What changes between the two schemes are just the choice of the error indicator  $\mathcal{E}[\cdot]$  and the way  $\Delta W(\mathbf{i})$  and  $\Delta E(\mathbf{i})$  are computed.

#### 3.4.1 An “a priori” strategy: quasi-optimal sparse grids

Here we introduce the *a priori* quasi-optimal sparse grid strategy [60, 12]; in this case the error between  $Q$  and its sparse grid interpolant  $\mathcal{S}_{\mathbf{I}(w)}^{\mathbf{m}}[Q]$  is measured in the  $L_\rho^2$ -norm, so that the bound (3.8) becomes

$$\|Q - \mathcal{S}_{\mathbf{I}(w)}^{\mathbf{m}}[Q]\|_{L_\rho^2} \leq \sum_{\mathbf{i} \notin \mathbf{I}(w)} \|\Delta^{\mathbf{m}(\mathbf{i})}[Q]\|_{L_\rho^2}. \quad (3.10)$$

In order to define the profits  $P(\mathbf{i})$  we need to provide an explicit bound of the  $L_\rho^2$ -norm of the  $\Delta^{\mathbf{m}(\mathbf{i})}$ -operator applied to  $Q$  and identify the error contribution operator  $\Delta E(\mathbf{i})$  with such a bound.

Following the approach in [60, 12, 11], we introduce the spectral expansion of  $Q$  over a  $N$ -variate  $\tilde{\rho}$ -orthonormal polynomial basis  $\varphi_{\mathbf{q}}(\mathbf{y})$ ,<sup>2</sup> with  $\tilde{\rho}$  not necessarily equal to  $\rho$ ; for example, in the case where  $y_n$  are uniform random variables,  $\rho_n(y_n) = 1/|\Gamma_n|$ , one is allowed to expand  $Q$  on tensorized Chebyshev polynomials, which are orthonormal with respect to  $\tilde{\rho} = \prod_{n=1}^N \tilde{\rho}_n$ , with  $\tilde{\rho}_n(y_n) = 1/\sqrt{1-y_n^2}$ . Next, let  $Q_{\mathbf{q}}$  be the  $\mathbf{q}$ -th coefficient of the  $\tilde{\rho}$ -expansion of  $Q$  and let the  $\mathbb{M}_n^{m(i_n)}$  be the “ $C_\pi^0 \rightarrow L_\rho^2$  Lebesgue constant” of the univariate interpolant operators  $\mathcal{W}_n^{m(i_n)}$ ,  $n = 1, \dots, N$ , for a suitable weight  $\pi$ , i.e.

$$\mathbb{M}_n^{m(i_n)} = \sup_{\|f\|_{C_\pi^0(\Gamma_n)}=1} \|\mathcal{W}_n^{m(i_n)}[f]\|_{L_\rho^2(\Gamma_n)}.$$

In order to obtain a rigorous bound of each term in (3.10) by following the procedure developed in [60], we assume that the coefficients  $Q_{\mathbf{q}}$  are at least exponentially decreasing in each  $y_n$ , i.e.  $|Q_{\mathbf{q}}| \leq C_1 \prod_n \exp(-g_n q_n)$ , and that  $\|\varphi_{q_n}\|_{C_\pi^0(\Gamma_n)} \leq C_2$ ,  $\forall n = 1, \dots, N$ . With such assumptions, by using the properties of the hierarchical surplus operator, it can be shown that for a suitable constant  $C$

<sup>2</sup>Here the  $n$ -th component of  $\mathbf{q}$  denotes the polynomial degree with respect to  $y_n$ .

there holds

$$\begin{aligned} \|\Delta^{\mathbf{m}(\mathbf{i})}[Q]\|_{L^2_\varrho} &= \left\| \Delta^{\mathbf{m}(\mathbf{i})} \left[ \sum_{\mathbf{q} \in \mathbb{N}^N} Q_{\mathbf{q}} \varphi_{\mathbf{q}} \right] \right\|_{L^2_\varrho} = \left\| \sum_{\mathbf{q} \geq \mathbf{m}(\mathbf{i}-1)} Q_{\mathbf{q}} \Delta^{\mathbf{m}(\mathbf{i})}[\varphi_{\mathbf{q}}] \right\|_{L^2_\varrho} \\ &\leq |Q_{\mathbf{m}(\mathbf{i}-1)}| \prod_{n=1}^N \mathbb{M}_n^{m(i_n)} \sum_{q_n \geq m(i_n-1)} \|\varphi_{q_n}\|_{C^0_\pi(\Gamma_n)} \leq \Delta E(\mathbf{i}) = C(N) |Q_{\mathbf{m}(\mathbf{i}-1)}| \prod_{n=1}^N \mathbb{M}_n^{m(i_n)}. \end{aligned} \quad (3.11)$$

To obtain computable formulas for  $\Delta E(\mathbf{i})$  we numerically estimate the Lebesgue constants  $\mathbb{M}_n^{m(i_n)}$  and we use a computable ansatz for the term  $|Q_{\mathbf{m}(\mathbf{i}-1)}|$ ; such an ansatz depends on the exponential coefficients  $g_1, \dots, g_N$  that can be estimated by convenient pre-computations. In practice, by denoting with  $Q_q^n$  the  $q$ -th coefficient of the one dimensional  $\tilde{\rho}_n$ -expansion of  $Q_{m(i_n)}(0, \dots, 0, y_n, 0, \dots, 0)$ , we estimate the  $g_n$  coefficients by further assuming  $|Q_{\mathbf{m}(\mathbf{i})}| < \prod_{n=1}^N |Q_{m(i_n)}^n|$  so that we require only  $\mathcal{O}(N)$  evaluations of  $Q$ , since we fix a priori the number of evaluations to compute any unidimensional coefficient  $|Q_{m(i_n)}(y_1, \dots, y_{n-1}, \cdot, y_{n+1}, \dots, y_N)|$ . In the next sections an explicit ansatz for the lognormal Darcy problem will be used, and other details on the numerical procedure needed to estimate  $g_1, \dots, g_N$  will be provided.

**Remark 3.4.1.** Concerning the Darcy problem, the solution  $p$  (and then  $Q$ ) depends analytically on the parameters  $\mathbf{y}$  [4, 35], therefore it seems a reasonable idea to use a sparse grid scheme to approximate it. In [49] it has been shown that  $p \in C^0_\pi(\Gamma, V)$  with  $\pi(\mathbf{y}) = \prod_{n=1}^N \exp(-|y_n| \sqrt{\lambda_n}) \|\psi_n\|_{L^\infty(D)}$ . Nonetheless, we will choose  $\varrho = \tilde{\rho} = \pi$  in equation (3.11). In particular, this means that we can use Hermite polynomials  $\varphi_{\mathbf{q}}$  in the quasi-optimal approach, for which indeed  $\|\varphi_{\mathbf{q}}\|_{C^0_\pi} \leq C$ , and, according to [12], we will use the following ansatz for the Hermite coefficient of  $Q$ :

$$|Q_{\mathbf{q}}| \approx C \prod_{n=1}^N \frac{e^{-g_n q_n}}{\sqrt{q_n!}}. \quad (3.12)$$

The computation of the work contributions  $\Delta W(\mathbf{i})$  is more straightforward. In particular different formulas can be applied depending on whether the nodes we are considering are nested or not (see [60] for details). In the former case, we can simply write

$$\Delta W(\mathbf{i}) = \prod_{n=1}^N (m(i_n) - m(i_n - 1)), \quad (3.13)$$

and there holds

$$W_{\mathbf{I}(w)}^{\mathbf{m}} = \sum_{\mathbf{i} \in \mathbf{I}(w)} \Delta W(\mathbf{i}),$$

i.e. the cardinality of the sparse grid is equal to the sum of the work contributions. On the other hand, when considering non-nested nodes, the number of new evaluations of  $Q$ , i.e. the number of new points added to the sparse grid, can not be precisely estimated a priori; in light of this fact we define the work contributions  $\Delta W(\mathbf{i})$  through the pessimistic estimate

$$\Delta W(\mathbf{i}) = \prod_{n=1}^N m(i_n) = \text{card}(\mathcal{T}_{\mathbf{i}}^{\mathbf{m}}[Q]), \quad (3.14)$$

### Chapter 3. Sparse Grid polynomial approximation for random PDEs in high dimensional probability spaces

---

i.e. the cardinality of the entire tensor grid associated to  $\mathbf{i}$ , which gives us the following bound

$$W_{\mathbf{I}(w)}^{\mathbf{m}} \leq \sum_{\mathbf{i} \in \mathbf{I}(w)} \Delta W(\mathbf{i}).$$

The choice of bounding the number of new points added to the sparse grid by including the surplus operator  $\Delta^{\mathbf{m}(\mathbf{i})}$  in the sparse grid construction with the number of all the points forming the tensor grid associated to  $\mathbf{i}$  can be also motivated by observing how, for non-nested nodes, the number of new points will depend in general on the set  $\mathbf{I}$  to which  $\mathbf{i}$  is added; consider for instance two sets  $\mathbf{I}, \mathbf{I}'$  downward closed and a multi-index  $\mathbf{i}$  such that also  $\mathbf{I} \cup \{\mathbf{i}\}$  and  $\mathbf{I}' \cup \{\mathbf{i}\}$  are downward closed; then it might happen that

$$\text{card}(\mathcal{S}_{\mathbf{I} \cup \{\mathbf{i}\}}^{\mathbf{m}}) \neq \text{card}(\mathcal{S}_{\mathbf{I}' \cup \{\mathbf{i}\}}^{\mathbf{m}}), \quad (3.15)$$

with some nodes belonging at the same time to  $\mathcal{S}_{\mathbf{I}}^{\mathbf{m}}$  and to the sparse grid built over  $\mathbf{I}$  but not necessarily to the one built over  $\mathbf{I}'$ , i.e.

$$\text{pts}(\mathcal{S}_{\mathbf{I}}^{\mathbf{m}}) \not\subset \text{pts}(\mathcal{S}_{\mathbf{I}' \cup \{\mathbf{i}\}}^{\mathbf{m}}). \quad (3.16)$$

With the numerical values of  $\Delta E(\mathbf{i})$  and  $\Delta W(\mathbf{i})$  at our disposal we have all the ingredients to start the actual sparse grid construction. Here we report a sketch of the algorithm used in the case of the Darcy problem:

1.  $\forall n = 0, \dots, N$  we consider Hermite polynomial approximations with increasing level  $w_1, w_2, \dots, w_K$  and look at the errors  $|Q_{w_k}^n - Q_{w_K}^n|$  for  $k = 1, \dots, K-1$ ; starting from those errors we compute the rates  $g_n, n = 1, \dots, N$ , by fitting the model (3.12).
2. We create a “multi-index universe”, i.e. a very large set of multi-indices  $\mathbf{i} \in \mathbb{N}^N$  and assign to each of them the corresponding profit  $P(\mathbf{i})$  by using the following computable formula:

$$P(\mathbf{i}) = \frac{\prod_{n=1}^N \frac{e^{-g_n m(i_n-1)}}{\sqrt{m(i_n-1)}} \mathbb{M}_n^{m(i_n)}}{\Delta W(\mathbf{i})}.$$

If we deal with KPN nodes, we select the work contribution  $\Delta W(\mathbf{i})$  according to (3.14) and we numerically estimate the Lebesgue constants  $\mathbb{M}_n^{m(i_n)} \forall n = 1, \dots, N$ , while if we deal with non-nested Hermite nodes we select  $\Delta W(\mathbf{i})$  according to (3.13) and set  $\mathbb{M}_n^{m(i_n)} = 1 \forall n = 1, \dots, N$  (see [60]).

3. We order all the profits in a decreasing order and denote  $\mathbf{i}_s$  the multi-index  $\mathbf{i}$  corresponding to the  $s$ -th highest profit; we keep the first  $S$  multi-indexes  $\mathbf{i}_1, \dots, \mathbf{i}_S$  such that a certain stopping condition is fulfilled: either on the number of sparse grid nodes associated to the multi-index set formed by  $\mathbf{I} = \{\mathbf{i}_s\}_{s=1}^S$ , i.e.  $W_{\mathbf{I}} \leq W_{\max}$ , or on the profits, i.e. by choosing  $S$  such that  $P(\mathbf{i}_S) \geq P_{\min}$  and  $P(\mathbf{i}_{S+1}) < P_{\min}$ .
4. We check the admissibility of the set  $\mathbf{I} = \{\mathbf{i}_s\}_{s=1}^S$ ; if it holds we move to the next step, otherwise we complete the multi-index set by adding the minimum number of indexes to make it downward closed.
5. Once the multi-index set  $\mathbf{I}$  is given the sparse grid approximation can be computed according to (3.5).

### 3.4.2 An “a posteriori” strategy: adaptive sparse grids

In this section we present the adaptive sparse grid construction algorithm [40, 22, 69], and its extension to non-nested points and unbounded intervals. In this case the exploration of the multi-index space  $\mathbb{N}^N$  becomes quite costly since the estimation of the  $\Delta^{\mathbf{m}(\mathbf{i})}$  contributions will require the actual solution of the PDE in the new candidate points belonging to  $\mathcal{T}_{\mathbf{i}}^{\mathbf{m}}$ . Starting from a current multi-index space  $\mathbf{I}$ , in order to deal with this adaptive exploration of  $\mathbb{N}^N$ , we introduce the concepts of *margin* and *reduced margin* of  $\mathbf{I}$ , and the concept of *neighbors* of  $\mathbf{I}$ .

**Definition 3.4.1.** *The margin of  $\mathbf{I}$ , denoted by  $\mathbf{M}_{\mathbf{I}}$ , is the multi-index set that contains all  $\mathbf{i} \in \mathbb{N}^N$  that can be reached within “one-step forward” from  $\mathbf{I}$ , i.e.*

$$\mathbf{M}_{\mathbf{I}} = \{\mathbf{i} \in \mathbb{N}_+^N \setminus \mathbf{I} : \exists \mathbf{j} \in \mathbf{I} : |\mathbf{i} - \mathbf{j}|_1 = 1\}.$$

**Definition 3.4.2.** *The reduced margin of  $\mathbf{I}$ , denoted by  $\mathbf{R}_{\mathbf{I}}$ , is the subset of the margin of  $\mathbf{I}$  containing only those indices  $\mathbf{i}$  such that “one-step backward” in any direction takes into  $\mathbf{I}$ , i.e.*

$$\mathbf{R}_{\mathbf{I}} = \{\mathbf{i} \in \mathbb{N}_+^N \setminus \mathbf{I} : \mathbf{i} - \mathbf{e}_j \in \mathbf{I}, \forall j = 1, \dots, N : i_j > 1\} \subset \mathbf{M}_{\mathbf{I}}.$$

Observe that the reduced margin of  $\mathbf{I}$  contains all indices  $\mathbf{i}$  such that  $\mathbf{I} \cup \{\mathbf{i}\}$  is downward closed, provided that  $\mathbf{I}$  itself is downward closed; we will see soon how this property is appealing for computational purposes.

**Definition 3.4.3.** *For any index  $\mathbf{i}$  on the boundary of  $\mathbf{I}$ , i.e. such that  $\mathbf{i} + \mathbf{e}_n \notin \mathbf{I}$  for some  $n = 1, \dots, N$ , we call neighbors of  $\mathbf{i}$  with respect to  $\mathbf{I}$ , and we denote them as  $\text{neigh}(\mathbf{i}, \mathbf{I})$ , the indices  $\mathbf{j}$  not included in  $\mathbf{I}$  that can be reached with “one step forward” from  $\mathbf{i}$ , so that  $\mathbf{M}_{\mathbf{I}} = \bigcup_{\mathbf{i} \in \mathbf{I}} \text{neigh}(\mathbf{i}, \mathbf{I})$ .*

Note that in this case, instead of computing profits and sets  $\mathbf{I}(w)$  all at once as in the quasi-optimal algorithm, the adaptive algorithm computes the profits and the sets  $\mathbf{I}(w)$  at run-time, proceeding iteratively in a greedy way. In order to compute the  $\Delta^{\mathbf{m}(\mathbf{i})}$  contribution  $\forall \mathbf{i} \in \mathbf{I}$  using (3.3), we need  $\mathbf{I}$  to be downward closed; moreover, since at each iteration of our adaptive scheme, we want to compute profits associated to multi-indices  $\mathbf{i}$  which are candidates to enter the sparse grid approximation, in our greedy strategy we will only look for candidate indices  $\mathbf{i}$  in the reduced margin  $\mathbf{R}_{\mathbf{I}}$ , and we select those with highest associated profit. In general, given a multi-index set  $\mathbf{I}$  and its reduced margin  $\mathbf{R}_{\mathbf{I}}$ , the adaptive algorithm operates as follows:

1. the profits of  $\mathbf{i} \in \mathbf{R}_{\mathbf{I}}$  are computed;
2. the index  $\mathbf{i}$  with the highest profit is moved from  $\mathbf{R}_{\mathbf{I}}$  to  $\mathbf{I}$ ;
3. the reduced margin is updated and the algorithm moves to the next iteration, until some stopping criterion is met (usually, a check on the number of evaluations of  $Q$  or on the values of the profits or error contributions of the multi-indices in  $\mathbf{R}_{\mathbf{I}}$ ).

**Remark 3.4.2.** *The profits of the indices  $\mathbf{i} \in \mathbf{R}_{\mathbf{I}}$  are computed by actually adding the hierarchical surpluses to the sparse grid operator, hence the definition of “a-posteriori”; therefore, the outcome of the algorithm at each iteration is the sparse grid approximation built on  $\mathbf{I} \cup \mathbf{R}_{\mathbf{I}}$  and not just on  $\mathbf{I}$ .*

The *a posteriori* adaptive strategy is rather flexible in the choice of possible error indicators  $\mathcal{E}[\cdot]$ ; in this thesis we focus on two of them, namely the absolute value of the expectation of  $Q - \mathcal{S}_{\mathbf{I}(w)}^{\mathbf{m}}[Q]$ ,

### Chapter 3. Sparse Grid polynomial approximation for random PDEs in high dimensional probability spaces

---

which is a semi-norm on  $L^1_\rho(\Gamma)$ , and the weighted  $C^0_\pi(\Gamma)$  norm, so that (3.8) becomes

$$\begin{aligned} \left| \mathbb{E}[Q] - \mathbb{E}[\mathcal{S}_I^m[Q]] \right| &\leq \sum_{\mathbf{i} \notin \mathbf{I}} \left| \mathbb{E}[\Delta^{m(\mathbf{i})}[Q]] \right|, \\ \|Q - \mathcal{S}_I^m[Q]\|_{C^0_\pi(\Gamma)} &\leq \sum_{\mathbf{i} \notin \mathbf{I}} \|\Delta^{m(\mathbf{i})}[Q]\|_{C^0_\pi(\Gamma)}. \end{aligned}$$

Now we discuss our approach to actually compute the error indicator  $\Delta E(\mathbf{i})$  for the quantity  $\mathcal{E}[\Delta^{m(\mathbf{i})}[Q]]$ . Given a current set  $\mathbf{I}$  downward closed we explore all the multi-indices  $\mathbf{i} \in \mathbf{R}_I$ : it is then preferable to introduce the set  $\mathbf{J} = \mathbf{I} \cup \{\mathbf{i}\}$ . Again we stress the fact that also  $\mathbf{J}$  is downward closed, making easy the estimation of our error indicator according to (3.5). Moreover it can be observed that  $\Delta^{m(\mathbf{i})}[Q] = \mathcal{S}_J^m[Q] - \mathcal{S}_I^m[Q]$ .

By operating in this way, for the  $L^1_\rho(\Gamma)$  seminorm we immediately obtain

$$\mathbb{E}[\Delta^{m(\mathbf{i})}[Q]] = \mathbb{E}[\mathcal{S}_J^m[Q] - \mathcal{S}_I^m[Q]] = \mathcal{Q}_J^m[Q] - \mathcal{Q}_I^m[Q],$$

which leads to the straightforward definition

$$\Delta E(\mathbf{i}) = |\mathbb{E}[\Delta^{m(\mathbf{i})}[Q]]| = |\mathcal{Q}_J^m[Q] - \mathcal{Q}_I^m[Q]|. \quad (3.17)$$

In the case of the  $C^0_\pi(\Gamma)$  norm, things are slightly more involved; in particular our definition of the error indicator  $\Delta E(\mathbf{i})$  depends on the nature of the nodes we are considering. It is well known that, for nested points, the sparse grid operator is interpolatory (see e.g [8, Prop. 6]); this means that  $\Delta^{m(\mathbf{i})}[Q](\mathbf{y}) = \mathcal{S}_J^m[Q](\mathbf{y}) - \mathcal{S}_I^m[Q](\mathbf{y}) = 0$  for any  $\mathbf{y} \in \text{pts}(\mathcal{S}_I^m)$  while  $\mathcal{S}_J^m[Q](\mathbf{y}) = Q(\mathbf{y}) \neq \mathcal{S}_I^m[Q](\mathbf{y})$  for any  $\mathbf{y} \in \mathcal{N}_{\text{ew}} = \text{pts}(\mathcal{S}_J^m) \setminus \text{pts}(\mathcal{S}_I^m)$ , cf. eq. (3.6). Therefore the estimation of the  $C^0_\pi(\Gamma)$  norm of  $\Delta^{m(\mathbf{i})}[Q]$  can be performed by looking only at the values on  $\mathbf{y} \in \mathcal{N}_{\text{ew}}$ , namely,

$$\|\Delta^{m(\mathbf{i})}[Q]\|_{C^0_\pi(\Gamma)} \approx \max_{\mathbf{y} \in \mathcal{N}_{\text{ew}}} |\Delta^{m(\mathbf{i})}[Q](\mathbf{y})\pi(\mathbf{y})|$$

which leads to the following definition of the error indicator  $\Delta E(\mathbf{i})$

$$\Delta E(\mathbf{i}) = \max_{\mathbf{y} \in \mathcal{N}_{\text{ew}}} |\Delta^{m(\mathbf{i})}[Q](\mathbf{y})\pi(\mathbf{y})| = \max_{\mathbf{y} \in \mathcal{N}_{\text{ew}}} |(Q(\mathbf{y}) - \mathcal{S}_I^m[Q](\mathbf{y}))\pi(\mathbf{y})|. \quad (3.18)$$

On the other hand, sparse grids built with non-nested points are not interpolatory, and the set of points added to a sparse grid by  $\Delta^{m(\mathbf{i})}$  depends on  $\mathbf{i}$  and  $\mathbf{I}$ : it can occur that points considered in the sparse grid are removed from the approximations when others are added, cf. eq. (3.15), as it depends on the current index set  $\mathbf{I}$  to which  $\mathbf{i}$  is added. Thus, we define  $\mathcal{N}_{\text{ew}} = \text{pts}(\mathcal{S}_I^m)$  and approximate the  $C^0_\pi(\Gamma)$  norm as

$$\Delta E(\mathbf{i}) = \max_{\mathbf{y} \in \mathcal{N}_{\text{ew}}} |\Delta^{m(\mathbf{i})}[Q](\mathbf{y})\pi(\mathbf{y})| = \max_{\mathbf{y} \in \mathcal{N}_{\text{ew}}} |(\mathcal{S}_J^m[Q](\mathbf{y}) - \mathcal{S}_I^m[Q](\mathbf{y}))\pi(\mathbf{y})|. \quad (3.19)$$

The right hand side in (3.17)-(3.19), with suitably chosen  $\mathbf{I}$  and  $\mathbf{J}$ , are the actual formulas we have used to compute  $\Delta E(\mathbf{i})$  (see Algorithm 1 and 2 for further details). However, we remark that these values do not depend on the set  $\mathbf{I}$  chosen for evaluation, since the quantities  $\Delta E(\mathbf{i})$  are defined starting from the hierarchical surplus operator  $\Delta^{m(\mathbf{i})}$ . This means that we can consider the indices of the reduced margin  $\mathbf{R}_I$  in any order, and that the values of  $\Delta E(\mathbf{i})$  do not need to be recomputed at

each iteration.

As for the work contribution  $\Delta W(\mathbf{i})$ , we consider the same indicators defined in the quasi-optimal case, i.e. (3.13) for nested points and (3.14) for non-nested points, which is equivalent to setting the work contributions equal to the cardinality of the sets  $\mathcal{N}_{\text{ew}}$  introduced above. A third option is to consider  $\Delta W(\mathbf{i}) = 1$ , i.e. driving the adaptivity only by the error contributions. This is the choice considered e.g. in [22, 69], while [40, 53] combine  $\Delta E(\mathbf{i})$  and  $\Delta W(\mathbf{i})$  in a different way. To summarize, we will drive the adaptive algorithm with any of the four profit definitions listed next, whose formulas differ depending on whether nested or non-nested points are used: To summarize, we will adapt the ideas coming from these works to numerically solve the log-normal Darcy problem and we will extend the methodology to cover also the case of non-nested nodes; we will drive the adaptive algorithm by proper selection  $P(\mathbf{i}) = \frac{\Delta E(\mathbf{i})}{\Delta W(\mathbf{i})}$  with any of the four profit definitions listed next, whose formulas differ depending on whether nested or non-nested points are used:

- **“deltaint”**: set  $\Delta E(\mathbf{i})$  as in (3.17) and  $\Delta W(\mathbf{i}) = 1$ ;
- **“deltaint / new points”** combine  $\Delta E(\mathbf{i})$  as in (3.17) with  $\Delta W(\mathbf{i})$  in (3.13) for nested points and in (3.14) for non-nested points;
- **“weighted Linf”** set  $\Delta E(\mathbf{i})$  as in (3.18) and  $\Delta W(\mathbf{i}) = 1$  for nested points, and  $\Delta E(\mathbf{i})$  as in (3.19) and  $\Delta W(\mathbf{i}) = 1$  for non-nested points;
- **“weighted Linf / new points”** combine  $\Delta E(\mathbf{i})$  in (3.18) with  $\Delta W(\mathbf{i})$  in (3.13) for nested points and  $\Delta E(\mathbf{i})$  in (3.19) with  $\Delta W(\mathbf{i})$  in (3.14) for non-nested points.

The pseudo-code of the algorithm is listed in Algorithm 1. Since nodes that are present in a given sparse grid are not necessarily present in the following ones when using non-nested points, cf. eq. (3.16), the full work count in this case is not simply  $\text{pts}(\mathcal{S})$  (as it would be for nested points), but should rather include all the points “visited” to reach that grid in the adaptive algorithm, which motivates lines L1-L2 in Algorithm 1. Observe however that all Gaussian quadrature rules associated to a symmetric weight (or probability density) are in a sense “partially nested”, meaning that rules with odd number of points place a quadrature node in the midpoint of the interval, implying that a non-negligible number of points can still be in common between two grids.



**Algorithm 1:** Adaptive sparse grids algorithm.

**Adaptive sparse grids** ( $MaxPts, ProfTol, \pi, <ProfitName>$ )

```

I = {1}, G = {1}, RI = ∅, i = 1 ;
 $\mathcal{S}_{old} = \mathcal{S}_1^m[Q]$ ,  $\mathcal{Q}_{old} = \mathcal{Q}_1^m[Q]$  ;
 $\mathcal{H} = \text{pts}(\mathcal{S}_{old})$ ,  $NbPts = \text{card}(\mathcal{S}_{old})$ ,  $ProfStop = \infty$  ;
while  $NbPts < MaxPts$  and  $ProfStop > ProfTol$  do
     $\mathcal{N}g = \text{neigh}(\mathbf{i}, \mathbf{I})$ 
    for  $\mathbf{j} \in \mathcal{N}g$  and  $\mathbf{I} \cup \{\mathbf{j}\}$  downward closed do
        G = G ∪ {j} ;
         $\mathcal{S} = \mathcal{S}_G^m[Q]$  ;
         $\mathcal{Q} = \mathcal{Q}_G^m[Q]$  ;
        if using nested points then
             $\mathcal{N}ew = \text{pts}(\mathcal{S}) \setminus \text{pts}(\mathcal{S}_{old})$  ;
             $NbPts = NbPts + \text{card}(\mathcal{N}ew)$  ;
            v = evaluations of  $Q$  on each  $\mathbf{y} \in \mathcal{N}ew$  ;
            else
                 $\mathcal{N}ew = \text{pts}(\mathcal{T}_1^m)$  ;
                 $\mathcal{H} = \mathcal{H} \cup \text{pts}(\mathcal{S})$  ;
                 $NbPts = \text{card}(\mathcal{H})$  ;
                v = evaluations of  $\mathcal{S}$  on each  $\mathbf{y} \in \mathcal{N}ew$  ;
             $\mathbf{v}_{old} = \text{evaluations of } \mathcal{S}_{old} \text{ on each } \mathbf{y} \in \mathcal{N}ew$  ;
             $\boldsymbol{\pi} = \text{evaluations of } \pi \text{ on each } \mathbf{y} \in \mathcal{N}ew$  ;
             $P(\mathbf{j}) = \text{Compute\_profit}(\mathcal{N}ew, \mathbf{v}, \mathbf{v}_{old}, \boldsymbol{\pi}, \mathcal{Q}, \mathcal{Q}_{old}, <ProfitName>)$ 
            RI = RI ∪ {j} ;
             $\mathcal{S}_{old} = \mathcal{S}$ ,  $\mathcal{Q}_{old} = \mathcal{Q}$  ;
        choose the i from RI with highest profit;
        I = I ∪ {i}, RI = RI \ {i}
        update  $ProfStop$  with a suitable criterion based on the values of  $P$ 
    return  $\mathcal{S}, \mathcal{Q}$ 

```

*at the end of the for loop,  $\mathbf{G} = \mathbf{I} \cup \mathbf{R}_I$   
**j** must be added to  $\mathcal{S}$  to evaluate its profit.  
*i.e. the points added by **j** to  $\mathcal{S}$*   
*add points of  $\mathcal{S}$  to  $\mathcal{H}$  (no repetitions)*  
*for non-nested points,  $\text{card}(\mathcal{H}) > \text{card}(\mathcal{S})$*   
*cf. eq. (3.18)*  
*cf. eq. (3.19)**

**Compute\_profit** ( $\mathcal{N}ew, \mathbf{v}, \mathbf{v}_{old}, \boldsymbol{\pi}, \mathcal{Q}, \mathcal{Q}_{old}, <ProfitName>$ )

```

switch  $ProfitName$  do
    case deltaint
        profit(i) =  $|\mathcal{Q} - \mathcal{Q}_{old}|$  ;
    case deltaint/new points
        profit(i) =  $\frac{|\mathcal{Q} - \mathcal{Q}_{old}|}{\text{card}(\mathcal{N}ew)}$  ;
    case Weighted Linf
        profit(i) =  $\max\{|\mathbf{v} - \mathbf{v}_{old}| \odot \boldsymbol{\pi}\}$  ;
    case Weighted Linf/new points
        profit(i) =  $\frac{\max\{|\mathbf{v} - \mathbf{v}_{old}| \odot \boldsymbol{\pi}\}}{\text{card}(\mathcal{N}ew)}$  ;
return profit(i)

```

*⊙ denotes element-wise multiplication*

### 3.5 A dimension-adaptive sparse grid algorithm

When dealing with high (infinite) dimensional spaces in probability, say hundreds of random variables, the computations become quickly very costly, since the generation and the exploration of the reduced margin  $\mathbf{R}_I$  becomes more and more demanding. In order to better address such



cases we propose an improved version of the classical sparse grid algorithm listed in Algorithm 1. In particular, the idea is to start with a moderate number  $\tilde{N}$  of random variables and then introduce a way to gradually add random variables on the fly.

To do this we assume a sort of “weak ordering” of the random variables  $y_n$ ; more precisely we assume that there exist  $N_b \in \mathbb{N}$  such that the variable  $y_n$  is guaranteed to affect the QoI  $Q$  more than  $y_{n+N_b}$ .

At this stage the sparse grid algorithm can be launched by considering  $\tilde{N} = N_b$  random variables only, i.e.  $\mathbf{y} = (y_1, \dots, y_{\tilde{N}})$ ; then, whenever one of these random variables is activated, we increase by 1 the counter  $N$  of random variables and we add the variable  $y_{\tilde{N}+1}$  to the sparse grid approximation, i.e. the multi-index  $\mathbf{i} = \mathbf{1} + \mathbf{e}_{\tilde{N}+1}$  is included in the reduced margin  $\mathbf{R}_1$ . In this way the integer  $N_b$  plays the role of a “dimensional buffer”, as it is needed to maintain unchanged the number of random variables ready to be activated to improve the sparse grid approximation.

Since we are not anymore requiring an a priori truncation of the input random field, It is important to notice that the algorithm is potentially able to work also with  $N = \infty$  random variables.

For the moment we have not been able to prove a convergence result for this algorithm, since dealing with not uniformly coercive PDEs makes the analysis involved; if we look among other adaptive schemes able to work with infinite random variables, in [26] the authors proved a convergence result for an algorithm based on Taylor expansions of the solution, but covering only the case of bounded random variables. To test the effectiveness of our strategy fully detailed in Algorithm 2, we will compare our results with the error bound given in [26], Eq. (7.46), which in our case would predict a convergence rate, in the  $C_\pi^0(\Gamma)$  norm, equal to  $\nu - 1$ , with respect to the number of terms in the Taylor approximation. Of course this represents a severe test since we are dealing with log-normal fields, but as we will see in the numerical results, the observed convergence rate is not far from this value  $\nu - 1$  we are comparing with.

### 3.6 Numerical Results

The methodology presented so far is very general and can be applied to a wide class of UQ problem characterized by different stochasticity. In this section we apply the methodology to the Darcy problem with log-normal permeability and study numerically its convergence properties. In particular we will focus on the case of an undisturbed flow from left to right modeled through the equations

$$\begin{cases} -\operatorname{div}(a(\mathbf{x}, \mathbf{y}) \nabla p(\mathbf{x}, \mathbf{y})) = 0 & \mathbf{x} \in (0, 1)^2, \\ p(\mathbf{x}, \mathbf{y}) = 1 - x_1 & \mathbf{x} \in x_1 = \{0\} \cup x_1 = \{1\}, \\ \nabla p(\mathbf{x}, \mathbf{y}) \cdot \mathbf{n} = 0 & \mathbf{x} \in x_2 = \{0\} \cup x_2 = \{1\}. \end{cases} \quad (3.20)$$

The spatial approximation of the Darcy problem is done by piecewise linear finite elements defined on a structured mesh with  $33 \times 33$  vertices, which has been verified to be sufficiently refined for our purposes. The input log-permeability  $\gamma$  has been parametrized with a KL expansion  $\gamma(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^{\infty} \sqrt{\lambda_n} \psi_n(\mathbf{x}) y_n$ .

### Chapter 3. Sparse Grid polynomial approximation for random PDEs in high dimensional probability spaces

---

#### Algorithm 2: Dimension Adaptive Algorithm

---

**Note:** To avoid ambiguities we write  $\mathbf{v}^N$  to make clear that the vector  $\mathbf{v}$  has  $N$  components; analogously  $\mathbf{I}^N$  indicates that the multi-index set  $\mathbf{I}$  is composed of  $N$ -dimensional vectors.

**Dimension adaptive sparse grids** ( $MaxPts$ ,  $ProfTol$ ,  $\pi$ ,  $\langle ProfitName \rangle$ ,  $N_b$ )

```

 $\tilde{N} = N_b$ ,  $\mathbf{A}^{\tilde{N}} = \mathbf{0}^{\tilde{N}}$ ;  $\mathbf{A}$  is a Boolean vector indicating which variables are active
 $\mathbf{I}^{\tilde{N}} = \{\mathbf{1}^{\tilde{N}}\}$ ,  $\mathbf{G}^{\tilde{N}} = \{\mathbf{1}^{\tilde{N}}\}$ ,  $\mathbf{R}_1^{\tilde{N}} = \emptyset$ ,  $\mathbf{i}^{\tilde{N}} = \mathbf{1}^{\tilde{N}}$ ,  $\mathcal{S}_{old} = \mathcal{S}_{\mathbf{I}^{\tilde{N}}}^m[Q]$ ,  $\mathcal{Q}_{old} = \mathcal{Q}_{\mathbf{I}^{\tilde{N}}}^m[Q]$ ;
 $\mathcal{H} = \text{pts}(\mathcal{S}_{old})$ ,  $NbPts = \text{card}(\mathcal{S}_{old})$ ,  $ProfStop = \infty$ ;
while  $NbPts < MaxPts$  and  $ProfStop > ProfTol$  do
     $\mathcal{N}g = \text{neigh}(\mathbf{i}^{\tilde{N}}, \mathbf{I}^{\tilde{N}})$ ;
    for  $\mathbf{j} \in \mathcal{N}g$  and  $\mathbf{I}^{\tilde{N}} \cup \{\mathbf{j}\}$  is downward closed do
         $\mathbf{G}^{\tilde{N}} = \mathbf{G}^{\tilde{N}} \cup \{\mathbf{j}\}$ ;
         $\mathcal{S} = \mathcal{S}_{\mathbf{G}^{\tilde{N}}}^m[Q]$ ;
         $\mathcal{Q} = \mathcal{Q}_{\mathbf{G}^{\tilde{N}}}^m[Q]$ ;
        if using nested points then
             $\mathcal{N}ew = \text{pts}(\mathcal{S}) \setminus \text{pts}(\mathcal{S}_{old})$ ,  $NbPts = NbPts + \text{card}(\mathcal{N}ew)$ ;
             $\mathbf{v} = \text{evaluations of } f \text{ on each } \mathbf{y} \in \mathcal{N}ew$ ; cf. eq. (3.18)
        else
             $\mathcal{N}ew = \text{pts}(\mathcal{T}_1^m)$ ,  $\mathcal{H} = \mathcal{H} \cup \text{pts}(\mathcal{S})$ ,  $NbPts = \text{card}(\mathcal{H})$ ;
             $\mathbf{v} = \text{evaluations of } \mathcal{S} \text{ on each } \mathbf{y} \in \mathcal{N}ew$ ; cf. eq. (3.19)
         $\mathbf{v}_{old} = \text{evaluations of } \mathcal{S}_{old} \text{ on each } \mathbf{y} \in \mathcal{N}ew$ ;
         $\boldsymbol{\pi} = \text{evaluations of } \pi \text{ on each } \mathbf{y} \in \mathcal{N}ew$ ;
         $P(\mathbf{j}) = \text{Compute\_profit}(\mathcal{N}ew, \mathbf{v}, \mathbf{v}_{old}, \boldsymbol{\pi}, \mathcal{Q}, \mathcal{Q}_{old}, \langle ProfitName \rangle)$ 
         $\mathbf{R}_1^{\tilde{N}} = \mathbf{R}_1^{\tilde{N}} \cup \{\mathbf{j}\}$ 
         $\mathcal{S}_{old} = \mathcal{S}$ ,  $\mathcal{Q}_{old} = \mathcal{Q}$ ;
    choose  $\mathbf{k}^{\tilde{N}}$  from  $\mathbf{R}_1^{\tilde{N}}$  with highest profit;  $\mathbf{i}^{\tilde{N}} = \mathbf{k}^{\tilde{N}}$ ;
    if  $\exists n = 1, \dots, \tilde{N}$  s.t.  $A_n = 0$  and  $k_n > 1$  then
         $A_n = 1$ ,  $\tilde{N} = \tilde{N} + 1$ ; activate  $n$ -th variable and update  $\tilde{N}$ 
        extend the containers  $\mathbf{I}, \mathbf{R}_1, \mathbf{G}, \mathbf{k}, \mathbf{A}$  by adding the new direction.
         $\mathbf{G}^{\tilde{N}} = \mathbf{G}^{\tilde{N}} \cup \{\mathbf{1}^{\tilde{N}} + \mathbf{e}_N^{\tilde{N}}\}$ ;  $\mathcal{S} = \mathcal{S}_{\mathbf{G}^{\tilde{N}}}^m[Q]$ ;  $\mathcal{Q} = \mathcal{Q}_{\mathbf{G}^{\tilde{N}}}^m[Q]$ ;
        if using nested points then
             $\mathcal{N}ew = \text{pts}(\mathcal{S}) \setminus \text{pts}(\mathcal{S}_{old})$ ,  $NbPts = NbPts + \text{card}(\mathcal{N}ew)$ ;
             $\mathbf{v} = \text{evaluations of } f \text{ on each } \mathbf{y} \in \mathcal{N}ew$ ; cf. eq. (3.18)
        else
             $\mathcal{N}ew = \text{pts}(\mathcal{T}_1^m)$ ,  $\mathcal{H} = \mathcal{H} \cup \text{pts}(\mathcal{S})$ ,  $NbPts = \text{card}(\mathcal{H})$ ;
             $\mathbf{v} = \text{evaluations of } \mathcal{S} \text{ on each } \mathbf{y} \in \mathcal{N}ew$ ; cf. eq. (3.19)
         $\mathbf{v}_{old} = \text{evaluations of } \mathcal{S}_{old} \text{ on each } \mathbf{y} \in \mathcal{N}ew$ ;
         $\boldsymbol{\pi} = \text{evaluations of } \pi \text{ on each } \mathbf{y} \in \mathcal{N}ew$ ;
         $P(\mathbf{1}^{\tilde{N}} + \mathbf{e}_N^{\tilde{N}}) = \text{Compute\_profit}(\mathcal{N}ew, \mathbf{v}, \mathbf{v}_{old}, \boldsymbol{\pi}, \mathcal{Q}, \mathcal{Q}_{old}, \langle ProfitName \rangle)$ 
         $\mathbf{R}_1^{\tilde{N}} = \mathbf{R}_1^{\tilde{N}} \cup \{\mathbf{1}^{\tilde{N}} + \mathbf{e}_N^{\tilde{N}}\}$ ,
         $\mathbf{i}^{\tilde{N}} = \text{argmax}(\max(P(\mathbf{1}^{\tilde{N}} + \mathbf{e}_N^{\tilde{N}}), P(\mathbf{k}^{\tilde{N}})))$ ; select  $\mathbf{i}^{\tilde{N}}$  with highest profit
         $\mathbf{I}^{\tilde{N}} = \mathbf{I}^{\tilde{N}} \cup \{\mathbf{i}^{\tilde{N}}\}$ ,  $\mathbf{R}_1^{\tilde{N}} = \mathbf{R}_1^{\tilde{N}} \setminus \{\mathbf{i}^{\tilde{N}}\}$ 
        update  $ProfStop$  with a suitable criterion based on the values of  $P(\mathbf{j})$ 
    return  $\mathcal{S}, \mathcal{Q}$ 

```

---

In particular we will be interested in approximating the expected value of the functional

$$Q(\mathbf{y}) = \int_0^1 a(1, x_2, \mathbf{y}) \frac{\partial p}{\partial x_1}(1, x_2, \mathbf{y}) dx_2, \quad \mathbf{y} \in \mathbb{R}^N \quad (3.21)$$

which represents the mass flow on the outlet. The aims of this section are:

1. establish whether using non-nested points in an adaptive sparse grid framework might be convenient or not;
2. verify the performances of adaptive sparse grids built with respect to different profit indicators;
3. compare such performances with the ones of the quasi-optimal sparse grids, also in light of the fact that such approaches behave similarly when applied to UQ problems depending on uniform random variables by using nested collocation points (see [11, 60]);
4. test the performances of the method with different scenarios in order to study its robustness: different covariance structure, variances of the input random field, in presence of permeability measurements.

In order to achieve such goals, we will solve then the problem (3.20) in three cases:

1. first a sort of reference case: we consider a smooth random field  $\gamma$  with relatively moderate variability  $\sigma$ , corresponding to the choice  $\nu = 2.5$  in (2.6) and (2.5). In this part we carefully detail the numerical procedure used to obtain the convergence results and provide further details on the sparse grid algorithm.
2. Then we consider higher variances, moving from  $\sigma^2 = 1$  to  $\sigma^2 = 4$ ; observe that, in terms of permeability, such values of  $\sigma$  generate variances roughly speaking equal to 5 and  $3 \times 10^3$  respectively, covering then a wide range of permeability values.
3. Finally we apply the sparse grid approximation in the case in which some measurements of the permeability are available all over the domain in order to understand how significant is the effect of conditioning the random field to available observations, on the convergence of the scheme.

#### 3.6.1 Reference case: $\nu = 2.5$ , $\sigma = 1$ , $L_c = 0.5$

In this case we deal with an input random field with twice differentiable realizations with two possible covariance structure given by (2.5) and (2.6); in both cases the eigenvalues of the Karhunen–Loève expansion decay quickly enough to justify the use of the  $N$ -adaptive sparse grid algorithm to approximate the QoI, more precisely they behave like  $n^{-(1+2\nu)}$  and  $n^{-(1+2\nu/d)}$  in the tensor and radial case respectively.

For this test we consider two families of covariance functions defined in (2.5) and (2.6); in both cases we consider as a reference solution the approximation of the QoI obtained with a quasi-optimal sparse grid with approximately 21900 and 30300 quadrature points base of Gauss–Hermite abscissas, for which, 322 (out of the first 350) for the radial case and 84 (out of 152) for the tensor case random variables of the KL expansion are active; as we will see later on, this is sufficient to take into account a fraction of the total variability of the permeability field large enough not to affect our convergence results; this essentially means that no KL truncation error occurs. We monitor the convergence of the error measured as

$$err(w) \approx |\mathcal{Q}_{\mathbf{I}(w)}^{\mathbf{m}}[Q] - \mathcal{Q}_{\mathbf{I}(w_{ref})}^{\mathbf{m}}[Q]|, \quad (3.22)$$

### Chapter 3. Sparse Grid polynomial approximation for random PDEs in high dimensional probability spaces

---

i.e. the absolute value of the sparse grid quadrature error, where  $\mathbf{I}(w)$ ,  $w = 0, 1, 2, \dots$ , are the sequences of multi-index sets generated either by the adaptive or the quasi-optimal sparse grid scheme and  $\mathbf{I}(w_{ref})$  is the multi-index set corresponding to the above-mentioned reference solution. More specifically, the sets  $\mathbf{I}(w)$  for the adaptive strategies are obtained by stopping the algorithm as soon as at least  $W_{max}(w)$  points have been added to the sparse grid (including the points needed for the exploration of the reduced margin), with  $W_{max}(w) = \{1, 20, 50, 100, 250, 500, 1000, 2000, 4000\}$ , for  $w = 0, \dots, 8$ . As for the quasi-optimal sparse grids, the sets  $\mathbf{I}(w)$  are defined as

$$\mathbf{I}(w) = \{\mathbf{i} \in \mathbb{N}_+^N : P(\mathbf{i}) \geq e^{-w}\} \quad (3.23)$$

with  $w = 0, 1, \dots, 5$ , the reference solution being obtained with  $w = 6$ . We recall that the profits  $P(\mathbf{i})$  are defined as the ratios between the error and work contributions,  $P(\mathbf{i}) = \Delta E(\mathbf{i}) / \Delta W(\mathbf{i})$ , where  $\Delta E(\mathbf{i})$  are estimated combining equations (3.12) and (3.11), and  $\Delta W(\mathbf{i})$  are defined either as (3.13) or (3.14).

The computational cost associated to each sparse grid is expressed in terms of number of linear system solves. For the adaptive sparse grids, this count also includes the cost of the exploration of the reduced margin. Moreover, when using non-nested points we also take into account the system solves related to the points that have been included and then excluded from the sparse grid, cf. equation (3.16). As for the quasi-optimal sparse grids, their construction requires some additional solves to estimate the parameters  $g_1, \dots, g_N$  in (3.12), cf. [60, 11]. More precisely, the  $n$ -th rate is estimated by fixing all variables but  $y_n$  to their expected value, computing the value of the QoI increasing the number of collocation points along  $y_n$  and then fitting the resulting interpolation error with respect to a reference solution with a fixed number  $K + 1$  of collocation points: in practice, this amounts to solve  $K(K + 1)/2$  linear systems per random variable. In this specific case we fixed  $K = 7$  to estimate the rates: this choice provided stable fits for all the random variables included in the simulation.

#### Dimension adaptive sparse grids

We start our discussion from Figure 3.1, where we show the convergence results obtained with the dimension-adaptive Algorithm 2 in the case of isotropic covariance function varying the choice of profit indicators (cf. Algorithm 1) and the choice of interpolation points, i.e. Genz–Keister versus Gauss–Hermite points, the latter denoted by a suffix NN in the plot, as per “non nested” (cf. Table 3.1); in this test, we have set the buffer size to  $N_b = 10$ . More specifically, we used the “deltaint-based” profit indicators in Figure 3.1-(a) (D/NP in the plots, where D stands for “deltaint” and NP for “divided by number of points”) and “weighted  $L^\infty$ -based” profit indicators in Figure 3.1-(b) (WLinf/NP in the plots). The curve labeled as “MC error” represents the expected MC decay that has been generated as  $\sqrt{\text{Var}(u) / M(w)}$ , with  $\text{Var}(u)$  estimated as  $\text{Var}(u) \approx \mathcal{Q}_{\mathbf{I}(w_{ref})}^m[u^2] - (\mathcal{Q}_{\mathbf{I}(w_{ref})}^m[u])^2$ ; such an estimate has been verified to be accurate enough by comparing with actual realizations of the MC estimator. We have also put, as a more advanced reference method to compare with, the convergence obtained by using a Quasi Monte Carlo (QMC) estimator built on Sobol points ([45, 44]). In both Figures 3.1-(a,b) we observe that there is not much difference between the adaptive scheme which uses nested points and the one that does not. The convergence obtained in this case is satisfying, since we obtain an improved convergence rate with respect to the MC case that almost matches the rate predicted in [26] for bounded random variables with a different adaptive scheme and also in [24] as anticipated in section 3.2. QMC and the sparse grid approximation perform

similarly in this case.

The numbers next to each point give information about the shape of the multi-index sets  $\mathbf{I}(w)$  generated by the adaptive algorithm, and consequently on the distribution of the sparse grid points on the  $\tilde{N}$ -dimensional parameter space. The first number (out of the brackets) indicates the number of active directions, while the second number (in the brackets) denotes the maximum number of directions that have been activated at the same time, i.e. the highest dimensionality among all the tensor grids composing the sparse grid, cf. equation (3.5). Here and in the following, green labels refer to grids with nested points, while red labels to grids with non-nested points. Observe that after  $\approx 20$  problem solves the algorithm has activated “only” 1 variable due to the fact that at the beginning of the algorithm  $N_b$  variables must be explored, requiring  $1 + 2N_b = 21$  solver calls, in order to decide which variable should be activated as second; moreover the number of “active” variables is always smaller than  $N = 322$ , which is the number of “active” variables for the reference solution, hence our error estimator (3.22) effectively measures also a truncation error.

In Figure 3.2 we repeat the procedure starting from a tensor covariance function as in (2.6): the faster decay of the eigenvalues of this covariance function does not generate an improved convergence rate with respect to the previous case, at least in this range of values, indeed we obtain very similar results also in terms of random variables activated. However, since in the two cases the decay of the eigenvalues is different, at least asymptotically, we claim that this is only a pre-asymptotic effect. In this regime we observe a good improvement even with respect to QMC.

### Dimension quasi-optimal sparse grids

In Figure 3.3 we show instead the errors obtained by using quasi-optimal sparse grid approximations of the QoI built on Genz–Keister and Gauss–Hermite knots (labeled OPT and OPT NN respectively), for both isotropic and tensor covariances. Observe also that the Lebesgue constant  $\mathbb{M}_n^{m(i_n)}$ , introduced in section 3.4.1 and needed for computations, has been proven in [60] to be identically equal to one in the case of Gauss–Hermite abscissas, while can be numerically estimated in the case of Genz–Keister points. Again, the labels next to each point represent the number of active variables (outside the brackets) and the number of variables activated at the same time (in the brackets). As in the previous case the errors obtained with tensor and isotropic covariance behave similarly; the labels suggest that, for the same work, the adaptive sparse grids seem to activate a slightly smaller number of variables than the quasi-optimal ones, while the tensor grids dimensionality seems to be comparable. Also for the quasi-optimal sparse grids the number of “active” variables is always smaller than the reference solutions’ ones. Again the QMC and sparse grid perform similarly in the case of isotropic input covariance while the latter remains more effective in the case of tensor input covariance.

Finally, we test numerically the pointwise approximation properties of the sparse grids methodology. In an unbounded domain context. The main difficulty encountered in performing pointwise approximation is that the Lagrangian polynomials on which the sparse grid construction is based are not uniformly bounded (regardless of the choice of collocation points). On the other hand, the solution of the Darcy problem is unbounded as well, and indeed it can only be shown that  $p \in C_\pi^0(\Gamma, V)$  with an exponentially decaying weight given by  $\pi(\mathbf{y}) = \prod_{n=1}^N \exp(-|y_n| \sqrt{\lambda_n} \|\psi_n\|_{L^\infty(D)})$ , see [3, 49]. Since such convergence also implies the convergence in norm  $C_\rho^0(\Gamma, V)$  we will verify numerically the convergence in such weighted norms for both, the exponential weight  $\pi$  and the

### Chapter 3. Sparse Grid polynomial approximation for random PDEs in high dimensional probability spaces

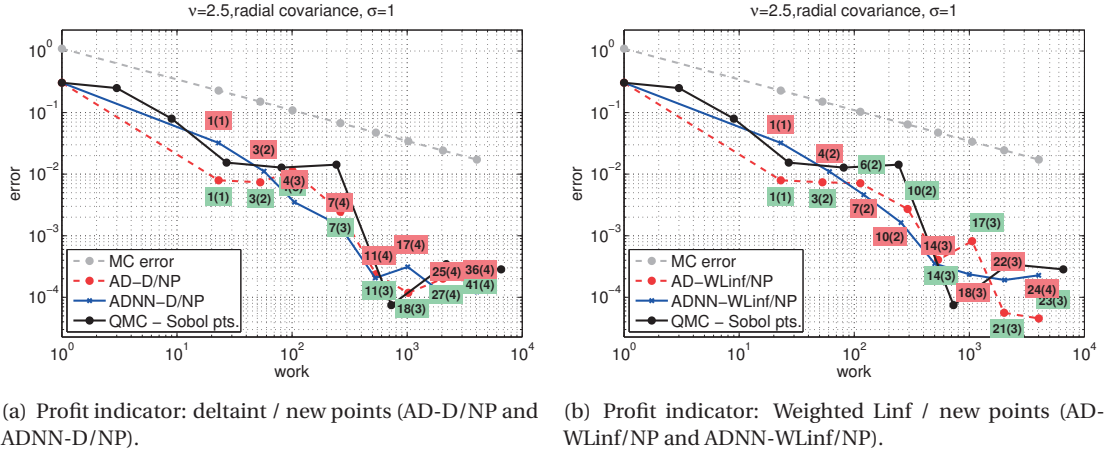


Figure 3.1 – Case  $\nu = 2.5$  radial covariance, adaptive sparse grids error.

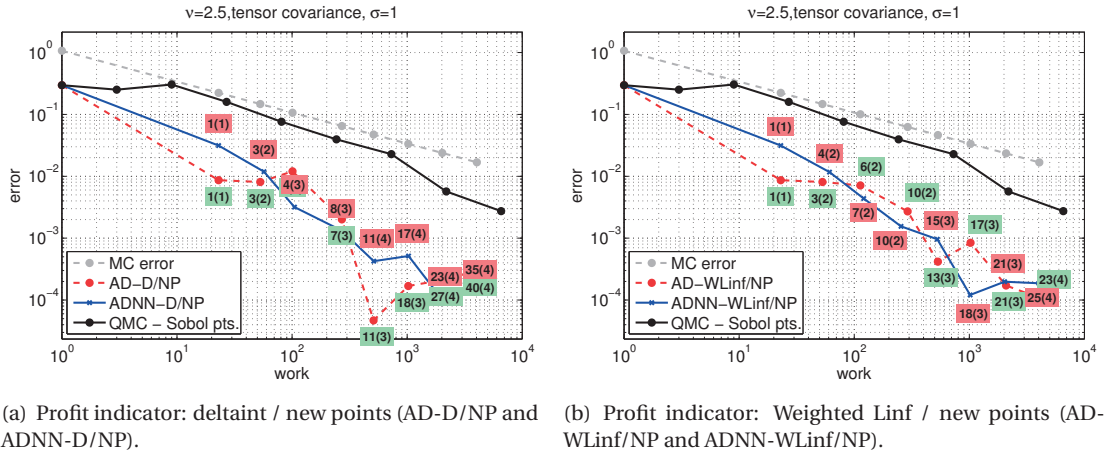


Figure 3.2 – Case  $\nu = 2.5$  tensor covariance, adaptive sparse grids error.

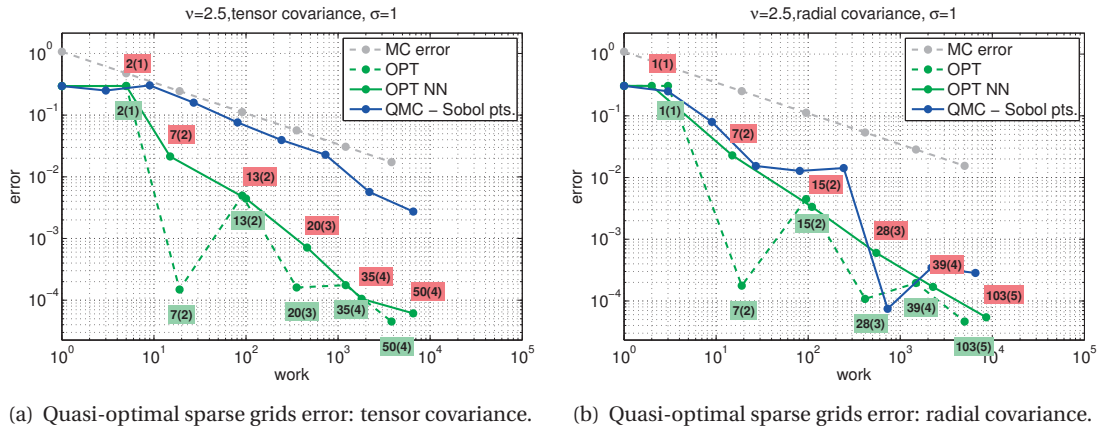


Figure 3.3 – Case  $\nu = 2.5$ , quasi-optimal sparse grids error obtained with tensor (left) and radial (right) covariance function.

gaussian weight  $\rho$ , by sampling the difference between the exact value of the Quantity of Interest  $Q$  and its sparse grid approximation over a set of 10000 points randomly sampled from a multivariate



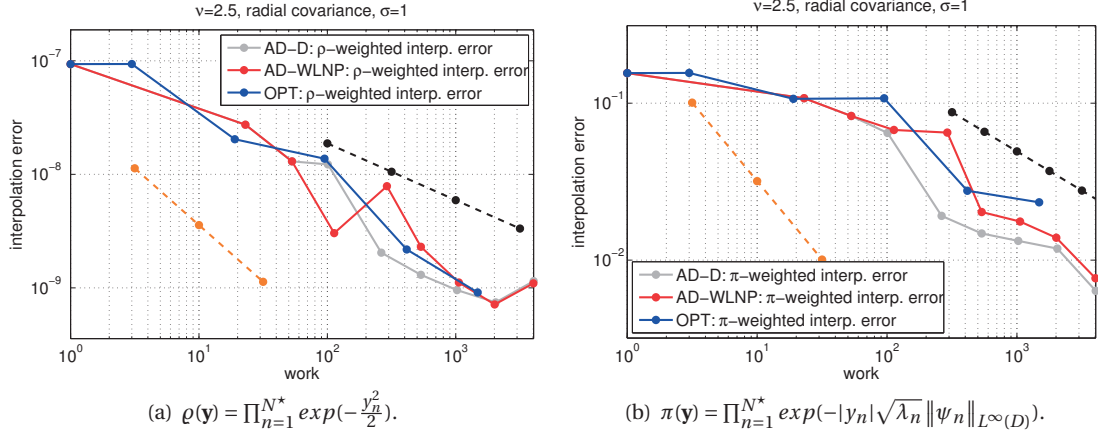


Figure 3.4 – Case  $v = 2.5$ , radial covariance function. Interpolation error with different weight functions for the Quasi Optimal (OPT) and Adaptive ( profit indicators Deltaint (AD-D) and Weighted Linf / new points (AD-WLNP) ) cases.  $work = W_{I,m}$ ,  $N^* = 60$ ,  $card(\mathcal{R}) = 10000$ . Dashed lines represent the slopes -0.5 (black) and -1 (orange).

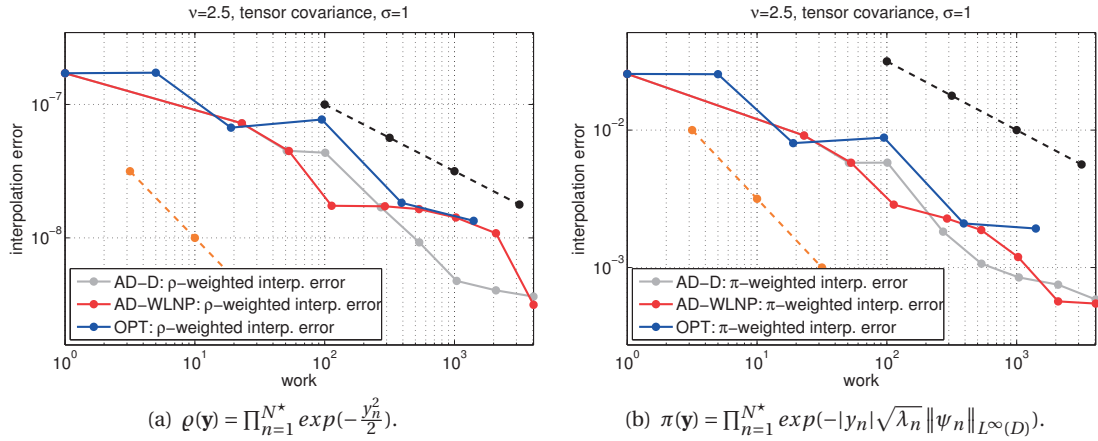


Figure 3.5 – Case  $v = 2.5$ , tensor covariance function. Interpolation error with different weight functions for the Quasi Optimal (OPT) and Adaptive ( profit indicators Deltaint (AD-D) and Weighted Linf / new points (AD-WLNP) ) cases.  $work = W_{I,m}$ ,  $N^* = 60$ ,  $card(\mathcal{R}) = 10000$ . Dashed lines represent the slopes -0.5 (black) and -1 (orange).

standard gaussian distribution:

$$\|Q - \mathcal{I}_I^m[Q]\|_{C_\pi^0(\Gamma)} \approx \max_{\mathbf{y} \in \mathcal{R}} |(Q(\mathbf{y}) - \mathcal{I}_I^m[Q](\mathbf{y}))\pi(\mathbf{y})|.$$

Observe that the number of random variables considered by  $\mathcal{I}_I^m[Q]$  is increasing as the sparse grids algorithm keeps running; thus, we choose  $\mathcal{R} \subset \mathbb{R}^{N^*}$ , with  $N^*$  sufficiently larger than the number of random variables activated by the most refined sparse grids. The results are shown in Figure 3.4 and 3.5, and indicate that the sparse grid pointwise weighted approximation error is indeed decreasing. Here we report only the results obtained with tensor covariance. These results suggest again that the various sparse grid construction techniques considered in this work behave similarly, and in particular “weighted  $L^\infty$ ” based sparse grids do not show particular gains with respect to the

“deltaint”-based ones.

The sparse grids considered in this test are the same ones used to obtain the results shown in Figures 3.1, 3.2 and 3.3; again we remark that the number of active variables remains significantly smaller than the number of variables  $N^* = 60$  used to compute our approximated sample space  $\mathcal{R}$ . We actually observe convergence in the norms  $\|\cdot\|_{C^0_b(\Gamma)}$  and  $\|\cdot\|_{C^0_\pi(\Gamma)}$ ; moreover the “weighted  $L^\infty$ ” norms of the differences between the exact sample  $Q$  and its sparse grid reconstruction  $\mathcal{S}_I^m[Q]$  converge with a similar rate when using the exponential weight  $\pi$  and the gaussian one  $\rho$ .

### Discussion

To conclude the presentation of this reference case in Figure 3.6-(a) we report a comparison between some of the errors previously presented; for the quasi-optimal case, we include in the definition of the work also the extra cost needed to compute the rates  $g_1, \dots, g_N$  in (3.12), cf. [60, 11]; more precisely, the  $n$ -th rate is estimated by fixing all variables but  $y_n$  to their expected value, computing the value of the QoI increasing the number of collocation points along  $y_n$  and then fitting the resulting interpolation error: in practice, this amounts to solving 25 linear systems per random variable, which are included in the work count. In Figure 3.6-(b) we show the convergence obtained

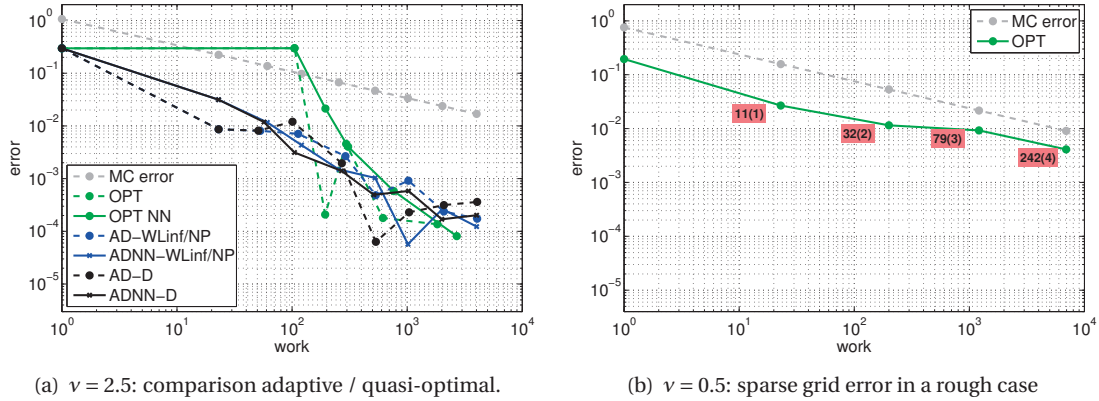


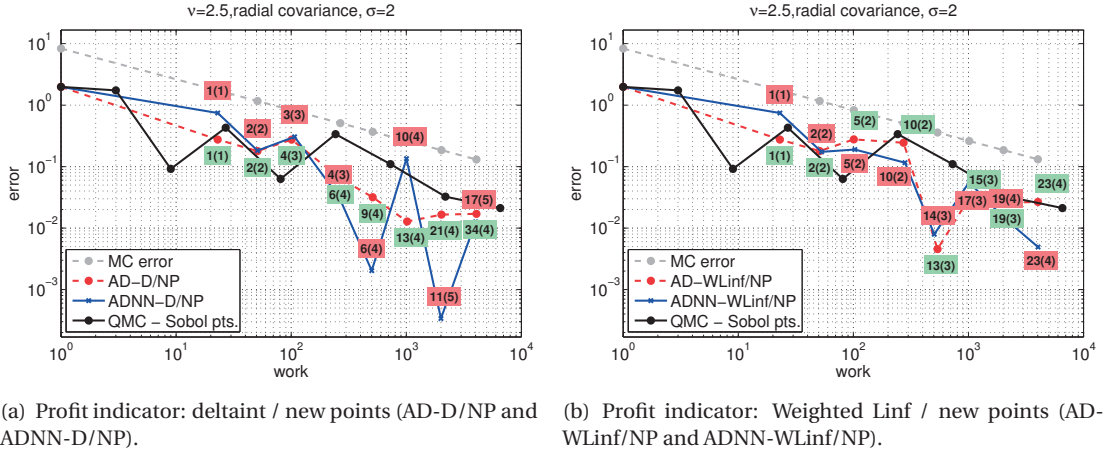
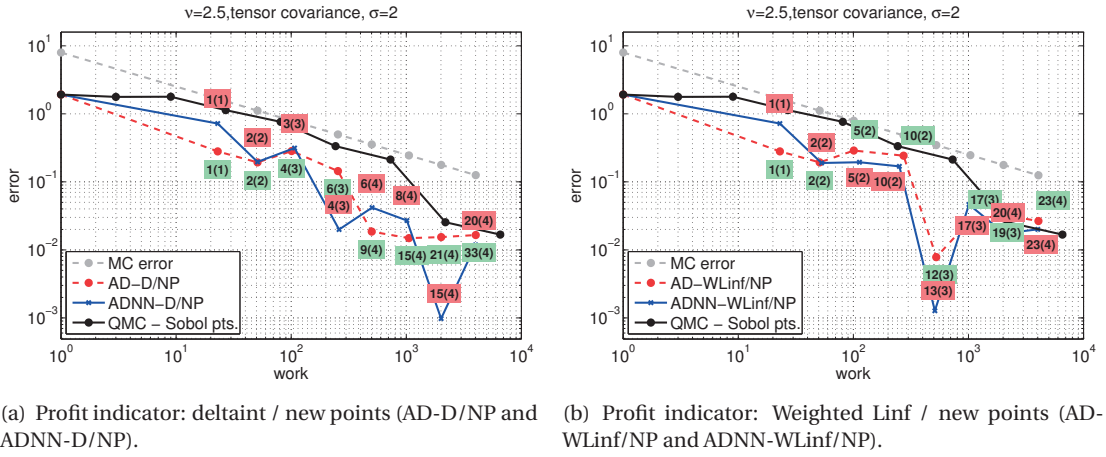
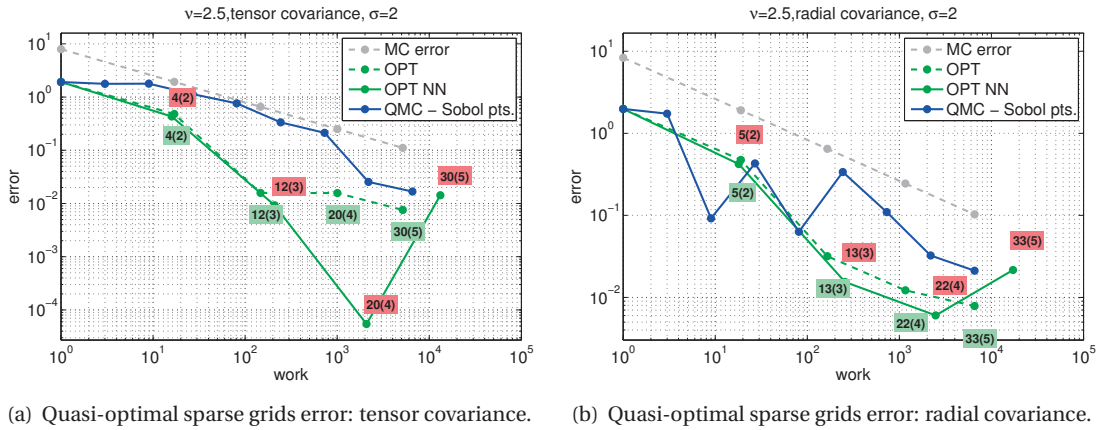
Figure 3.6 – Tensor covariance: sparse grids error for different regularities of the input random field.

by applying the sparse grid algorithm to a problem depending on a rough input random field: it is clear that in such a case the sparse grid approximation does not bring any significant improvement with respect to a standard Monte Carlo scheme and therefore its use is not recommended.

### 3.6.2 Permeabilities with high variability: $\sigma^2 = 4$

Here we show how the sparse grid approximation suffers the effects of a larger variability of the input. In Figures 3.7-3.9 we repeat the previous computations by changing  $\sigma$  in  $\sigma = 2$ . By looking at the results it appears evident that the convergence properties of the adaptive schemes deteriorate when considering larger values of  $\sigma$ ; however if from the one hand the quasi-optimal convergence looks pretty bad, on the other hand the adaptive scheme seems at least to be more robust; in fact in the case  $\sigma = 2$ , all errors coming from isotropic and tensor covariances and obtained with “deltaint-based” and “weighted  $L^\infty$ -based” profits, present a better constant and a slightly better rate than both, standard MC and QMC.




 Figure 3.7 – Case  $v = 2.5, \sigma = 2$ , adaptive sparse grids error.

 Figure 3.8 – Case  $v = 2.5, \sigma = 2$ , adaptive sparse grids error.

 Figure 3.9 – Case  $v = 2.5, \sigma = 2$ , quasi-optimal sparse grids error obtained with tensor (left) and radial (right) covariance function.

If we look at the number of random variables activated by the algorithm it is clear what is happening: since the variability of the input field is large, the algorithm before activating new random variables

tends to add points to the first directions; again we claim that this is a pre-asymptotic effect and that at a certain point, once a sufficiently large number of points has been added to the sparse grid along the first directions, some sudden drop of the error should occur.

### 3.6.3 Conditioned Darcy problem

Here we want to estimate the effects of including in the model some additional information that could be available in practical cases. In particular we imagine to have at our disposal 9 measurements  $a_1, \dots, a_9$ , listed in the vector  $\mathbf{a} \in \mathbb{R}^{9 \times 1}$ , of the permeability  $a$  located in the points  $\{0, 1/2, 1\}^2$  sorted in the list  $\mathbf{x}^m$ . Since the log-permeability is Gaussian, it is straightforward to obtain the law of the conditioned log-permeability  $\gamma_{cond}(\mathbf{x}, \omega) = (\gamma(\mathbf{x}, \omega) | \gamma(\mathbf{x}_1^m, \omega) = \log(a_1), \dots, \gamma(\mathbf{x}_9^m, \omega) = \log(a_9))$ , in fact the conditioned mean and covariance functions are given by

$$\begin{aligned}\mu_{cond}(\mathbf{x}) &= \mu - \text{cov}_\gamma(\mathbf{x}, \mathbf{x}^m) \text{cov}_\gamma(\mathbf{x}^m, \mathbf{x}^m)^{-1} \mathbf{a}, \\ \text{cov}_{cond}(\mathbf{x}, \mathbf{y}) &= \text{cov}_\gamma(\mathbf{x}, \mathbf{y}) - \text{cov}_\gamma(\mathbf{x}, \mathbf{x}^m) \text{cov}_\gamma(\mathbf{x}^m, \mathbf{x}^m)^{-1} \text{cov}_\gamma(\mathbf{x}^m, \mathbf{y}),\end{aligned}$$

where  $\text{cov}_\gamma(\mathbf{x}, \mathbf{x}^m) \in \mathbb{R}^{1 \times 9}$ ,  $\text{cov}_\gamma(\mathbf{x}^m, \mathbf{y}) \in \mathbb{R}^{9 \times 1}$  and  $\text{cov}_\gamma(\mathbf{x}^m, \mathbf{x}^m) \in \mathbb{R}^{9 \times 9}$ . Since the conditioned covariance does not depend on the measurements  $\mathbf{a}$ , the KL basis depends only on the number and the location of the measurements, which is fixed in our case. In Figures 3.10-3.12 we can observe how, for  $\sigma = 2$ , the performance of the adaptive / quasi-optimal sparse grid algorithms is incomparably better in the conditioned case than in the unconditioned one (cf. Figures 3.7-3.12). we observe very good convergence properties for all schemes, especially if compared with the corresponding unconditioned case in which there was no significant improvement with respect to the standard MC case.

Observe that in the conditioned case, if we look at the number of random variables activated, there is a significant difference between the adaptive and the quasi-optimal scheme: the former activates a moderate number of random variables, comparable with the numbers we obtained in the unconditioned case, while the quasi-optimal schemes tend to activate many random variables at the beginning; at the moment we are not able to provide a valid explanation to this behavior.

## 3.7 Conclusions

In this chapter we presented an improved version of the standard adaptive sparse grid algorithm; in particular we set up a general framework based on the choice of suitable error indicators to measure the overall sparse grid error. The methods proposed are able to handle several issues: they can be applied to solve problems on both, bounded and unbounded domains, by using nested or non-nested nodes; moreover, whenever dealing with high dimensional problems and assuming that a “weak ordering” of the random variables holds, they can be used in their dimension-adaptive version, being therefore able to work with an infinite number of random variables. Several indicators have been introduced to drive the construction of the sparse grid in the “a posteriori” case.

For our purposes, we have then used this algorithm to solve the Darcy problem with random log-normal permeability, and compared the results obtained by changing collocation points and adaptivity indicators against those obtained by the quasi-optimal sparse grids algorithm. The computational analysis has been performed on a case with smooth permeability realizations, first by considering a reasonably small variance of the input random field, then taking into account

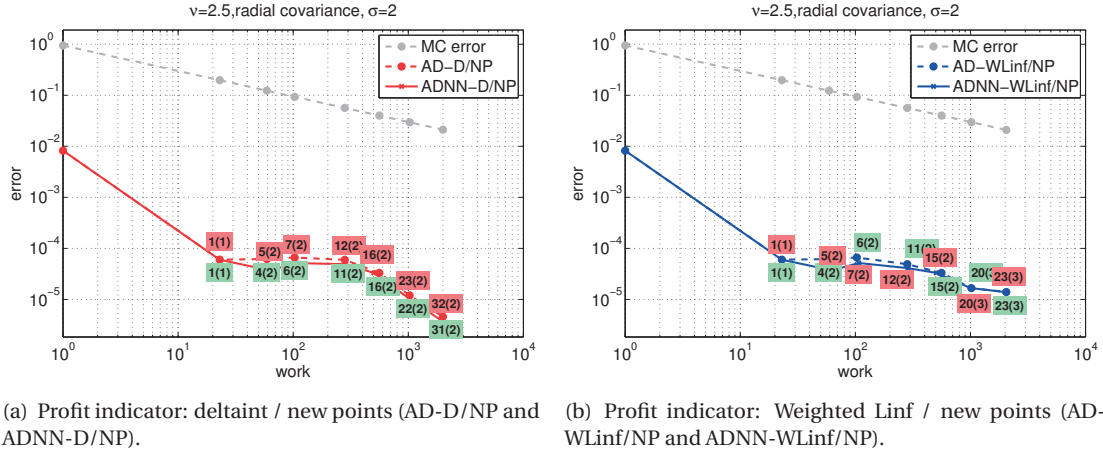


Figure 3.10 – Conditioned case:  $v = 2.5$  radial covariance, adaptive sparse grids error.

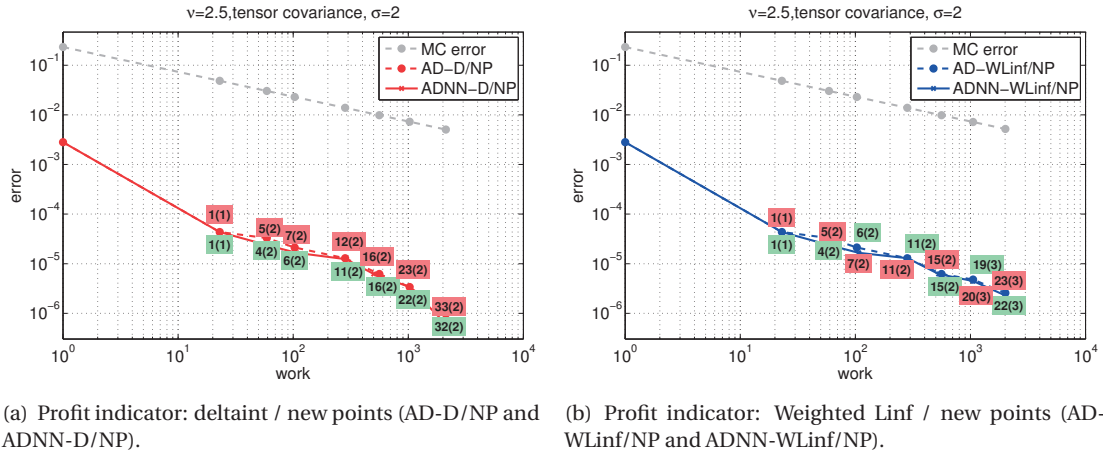


Figure 3.11 – Conditioned case:  $v = 2.5$  tensor covariance, adaptive sparse grids error.

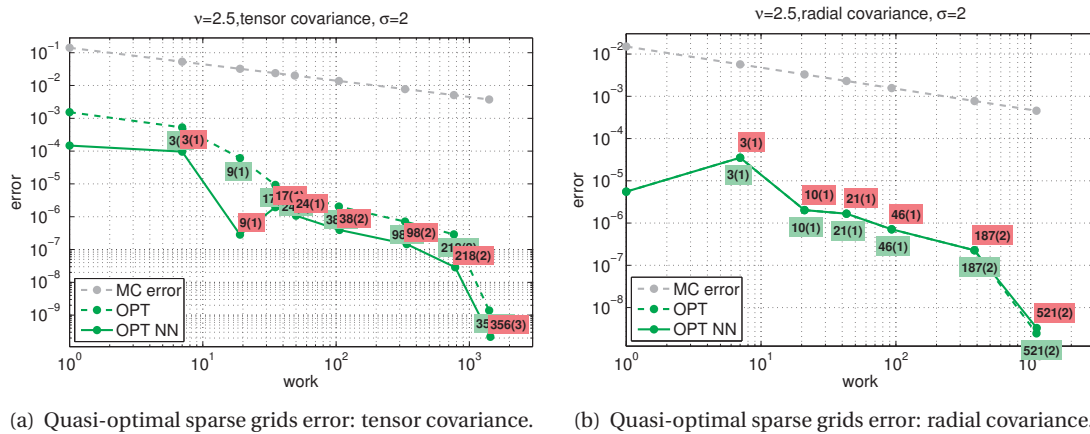


Figure 3.12 – Conditioned case:  $v = 2.5$ ,  $\sigma = 2$ , quasi-optimal sparse grids error obtained with tensor (left) and radial (right) covariance function.

highly varying permeabilities, and finally by including into the model the information coming from some measurements of the input random field, as it could happen in a realistic case. In all the

### Chapter 3. Sparse Grid polynomial approximation for random PDEs in high dimensional probability spaces

---

these cases the sparse grid has been tested by monitoring the quadrature error committed in order to approximate the mean of our QoI. Only in the first case, we have also tested numerically the convergence of the sparse grid approximation in weighted maximum norm.

From this numerical investigation, we can draw the following conclusions:

1. using non-nested points in an adaptive sparse grid framework yields results that are comparable to those obtained by nested points, at least in the log-normal context;
2. changing the indicator driving the adaptivity process does not have a dramatic impact on the quality of the solution; this however may be due to the specific choice of the QoI considered here, and more testing should be performed;
3. the sparse grid approximation will also converge with respect to suitably weighted maximum norms;
4. in the case of smooth log-permeability fields with moderate variances the adaptive and the quasi-optimal sparse grids perform similarly, in agreement with our previous findings on uniform random variables, and give quite satisfactory results;
5. in the case of highly varying log-permeability fields the adaptive and the quasi-optimal sparse grids perform again quite similarly but the convergence properties highly degenerate when considering higher and higher variances;
6. in the case of rough fields the adaptive / quasi-optimal sparse grids *alone* do not significantly improve the MC convergence and we do not advocate their use in such a case;
7. the convergence rates observed by using the dimension adaptive scheme are comparable with the one already available in literature for problems involving bounded, e.g. uniform, random variables.

## 4 Monte Carlo-type schemes with Control Variate

### 4.1 Introduction

In the previous Chapter we introduced different sparse grid approaches to deal with numerical polynomial approximation of quantities of interest related to the solution of partial differential equations, and applied this methodology to solve the Darcy problem with random permeability. We recall that we focus on the case of a lognormal permeability  $a = e^\gamma$  having a covariance function belonging to one of the two Matérn-type families (2.5) and (2.6); since such families cover a wide range of possible regularities, cf. Section 2.2, the numerical method to be used has to be adapted to the nature of the problem.

We showed in Chapter 3 how the sparse grid performs well in the presence of a smooth input field  $a$  while it suffers the effects of *the curse of dimensionality* when dealing with a rough input field. More generally, in the case of smooth random fields, different authors focused on developing (possibly sparse) polynomial chaos expansions of the solution, either in their Galerkin or Collocation versions (see e.g. [3, 2, 5, 64, 77]), and applied such strategies to solve the stochastic Darcy problem [35, 13].

On the other hand, as it can be seen for instance in Figure 3.6, the performance of polynomial approximations deteriorate when dealing with problems involving rough input permeability  $a$ ; this is due to a slower decay of the KL (or Fourier) coefficients that makes slower the convergence of the truncated field  $a_N$  to  $a$  with respect to the case of a smooth permeability. Hence, a very large number of terms (and then of random variables) has to be retained in the expansion of the random permeability  $a$ , slowing down the convergence of polynomial approximations, which may not be advantageous anymore with respect to standard MC-type schemes. Very popular because of their straightforward implementation and robustness with respect to the dimensionality of the probability space, MC-type schemes have been widely adopted to solve problems depending on rough inputs [16, 25, 44, 45]; however, because of the poor convergence properties of the standard MC estimator, the computational cost needed to obtain an accurate result could become easily unaffordable. This is also due to the fact that in general, when dealing with rough coefficients, a fine mesh is required to properly represent the roughness of the coefficient.

A way of dealing with this issue is provided by the so called Multi Level Monte Carlo method (see e.g. [19, 7, 74, 7, 41]). In this case, instead of using a fixed triangulation having sufficiently small mesh size  $h$  as in MC, a sequence of increasingly fine triangulations having mesh sizes  $h_0, \dots, h_L = h$  is considered; the idea is then to perform many computations on the coarser levels, computationally much less expensive, and only a few on the finest level, in order to obtain the same overall accuracy

of the standard MC estimator with a considerably smaller computational cost.

In this Chapter we focus on MC and MLMC schemes and we propose a way of combining such sampling schemes with the methodology developed in the previous Chapter 3, that has been verified to be very effective in the presence of smooth input random fields. More specifically we aim at using the sparse grid as a “preconditioner” for the MC sampling, in such a way to obtain an estimator with reduced variance. This is done thanks to the introduction of a suitable *control variate*, defined starting from an auxiliary Darcy problem with a regularized version of the lognormal random field  $a$ , namely  $a^\epsilon$ , as input random datum; notice that  $a^\epsilon$  will be chosen in such a way to be highly correlated with  $a$ , implying that also the two QoI related to the original and auxiliary problems will also be highly correlated; moreover since the random field  $a^\epsilon$  is smooth the mean of the QoI related to the auxiliary problem can be successfully approximated with a sparse grid scheme, and the extra bias term due to the approximation of the mean of the exact control variate through sparse grids can be easily kept under control. In general this strategy, combined with our choice of using a regularized log-permeability field obtained via convolution of the original random field  $\gamma$  with a Gaussian kernel with properly tuned variance  $\epsilon^2$ , produces significant variance reduction with respect to the standard MC corresponding schemes. We analyze the effectiveness of this strategy in both a single level and a multi level framework. We name our schemes respectively as Monte Carlo method with control variate (MCCV) and Multi Level Monte Carlo method with control variate (MLMCCV).

In particular we focus on the mean square error of the MCCV and MLMCCV estimators and provide rigorous bounds for the associated statistical errors. Then, we optimally choose the number of solves to be assigned to the MC (or MLMC) sampler and to the sparse grid scheme to approximate the mean of the regularized QoI, needed to achieve a prescribed tolerance in terms of overall mean square error. We also propose possible choices of the regularization parameter to equilibrate the space discretization error, the statistical error and the error in the computation of the expected value of the control variate by sparse grids.

The outline of the Chapter is the following: in Section 4.2 we introduce the idea of the control variate approach, the MCCV estimator and the corresponding mean square error; then in Section 4.3 we proceed with a rigorous analysis of the statistical error as a function of the discretization parameter  $h$  and the regularization parameter  $\epsilon$ ; in Section 4.4 we introduce the MLMC approach and rewrite the MCCV scheme in this framework. In Section 4.5 we derive a bound on the mean square error (MSE) of the MLMCCV estimator, the main result being in Theorem 4.2. In Section 4.6 we give some specifics about the algorithm that we used to choose the number of samples on each level as well as the sparse grid to be used to approximate the expected value of the control variate. In Section 4.8 we present some numerical results: in particular we compare the MCCV and the MLMCCV schemes with the classical MC and MLMC versions to show the effectiveness of such control variate construction. Finally, we draw some conclusions in Section 4.9.

In the Chapter there are some new aspects with respect to the paper *A Multi Level Monte Carlo method with control variate for elliptic PDEs with lognormal diffusion coefficient* [75]. From the theoretical point of view at the beginning we present an analysis comparing MCCV and MC; then we provide two complexity results in Theorems 4.5 and 4.6 concerning the two proposed algorithms to address cases with smooth and rough input permeabilities. On the other hand, among the numerical tests, we present some results obtained by using a MLMCCV estimator with level dependent truncation of the random field.

## 4.2 Monte Carlo method with control variate (MCCV)

In this section we review briefly the idea of standard Monte Carlo sampling and then introduce our variance reduction technique. We start by denoting as  $Q_h(\omega)$  the quantity of interest computed by finite elements on the structured triangulation  $\mathcal{T}_h$  for the random elementary event  $\omega \in \Omega$ . The MC idea is to approximate the mean of the quantity of interest by generating a sufficiently large random sample of size  $M$  by solving the truncated Darcy problem (2.12), possibly with a very large  $N$ , being  $N$  the number of random variables kept in the truncation of the input permeability, and by computing the corresponding QoI  $Q(\mathbf{y}_j)$ , for  $j = 1, \dots, M$ . In the following we will assume always  $N$  large enough to obtain a negligible truncation error, and we will always write  $Q$  and  $Q_h$  instead than  $Q_N$  and  $Q_{N,h}$ . Then the Monte Carlo (MC) estimator  $\hat{Q}_{h,M}^{MC}$  of the mean of  $Q$  associated to a particular spatial mesh of size  $h$  and a sample size  $M$  is defined as:

$$\hat{Q}_{h,M}^{MC} = \frac{1}{M} \sum_{i=1}^M Q_h(\mathbf{y}_i), \quad (4.1)$$

where  $Q_h(\mathbf{y}_i)$  are independent random variables all distributed as  $Q_h(\mathbf{y})$ . The mean square error of this estimator is given by:

$$e(\hat{Q}_{h,M}^{MC})^2 := \mathbb{E}[(\hat{Q}_{h,M}^{MC} - \mathbb{E}[Q])^2] = \frac{\text{Var}(Q_h)}{M} + (\mathbb{E}[Q_h - Q])^2.$$

Hence the error naturally splits in two terms: a statistical error given by the variance of the estimator and a bias term related to the finite element approximation of the PDE and the quantity of interest. The Monte Carlo approach is straightforward to implement, but unfortunately presents a rather slow convergence rate with respect to the sample size  $M$  which makes the computation of accurate solutions very demanding.

As mentioned in the introduction, MC methods are preferable to deterministic schemes such as stochastic collocation and stochastic Galerkin for their robustness also in case of rough input random fields, while sparse grid collocation methods perform significantly better in case of smooth input data. Here we want to combine the advantages coming from these two classes of schemes to further improve the performance of the MC method in case of rough coefficients by exploiting the well known control variate variance reduction technique (see e.g. [68]). The idea is to introduce an auxiliary Darcy problem having a smoothed coefficient, which can be effectively approximated by using the methodology developed in Chapter 3, and use the quantity of interest, computed from the corresponding solution, as control variate in the Monte Carlo sampling.

Let  $\gamma(\mathbf{x}, \omega)$  and  $\gamma^\epsilon(\mathbf{x}, \omega)$  be the input random fields obtained, respectively, by considering a covariance function of the Matérn family and the convolution of  $\gamma(\mathbf{x}, \omega)$  with a smooth kernel (e.g. Gaussian), namely:

$$\gamma^\epsilon(\cdot, \omega) = \gamma(\cdot, \omega) * \phi_\epsilon(\cdot) \quad \text{where} \quad \phi_\epsilon(\mathbf{x}) = e^{-\frac{|\mathbf{x}|^2}{2\epsilon^2}} / (2\pi\epsilon^2)^{\frac{d}{2}}, \quad (4.2)$$

analogously let  $a^\epsilon(\mathbf{x}, \omega) = e^{\gamma^\epsilon(\mathbf{x}, \omega)}$  and let  $p(\mathbf{x}, \omega)$  and  $p^\epsilon(\mathbf{x}, \omega)$  denote the solutions corresponding to the two (highly correlated) input random fields. Let us assume for the moment that we know exactly the mean of the control variate  $Q^\epsilon(\omega) = Q(p^\epsilon(\cdot, \omega))$  obtained starting from the solution of the



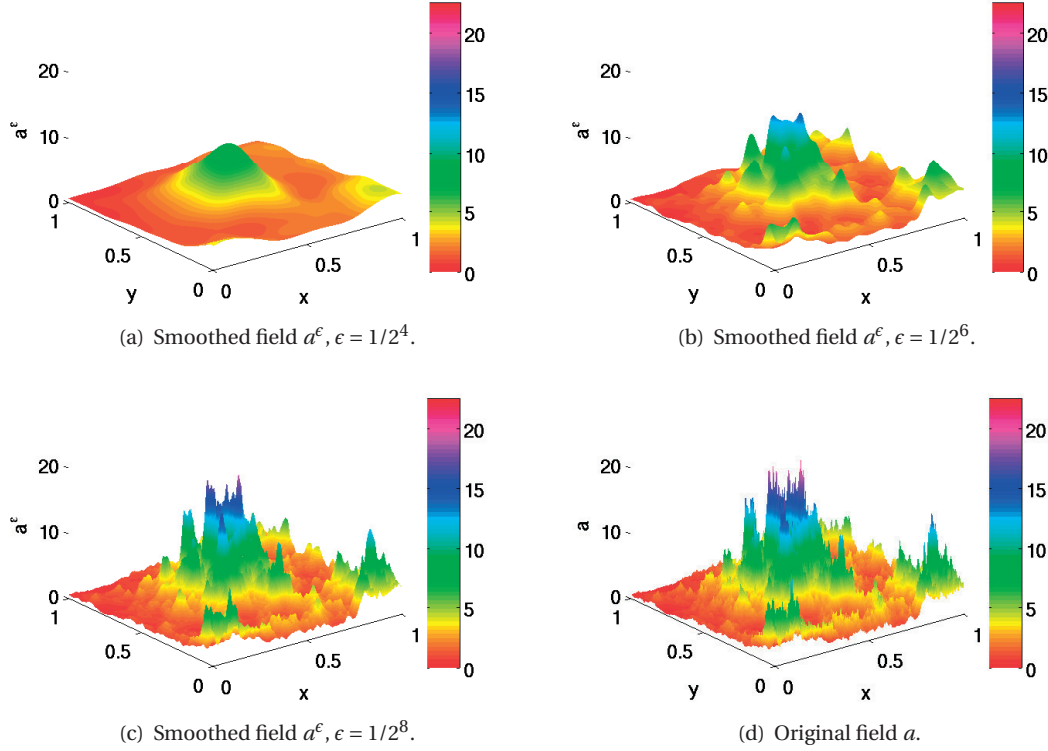


Figure 4.1 – Three different regularizations of the same realization of  $a$ .  $\nu = 0.5$ ,  $L_c = 0.5$ ,  $\sigma = 1$ .

auxiliary problem having  $a^\epsilon$  as input datum. We define

$$Q^{CV}(\omega) := Q(\omega) - Q^\epsilon(\omega) + \mathbb{E}[Q^\epsilon].$$

This new random variable is such that  $\mathbb{E}[Q^{CV}] = \mathbb{E}[Q]$  and

$$\text{Var}(Q^{CV}) = \text{Var}(Q) + \text{Var}(Q^\epsilon) - 2\text{cov}(Q, Q^\epsilon),$$

showing that the more positively correlated the two random fields are the more positively correlated the corresponding quantities of interest are and the larger the variance reduction achievable. See for instance in Figure 4.1 how a single realization of the smoothed random field  $a^\epsilon(\omega_i)$  gets closer and closer to the original one  $a(\omega_i)$ , and therefore more and more correlated, when the parameter  $\epsilon$  gets smaller and smaller.

Observe that, in order to apply this strategy to solve the stochastic Darcy problem, we should know the exact mean of  $Q^\epsilon(\omega)$ . Actually we do not have this information available but, since  $a^\epsilon(x, \omega)$  has smooth realizations, we can successfully use one of the previous sparse grid schemes to compute it accurately as long as the smoothing parameter  $\epsilon$  remains sufficiently large. Denoting by  $\mathcal{Q}_W[Q^\epsilon]$  the sparse grid quadrature used to approximate the mean of  $Q^\epsilon$  with  $W = W_1^m$  linear solves, the final variable to which we apply the MC sampling is

$$Q_W^{CV}(\mathbf{y}) := Q(\mathbf{y}) - Q^\epsilon(\mathbf{y}) + \mathcal{Q}_W[Q^\epsilon]; \quad (4.3)$$

by using linear finite elements for the solution of the truncated Darcy problem (2.12), the MCCV



estimator is then defined as

$$\hat{Q}_{h,M,W}^{MCCV} = \frac{1}{M} \sum_{i=1}^M (Q_h(\mathbf{y}_i) - Q_h^\epsilon(\mathbf{y}_i)) + \mathcal{Q}_W[Q_h^\epsilon], \quad (4.4)$$

where  $\mathbf{y}_i$ ,  $i = 1, \dots, M$ , are independent identically distributed draws from  $\rho(\mathbf{y})$ . Concerning the mean square error associated with the estimator (4.4), the following result holds:

**Lemma 4.2.1.** *The mean square error of the estimator (4.4) can be bounded as*

$$e(\hat{Q}_{h,M,W}^{MCCV})^2 \leq \frac{\mathbb{V}ar(Q_h - Q_h^\epsilon)}{M} + 2(\mathbb{E}[Q_h^\epsilon] - \mathcal{Q}_W[Q_h^\epsilon])^2 + 2\mathbb{E}[Q_h - Q]^2. \quad (4.5)$$

*Proof.* The mean square error associated with this estimator naturally splits into a variance and a bias term as

$$e(\hat{Q}_{h,M,W}^{MCCV})^2 = \mathbb{E}[(\hat{Q}_{h,M,W}^{MCCV} - \mathbb{E}[Q])^2] = \frac{\mathbb{V}ar(Q_{W,h}^{CV})}{M} + \mathbb{E}[Q_{W,h}^{CV} - Q]^2 = \frac{\mathbb{V}ar(Q_h - Q_h^\epsilon)}{M} + \mathbb{E}[Q_{W,h}^{CV} - Q]^2.$$

The second term on the right hand side represents the bias and can further be bounded as

$$\mathbb{E}[Q_{W,h}^{CV} - Q]^2 = \mathbb{E}[Q_h - Q_h^\epsilon + \mathcal{Q}_W[Q_h^\epsilon] - Q]^2 \leq 2(\mathbb{E}[Q_h^\epsilon] - \mathcal{Q}_W[Q_h^\epsilon])^2 + 2\mathbb{E}[Q_h - Q]^2.$$

□

The first term on the right hand side of (4.5), represents the variance of the estimator  $\hat{Q}_{h,M,W}^{MCCV}$ , and is expected to be smaller than the variance of the standard MC estimator thanks to the presence of the control variate. The second term represents the error due to the approximation of the mean of the smoothed quantity of interest via sparse grid quadrature; the third one represents the mean of the finite element error of the original quantity of interest. When  $\epsilon$  goes to 0 the regularized input random field tends to the original one. Consequently the solution  $p^\epsilon$  tends to the solution  $p$ ; this means that the variance associated to the estimator tends to 0. On the other hand, according to the previous considerations, an accurate approximation of the mean of the quantity of interest by a sparse grid scheme could become extremely costly and practically unfeasible for rough random fields. The parameter  $\epsilon$  should therefore be chosen so as to yield a good variance reduction while still keeping a manageable sparse grid approximation problem.

**Remark 4.2.1.** *The numerical results obtained in Chapter 3 fit into this framework; by choosing a vanishing  $\epsilon$ , since in the previous Chapter we monitored the convergence as the weak error  $|\mathcal{Q}_{\mathbf{1}(w)}^{\mathbf{m}}[Q] - \mathcal{Q}_{\mathbf{1}(w_{ref})}^{\mathbf{m}}[Q]|$ , those results represent exactly the square root of the second term on the right hand side of (4.5); on the other hand, when considering a positive value of  $\epsilon$  and a rough random field  $a$ , the regularization procedure makes the input permeability smooth and therefore similar considerations apply to the mean of the control variate.*

### 4.3 Error analysis of the MCCV method

In this section we study the statistical error of the MCCV estimator and give bounds on the variance terms  $\mathbb{V}ar(Q_{W,h}^{CV})$  in the case of the Stochastic Darcy Problem (2.2) with log-normal permeability. This, in particular, implies the study of the difference  $Q - Q^\epsilon$  as a function of the regularization

parameter  $\epsilon$ . The main result of this section is given in Theorem 4.3. Before proceeding with the analysis, we define the random variables

$$a_{\max}^\epsilon(\omega) = \max_{\mathbf{x} \in \bar{D}} a^\epsilon(\mathbf{x}, \omega), \quad a_{\min}^\epsilon(\omega) = \min_{\mathbf{x} \in \bar{D}} a^\epsilon(\mathbf{x}, \omega).$$

and restrict ourselves, in this section, to the case of a fully homogeneous Dirichlet problem ( $\Gamma_D = \partial D$ ,  $g = 0$ ) defined on a convex  $\mathcal{C}^2$  bounded domain  $D$ . Moreover, we assume that the Gaussian random field  $\gamma$  is defined in  $\mathbb{R}^d$  and for technical reasons, we consider the following smoothed version  $\gamma^\epsilon$ , slightly different from the one given in (4.2). Let  $D_\eta = \{\mathbf{x} \in \mathbb{R}^d \text{ s.t. } \text{dist}(\mathbf{x}, \bar{D}) \leq \eta\}$  and consider a function  $\varphi \in \mathcal{C}^\infty(\mathbb{R}^d)$ ,  $0 \leq \varphi \leq 1$ ,  $\varphi = 1$  on  $D_{n\epsilon}$  and  $\varphi = 0$  on  $D_1^c$  for some  $n \in \mathbb{N}$  such that  $n\epsilon < 1$ . Then, we define the smoothed field  $\gamma^\epsilon$  as

$$\gamma^\epsilon(\mathbf{x}) = (\tilde{\gamma} * \phi_\epsilon)(\mathbf{x}), \quad \text{where } \tilde{\gamma}(\mathbf{x}) = \varphi(\mathbf{x})\gamma(\mathbf{x}) \quad \text{and} \quad \phi_\epsilon(\mathbf{x}) = \frac{e^{-\frac{|\mathbf{x}|^2}{2\epsilon^2}}}{(2\pi\epsilon^2)^{\frac{d}{2}}}. \quad (4.6)$$

Essentially,  $D_{n\epsilon}$  represents the domain upon which the convolution integral involved in the definition (4.6) is computed up to an error that can be made arbitrarily small by increasing  $n$ . By taking  $n > 3$  we will have  $\gamma^\epsilon = \tilde{\gamma} * \phi_\epsilon \approx \gamma * \phi_\epsilon$  in  $D$  up to a very small error. On the other hand,  $\tilde{\gamma}$  has a compact support so that the quantities  $\tilde{\gamma}_{\max}(\omega) = \max_{\mathbf{x} \in \bar{D}} \tilde{\gamma}(\mathbf{x}, \omega)$ ,  $\tilde{\gamma}_{\min}(\omega) = \min_{\mathbf{x} \in \bar{D}} \tilde{\gamma}(\mathbf{x}, \omega)$ ,  $\|\tilde{\gamma}(\cdot, \omega)\|_{C^\alpha(\mathbb{R}^d)}$  are all  $L_{\mathbb{P}}^q(\Omega)$  functions,  $\forall q \in \mathbb{R}_+$  and can be bounded by the corresponding quantities on  $\gamma$  evaluated in the extended domain  $\bar{D}_1$ . In particular, we have that  $|\tilde{\gamma}|_{C^\alpha(\mathbb{R}^d)} \leq \|\varphi\|_{C^\alpha(\mathbb{R}^d)} \|\gamma\|_{C^\alpha(\bar{D}_1)}$ . We start with the following observation:

**Lemma 4.3.1.** *Let  $\gamma(\mathbf{x}, \omega)$  be a Gaussian random field with realizations a.s. in  $C^\alpha(\bar{D}_1)$  and  $\gamma^\epsilon(\mathbf{x}, \omega)$  a smoothed version of  $\gamma(\mathbf{x}, \omega)$  as introduced in (4.6). Moreover set  $a = e^\gamma$  and  $a^\epsilon = e^{\gamma^\epsilon}$ . For all  $0 \leq \beta \leq \min(\alpha, 1)$  it holds:*

$$\|a(\cdot, \omega) - a^\epsilon(\cdot, \omega)\|_{C^\beta(\bar{D})} \lesssim C_{4.3.1}(\omega, \alpha) \epsilon^{\min(\alpha-\beta, 2)}, \quad \text{a.s. in } \Omega,$$

where

$$C_{4.3.1}(\omega, \alpha) = a_{\max}(\omega) \left\| 1 + e^{(\gamma^\epsilon - \gamma)(\cdot, \omega)} \right\|_{\mathcal{C}^0(\bar{D})} \left( 1 + |\gamma(\cdot, \omega)|_{C^{\min(\alpha, 1)}(\bar{D})} \right) \|\gamma(\cdot, \omega)\|_{C^{\min(\alpha, 3)}(\bar{D}_1)} \|\varphi(\cdot, \omega)\|_{C^{\min(\alpha, 3)}(\bar{D}_1)}.$$

Moreover, if  $\alpha > 1$ , there holds also the following bound for the standard  $\mathcal{C}^1$  norm

$$\|a(\cdot, \omega) - a^\epsilon(\cdot, \omega)\|_{\mathcal{C}^1(\bar{D})} \lesssim \tilde{C}_{4.3.1}(\omega, \alpha) \epsilon^{\min(\alpha-1, 2)}, \quad \text{a.s. in } \Omega,$$

where

$$\tilde{C}_{4.3.1}(\omega, \alpha) = a_{\max}(\omega) \left\| 1 + e^{(\gamma^\epsilon - \gamma)(\cdot, \omega)} \right\|_{\mathcal{C}^0(\bar{D})} \left( 1 + |\gamma(\cdot, \omega)|_{\mathcal{C}^1(\bar{D})} \right) \|\gamma(\cdot, \omega)\|_{C^{\min(\alpha, 3)}(\bar{D}_1)} \|\varphi(\cdot, \omega)\|_{C^{\min(\alpha, 3)}(\bar{D}_1)}.$$

The constants  $C_{4.3.1}(\omega, \alpha)$  and  $\tilde{C}_{4.3.1}(\omega, \alpha)$  are both  $L_{\mathbb{P}}^q$  integrable  $\forall q \in \mathbb{R}_+$ .

*Proof.* See appendix 4.A. □

Next we use this result to estimate how the distance between the solution of the original problem (2.3) and that of the auxiliary one with smoothed coefficient  $a^\epsilon$  in a given Sobolev norm, namely  $\|p - p^\epsilon\|_{H^{1+\beta}(D)}$ , depends on the regularization parameter  $\epsilon$ . The following result holds:

**Lemma 4.3.2.** *Let  $a(\mathbf{x}, \omega)$  and  $a^\epsilon(\mathbf{x}, \omega)$  be as in Lemma 4.3.1 and  $f \in H^{r-1}(D)$  with  $r = \min(\alpha, 1)$ . A.s. in  $\Omega$  it holds:*

$$\|p(\cdot, \omega) - p^\epsilon(\cdot, \omega)\|_{H_0^1(D)} \lesssim \tilde{C}_{4.3.2}(\omega, \alpha) \|f\|_{H^{r-1}(D)} \epsilon^{\min(\alpha, 2)},$$

$$\|p(\cdot, \omega) - p^\epsilon(\cdot, \omega)\|_{H^{1+\beta}(D)} \lesssim \frac{C_{4.3.2}(\omega, \alpha)}{(\alpha - \beta)^2 \sqrt{\eta}} \|f\|_{H^{r-1}(D)} \epsilon^{\min(\alpha - \beta - \eta, 2)}, \quad \forall 0 \leq \beta < r \text{ and } \forall 0 < \eta \leq \alpha - \beta,$$

where

$$\tilde{C}_{4.3.2}(\omega, \alpha) = \frac{C_{4.3.1}(\omega, \alpha)}{a_{\min}(\omega) a_{\min}^\epsilon(\omega)},$$

$$C_{4.3.2}(\omega, \alpha) = \begin{cases} C_{4.3.1}(\omega, \alpha) C_{2.2.1}(\omega, \alpha) C_{2.2.1}^\epsilon(\omega, \alpha) & \alpha \leq 1, \\ \tilde{C}_{4.3.1}(\omega, \alpha) C_{2.2.1}(\omega, \alpha) C_{2.2.1}^\epsilon(\omega, \alpha) & \alpha > 1. \end{cases}$$

and  $C_{2.2.1}^\epsilon(\omega, \alpha)$  as in (2.4) with  $a$  replaced by  $a^\epsilon$ . If the assumptions hold also for  $\alpha > 1$  then it is also valid the bound

$$\|p(\cdot, \omega) - p^\epsilon(\cdot, \omega)\|_{H^2(D)} \lesssim C_{4.3.2}(\omega, \alpha) \|f\|_{L^2(D)} \epsilon^{\min(\alpha - 1, 2)},$$

Moreover, the constants  $\tilde{C}_{4.3.2}(\omega, \alpha)$  and  $C_{4.3.2}(\omega, \alpha)$  are  $L_{\mathbb{P}}^q$  integrable  $\forall q \in \mathbb{R}_+$  since a.s. in  $\Omega$  it holds  $a_{\max}^\epsilon(\omega) \leq a_{\max}(\omega)$  and  $a_{\min}^\epsilon(\omega) \geq a_{\min}(\omega)$ .

*Proof.* We start by noticing that the original problem (2.2) satisfies the bound

$$\|p\|_{H_0^1(D)} \leq \frac{\|f\|_{H^{r-1}(D)}}{a_{\min}}.$$

By considering the difference between the original and the regularized problem we get:

$$\int_D a^\epsilon \nabla(p - p^\epsilon) \nabla v d\mathbf{x} = - \int_D (a - a^\epsilon) \nabla p \nabla v d\mathbf{x}, \quad \forall v \in V_0; \quad (4.7)$$

then, in order to prove the first bound, by choosing  $v = p - p^\epsilon$  in (4.7), we directly get

$$\|p - p^\epsilon\|_{H_0^1(D)} \leq \|a - a^\epsilon\|_{\mathcal{C}^0(D)} \frac{\|p\|_{H_0^1(D)}}{a_{\min}^\epsilon},$$

which, from Lemma 4.3.1 and the above bound on  $\|p\|_{H_0^1(D)}$ , implies the desired result. In order to complete the proof we will use the result in Lemma 2.2.2 in Chapter 2 that states that,  $\forall b \in C^\alpha(\bar{D})$ , and  $\forall v \in H^\beta(D)$  for some  $0 < \beta < \min(\alpha, 1)$ , the following bound holds

$$\|bv\|_{H^\beta(D)} \lesssim \frac{1}{\sqrt{\eta}} \|b\|_{C^{\beta+\eta}(\bar{D})} \|v\|_{H^\beta(D)} \quad \forall \eta \leq \alpha - \beta.$$

By integrating by parts (4.7) we obtain

$$\int_D a^\epsilon \nabla(p - p^\epsilon) \nabla v d\mathbf{x} = \int_D \operatorname{div}((a - a^\epsilon) \nabla p) v d\mathbf{x} = \int_D \tilde{f} v d\mathbf{x}.$$

In order to use the result given in Lemma 2.2.1 we need to ensure that  $\tilde{f}$  is in  $H^{\beta-1}$ . Indeed

$$\begin{aligned} \|\operatorname{div}((a - a^\epsilon) \nabla p)\|_{H^{\beta-1}(D)} &\lesssim \|(a - a^\epsilon) \nabla p\|_{H^\beta(D)} \\ &\lesssim \frac{1}{\sqrt{\eta}} \|a - a^\epsilon\|_{C^{\beta+\eta}(\bar{D})} \|p\|_{H^{\beta+1}(D)} \quad \forall 0 < \eta \leq \alpha - \beta. \end{aligned}$$

Hence for the difference  $p - p^\epsilon$  the following estimate holds

$$\begin{aligned} \|p - p^\epsilon\|_{H^{1+\beta}(D)} &\lesssim \frac{1}{\alpha - \beta} \frac{a_{\max}^\epsilon(\omega) \|a^\epsilon(\cdot, \omega)\|_{C^\alpha(\bar{D})}}{(a_{\min}^\epsilon)^3(\omega)} \|\tilde{f}(\cdot, \omega)\|_{H^{\beta-1}(D)} \\ &\lesssim \frac{1}{(\alpha - \beta)\sqrt{\eta}} \frac{a_{\max}^\epsilon(\omega) \|a^\epsilon(\cdot, \omega)\|_{C^\alpha(\bar{D})}}{(a_{\min}^\epsilon)^3(\omega)} C_{4.3.1}(\omega, \alpha) \|p\|_{H^{\beta+1}(D)} \epsilon^{\min(\alpha-\beta-\eta, 2)} \\ &\lesssim \frac{a_{\max}^\epsilon(\omega) \|a^\epsilon(\cdot, \omega)\|_{C^\alpha(\bar{D})} a_{\max}(\omega) \|a(\cdot, \omega)\|_{C^\alpha(\bar{D})}}{(\alpha - \beta)^2 \sqrt{\eta} (a_{\min}^\epsilon)^3(\omega) (a_{\min})^3(\omega)} C_{4.3.1}(\omega, \alpha) \|f\|_{H^{\beta-1}(D)} \epsilon^{\min(\alpha-\beta-\eta, 2)}. \end{aligned}$$

In the case  $\alpha > 1$  we use again the result given in Lemma 2.2.1

$$\|p(\cdot, \omega)\|_{H^2(D)} \lesssim \frac{a_{\max}(\omega) \|a(\cdot, \omega)\|_{C^1(\bar{D})}}{a_{\min}^3(\omega)} \|f(\cdot, \omega)\|_{L^2(D)};$$

analogously we need to ensure that  $\tilde{f}$  is in  $L^2(D)$ . Indeed

$$\begin{aligned} \|\operatorname{div}((a - a^\epsilon)\nabla p)\|_{L^2(D)} &\leq \|\nabla(a - a^\epsilon) \cdot \nabla p\|_{L^2(D)} + \|(a - a^\epsilon)\operatorname{div}(\nabla p)\|_{L^2(D)} \\ &\lesssim \|\nabla(a - a^\epsilon)\|_{C^0(\bar{D})} \|\nabla p\|_{L^2(D)} + \|a - a^\epsilon\|_{C^0(\bar{D})} \|\nabla p\|_{H_0^1(D)} \lesssim \|a - a^\epsilon\|_{C^1(\bar{D})} \|p\|_{H^2}. \end{aligned}$$

Therefore it holds

$$\|p - p^\epsilon\|_{H^2(D)} \lesssim \frac{a_{\max}^\epsilon(\omega) \|a^\epsilon(\cdot, \omega)\|_{C^1(\bar{D})}}{(a_{\min}^\epsilon)^3(\omega)} \frac{a_{\max}(\omega) \|a(\cdot, \omega)\|_{C^1(\bar{D})}}{(a_{\min})^3(\omega)} \tilde{C}_{4.3.1}(\omega, \alpha) \|f\|_{L^2(D)} \epsilon^{\min(\alpha-1, 2)}.$$

To conclude the proof we show that  $a_{\max}^\epsilon(\omega) \leq a_{\max}(\omega)$  and  $a_{\min}^\epsilon(\omega) \geq a_{\min}(\omega)$ ; by periodically extending the random field  $a$  in  $\mathbb{R}^d$  of course there holds  $\max_{x \in \mathbb{R}^d} a(x, \omega) = \max_{x \in \bar{D}} a(x, \omega)$  and  $\min_{x \in \mathbb{R}^d} a(x, \omega) = \min_{x \in \bar{D}} a(x, \omega)$ ; moreover, we obtain

$$\begin{aligned} a_{\max}^\epsilon &= \max_{x \in \mathbb{R}^d} a^\epsilon = \max_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} a(x - y) \phi_\epsilon(y) dy \leq \max_{x \in \mathbb{R}^d} a_{\max} \int_{\mathbb{R}^d} \phi_\epsilon(y) dy = a_{\max}. \\ a_{\min}^\epsilon &= \min_{x \in \mathbb{R}^d} a^\epsilon = \min_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} a(x - y) \phi_\epsilon(y) dy \geq \min_{x \in \mathbb{R}^d} a_{\min} \int_{\mathbb{R}^d} \phi_\epsilon(y) dy = a_{\min}. \end{aligned}$$

□

The next theorem extends the previous result to a linear quantity of interest  $Q$ .

**Theorem 4.1.** *Let  $a(x, \omega)$  and  $a^\epsilon(x, \omega)$  be as in Lemma 4.3.1,  $f \in H^{r-1}(D)$  and let  $Q(\cdot)$  be a functional on  $H^{1-r}(D)$ , i.e.  $Q \in H^{r-1}(D)$  with  $r = \min(\alpha, 1)$ , representing our QoI. Then a.s. in  $\Omega$  it holds:*

$$|Q(p)(\omega) - Q(p^\epsilon)(\omega)| \lesssim C_{4.1}(\omega, \alpha) \|f\|_{H^{r-1}(D)} \|Q\|_{H^{r-1}(D)} \epsilon^{\min(\alpha, 2)},$$

where

$$C_{4.1}(\omega, \alpha, \beta) = C_{2.2.1}(\omega, \alpha) \left( C_{4.3.1}(\omega, \alpha) C_{2.2.1}(\omega, \alpha) + 2a_{\max}^\epsilon(\omega) C_{2.3.1}(\omega, \alpha) \right)$$

*Proof.* Let us consider the adjoint problems related to the original and the auxiliary problems having  $Q(\cdot)$  as right hand side

$$\int_D a \nabla v \nabla \Phi d\mathbf{x} = Q(v), \quad \int_D a^\epsilon \nabla v \nabla \Phi^\epsilon d\mathbf{x} = Q(v), \quad \forall v \in H_0^1(D)$$

### 4.3. Error analysis of the MCCV method

where  $\Phi$  and  $\Phi^\epsilon$  are respectively the solutions of the two adjoint problems. By choosing  $v = p$  in the first problem and  $v = p^\epsilon$  in the second problem and by taking the difference we get:

$$Q(p) - Q(p^\epsilon) = \int_D a \nabla p \nabla \Phi d\mathbf{x} - \int_D a^\epsilon \nabla p^\epsilon \nabla \Phi^\epsilon d\mathbf{x}.$$

By adding and subtracting some mixed terms we get:

$$Q(p) - Q(p^\epsilon) = \int_D a \nabla p \nabla \Phi d\mathbf{x} - \int_D a^\epsilon \nabla p^\epsilon \nabla \Phi^\epsilon d\mathbf{x} \pm \int_D a^\epsilon \nabla p \nabla \Phi d\mathbf{x} \pm \int_D a^\epsilon \nabla p^\epsilon \nabla \Phi d\mathbf{x};$$

by properly grouping the terms above we obtain:

$$\begin{aligned} |Q(p) - Q(p^\epsilon)| &\leq \left| \int_D (a - a^\epsilon) \nabla p \nabla \Phi d\mathbf{x} \right| + \left| \int_D a^\epsilon \nabla (p - p^\epsilon) \nabla \Phi d\mathbf{x} \right| + \left| \int_D a^\epsilon \nabla p^\epsilon \nabla (\Phi - \Phi^\epsilon) d\mathbf{x} \right| \\ &\leq \|a - a^\epsilon\|_{L^\infty(D)} \|p\|_{H_0^1(D)} \|\Phi\|_{H_0^1(D)} + \|a^\epsilon\|_{L^\infty(D)} \|p - p^\epsilon\|_{H_0^1(D)} \|\Phi\|_{H_0^1(D)} + \\ &\quad + \|a^\epsilon\|_{L^\infty(D)} \|p^\epsilon\|_{H_0^1(D)} \|\Phi - \Phi^\epsilon\|_{H_0^1(D)}. \end{aligned}$$

Since for the solutions of the smoothed and of the adjoint problems we have identical error bounds as for the the primal one, by using Lemmas 2.2.1, 2.3.1 and 4.3.1 all with  $\beta = 0$  we obtain the bound

$$\begin{aligned} |Q(p) - Q(p^\epsilon)| &\lesssim C_{4.3.1}(\omega, \alpha) C_{2.2.1}^2(\omega, \alpha) \|f\|_{H^{r-1}(D)} \|Q\|_{H^{r-1}(D)} \epsilon^{\min(\alpha, 2)} \\ &\quad + a_{\max}^\epsilon(\omega) C_{2.2.1}(\omega, \alpha) C_{2.3.1}(\omega, \alpha) \|f\|_{H^{r-1}(D)} \|Q\|_{H^{r-1}(D)} \epsilon^{\min(\alpha, 2)} \\ &\quad + a_{\max}^\epsilon(\omega) C_{2.3.1}(\omega, \alpha) C_{2.2.1}(\omega, \alpha) \|f\|_{H^{r-1}(D)} \|Q\|_{H^{r-1}(D)} \epsilon^{\min(\alpha, 2)}, \end{aligned}$$

which proves the desired result.  $\square$

With such bound on the difference between the original and the smoothed quantities of interest we can control the mean square error associated to the MCCV estimator; it holds the following theorem:

**Theorem 4.2.** *Let  $a(x, \omega)$  and  $a^\epsilon(x, \omega)$  be as in Lemma 4.3.1, and let  $f$ ,  $Q$  and  $r$  be as in Theorem 4.1. The mean square error related to the estimator (4.4) can be bounded as*

$$e(\hat{Q}_{h,M,W}^{MCCV})^2 \lesssim c_{4.1}^2(\alpha, 2) \|f\|_{H^{r-1}(D)}^2 \|Q\|_{H^{r-1}(D)}^2 \frac{\epsilon^{\min(\alpha, 2)}}{M} + 2(\mathbb{E}[Q_h^\epsilon] - \mathcal{Q}_W[Q_h^\epsilon])^2 + 2\mathbb{E}[Q_h - Q]^2 \quad (4.8)$$

where  $c_{4.1}(\alpha, q) = \|C_{4.1}(\cdot, \alpha)\|_{L_p^q(\Omega)}$ .

*Proof.* The formula of the mean square error related to the estimator (4.4) is

$$e(\hat{Q}_{h,M,W}^{MCCV})^2 \leq \frac{\text{Var}(Q_h - Q_h^\epsilon)}{M} + 2(\mathbb{E}[Q_h^\epsilon] - \mathcal{Q}_W[Q_h^\epsilon])^2 + 2\mathbb{E}[Q_h - Q]^2.$$

We obtain

$$\begin{aligned} \text{Var}(Q_h - Q_h^\epsilon) &\leq \|Q_h - Q_h^\epsilon\|_{L_p^2(\Omega)}^2 \leq \|C_{4.1}(\cdot, \alpha)\|_{L_p^q(\Omega)}^2 \|f\|_{H^{r-1}(D)}^2 \|Q\|_{H^{r-1}(D)}^2 \epsilon^{\min(\alpha, 2)} \\ &= c_{4.1}^2(\alpha, 2) \|f\|_{H^{r-1}(D)}^2 \|Q\|_{H^{r-1}(D)}^2 \epsilon^{\min(\alpha, 2)} \end{aligned}$$

By replacing in the inequality of the mean square error we get the desired result.  $\square$

The previous result gives a bound on the variance reduction we can achieve by using this control

variate approach. Notice that the actual variance reduction could be quite larger than predicted by this estimate since, in its derivation, we conservatively bounded the variance of the difference  $Q_h - Q_h^\epsilon$  with its second moment. Nevertheless this estimate links the variance reduction to the regularity  $\alpha < \nu$  of the input permeability field: the higher the regularity the larger the variance reduction, up to a maximal factor of  $\epsilon^2$ , for  $\nu \geq 2$ .

#### 4.4 Multi Level Monte Carlo method with control variate (MLMCCV)

In this section we extend the control variate approach previously introduced to a Multi Level framework. MLMC methods [41, 19, 25, 7, 74] have been recently proposed and studied in order to reduce the variance of the MC estimator and its overall computational cost needed to meet a given tolerance. The basic idea is to consider a sequence of increasingly fine spatial meshes having mesh size  $h_0 > \dots > h_\ell > \dots > h_L$  and to use the linearity of the expectation operator to write the mean of the quantity of interest on the finest grid  $h_L$  as a telescopic sum of the mean of the quantity of interest on the coarsest level plus a sum of correcting terms given by the difference on two consecutive levels:

$$\mathbb{E}[Q_{h_L}] = \mathbb{E}[Q_{h_0}] + \sum_{\ell=1}^L \mathbb{E}[Q_{h_\ell} - Q_{h_{\ell-1}}].$$

Hence, the idea of independently estimating via standard MC estimators the terms on each level, with suitably chosen sample sizes, in order to minimize the overall complexity. Given a sequence  $\{M_\ell\}_{\ell=0}^L$  of sample sizes to be used on each level, the Multi Level Monte Carlo (MLMC) estimator is

$$\hat{Q}_{\{h_\ell\}, \{M_\ell\}}^{MLMC} = \sum_{\ell=0}^L \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} (Q_{h_\ell}(\mathbf{y}_{\ell,i}) - Q_{h_{\ell-1}}(\mathbf{y}_{\ell,i})), \text{ where } Q_{h_{-1}} = 0, \quad (4.9)$$

where  $\mathbf{y}_{\ell,i}$  are independent identically distributed draws from  $\rho(\mathbf{y})$ ; the mean square error associated to this estimator is

$$e(\hat{Q}_{\{h_\ell\}, \{M_\ell\}}^{MLMC})^2 = \mathbb{E}[(\hat{Q}_{\{h_\ell\}, \{M_\ell\}}^{MLMC} - \mathbb{E}[Q])^2] = \sum_{\ell=0}^L \frac{\mathbb{V}\text{ar}(Q_{h_\ell} - Q_{h_{\ell-1}})}{M_\ell} + (\mathbb{E}[Q_{h_L} - Q])^2. \quad (4.10)$$

In the construction of the method a key point is the choice of the sample sizes  $M_\ell$  on each level and the choice of the mesh sizes  $h_\ell$ . Several strategies have been proposed by different authors. Just to mention a few, Giles, Scheichl et al. in their works [25, 74] consider a continuous minimization problem in  $M_0, \dots, M_L$  to determine the sample sizes on each level given the mesh hierarchy  $\{h_\ell\}_{\ell=0}^L$ : the finest level of the hierarchy is chosen so that the bias term meets half of the prescribed tolerance. Alternatively Schwab et al. in [7] select the samples sizes to equilibrate all the  $L+1$  terms in the right hand side of (4.10). A global optimization of the MLMC strategy has been investigated in [28] where the authors show that geometric sequences of  $h_\ell = h_0 \beta^{-\ell}$ , with  $\beta > 1$ , lead to nearly optimal MLMC samples; however, the corresponding meshes are in general not nested. Also, optimal strategies might not split equally the prescribed tolerance into the discretization error and the statistical one.

Here we analyze the effects of combining this multi level strategy with the control variate approach. To do this, we write the mean of  $Q^{CV}$  on the finest mesh as a telescopic sum following the multi level idea:

$$\mathbb{E}[Q_{W, h_L}^{CV}] = \mathbb{E}[Q_{W, h_0}^{CV}] + \sum_{\ell=1}^L \mathbb{E}[Q_{W, h_\ell}^{CV} - Q_{W, h_{\ell-1}}^{CV}].$$

#### 4.4. Multi Level Monte Carlo method with control variate (MLMCCV)

Then, by independently estimating through MC samplers all the terms coming from different levels, we define the Multi Level Monte Carlo estimator with control variate (MLMCCV) as

$$\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV} = \sum_{\ell=0}^L \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} \left( Q_{W, h_\ell}^{CV}(\omega_{\ell, i}) - Q_{W, h_{\ell-1}}^{CV}(\omega_{\ell, i}) \right),$$

with  $Q_{h_{-1}}^{CV} = 0$ . By replacing the terms  $Q_{W, h_\ell}^{CV}(\omega_{\ell, i}) = Q_{h_\ell}(\omega_{\ell, i}) - Q_{h_\ell}^\epsilon(\omega_{\ell, i}) + \mathcal{Q}_W[Q_{h_\ell}^\epsilon]$  and by noticing that all the sparse grid quadratures except the one on the finest level cancel out, the estimator can be equivalently rewritten as

$$\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV} = \sum_{\ell=0}^L \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} \left( Q_{h_\ell}(\omega_{\ell, i}) - Q_{h_{\ell-1}}(\omega_{\ell, i}) - (Q_{h_\ell}^\epsilon(\omega_{\ell, i}) - Q_{h_{\ell-1}}^\epsilon(\omega_{\ell, i})) \right) + \mathcal{Q}_W[Q_{h_L}^\epsilon], \quad (4.11)$$

where again  $Q_{h_{-1}}, Q_{h_{-1}}^\epsilon = 0$ . Concerning the mean square error associated to the estimator (4.11), the following result generalizes (4.10):

**Lemma 4.4.1.** *The mean square error of the estimator (4.11) can be bounded as*

$$e(\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV})^2 \leq \sum_{\ell=0}^L \frac{\mathbb{V}ar(Q_{h_\ell} - Q_{h_{\ell-1}} - (Q_{h_\ell}^\epsilon - Q_{h_{\ell-1}}^\epsilon))}{M_\ell} + 2 \left( \mathbb{E}[Q_{h_L}^\epsilon] - \mathcal{Q}_W[Q_{h_L}^\epsilon] \right)^2 + 2\mathbb{E}[Q_{h_L} - Q]^2. \quad (4.12)$$

*Proof.* The mean square error associated to this estimator naturally splits into a variance and a bias term as

$$e(\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV})^2 = \mathbb{E}[(\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV} - \mathbb{E}[Q])^2] = \sum_{\ell=0}^L \frac{\mathbb{V}ar(Q_{W, h_\ell}^{CV} - Q_{W, h_{\ell-1}}^{CV})}{M_\ell} + \mathbb{E}[Q_{W, h_L}^{CV} - Q]^2.$$

The second term on the right hand side represents the bias and can further be bounded as

$$\mathbb{E}[Q_{W, h_L}^{CV} - Q]^2 = \mathbb{E}[Q_{h_L} - Q_{h_L}^\epsilon + \mathcal{Q}_W[Q_{h_L}^\epsilon] - Q]^2 \leq 2 \left( \mathbb{E}[Q_{h_L}^\epsilon] - \mathcal{Q}_W[Q_{h_L}^\epsilon] \right)^2 + 2\mathbb{E}[Q_{h_L} - Q]^2.$$

The result is obtained by observing that  $\mathbb{V}ar(Q_{W, h_\ell}^{CV} - Q_{W, h_{\ell-1}}^{CV}) = \mathbb{V}ar(Q_{h_\ell} - Q_{h_{\ell-1}} - (Q_{h_\ell}^\epsilon - Q_{h_{\ell-1}}^\epsilon))$ .  $\square$

As in the single level case, the first term of the right hand side of (4.12) represents the variance of the estimator (4.11) while the second and the third represent respectively the sparse grid quadrature and the discretization errors; observe that the statistical error is given by the sum over all levels of the variances of the double difference  $Q_{h_\ell} - Q_{h_{\ell-1}} - (Q_{h_\ell}^\epsilon - Q_{h_{\ell-1}}^\epsilon)$ : it is therefore interesting to study how the spatial and the regularization parameters  $h$  and  $\epsilon$  interact and how their effects combine; in particular we want to understand if some sort of “multiplicative effect” holds, i.e. a scaling of the type

$$|Q_h - Q - (Q_h^\epsilon - Q^\epsilon)| \lesssim h^\beta \epsilon^{\min(2, \alpha)}, \quad \forall \beta < \alpha. \quad (4.13)$$

These issues will be investigated in the next section in which we will see that indeed only a slightly weaker results than (4.13) holds.

## 4.5 Error analysis of the MLMCCV method

In this section we want to derive a rigorous bound on the statistical error related to the MLMCCV estimator introduced in the previous section. In order to derive such a result, we will need first a bound on the  $H^{1+\beta}$ -norm of the “double difference”  $p - p_h - (p^\epsilon - p_h^\epsilon)$  depending on both the mesh parameter  $h$  and the regularization parameter  $\epsilon$ .

**Lemma 4.5.1.** *Let  $a(x, \omega)$  and  $a^\epsilon(x, \omega)$  be as in Lemma 4.3.1 and  $f \in H^{r-1}(D)$  with  $r = \min(\alpha, 1)$ . By using linear finite elements for the spatial discretization, a.s. in  $\Omega$  it holds:*

$$\|p(\cdot, \omega) - p_h(\cdot, \omega) - (p^\epsilon(\cdot, \omega) - p_h^\epsilon(\cdot, \omega))\|_{H^1(D)} \lesssim C_{4.5.1}(\omega, \alpha) \|f\|_{H^{r-1}(D)} \inf_{\substack{0 \leq \beta < 1 \\ 0 < \eta + \beta \leq \alpha}} \frac{h^\beta \epsilon^{\min(\alpha - \beta - \eta, 2)}}{(\alpha - \beta)^2 \sqrt{\eta}}$$

where

$$C_{4.5.1}(\omega, \alpha) = \begin{cases} \frac{1}{a_{\min}^\epsilon} (C_{4.3.1}(\omega, \alpha) C_{2.3.1}(\omega, \alpha) + (a_{\min}^\epsilon + a_{\max}^\epsilon) C_{4.3.2}(\omega, \alpha)) & \alpha \leq 1, \\ \frac{1}{a_{\min}^\epsilon} (\tilde{C}_{4.3.1}(\omega, \alpha) C_{2.3.1}(\omega, \alpha) + (a_{\min}^\epsilon + a_{\max}^\epsilon) C_{4.3.2}(\omega, \alpha)) & \alpha > 1. \end{cases}$$

If the assumptions hold also for  $\alpha > 1$  then the following bound holds as well

$$\|p(\cdot, \omega) - p_h(\cdot, \omega) - (p^\epsilon(\cdot, \omega) - p_h^\epsilon(\cdot, \omega))\|_{H^1(D)} \lesssim C_{4.5.1}(\omega, \alpha) \|f\|_{L^2(D)} h \epsilon^{\min(\alpha-1, 2)}.$$

*Proof.* Let us consider the difference between the original problem and the auxiliary one in the continuous and in the discretized case:

$$\begin{aligned} \int_D a^\epsilon \nabla(p - p^\epsilon) \nabla v dx &= - \int_D (a - a^\epsilon) \nabla p \nabla v dx \quad \forall v \in H_0^1(D); \\ \int_D a^\epsilon \nabla(p_h - p_h^\epsilon) \nabla v_h dx &= - \int_D (a - a^\epsilon) \nabla p_h \nabla v_h dx \quad \forall v_h \in V_{h,0} \end{aligned}$$

By taking the difference between these two equations we get

$$\int_D a^\epsilon \nabla(p - p^\epsilon - p_h + p_h^\epsilon) \nabla v_h dx = - \int_D (a - a^\epsilon) \nabla(p - p_h) \nabla v_h dx \quad \forall v_h \in V_{h,0}; \quad (4.14)$$

by using this equality,  $\forall v_h \in V_{h,0}$ , we can bound the term  $\|p_h^\epsilon - p_h - v_h\|_{H_0^1(D)}$  as

$$\begin{aligned} \|p_h^\epsilon - p_h - v_h\|_{H_0^1(D)}^2 &\leq \frac{1}{a_{\min}^\epsilon} \int_D a^\epsilon \nabla(p_h^\epsilon - p_h - v_h \pm (p^\epsilon - p)) \nabla(p_h^\epsilon - p_h - v_h) dx \\ &= \frac{1}{a_{\min}^\epsilon} \left( \int_D a^\epsilon \nabla(p_h^\epsilon - p_h - (p^\epsilon - p)) \nabla(p_h^\epsilon - p_h - v_h) dx + \int_D a^\epsilon \nabla(p^\epsilon - p - v_h) \nabla(p_h^\epsilon - p_h - v_h) dx \right) \\ &= \frac{1}{a_{\min}^\epsilon} \left( - \int_D (a - a^\epsilon) \nabla(p - p_h) \nabla(p_h^\epsilon - p_h - v_h) dx + \int_D a^\epsilon \nabla(p^\epsilon - p - v_h) \nabla(p_h^\epsilon - p_h - v_h) dx \right) \\ &\leq \frac{1}{a_{\min}^\epsilon} \left( \|a - a^\epsilon\|_{L^\infty(D)} \|p - p_h\|_{H_0^1(D)} + a_{\max}^\epsilon \|p - p^\epsilon + v_h\|_{H_0^1(D)} \right) \|p_h^\epsilon - p_h - v_h\|_{H_0^1(D)}, \end{aligned}$$

so we finally get

$$\|p_h^\epsilon - p_h - v_h\|_{H_0^1(D)} \leq \frac{\|a - a^\epsilon\|_{L^\infty(D)}}{a_{\min}^\epsilon} \|p - p_h\|_{H_0^1(D)} + \frac{a_{\max}^\epsilon}{a_{\min}^\epsilon} \|p^\epsilon - p - v_h\|_{H_0^1(D)};$$



using now the triangular inequality

$$\|p^\epsilon - p - (p_h^\epsilon - p_h)\|_{H_0^1(D)} \leq \|p_h^\epsilon - p_h - v_h\|_{H_0^1(D)} + \|p^\epsilon - p - v_h\|_{H_0^1(D)} \quad \forall v_h \in V_{h,0},$$

and the arbitrariness of  $v_h$ , for any  $0 \leq \beta < r$ ,  $0 < \eta + \beta \leq \alpha$ , we obtain

$$\begin{aligned} \|p^\epsilon - p - (p_h^\epsilon - p_h)\|_{H_0^1(D)} &\leq \frac{\|a - a^\epsilon\|_{L^\infty(D)}}{a_{min}^\epsilon} \|p - p_h\|_{H_0^1(D)} + \left(1 + \frac{a_{max}^\epsilon}{a_{min}^\epsilon}\right) \inf_{v_h \in V_{h,0}} \|p - p^\epsilon - v_h\|_{H_0^1(D)} \\ &\lesssim \frac{\|a - a^\epsilon\|_{L^\infty(D)}}{a_{min}^\epsilon} \|p - p_h\|_{H_0^1(D)} + \left(1 + \frac{a_{max}^\epsilon}{a_{min}^\epsilon}\right) \|p - p^\epsilon\|_{H^{1+\beta}(D)} h^\beta \\ &\lesssim \frac{\|f\|_{H^{r-1}(D)}}{(\alpha - \beta) a_{min}^\epsilon} C_{4.3.1}(\omega, \alpha) C_{2.3.1}(\omega, \alpha) \epsilon^\alpha h^\beta + \|f\|_{H^{r-1}(D)} \left(1 + \frac{a_{max}^\epsilon}{a_{min}^\epsilon}\right) C_{4.3.2}(\omega, \alpha) \frac{h^\beta \epsilon^{\min(\alpha - \beta - \eta, 2)}}{(\alpha - \beta)^2 \sqrt{\eta}} \\ &\lesssim \frac{\|f\|_{H^{r-1}(D)}}{a_{min}^\epsilon} (C_{4.3.1}(\omega, \alpha) C_{2.3.1}(\omega, \alpha) + (a_{min}^\epsilon + a_{max}^\epsilon) C_{4.3.2}(\omega, \alpha)) \frac{h^\beta \epsilon^{\min(\alpha - \beta - \eta, 2)}}{(\alpha - \beta)^2 \sqrt{\eta}} \\ &= C_{4.5.1}(\omega, \alpha) \|f\|_{H^{r-1}(D)} \frac{h^\beta \epsilon^{\min(\alpha - \beta - \eta, 2)}}{(\alpha - \beta)^2 \sqrt{\eta}}. \end{aligned}$$

By taking the infimum over  $\eta$  we get the desired result. If the assumptions hold also for  $\alpha > 1$  then it can be analogously shown that

$$\|p(\cdot, \omega) - p_h(\cdot, \omega) - (p^\epsilon(\cdot, \omega) - p_h^\epsilon(\cdot, \omega))\|_{H^1(D)} \lesssim C_{4.5.1}(\omega, \alpha) \|f\|_{L^2(D)} h \epsilon^{\min(\alpha - 1, 2)}$$

□

The next theorem extends the previous result to a linear quantity of interest.

**Theorem 4.3.** *Let  $a(x, \omega)$  and  $a^\epsilon(x, \omega)$  be as in Lemma 4.3.1 and  $f$ ,  $Q$  and  $r = \min(\alpha, 1)$  as in Theorem 4.1. Then, by using linear finite elements for the spatial discretization, a.s. in  $\Omega$  it holds:*

$$|Q(p - p_h)(\omega) - Q(p^\epsilon - p_h^\epsilon)(\omega)| \lesssim C_{4.3}(\omega, \alpha) \|f\|_{H^{r-1}(D)} \|Q\|_{H^{r-1}(D)} \inf_{0 \leq t < r} \frac{h^t}{\alpha - t} \inf_{\substack{0 \leq \beta < r \\ 0 < \eta + \beta \leq \alpha}} \frac{h^\beta \epsilon^{\alpha - \beta - \eta}}{(\alpha - \beta)^2 \sqrt{\eta}},$$

where

$$C_{4.3}(\omega, \alpha, \beta) = \begin{cases} C_{2.3.1}(\omega, \alpha) (C_{4.3.1}(\omega, \alpha) C_{2.3.1}(\omega, \alpha) + 2a_{max}^\epsilon(\omega) C_{4.5.1}(\omega, \alpha)) & \alpha \leq 1, \\ C_{2.3.1}(\omega, \alpha) (\tilde{C}_{4.3.1}(\omega, \alpha) C_{2.3.1}(\omega, \alpha) + 2a_{max}^\epsilon(\omega) C_{4.5.1}(\omega, \alpha)) & \alpha > 1. \end{cases}$$

If the assumptions hold also for  $\alpha > 1$  then, the following bound holds as well

$$|Q(p - p_h)(\omega) - Q(p^\epsilon - p_h^\epsilon)(\omega)| \lesssim C_{4.3}(\omega, \alpha) \|f\|_{L^2(D)} \|Q\|_{L^2(D)} h^2 \epsilon^{\min(\alpha - 1, 2)}.$$

*Proof.* Let us consider the adjoint problems related to the original and the auxiliary problems having  $Q(\cdot)$  as right hand side

$$\int_D a \nabla v \nabla \Phi dx = Q(v), \quad \int_D a^\epsilon \nabla v \nabla \Phi^\epsilon dx = Q(v), \quad \forall v \in H_0^1(D)$$

where  $\Phi$  and  $\Phi^\epsilon$  are respectively the solutions of the two adjoint problems. Moreover, we denote

by  $\Phi_h$  and  $\Phi_h^\epsilon$  their respective finite element approximation. By choosing  $v = p - p_h$  in the first problem and  $v = p^\epsilon - p_h^\epsilon$  in the second problem and by taking the difference we get:

$$Q(p - p_h) - Q(p^\epsilon - p_h^\epsilon) = \int_D a \nabla(p - p_h) \nabla \Phi dx - \int_D a^\epsilon \nabla(p^\epsilon - p_h^\epsilon) \nabla \Phi^\epsilon dx.$$

Using the Galerkin orthogonality and adding and subtracting some mixed terms we get:

$$\begin{aligned} Q(p - p_h) - Q(p^\epsilon - p_h^\epsilon) &= \int_D a \nabla(p - p_h) \nabla(\Phi - \Phi_h) dx - \int_D a^\epsilon \nabla(p^\epsilon - p_h^\epsilon) \nabla(\Phi^\epsilon - \Phi_h^\epsilon) dx \\ &\quad \pm \int_D a^\epsilon \nabla(p - p_h) \nabla(\Phi - \Phi_h) dx \pm \int_D a^\epsilon \nabla(p - p_h) \nabla(\Phi^\epsilon - \Phi_h^\epsilon) dx; \end{aligned}$$

By properly grouping the terms above we obtain:

$$\begin{aligned} |Q(p - p_h) - Q(p^\epsilon - p_h^\epsilon)| &\leq \left| \int_D (a - a^\epsilon) \nabla(p - p_h) \nabla(\Phi - \Phi_h) dx \right| \\ &\quad + \left| \int_D a^\epsilon \nabla(p - p_h - (p^\epsilon - p_h^\epsilon)) \nabla(\Phi^\epsilon - \Phi_h^\epsilon) dx \right| \\ &\quad + \left| \int_D a^\epsilon \nabla(p - p_h) \nabla(\Phi - \Phi_h - (\Phi^\epsilon - \Phi_h^\epsilon)) dx \right| \\ &\leq \|a - a^\epsilon\|_{L^\infty(D)} \|p - p_h\|_{H_0^1(D)} \|\Phi - \Phi_h\|_{H_0^1(D)} \\ &\quad + \|a^\epsilon\|_{L^\infty(D)} \|p - p_h - (p^\epsilon - p_h^\epsilon)\|_{H_0^1(D)} \|\Phi^\epsilon - \Phi_h^\epsilon\|_{H_0^1(D)} + \\ &\quad + \|a^\epsilon\|_{L^\infty(D)} \|p - p_h\|_{H_0^1(D)} \|\Phi - \Phi_h - (\Phi^\epsilon - \Phi_h^\epsilon)\|_{H_0^1(D)}. \end{aligned}$$

Since for the solutions of the adjoint problems we have identical error bounds as for the the primal ones we obtain the bound

$$\begin{aligned} |Q(p - p_h) - Q(p^\epsilon - p_h^\epsilon)| &\lesssim C_{4.3.1}(\omega, \alpha) C_{2.3.1}^2(\omega, \alpha) \|f\|_{H^{\beta_1-1}(D)} \|Q\|_{H^{\beta_2-1}(D)} \frac{h^{\beta_1+\beta_2} \epsilon^{\min(\alpha, 2)}}{(\alpha - \beta_1)(\alpha - \beta_2)} \\ &\quad + a_{\max}^\epsilon(\omega) C_{4.5.1}(\omega, \alpha) C_{2.3.1}(\omega, \alpha) \|f\|_{H^{\beta_3-1}(D)} \|Q\|_{H^{\beta_4-1}(D)} \frac{h^{\beta_3+\beta_4} \epsilon^{\min(\alpha - \beta_3 - \eta_3, 2)}}{(\alpha - \beta_3)^2 (\alpha - \beta_4) \sqrt{\eta_3}} \\ &\quad + a_{\max}^\epsilon(\omega) C_{4.5.1}(\omega, \alpha) C_{2.3.1}(\omega, \alpha) \|f\|_{H^{\beta_5-1}(D)} \|Q\|_{H^{\beta_6-1}(D)} \frac{h^{\beta_5+\beta_6} \epsilon^{\min(\alpha - \beta_6 - \eta_6, 2)}}{(\alpha - \beta_5)(\alpha - \beta_6)^2 \sqrt{\eta_6}} \end{aligned}$$

where  $\beta_i, i = 1, \dots, 6$  and  $\eta_j, j = 3, 6$  are the parameters coming from the bounds of the primal and adjoint problems. Since the bound is valid  $\forall \beta_i < r$ , it is possible to choose  $\beta_1 = \beta_2 = \beta_4 = \beta_5 = t$ ; moreover the remaining part of the bound assumes its maximum value when  $\beta_3 = \beta_6 = \beta$  and  $\eta_3 = \eta_6 = \eta$ ; hence,  $\forall 0 < t, \beta < r$  and  $\eta > 0$  such that  $\beta + \eta \leq \alpha$ , the bound can be rewritten as

$$|Q(p - p_h) - Q(p^\epsilon - p_h^\epsilon)| \lesssim C_{4.3}(\omega, \alpha) \|f\|_{H^{t-1}(D)} \|Q\|_{H^{t-1}(D)} \frac{h^t}{\alpha - t} \frac{h^\beta \epsilon^{\min(\alpha - \beta - \eta, 2)}}{(\alpha - \beta)^2 \sqrt{\eta}}.$$

By taking the first infimum over  $t$  and the second infimum over  $\beta$  and  $\eta$  we obtain the desired result. If the assumptions hold also for  $\alpha > 1$  then it can be analogously shown that

$$|Q(p - p_h)(\omega) - Q(p^\epsilon - p_h^\epsilon)(\omega)| \lesssim C_{4.3}(\omega, \alpha) \|f\|_{L^2(D)} \|Q\|_{L^2(D)} h^2 \epsilon^{\min(\alpha-1, 2)}.$$

□

**Proposition 4.5.1.** *Up to logarithmic terms, the infima in the previous estimates can be bound as*

follows:

$$\inf_{0 \leq t \leq \min(\alpha, 1)} \frac{h^t}{\alpha - t} \lesssim h^{\min(\alpha, 1)} \quad \forall h \leq e^{-\frac{1}{\alpha}}$$

$$\inf_{\substack{0 \leq \beta \leq \min(\alpha, 1) \\ 0 < \eta + \beta \leq \alpha}} \frac{h^\beta \epsilon^{\min(\alpha - \beta - \eta, 2)}}{(\alpha - \beta)^2 \sqrt{\eta}} \lesssim \begin{cases} h^{\min(\alpha, 1)} \epsilon^{\min(\max(0, \alpha - 1), 2)}, & h \leq e^{-\frac{2}{\alpha}} \epsilon \\ h^{\max(\min(\alpha - 2, 1), 0)} \epsilon^{\min(\alpha, 2)}, & h \geq e^{-\frac{2}{\alpha}} \epsilon \end{cases}$$

The complete result can be found in Appendix 4.B.

**Corollary 4.5.1.** *Up to logarithmic terms, the bound in Theorem 4.3 becomes*

$$|Q(p - p_h) - Q(p^\epsilon - p_h^\epsilon)| \lesssim C_{4.3}(\omega, \alpha) \|f\|_{H^{r-1}(D)} \|Q\|_{H^{r-1}(D)} \begin{cases} h^{2\min(\alpha, 1)} \epsilon^{\min(\max(0, \alpha - 1), 2)}, & h \leq e^{-\frac{2}{\alpha}} \epsilon \\ h^{\min(\alpha, 1)} h^{\max(\min(\alpha - 2, 1), 0)} \epsilon^{\min(\alpha, 2)}, & e^{-\frac{2}{\alpha}} \epsilon \leq h \leq e^{-\frac{1}{\alpha}}. \end{cases}$$

If we look at the bound given in Theorem 4.3 and combine it with the estimates presented in Proposition 4.5.1, since  $\epsilon$  is fixed and  $h$  changes on different levels, we can observe what follows:

- for course levels, namely  $h \geq e^{-\frac{2}{\alpha}} \epsilon$ , the infimum is achieved for  $\beta, \eta$  close to 0 and  $t$  close to  $r = \min(\alpha, 1)$ ;
- for fine levels, namely  $h < e^{-\frac{2}{\alpha}} \epsilon$ , the infimum is achieved for  $\beta, t$  close to  $r$  and  $\eta$  close to 0.

This fact can be restated by saying that, in practice, there are two convergence regimes:

- the first is obtained for sufficiently large mesh sizes  $h$ : in such a regime we get a suboptimal  $h$ -convergence rate (with respect to the MLMC case); however, this effect is compensated by the fact that here the variance reduction is maximal.
- the second is obtained for sufficiently small mesh sizes  $h$ : in this regime we recover the same  $h$ -convergence rate of the standard MLMC case and, at the same time, we have a further variance reduction, given by the factor  $\epsilon^{\min(\alpha - 1, 2)}$ ; notice however that the variance reduction appears only if the input random field is sufficiently smooth ( $\alpha > 1$ ).

In general we can state that we get always an overall variance reduction with respect to the standard MLMC case and that this variance reduction affects all the possible levels if the input random field is sufficiently smooth ( $\alpha > 1$ ) and only the levels for which  $h_\ell > e^{-\frac{2}{\alpha}} \epsilon$  otherwise. In light of these considerations, as we will see in the next Section, different strategies can be adopted so as to activate the control variate only on those levels in which we actually expect some variance reduction with respect to MLMC thus avoiding waste of computational resources whenever we know that the control variate is not capable to produce any further variance reduction.

In light of these results and considerations the mean square error associated to the MLCV estimator satisfies the following

**Theorem 4.4.** *Let  $a(x, \omega)$  and  $a^\epsilon(x, \omega)$  be as in Lemma 4.3.1, and let  $f, Q$  and  $r$  be as in Theorem 4.3.*

The mean square error related to the estimator (4.11) can be bounded as

$$e(\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV})^2 \lesssim c_{4.3}^2(\alpha, 2) \|f\|_{H^{r-1}(D)}^2 \|Q\|_{H^{r-1}(D)}^2 \sum_{\ell=0}^L \frac{\left( \inf_{0 \leq t < r} \frac{h_\ell^t}{\alpha - t} \inf_{\substack{0 \leq \beta < r \\ 0 < \beta + \eta \leq \alpha}} \frac{h_\ell^\beta \epsilon^{\min(\alpha - \beta - \eta, 2)}}{(\alpha - \beta)^2 \sqrt{\eta}} \right)^2}{M_\ell} \quad (4.15)$$

$$+ 2 \left( \mathbb{E}[Q_{h_L}^\epsilon] - \mathcal{Q}_W[Q_{h_L}^\epsilon] \right)^2 + 2 \mathbb{E}[Q_{h_L} - Q]^2$$

where  $c_{4.3}(\alpha, q) = \|C_{4.3}(\omega, \alpha)\|_{L_p^q(\Omega)}$ . If the assumptions hold also for  $\alpha > 1$  then it is valid also the bound

$$e(\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV})^2 \lesssim c_{4.3}^2(\alpha, 2) \|f\|_{L^2(D)}^2 \|Q\|_{L^2(D)}^2 \sum_{\ell=0}^L \frac{h_\ell^4 \epsilon^{2 \min(\alpha - 1, 2)}}{M_\ell} \quad (4.16)$$

$$+ \left( \mathbb{E}[Q_{h_L}^\epsilon] - \mathcal{Q}_W[Q_{h_L}^\epsilon] \right)^2 + 2 \mathbb{E}[Q_{h_L} - Q]^2.$$

*Proof.* The formula of the mean square error related to the estimator (4.11) is

$$e(\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV})^2 \leq \sum_{\ell=0}^L \frac{\text{Var}(Q_{W, h_\ell}^{CV} - Q_{W, h_{\ell-1}}^{CV})}{M_\ell} + 2 \left( \mathbb{E}[Q_{h_L}^\epsilon] - \mathcal{Q}_W[Q_{h_L}^{\epsilon, SC}] \right)^2 + 2 \mathbb{E}[Q_{h_L} - Q]^2.$$

We get

$$\begin{aligned} \text{Var}(Q_{W, h_\ell}^{CV} - Q_{W, h_{\ell-1}}^{CV}) &\leq 2 \text{Var}(Q_{W, h_\ell}^{CV} - Q_W^{CV}) + 2 \text{Var}(Q_{W, h_{\ell-1}}^{CV} - Q_W^{CV}) \\ &\leq 2 \|Q - Q_{h_\ell} - (Q^\epsilon - Q_{h_\ell}^\epsilon)\|_{L_p^2(\Omega)}^2 + 2 \|Q - Q_{h_{\ell-1}} - (Q^\epsilon - Q_{h_{\ell-1}}^\epsilon)\|_{L_p^2(\Omega)}^2 \\ &\lesssim c_{4.3}^2(\alpha, 2) \|f\|_{H^{r-1}(D)}^2 \|Q\|_{H^{r-1}(D)}^2 \left( \min_{0 \leq t < r} \frac{h_\ell^t}{\alpha - t} \min_{\substack{0 \leq \beta < r \\ 0 < \beta + \eta \leq \alpha}} \frac{h_\ell^\beta \epsilon^{\alpha - \beta - \eta}}{(\alpha - \beta)^2 \sqrt{\eta}} \right)^2. \end{aligned}$$

By replacing this result in the inequality for the mean square error, we get the desired result. The result concerning the case  $\alpha > 1$  can be shown analogously.  $\square$

## 4.6 The Algorithm

In this section we present the algorithm used to guarantee a mean square error smaller than a prescribed tolerance  $tol$ . In order to be efficient, the method requires a strategy to properly select the number of levels  $L$  and the number of samples  $M_\ell$  to be taken on each level  $\ell = 0, 1, \dots, L$  as well as the number of sparse grid quadrature points  $W$  to get a sufficiently accurate approximation of the mean of the control variate  $\mathbb{E}[Q^\epsilon]$ . In the following we consider a given hierarchy of structured meshes such that  $h_{\ell-1} = 2h_\ell$  with a suitable coarsest mesh size  $h_0$ , so that we do not include the optimal mesh scaling in our optimization procedure. Moreover observe that the parameter  $\epsilon$  has been tuned and selected “ad hoc”.

Before going into the details, observe that, in light of what is stated in Theorem 4.4, we do not expect a variance reduction on the finest levels for non smooth input random fields  $a$ , i.e. for  $a \in C^\alpha(\bar{D})$  with  $\alpha < \nu < 1$ . Hence, depending on the smoothness of the input random field  $a$ , two possible strategies can be followed, namely:

- **Strategy 1:** if the input random field is smooth enough, i.e. if  $\alpha > 1$ , then our theoretical results

predict a variance reduction on each level; therefore  $\mathbb{V}\text{ar}(Q_{W,h_\ell}^{CV} - Q_{W,h_{\ell-1}}^{CV}) < \mathbb{V}\text{ar}(Q_{h_\ell} - Q_{h_{\ell-1}})$   $\forall \ell = 0, \dots, L$  and it is advantageous to apply the MLMCCV scheme as presented in (4.11), keeping the control variate on each level;

- **Strategy 2:** if the random field is rough, i.e. if  $\alpha \leq 1$ , then our theoretical results predict a variance reduction only on the coarsest levels and more precisely for all the levels such that  $h_\ell > e^{-\frac{2}{\alpha}\epsilon}$ ; therefore the most effective strategy consists in selecting as coarsest level the grid of mesh size  $h_{\ell_0}$  such that  $h_{\ell_0} \approx e^{-\frac{2}{\alpha}\epsilon}$ , keep the control variate only on this level and use a standard MLMC method on the subsequent levels; by doing this, we need to compute the approximate mean of the control variate on the coarsest mesh and not anymore on the finest one as in Strategy 1, namely:

$$\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV} = \frac{1}{M_{\ell_0}} \sum_{i=1}^{M_{\ell_0}} (Q_{h_{\ell_0}}^i - Q_{h_{\ell_0}}^{\epsilon, i}) + \sum_{\ell=\ell_0+1}^L \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} (Q_{h_\ell}^i - Q_{h_{\ell-1}}^i) + \mathcal{Q}_W[Q_{h_{\ell_0}}^\epsilon]. \quad (4.17)$$

The general algorithm used to properly choose the parameters  $\{M_\ell\}$ ,  $L$  and the number of sparse grid quadrature points  $W$  is the following:

1. We run the original Darcy problem, for different mesh sizes, by using the same small number of samples to have an estimate of the discretization error, or in other terms the weak error, from which we fit the constants  $c_w, r_w \in \mathbb{R}_+$  of the error model  $|\mathbb{E}[Q_{h_\ell} - Q]|^2 = c_w h_\ell^{r_w}$ .
2. Given a prescribed tolerance  $2tol^2$  on the mean square error, we select the finest grid having mesh size  $h_L$  in such a way to guarantee the discretization error  $2\mathbb{E}[Q_{h_\ell} - Q]^2$ , which does not depend on  $\epsilon$ , to be smaller than  $tol^2$ . This, according to (4.15), implies to choose the largest  $h_L$  such that  $c_w h_L^{r_w} \leq tol^2/2$ .
3. We set  $h_0 = \mathcal{O}(L_c)$ , being  $L_c$  the correlation length of the random field  $a$ , and evaluate, again by taking a few samples on each level,  $\mathbb{V}\text{ar}(Q_{W,h_\ell}^{CV} - Q_{W,h_{\ell-1}}^{CV})$ ,  $\mathbb{V}\text{ar}(Q_{h_\ell} - Q_{h_{\ell-1}})$  and  $\mathbb{V}\text{ar}(Q_{W,h_\ell}^{CV})$  on the levels in which, according to the selected strategy, these quantities are needed. Based on the estimate (4.15) we fit the statistical error in two different regions, namely  $\mathbb{V}\left(Q_{W,h_\ell}^{CV} - Q_{W,h_{\ell-1}}^{CV}\right) \approx c_{s_1} h_\ell^{r_{s_1}}$  for  $\ell = 0, \dots, \ell^*$  and  $\mathbb{V}\left(Q_{W,h_\ell}^{CV} - Q_{W,h_{\ell-1}}^{CV}\right) \approx c_{s_2} h_\ell^{r_{s_2}}$  for  $\ell = \ell^* + 1, \dots, L$ .
4. We run the sparse grid on an increasing number of quadrature points  $W_k$ ,  $k = 1, \dots, K$ , to estimate the (weak) sparse grid error. In particular, we fit the constants  $c, \delta \in \mathbb{R}_+$  of the error model  $|\mathbb{E}[Q^\epsilon] - \mathcal{Q}_W[Q^\epsilon]|^2 = c W^{-\delta}$ .
5. According to the selected strategy and the previously estimated convergence rates, we compute the number of samples  $M_\ell$  for  $\ell = 0, \dots, L$  and the number of quadrature points  $W$  to be used in the sparse grid approximation of the expected value of the control variate by solving an optimization problem in such a way to have, according to (4.15), the sum of the sampling error and twice the sparse grid quadrature error smaller than  $tol^2$ . This optimization procedure is described in the next subsection.
6. Once all the parameters appearing in the equations have been estimated the method can be run.

We remark that this algorithm requires a certain number of samples per level to estimate all the parameters of the different error models. These extra samples might have actually an impact on the

overall complexity of the algorithm, at least for reasonably large tolerances. A more efficient way of fitting the error models in a standard MLMC algorithm has been proposed in [28].

#### 4.6.1 Optimization Problem

Once the sequence of increasingly fine grids  $\mathcal{T}_{h_0}, \dots, \mathcal{T}_{h_L}$  has been fixed and the parameter  $\epsilon$  suitably tuned, we solve an optimization problem to find the optimal number of samples  $M_\ell$  for  $\ell = 0, \dots, L$  and the optimal number of quadrature points  $W$  forming the sparse grid upon which the expected value of the control variate is computed. The optimization problem minimizes the computational cost needed to achieve a prescribed tolerance in terms of mean square error. The computational cost needed to solve a single deterministic problem on level  $\ell$  is assumed to be of the form  $C_\ell = kh_\ell^{-d\rho}$  where  $\rho$  is a factor typically larger than 1 and smaller than 3/2 for optimal solvers. The model for the computational cost associated to the MLMCCV estimator for the two strategies is:

- strategy 1:

$$C(M_1, \dots, M_L, W) = 2M_0C_0 + 2 \sum_{\ell=1}^L M_\ell(C_\ell + C_{\ell-1}) + WC_L, \quad (4.18)$$

- strategy 2:

$$C(M_1, \dots, M_L, W) = 2M_{\ell_0}C_{\ell_0} + \sum_{\ell=\ell_0+1}^L M_\ell(C_\ell + C_{\ell-1}) + WC_{\ell_0}, \quad (4.19)$$

where the factor 2 in the above estimates comes from the fact that on each level on which the control variate is used, we have to solve both the original problem and the regularized one. The associated error, according to the selected strategy is:

- strategy 1:  $e(M_1, \dots, M_L, W)^2 = \frac{c_{s_1} \sum_{\ell=0}^{\ell^*} h_\ell^{r_{s_1}} + c_{s_2} \sum_{\ell=\ell^*+1}^L h_\ell^{r_{s_2}}}{M_\ell} + cW^{-\delta}$ , where  $\ell^* \approx -\log_2(\epsilon)$
- strategy 2:  $e(M_1, \dots, M_L, W)^2 = \sum_{\ell=\ell^*}^L \frac{c_{s_2} h_\ell^{r_{s_2}}}{M_\ell} + cW^{-\delta}$

**Remark 4.6.1.** In the estimate of the  $\mathbb{V}ar(Q_{W,h_\ell}^{CV} - Q_{W,h_{\ell-1}}^{CV})$ , according to the selected strategy, we have considered only the two extreme regimes  $\beta = 0$ , leading to  $\mathbb{V}ar(Q_{W,h_\ell}^{CV} - Q_{W,h_{\ell-1}}^{CV}) \sim \mathcal{O}(h_l^{r_{s_1}})$ , and  $\beta = \min(\alpha, 1)$ , leading to  $\mathbb{V}ar(Q_{W,h_\ell}^{CV} - Q_{W,h_{\ell-1}}^{CV}) \sim \mathcal{O}(h_l^{r_{s_2}})$ .

**Remark 4.6.2.** The parameters  $c_{s_1}$  and  $c_{s_2}$  strongly depend on the choice of  $\epsilon$ .

Once all the constants appearing in the definitions of the cost and error models have been estimated, according to the strategy selected, we perform a Lagrangian optimization by considering the Lagrange function

$$\mathcal{L}(M_1, \dots, M_L, W, \lambda) = C(M_1, \dots, M_L, W) + \lambda(e(M_1, \dots, M_L, W)^2 - tol^2).$$

Here below we report the results of such optimization procedure in the case of Strategy 1:

- $M_0 = \sqrt{\lambda} \sqrt{\frac{c_{s_1} h_0^{r_{s_1}}}{2C_0}}$

#### 4.7. Complexity of the Multi Level Monte Carlo with Control Variate method

- $M_\ell = \sqrt{\lambda} \sqrt{\frac{c_{s_1} h_\ell^{r_{s_1}}}{2(C_\ell + C_{\ell-1})}}$  for  $\ell = 1, \dots, \ell^*$
- $M_\ell = \sqrt{\lambda} \sqrt{\frac{c_{s_2} h_\ell^{r_{s_2}}}{2(C_\ell + C_{\ell-1})}}$  for  $\ell = \ell^* + 1, \dots, L$
- $W = (\lambda)^{\frac{1}{1+\delta}} \left( \delta \frac{c}{C_L} \right)^{\frac{1}{1+\delta}}$

where  $\lambda$  has to be computed from:

$$\frac{1}{\sqrt{\lambda}} = \frac{tol^2}{\sqrt{2C_0 v_0} + \sum_{\ell=1}^L \sqrt{2(C_\ell + C_{\ell-1}) v_\ell} + c \left( \frac{1}{\sqrt{\lambda}} \right)^{\frac{-1+\delta}{1+\delta}} \left( \delta \frac{c}{C_L} \right)^{\frac{-\delta}{1+\delta}}},$$

where  $v_\ell =, \ell = 0, \dots, L$ , are the fitted variances previously introduced, i.e.  $v_\ell = c_{s_1} \sum_{\ell=0}^{\ell^*} h_\ell^{r_{s_1}} + c_{s_2} \sum_{\ell=\ell^*+1}^L h_\ell^{r_{s_2}}$ .

**Remark 4.6.3.** *This problem admits an optimal solution for any  $\delta > 0$ ; nevertheless, since we are interested in using deterministic schemes able to speed up the convergence of our MLMCV estimator, our goal is to obtain a sparse grid convergence for the mean of the control variate that significantly outperforms the MC one (this restricts then to  $\delta > 1$ ). This requirement will be verified later on in Section 4.8.*

### 4.7 Complexity of the Multi Level Monte Carlo with Control Variate method

In the Section 4.6.1 we showed that, depending on the regularity of the input permeability, different algorithms are preferable. In particular, whenever we deal with smooth input fields  $a$ , our theoretical estimates predict a variance reduction on all levels, as will be verified numerically later on; on the other hand, when dealing with rough permeabilities  $a$ , our estimates predict a variance reduction only on the coarsest levels. These considerations of course influence the choice of the algorithm: in the former case we prefer to keep the control variate on each level and consider the MLMCCV estimator as in 4.11; in the latter we properly select the coarsest level  $\ell_0$ , we keep the control variate only on this level and then continue with a standard MLMC scheme, as in the estimator defined in (4.17).

In this Section we provide two complexity results for each one of these two strategies. Let us start with the result concerning the smooth case. Observe that, since we are providing an asymptotic result, we are not considering the two regimes for the contribution concerning the variance term, cf. (4.15).

**Theorem 4.5.** *Let  $c_1, c_2, c_3, c, \alpha, \eta, \beta, \gamma$  and  $\delta$  be positive constants such that  $\delta \geq \frac{2\alpha}{2\alpha-\gamma}$ ,  $2\alpha > \gamma$  and:*

- (i)  $|\mathbb{E}[Q_{h_\ell} - Q_{h_{\ell-1}}]| \leq c_1 h_\ell^\alpha,$
- (ii)  $\text{Var}(Q_{W, h_\ell}^{CV} - Q_{W, h_{\ell-1}}^{CV}) \leq c_2 \epsilon^\eta h_\ell^\beta,$
- (ii)  $C_\ell \leq c_3 h_\ell^{-\gamma},$
- (iv)  $|\mathbb{E}[Q^\epsilon] - \mathcal{Q}_W[Q^\epsilon]|^2 \leq c W^{-\delta}.$

Then, asymptotically with respect to our prescribed tolerance  $tol$ , i.e. when  $tol \rightarrow 0$ , there exists a finest level  $L$ , a sequence  $\{M_\ell\}_{\ell=0}^L$  and a number of sparse grid points  $W$  such that  $e(\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV})^2 \lesssim tol^2$  with an associated cost that scales as

$$C_{tol}(\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV}) \lesssim \epsilon^\eta \begin{cases} tol^{-2} & \text{if } \beta > \gamma, \\ tol^{-2} \log(tol)^2 & \text{if } \beta = \gamma, \end{cases} \quad (4.20)$$

where all the hidden constants do not depend on  $tol$  and depend only on  $c_1, c_2, c_3$  and  $c$ .

*Proof.* Since the result we are providing is an asymptotic one we do not track rigorously all the constants in its derivation. We assign half of the tolerance to the bias so that  $|\mathbb{E}[Q_{h_L} - Q]| \leq tol$  implies

$$L \lesssim \log_2(tol^{-\frac{1}{\alpha}});$$

To prove this result we consider  $h_\ell = 2^{-\ell}$  and use again a Lagrangian optimization by minimizing the computational cost needed to get a prescribed MSE:

$$\mathcal{L}(M_1, \dots, M_L, W, \lambda) = \sum_{\ell=0}^L M_\ell 2^{\ell\gamma} + W 2^{L\gamma} + \lambda \left( \sum_{\ell=0}^L \frac{\epsilon^\eta 2^{-\ell\beta}}{M_\ell} + c W^{-\delta} - tol^2 \right).$$

By readapting the previous results we have

$$M_\ell = \sqrt{\lambda} \sqrt{\epsilon^\eta 2^{\ell(-\gamma-\beta)}}, \text{ for } \ell = 0, \dots, L; \quad W = (\lambda)^{\frac{1}{1+\delta}} (2^{-L\gamma})^{\frac{1}{1+\delta}}.$$

Denote  $x = \frac{1}{\sqrt{\lambda}}$ . Deriving with respect to the Lagrange multiplier generates an equation for  $x$ : for suitable positive constants  $a(tol, \epsilon), b(tol)$  we get:

$$x = \frac{tol^2}{\sum_{\ell=0}^L \sqrt{2^{\ell(\gamma-\beta)} \epsilon^\eta} + x^{\frac{-1+\delta}{1+\delta}} 2^{L\frac{\delta\gamma}{1+\delta}}} = \frac{tol^2}{a(tol, \epsilon) + b(tol) x^{\frac{-1+\delta}{1+\delta}}}. \quad (4.21)$$

**Case  $\beta = \gamma$**

Observe that when  $tol \rightarrow 0$  then  $a(tol, \epsilon) \sim \sqrt{\epsilon^\eta} |\log_2(tol)|$  and  $b(tol) \sim tol^{-\frac{\gamma\delta}{\alpha(1+\delta)}}$ . Hence,  $tol \rightarrow 0$  implies  $x \rightarrow 0$ . Since  $x \rightarrow 0$ , by using  $\delta \geq \frac{2\alpha}{2\alpha-\gamma}$  we deduce that  $x$  scales as  $x \sim \frac{tol^2}{\sqrt{\epsilon^\eta} |\log_2(tol)|}$ . Replacing  $\lambda = \epsilon^\eta |\log_2(tol)|^2 tol^{-4}$  in the expression of  $M_\ell$  and  $W$ , and then evaluating the computational cost, up to constants term we get:

$$\begin{aligned} C_{tol}(\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV}) &\approx \sum_{\ell=0}^L tol^{-2} \epsilon^\eta |\log_2(tol)| + \epsilon^{\frac{\eta}{1+\delta}} |\log_2(tol)|^{\frac{1}{1+\delta}} tol^{-\frac{4}{1+\delta}} tol^{\frac{\gamma}{\alpha(1+\delta)}} tol^{-\frac{\gamma}{\alpha}} \\ &\approx tol^{-2} \epsilon^\eta |\log_2(tol)|^2 + \epsilon^{\frac{\eta}{1+\delta}} |\log_2(tol)|^{\frac{1}{1+\delta}} tol^{-\frac{1}{\delta+1} (4 + \frac{\gamma}{\alpha} \delta)}. \end{aligned} \quad (4.22)$$

By comparing the two terms we have that, under the assumption  $\delta \geq \frac{2\alpha}{2\alpha-\gamma}$ , the first term of the above right hand side dominates the second when  $tol$  tends to 0, so that the desired bound in the case  $\beta = \gamma$  is shown.



#### 4.7. Complexity of the Multi Level Monte Carlo with Control Variate method

**Case  $\beta > \gamma$**

By looking at (4.21) we can see how the term  $a(tol, \epsilon)$  is bounded for any  $L$  while  $b(tol)$  is as before; when  $tol \rightarrow 0$   $x$  again tends to 0 but now it scales as  $x = \frac{tol^2}{\sqrt{\epsilon^\eta}}$ ; replacing  $\lambda = \epsilon^\eta tol^{-4}$  in the expression of  $M_\ell$  and  $W$ , and then evaluating the computational cost, up to constants term we get:

$$\begin{aligned} C_{tol}(\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV}) &\approx \sum_{\ell=0}^L tol^{-2} \epsilon^\eta 2^{\frac{\ell(\gamma-\beta)}{2}} + \epsilon^{\frac{\eta}{1+\delta}} tol^{-\frac{4}{1+\delta}} tol^{\frac{\gamma}{\alpha(1+\delta)}} tol^{-\frac{\gamma}{\alpha}} \\ &\approx tol^{-2} \epsilon^\eta |\log_2(tol)| + \epsilon^{\frac{\eta}{1+\delta}} tol^{-\frac{1}{\delta+1}(4+\frac{\gamma}{\alpha}\delta)}. \end{aligned} \quad (4.23)$$

With analogous computations and considerations it is easy to prove the bound also in this case.  $\square$

**Remark 4.7.1.** For the moment we have been able to prove an analogous result also in the case  $\beta < \gamma$  only under a very restrictive condition that in many cases would not hold. However, since Theorem 4.6 refers to the case of smooth input permeabilities, the condition  $\beta \geq \gamma$  is verified.

In the rough case where the control variate is applied only on level  $\ell_0$ , we have

**Theorem 4.6.** Let  $\ell_0$  be such that  $\mathbb{V}ar(Q_{W, h_{\ell_0}}^{CV}) \approx \mathbb{V}ar(Q_{h_{\ell_0}} - Q_{h_{\ell_0-1}})$ . Let  $c_1, c_2, c_3, c, \alpha, \eta, \beta, \gamma$  and  $\delta$  be positive constant such that:

- (i)  $|\mathbb{E}[Q_{h_\ell} - Q_{h_{\ell-1}}]| \leq c_1 h_\ell^\alpha,$
- (ii)  $\mathbb{V}ar(Q_{h_\ell} - Q_{h_{\ell-1}}) \leq c_2 h_\ell^\beta,$
- (ii)  $C_\ell \leq c_3 h_\ell^{-\gamma},$
- (iv)  $|\mathbb{E}[Q^\epsilon] - \mathcal{Q}_W[Q^\epsilon]|^2 \leq c W^{-\delta}.$

Then, asymptotically with respect to our prescribed tolerance  $tol$ , i.e. when  $tol \rightarrow 0$ , by using the Strategy 2 detailed in Sections 4.6 and 4.6.1, when  $\beta > \gamma$ , there exists a finest level  $L$ , a sequence  $\{M_\ell\}_{\ell=0}^L$  and a number of sparse grid points  $W$  such that  $e(\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV})^2 \lesssim tol^2$  and its associated cost  $C_{tol}(\hat{Q}_{\{h_\ell\}_{\ell=\ell_0}^L, \{M_\ell\}_{\ell=\ell_0}^L, W}^{MLMCCV})$  is asymptotically proportional to a fraction of the cost  $C_{tol}(\hat{Q}_{\{h_\ell\}_{\ell=0}^L, \{M_\ell\}_{\ell=0}^L, W}^{MLMC})$  of the corresponding MLMC estimator, namely:

$$C_{tol}(\hat{Q}_{\{h_\ell\}_{\ell=\ell_0}^L, \{M_\ell\}_{\ell=\ell_0}^L, W}^{MLMCCV}) \sim 2^{\ell_0(\gamma-\beta)} C_{tol}(\hat{Q}_{\{h_\ell\}_{\ell=0}^L, \{M_\ell\}_{\ell=0}^L, W}^{MLMC}). \quad (4.24)$$

*Proof.* We assign again half of the tolerance to the bias so that  $|\mathbb{E}[Q_{h_L} - Q]| \leq tol$  implies

$$L \lesssim \log_2(tol^{-\frac{1}{\alpha}});$$

The Lagrangian optimization in this case becomes:

$$\mathcal{L}(M_{\ell_0}, \dots, M_L, W, \lambda) = \sum_{\ell=\ell_0}^L M_\ell 2^{\ell\gamma} + W 2^{\ell_0\gamma} + \lambda \left( \sum_{\ell=\ell_0}^L \frac{2^{-\ell\beta}}{M_\ell} + c W^{-\delta} - tol^2 \right).$$

By readapting the previous results we have

$$M_\ell = \sqrt{\lambda} \sqrt{2^{\ell(-\gamma-\beta)}}, \text{ for } \ell = \ell_0, \dots, L; \quad W = (\lambda)^{\frac{1}{1+\delta}} \left( 2^{-\ell_0\gamma} \right)^{\frac{1}{1+\delta}}.$$

Denote again  $x = \frac{1}{\sqrt{\lambda}}$ . Deriving with respect to the Lagrange multiplier generates an equation for  $x$ : for suitable positive constants  $a(tol, \ell_0), b(\ell_0)$  we get:

$$x = \frac{tol^2}{\sum_{\ell=\ell_0}^L \sqrt{2^{\ell(\gamma-\beta)}} + x^{\frac{-1+\delta}{1+\delta}} 2^{\ell_0 \frac{\delta\gamma}{1+\delta}}} = \frac{tol^2}{a(tol) + bx^{\frac{-1+\delta}{1+\delta}}}. \quad (4.25)$$

When  $tol \rightarrow 0$   $x$  again tends to 0 scaling as  $x = \frac{tol^2}{\sum_{\ell=\ell_0}^L \sqrt{2^{\ell(\gamma-\beta)}}}$ ; replacing  $\lambda = tol^{-4} \left( \sum_{\ell=\ell_0}^L \sqrt{2^{\ell(\gamma-\beta)}} \right)^2$  in the expression of  $M_\ell$  and  $W$ , and then evaluating the computational cost we get:

$$\begin{aligned} C_{tol}(\hat{Q}_{\{h_\ell\}_{\ell=\ell_0}^L, \{M_\ell\}_{\ell=\ell_0}^L, W}^{MLMCCV}) &\approx \sum_{\ell=\ell_0}^L 2^{\frac{\ell(\gamma-\beta)}{2}} \sum_{\ell=\ell_0}^L tol^{-2} 2^{\frac{\ell(\gamma-\beta)}{2}} + tol^{-\frac{4}{1+\delta}} + \left( \sum_{\ell=\ell_0}^L 2^{\frac{\ell(\gamma-\beta)}{2}} \right)^{\frac{2}{1+\delta}} \\ &\approx \left( \sum_{\ell=\ell_0}^L 2^{\frac{\ell(\gamma-\beta)}{2}} \right)^2 tol^{-2}. \end{aligned} \quad (4.26)$$

Since asymptotically the MLMC estimator has cost given by

$$C_{tol}(\hat{Q}_{\{h_\ell\}_{\ell=0}^L, \{M_\ell\}_{\ell=0}^L}^{MLMC}) \approx \left( \sum_{\ell=0}^L 2^{\frac{\ell(\gamma-\beta)}{2}} \right)^2 tol^{-2},$$

by taking the ratio of the two we conclude by saying that the control variate generates an asymptotic gain of a factor  $2^{\ell_0(\gamma-\beta)}$ .  $\square$

**Remark 4.7.2.** *The previous Theorems provided asymptotic results that actually give rather pessimistic bounds. Indeed, in the smooth case ( $\nu \geq 1$ ), by only looking at the asymptotic behavior of the variances, we completely ignore the additional gain given by the control variate on the coarsest levels. Moreover, we predict an asymptotic gain with respect to the standard MLMC complexity results only in the case  $\beta \geq \gamma$ . However, notice as we want to keep the control variate on each level only when we have a smooth random field as input; this automatically implies that the value  $\beta$  should be large enough to include a wide class of problems. In the rough case we have not been able to prove asymptotic gain with respect to the standard MLMC in the case  $\gamma \geq \beta$ . Nevertheless, numerical experiments show considerable improvements also in these cases as we will see later on.*

## 4.8 Numerical Results

In this section we present some numerical results obtained using the proposed control variate approach. We focus on results concerning the single-level MCCV and the MLMCCV approach. Consistently, in the first case we compare the results with the ones obtained with a standard MC scheme, while in the second one the comparison is done with the classical MLMC scheme [25]. We recall that all the results presented hereafter are obtained in the case of a lognormally distributed random field  $a(x, \omega)$ . We assume the log-permeability  $\log(a(x, \omega))$  to have a covariance function belonging to the Matérn family. Since there is not a remarkable difference between the numerical results obtained starting from a tensor and a radial covariance function, we will consider in this section only isotropic covariances of the form (2.5).

To compute the MC part of our control variate-based estimators (4.4) and (4.11) we perform the sampling from the random field  $a(\mathbf{x}, \omega)$  with FFT, as described in Subsection 2.5 (similar to a circulant embedding [31, 54, 55]), e.g. by expanding the field in a Fourier basis. More precisely, to

avoid boundary effects that could occur when replicating periodically the covariance function, we build the Fourier expansion on a sufficiently large domain, and then restrict the generated field to our physical domain  $D$ . In this way we sample by FFT the original random field  $a$  on a sufficiently fine grid so that the truncation error is negligible.

On the other hand, in order to compute the expected value of the control variate via sparse grid approximation, we consider a KL expansion of the input random field in order to keep the number of random variables in the expansion as small as possible. In the previous chapter we deeply investigated the performances of such sparse grid approximations and showed how they perform well in presence of smooth input permeabilities. Here we apply such strategies to the smoothed problem with regularized input field  $a^\epsilon$  which is  $\mathcal{C}^\infty(\bar{D})$  for any  $\epsilon > 0$  so we expect again good performances. In all the numerical tests we used a quasi-optimal scheme (see Subsection 3.4.1) to take care of the sparse grid part of the “control variate”-based algorithms. More details will be provided hereafter.

We show some results obtained in the 2D case. We consider the physical domain  $D$  to be  $(0, 1)^2$  and we are interested in the simulation of an undisturbed flow from left to right which is represented by the equations

$$\begin{cases} -\operatorname{div}(a(x, y, \omega) \nabla p(x, y, \omega)) = 0 & \text{in } D, \text{ a.s. in } \Omega, \\ p(0, y, \omega) = 1, \quad p(1, y, \omega) = 0, & y \in [0, 1], \\ p_y(x, 0, \omega) = p_y(x, 1, \omega) = 0, & x \in (0, 1). \end{cases} \quad (4.27)$$

In all the simulations presented hereafter we have considered a log-permeability field with correlation length  $L_c = 0.5$  and variance  $\sigma^2 = 1$ . Also, the level  $\ell$  in the MLMCCV and MLMC approaches is chosen as  $h_\ell = 2^{-\ell-1}$ ,  $\ell \geq 0$ .

### Case 1: a smooth functional

We start focusing on the approximation of the following smooth functional  $Q^1 \in L^2(D)$ :

$$Q^1(u) = \int_D \kappa(x) u(x) dx, \quad \text{with } u \in H_{\Gamma_D}^1(D), \quad \kappa(x) = \frac{e^{-\frac{(x_1 - \mu_Q)^2 + (x_2 - \mu_Q)^2}{2\sigma_Q^2}}}{2\pi\sigma_Q^2}, \quad \mu_Q = \frac{1}{4}, \quad \sigma_Q = \frac{1}{8},$$

and look at the quantity of interest  $Q^1(\omega) = Q^1(p(\cdot, \omega))$ . Figure 4.2 shows the variance of  $Q_\ell^1$  and the variance of the difference  $Q_\ell^1 - Q_{\ell-1}^1$  for the standard MLMC approach as well as  $Q_\ell^1 - (Q^1)_{\ell-1}^\epsilon$  and  $Q_\ell^1 - (Q^1)_{\ell-1}^\epsilon - (Q_\ell^1 - (Q^1)_{\ell-1}^\epsilon)$ , in the case of a rough field ( $\nu = 1/2$ ). As predicted by the theory, the difference between the MCCV and MC variances becomes larger and larger when considering smaller values of  $\epsilon$ ; at the same time, when looking at differences between consecutive levels, the variance reduction in MLMCCV appears only on the coarsest levels and, starting from a certain level  $\ell^* \approx -\log_2 \epsilon$ , the MLMC and MLMCCV variances are almost identical.

The second comparison, shown in Figure 4.3, demonstrates that, by considering a smoother random field as input, the variance reduction with respect to the standard MLMC case appears on each level, and, again, the smaller  $\epsilon$  the more significant the variance reduction.

In Figure 4.4 and Figure 4.5 we show convergence plots for the two other sources of errors, namely

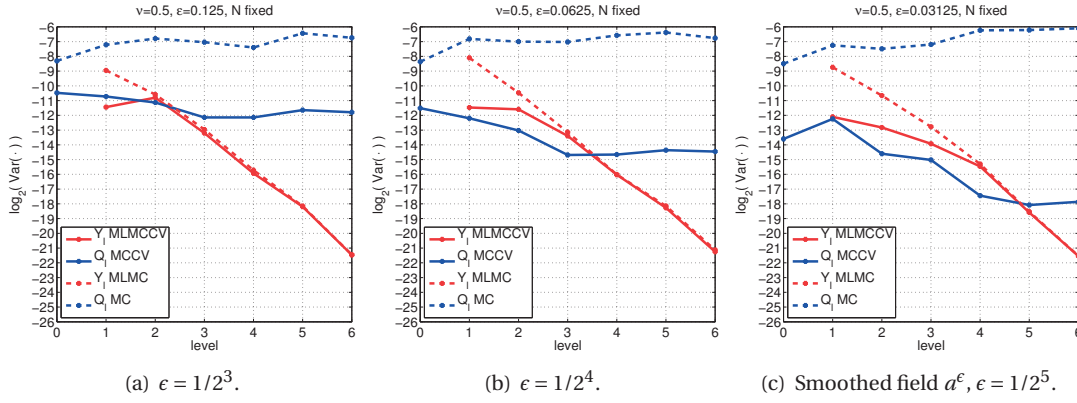


Figure 4.2 – Variance of  $Q_\ell^1(\omega) = \int_D \kappa(x) p_\ell(x, \omega) dx$  and  $Y_\ell^1(\omega) = Q_\ell^1(\omega) - Q_{\ell-1}^1(\omega)$  on each level, for  $v = 0.5$  and different values of  $\epsilon$ .

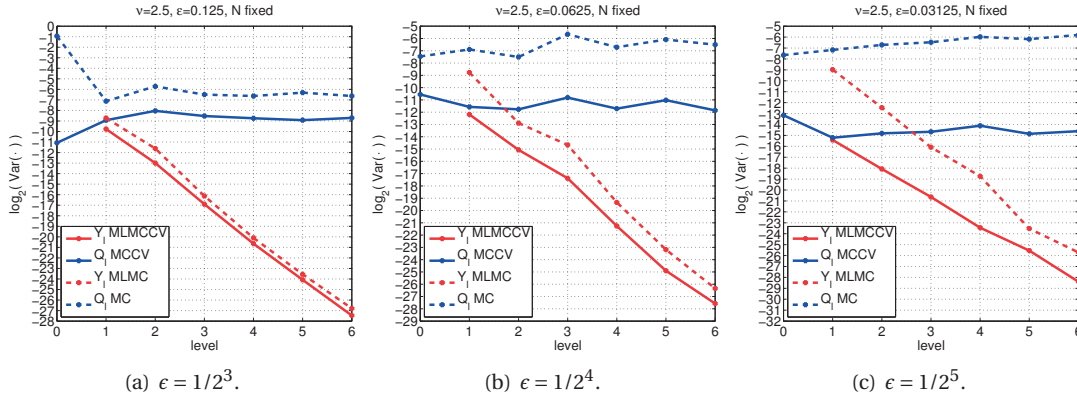


Figure 4.3 – Variance of  $Q_\ell^1(\omega) = \int_D \kappa(x) p_\ell(x, \omega) dx$  and  $Y_\ell^1(\omega) = Q_\ell^1(\omega) - Q_{\ell-1}^1(\omega)$  on each level, for  $v = 2.5$  and different values of  $\epsilon$ .

the Stochastic Collocation error committed when approximating the mean of the control variate by sparse grid, and the error coming from the spatial discretization, which determines the finest mesh of our sequence of meshes.

In Figure 4.4 we have the convergence of the sparse grid approximation of the mean of the control variate  $(Q^1)_{h_L}^\epsilon$ ; we compute these errors for a sequence of sparse grid quadratures using non nested Hermite knots and having increasing number of points  $W$ . Such values of solves correspond to different profit levels  $w_k$ ,  $k = 1, \dots, K$ , obtained by considering the quasi optimal sparse grid built starting from the multi-indices  $\mathbf{i}$  such that  $P(\mathbf{i}) > e^{-w_k}$ ; as reference solution we used the solution corresponding to  $w^* = w_{K+1}$ ; here we used  $K = 5$  for the cases  $v = 0.5$  and  $v = 2.5$ . The labels appearing next to each point contain the number of “active variables” used to obtain that particular approximation and, in brackets, the number of variables that have been activated at the same time. These numbers somehow give us an idea of the shape of the underlying multi-index set  $\mathbf{I}$  used to compute the sparse grid approximation of the mean of the control variate and, most importantly, tell us that these errors can be considered *dimension*-independent; in fact, we built our underlying  $N$ -dimensional sparse grid by considering  $N = 800$  variables in both cases  $v = 0.5$  and  $v = 2.5$ ;

## 4.8. Numerical Results

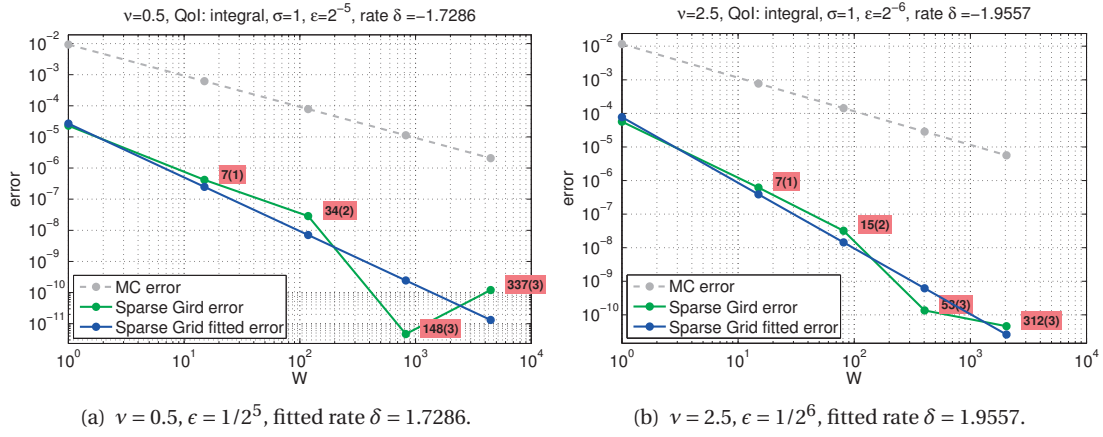


Figure 4.4 – SC error computed as  $\left(\mathbb{E}[(Q^1)_{h_L}^\epsilon] - \mathcal{Q}_W[(Q^1)_{h_L}^\epsilon]\right)^2$  versus computational cost, computed as number of sparse grid knots  $W$  for  $v = 0.5$  (left) and  $v = 2.5$  (right).

from the labels in Figure 4.4 we can see that the number of active variables included in the sparse grid approximations at all levels  $k = 1, \dots, K$  is always significantly smaller than  $N = 800$ . The same holds for the reference solution, built respectively over a sparse grid of approximately 23000 and 10000 nodes, presenting only 617 (respectively 539) active variables. In light of these considerations we consider the truncation error coming from such approximation negligible, at least up to the tolerances reached in Figure 4.4. The convergence obtained, for the square error, is significantly faster than the one of a standard MC scheme. Moreover, the fitted algebraic error model  $cW^{-\delta}$ , describes well the actual decay of the error.

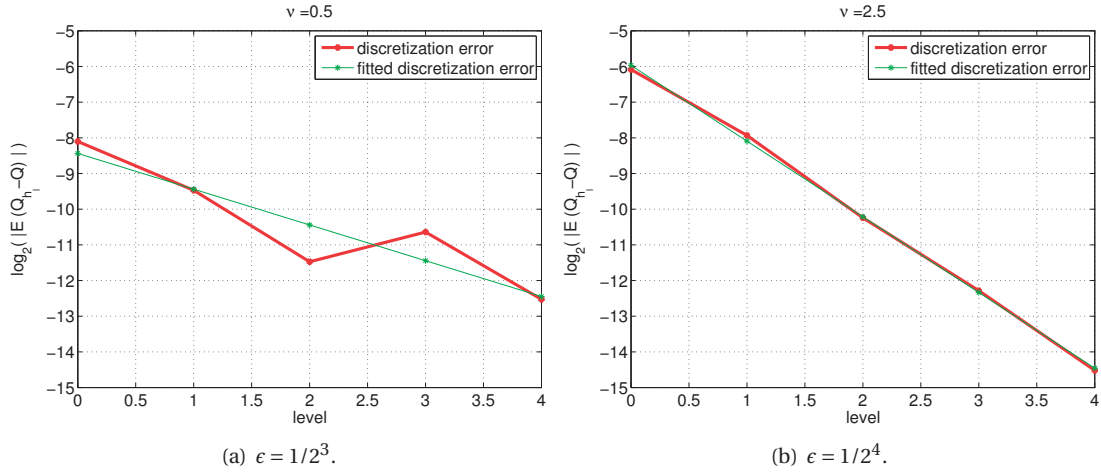


Figure 4.5 – Discretization error computed as  $\mathbb{E}[Q_{h_\ell}^1 - Q_{h_L}^1]^2$  for  $v = 0.5$  (left) and  $v = 2.5$  (right).

Concerning the spatial discretization error we see in Figure 4.5 that such error presents different slopes depending on the smoothness of the input random field: in particular when  $a$  is rough (in this case  $v = 0.5$ ) the convergence rate is close to 1; on the other hand, when the random field  $a$  has smooth realizations (in this case  $v = 2.5$  so the realizations are twice differentiable) the convergence rate is close to 2.

To conclude the presentation of the numerical results we show in Figure 4.6 a plot of the overall error; we compare here all at once the performances of the MCCV and MLMCCV methods with the corresponding MC and MLMC schemes. Observe that the finite element error appearing in the mean square error associated to these estimators is the same (up to a constant, set to 2 in our particular test); then, in order to make the comparison between the methods with and without control variate as clear as possible, we select the finest mesh size  $h_L$  in such a way to have the two finite element errors comparable, i.e. a half of the prescribed tolerance  $2tol^2$ . Then we compare the other terms coming from the mean square error associated to the MLMC and MLMCCV (respectively MC and MCCV) estimators, by forcing them to be smaller than  $tol^2$ . Of course, other choices are possible concerning the algorithm (see e.g. [28, 48]) which will not be discussed here. The error obtained with the control variate approach after having optimized all the parameters, given as sum of the sparse grid error and the statistical error, is plotted against the computational cost model, measured as estimated CPU time, according to the computational cost model defined in (4.18) and (4.19). Figure 4.6 shows a remarkable improvement in terms of prescribed tolerance on the mean square error versus corresponding computational cost. Indeed the control variate approach improves the performance of the classical schemes. Let us start from the “single-level” schemes: we can see how the MCCV sampler presents a remarkable gain with respect to standard MC due to the variance reduction induced by the control variate which is roughly speaking  $O(e^{\min(v,2)})$ . According with the statement of Theorem 4.5 the MCCV rate degenerates as the MC one (Theorem applies with  $\beta = 0$ ) when moving towards small tolerances; this is why, although for small tolerances (and hence coarse levels  $\ell$ ) the two schemes converge as  $tol^{-2}$ , starting from a certain point they recover the asymptotic behavior  $tol^{-2-\frac{\gamma}{\alpha}}$ . Asymptotically the MCCV scheme is able to achieve the same tolerances of the MC scheme with a saving in terms of computational cost of approximatively 2 order of magnitude in the rough case ( $v = 0.5$ ) and 3 in the smooth case ( $v = 2.5$ ). Concerning the two “multi-level” schemes, namely MLMCCV and MLMC, we do not expect to obtain, at least asymptotically, such a big difference; in fact, according to our theoretical estimates, the double differences  $Q_\ell - Q_\ell^e - (Q_{\ell-1} - Q_{\ell-1}^e)$  scales as predicted in Lemma 4.3 and therefore we obtain a less significant variance reduction which induces a less significant computational saving. However, according to the theory of MLMC (see for instance [25]), the variances of QoI computed on consecutive meshes decay sufficiently fast to make the overall computational cost proportional to the cost of a MC sampler on the coarsest level considered, i.e. proportional to  $tol^{-2}$ . In this case to achieve the same accuracy of the MLMC method, roughly speaking, the proposed MLMCCV method presents a gain in terms of computational time of about 1 order of magnitude in the rough case ( $v = 0.5$ ) and 2 in the smooth case ( $v = 2.5$ ); despite these factors are smaller than the ones in the “single-level”, the asymptotic benefit of a “multi-level” strategy appears evident. Notice that for low tolerances there is no significant difference between the performance of MLMCCV and MCCV because this particular QoI  $Q^1$  has a small variance.

#### Case 2: a less smooth functional

Now we focus on a less smooth functional  $Q^2$  defined as

$$Q^2(u, \omega) = - \int_D a(x, \omega) \nabla u(x) \cdot \nabla v(x) dx, \quad \text{with } u \in \mathcal{W}; \quad (4.28)$$

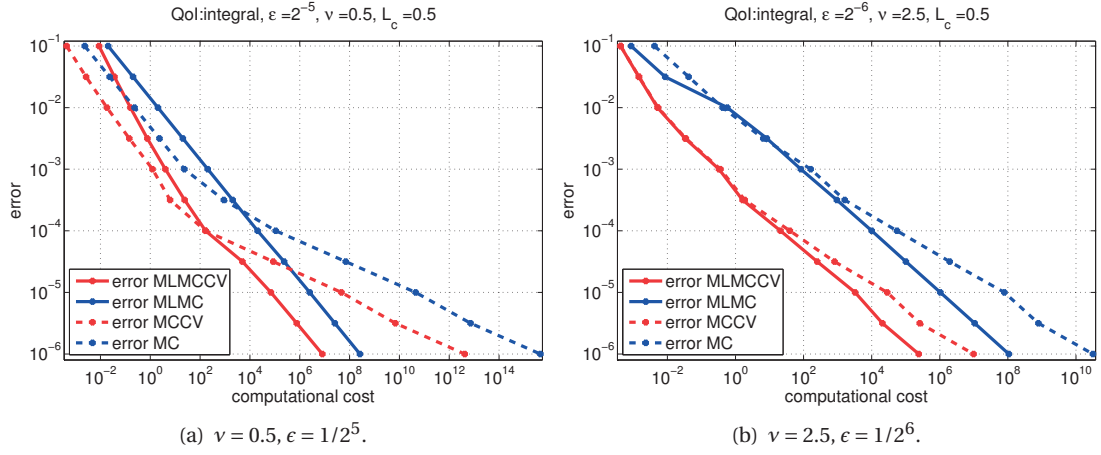


Figure 4.6 – Error vs Computational cost for  $\nu = 0.5$  (left) and  $\nu = 2.5$  (right). Error = Sparse Grid error + Statistical error, computed as root mean square error. Computational cost estimated according to (4.18) and (4.19), with  $C_\ell = h_\ell^{-d\rho}$  and  $\rho = 1.05$ , fitted by using a Matlab backslash command.

being  $\mathcal{W}$  a suitable functional space and  $\nu = 1$  on  $\Gamma_{out} \equiv \{y \in (0, 1), x = 1\}$  and  $\nu = 0$  on  $\Gamma_{in} \equiv \{y \in (0, 1), x = 0\}$ . Again we look at the quantity of interest  $Q^2(\omega) = Q^2(p(\cdot, \omega), \omega)$  that represents, in a distributional sense, the mass flux through the outflow boundary. Indeed, for smooth  $a$  and  $p$  satisfying problem (4.27), we have

$$Q^2(\omega) = - \int_{\Gamma_{out}} a(x, \omega) \nabla p(x, \omega) \cdot \mathbf{n} dx. \quad (4.29)$$

In the case of a fully Dirichlet problem with  $a \in C^\alpha(\overline{D})$ , by choosing  $\nu \approx 1$  on the boundary  $\Gamma_{out}$  and  $\nu \approx 0$  on  $\partial D \setminus \Gamma_{out}$ , this functional has been proven in [74] to be an  $H^{\beta-1}(D)$  approximation of the outflow through the boundary  $\Gamma_{out}$  for any  $0 < \beta < \alpha$ .

In the case of a problem with mixed boundary conditions as in (4.27), the analysis becomes slightly more involved; a similar result can be recovered by considering the Bochner space  $L^2((0, 1); H_0^s(0, 1))$  (in short  $L^2(H_0^s)$ ) as the space of the  $H_0^s$ -valued functions along the horizontal direction that are  $L^2$  integrable along the vertical direction, being  $H_0^s$  the closure of  $\mathcal{C}_0^\infty$  in the  $H^s$  norm. This space is equipped with the norm

$$\|v\|_{L^2(H^s)} = \left( \int_0^1 \left( \int_{(0,1)^2} \frac{(v(x_1, x_2) - v(x'_1, x_2))^2}{|x_1 - x'_1|^{1+2s}} dx_1 dx'_1 \right) dx_2 \right)^{\frac{1}{2}};$$

Notice that  $\mathcal{C}_0^\infty(D)$  is dense in this space. By choosing for instance  $v(x_1, x_2) = x_1$  in (4.28) then we obtain

$$Q^2(\omega) = - \int_{(0,1)^2} a(x, \omega) \frac{\partial p(x, \omega)}{\partial x_1} dx, \quad (4.30)$$

and therefore, for any  $p \in C_0^\infty(D)$ , it holds

$$Q^2 = - \int_D a \frac{\partial p}{\partial x_1} dx \lesssim \left\| \frac{\partial a}{\partial x_1} \right\|_{L^2(H^{\beta-1})} \|p\|_{L^2(H^{1-\beta})} \lesssim \|a\|_{L^2(H^\beta)} \|p\|_{L^2(H^{1-\beta})} \lesssim \frac{\|a\|_{L^2(C^\alpha)}}{\sqrt{\alpha - \beta}} \|p\|_{L^2(H^{1-\beta})}$$



and the functional can be extended continuously on the whole space  $L^2(H_0^{1-\beta})$ . Therefore,  $Q^2$  is an  $L^2(H^{\beta-1})$  functional  $\forall 0 < \beta < \alpha$ .

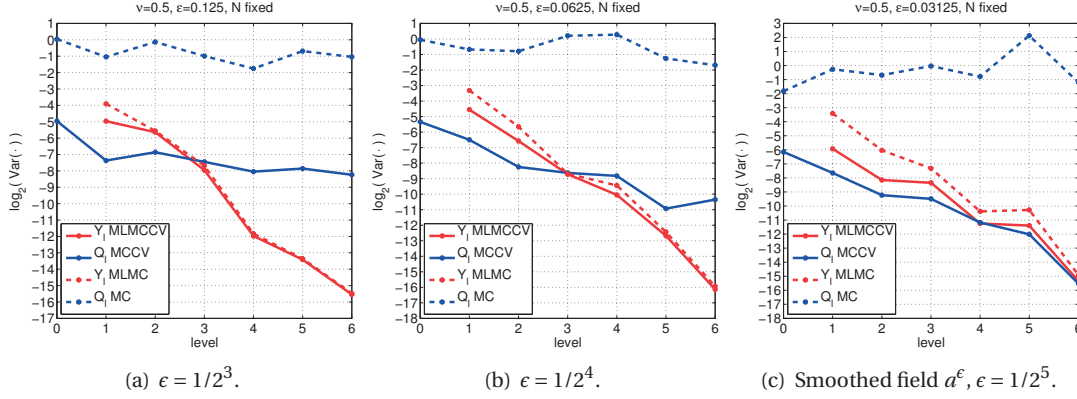


Figure 4.7 – Variance of  $Q_\ell^2(\omega) = -\int_D a(x, \omega) \frac{\partial p_\ell}{\partial x_1}(x, \omega) dx$  and of  $Y_\ell^2 = Q_\ell^2 - Q_{\ell-1}^2$ , for  $\nu = 0.5$  and different  $\epsilon$ .

In Figure 4.7 and 4.8 we show the variance reduction obtained in this case. Again in the rough case  $\nu = 0.5$  we obtain a variance reduction only on the coarsest levels; on the other hand, for the smooth case  $\nu = 2.5$ , although we obtain a more significant variance reduction with respect to the rough case, it seems that this reduction does not appear for all levels; therefore, from a numerical point of view, it is preferable to keep the control variate only on the first levels  $\ell = 0, \dots, \ell^*$  and then continue with a standard MLMC scheme. From a theoretical point of view for the moment we are not able to explain why there is no variance reduction on all levels; our numerical tests show that computing the flux as a boundary integral as in (4.29) leads to suboptimal convergence rates with respect to the discretization mesh size  $h$ , but produces variance reduction on all levels. On the other hand computing the flux as an integral on the whole domain  $D$  as in (4.30) produces optimal convergence rates in  $h$  but variance reduction only on the coarsest levels. This issue is still under investigation. In Figures 4.9 and 4.10 we show the errors coming from the approximation of the mean of the control

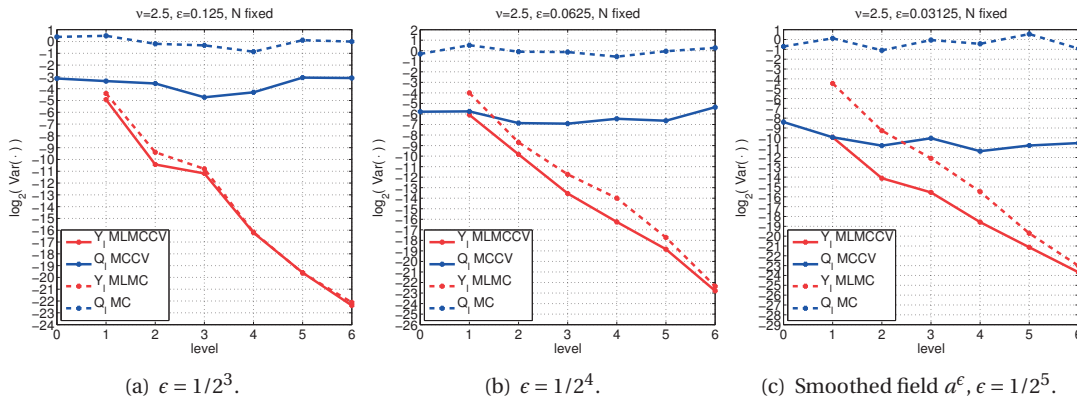


Figure 4.8 – Variance of  $Q_\ell^2(\omega) = -\int_D a(x, \omega) \frac{\partial p_\ell}{\partial x_1}(x, \omega) dx$  and of  $Y_\ell^2 = Q_\ell^2 - Q_{\ell-1}^2$ , for  $\nu = 2.5$  and different  $\epsilon$ .

variate and from the physical discretization on a grid of mesh size  $h_l$ . Analogous considerations to the ones made in the case of the smooth functional can be done.



To conclude we show again in Figure 4.11 the overall error given by the sum of statistical and sparse

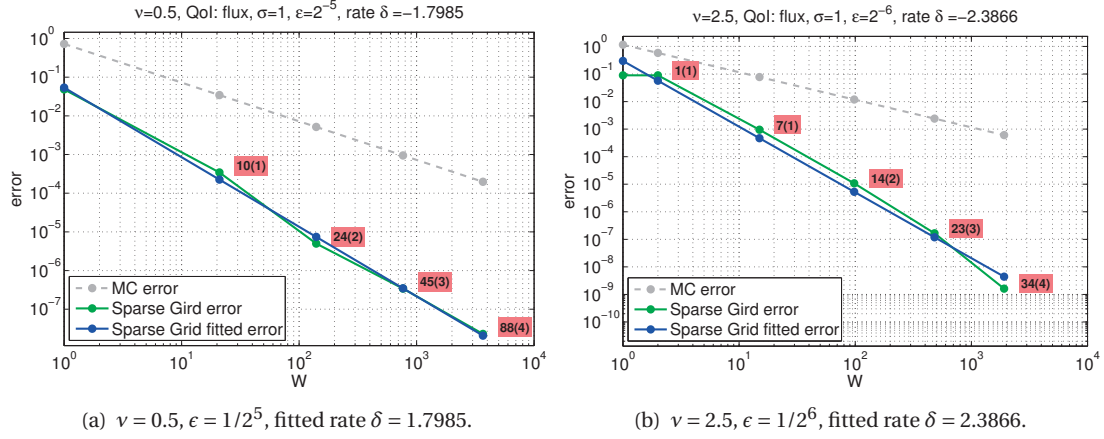


Figure 4.9 – SC error computed as  $\left(\mathbb{E}[(Q^2)_{h_L}^\epsilon] - \mathcal{Q}_W[(Q^2)_{h_L}^\epsilon]\right)^2$  versus computational cost, computed as number of sparse grid knots  $W$  for  $v = 0.5$  (left) and  $v = 2.5$  (right).

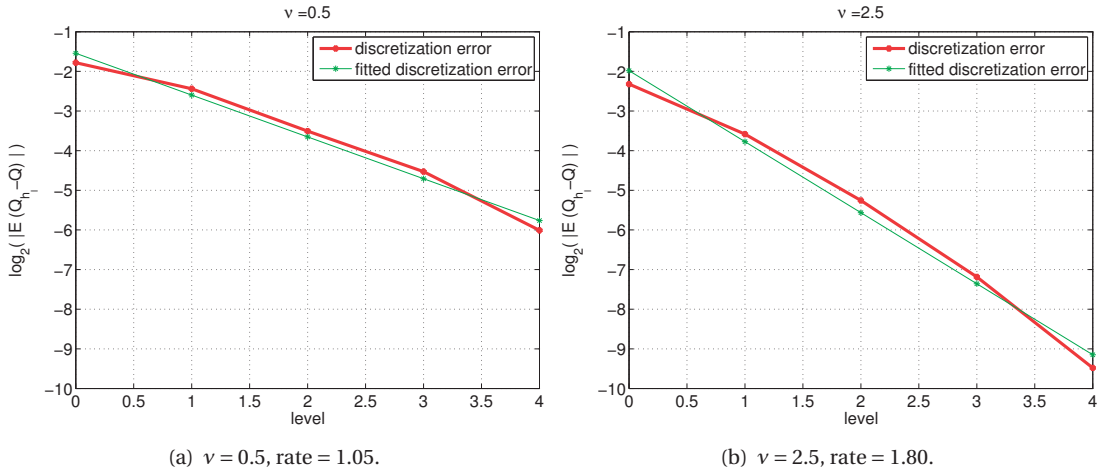


Figure 4.10 – Discretization error computed as  $\mathbb{E}[Q_{h_c}^1 - Q_{h_L}^1]^2$  for  $v = 0.5$  (left) and  $v = 2.5$  (right).

grid error; in this case the gain with respect to the standard MLMC scheme is slightly smaller than the one previously shown in Figure 1.2; again asymptotically the difference between MCCV and MC, around 3 and 4 orders of magnitudes respectively in the rough ( $v = 0.5$ ) and smooth case ( $v = 2.5$ ), is more remarkable than the one between MLMCCV and MLMC, around 1 and 2 orders of magnitudes respectively in the rough ( $v = 0.5$ ) and smooth case ( $v = 2.5$ ); however the “multi-level” schemes present a much better rate, which is again  $tol^{-2}$ ; observe how in this case the separation between the “single-level” and the “multi-level” schemes occur earlier than in the previous case since the flux functional is affected by a larger variance than the previous one.

#### 4.8.1 A level-dependent approach

As briefly anticipated in section 2.5, since we use the tools coming from the FFT to generate the realizations of the underlying permeability  $a$  on a Fourier basis, the number of random variables

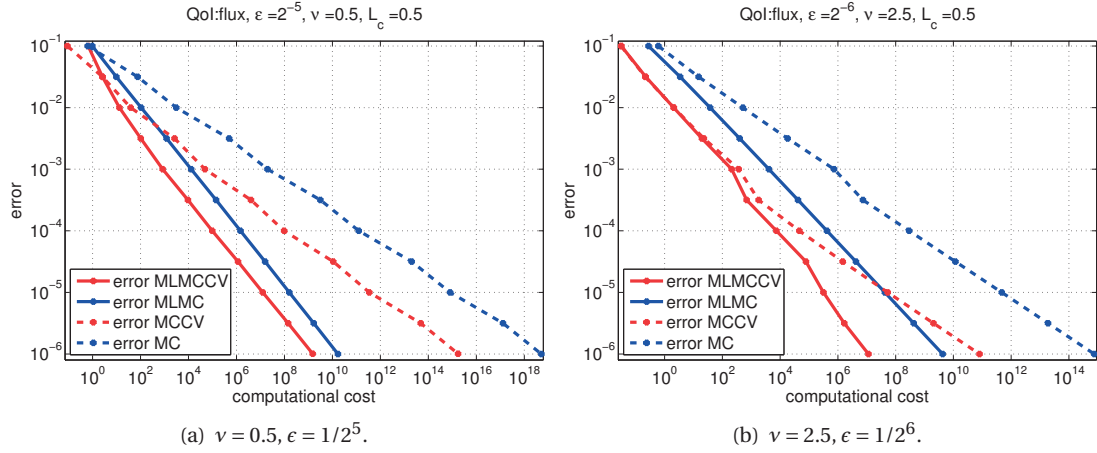


Figure 4.11 – Error vs Computational cost for  $v = 0.5$  (left) and  $v = 2.5$  (right). Error = Sparse Grid error + Statistical error, computed as root mean square error. Computational cost estimated according to (4.18) and (4.19), with  $C_\ell = h_\ell^{-d\rho}$  and  $\rho = 1.05$ , fitted by using a Matlab backslash command.

kept in the truncated field is proportional to the number of degrees of freedom of the corresponding spatial discretization. This fact does not have an actual impact on the computational cost when considering a MC or a MCCV scheme since the generation of the random field will be always considerably less costly than solving the PDE on the current mesh; however, in a multi-level framework, if we want to keep a fixed number of random variables on all levels, as we have done indeed so far, this could have an impact on the computational cost on coarse levels; indeed, the cost of generating a sample on each level would be proportional, up to logarithmic terms, to the number of random variables included in the random field, which is  $O(h_L^{-2})$ , i.e. the cost of generating a sample on a coarse level is proportional to the number of degrees of freedom in the finest one. In fact in 2D the truncated Fourier expansion of the log-permeability  $\gamma$  on the expanded domain  $(0, L')^2$  is

$$\gamma(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{n} \in \{0, \dots, N\}^2} c_{\mathbf{n}} \sum_{\ell \in \{0, 1\}^2} y_{\mathbf{n}}^\ell(\omega) \prod_{i=1}^2 \cos\left(\frac{\pi n_i}{L'} x_i\right)^{\ell_i} \sin\left(\frac{\pi n_i}{L'} x_i\right)^{1-\ell_i}. \quad (4.31)$$

As soon as we fix the finest mesh size  $h_L$ , we implicitly fix  $N$  so that the number of random variables involved in the truncation is linked to the spatial resolution of the finest grid. If we want to generate a field on the coarsest mesh with the same number of terms in the truncation we have to pay then a cost proportional to  $O(h_L^{-2})$ . Since the use of FFT naturally links the parameters  $h_L$  and  $N$ , it is then possible to follow the ideas in [74] to properly define a new “multi-level”-type estimator that considers a level-dependent truncation of the random field  $\gamma$ , namely  $\gamma_{N_\ell}$ . By denoting with  $Q_\ell^{LD}$  the QoI  $Q$  approximated with mesh size  $h_\ell$  and starting with a truncated field  $\gamma$  with  $N = N_\ell$  in (4.31), the sum  $\mathbb{E}[Q_L] = \sum_{\ell=0}^L \mathbb{E}[Q_\ell^{LD}] - \mathbb{E}[Q_{\ell-1}^{LD}]$  is still telescopic. Then by replacing  $Q_{h_\ell}$  with  $Q_\ell^{LD}$  in the definition of the MLMC and MLMCCV estimators, we obtain our level dependent MLLD and MLLDCV estimators:

$$\hat{Q}_{\{h_\ell\}, \{M_\ell\}}^{MLLD} = \sum_{\ell=0}^L \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} (Q_\ell^{LD}(\mathbf{y}_{\ell,i}) - Q_{\ell-1}^{LD}(\mathbf{y}_{\ell,i})), \text{ where } Q_{-1} = 0,$$

$$\hat{Q}_{\{h_\ell\}, \{M_\ell\}, W}^{MLLDCV} = \sum_{\ell=0}^L \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} \left( Q_\ell^{LD}(\omega_{\ell,i}) - Q_{\ell-1}^{LD}(\omega_{\ell,i}) - (Q_\ell^{LD,\epsilon}(\omega_{\ell,i}) - Q_{\ell-1}^{LD,\epsilon}(\omega_{\ell,i})) \right) + \mathcal{Q}_W[Q_{h_L}^\epsilon].$$

The derivation of the corresponding mean square error is straightforward. In Figure 4.12 and 4.13 we consider again the flux  $Q^2$  as QoI and we report the decay of the variance of the difference on consecutive levels of these level-dependent estimators.

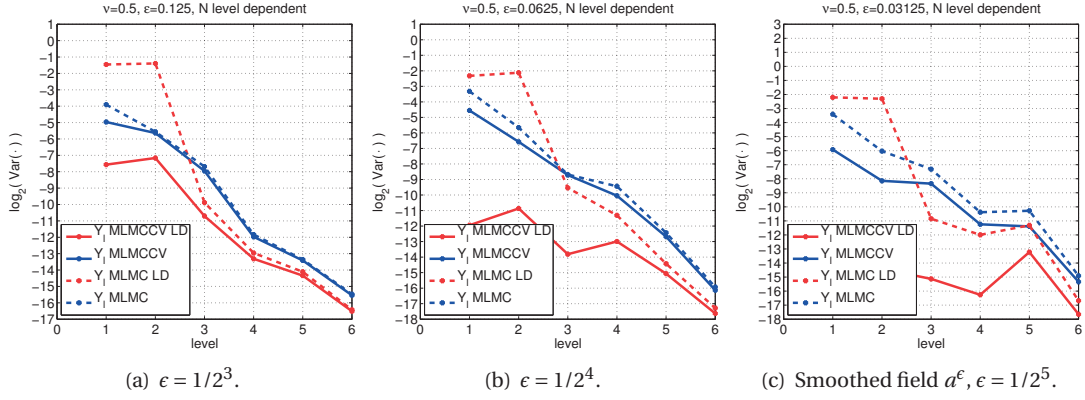


Figure 4.12 – Variance of  $Y_\ell^{LD} = Q_\ell^{LD} - Q_{\ell-1}^{LD}$ , with  $Q_\ell^{LD}(\omega) = -\int_D a_{N_\ell}(x, \omega) \frac{\partial p_\ell^{LD}}{\partial x_1}(x, \omega) dx$  for  $v = 0.5$  and different  $\epsilon$ .

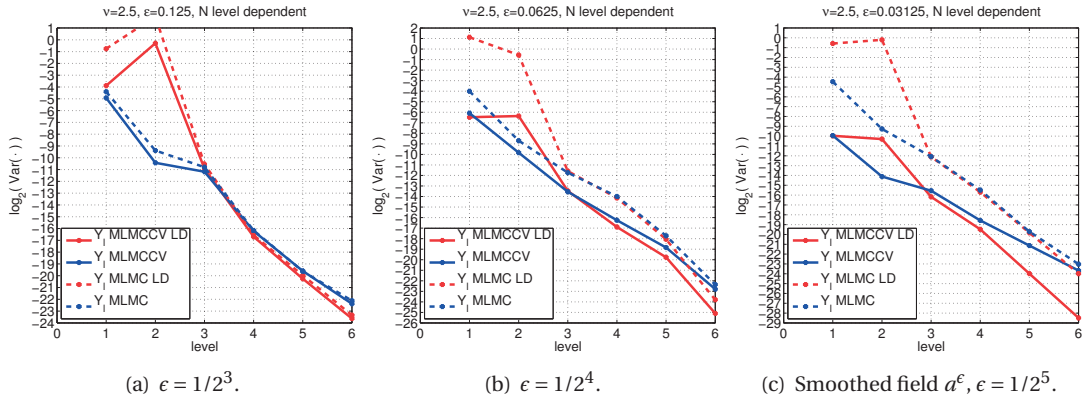


Figure 4.13 – Variance of  $Y_\ell^{LD} = Q_\ell^{LD} - Q_{\ell-1}^{LD}$ , with  $Q_\ell^{LD}(\omega) = -\int_D a_{N_\ell}(x, \omega) \frac{\partial p_\ell^{LD}}{\partial x_1}(x, \omega) dx$  for  $v = 2.5$  and different  $\epsilon$ .

In [74] the authors pointed out how the variance reduction desirable with such level-dependent truncation is highly sensitive to the choice of  $N_\ell$ . Here, for the sake of simplicity, we choose the  $N_\ell$  “induced by the the FFT”, i.e.  $N_\ell = L' h_\ell^{-1}$ ; such choice turns out to be a good one: in fact the numerical experiments show how, by using such level-dependent truncation, the MLMCCV estimator is affected by a smaller variance than the MLMC one; moreover we have variance reduction with respect to the corresponding MLMC estimator also in those situations in which MLMCCV did not feature any variance reduction with respect to MLMC.

In Figure 4.14 we show the overall error given by the sum of statistical and sparse grid errors. Observe how we could significantly improve the gain in terms of computational cost with respect to the previous case in which we were keeping the same truncation on all levels.

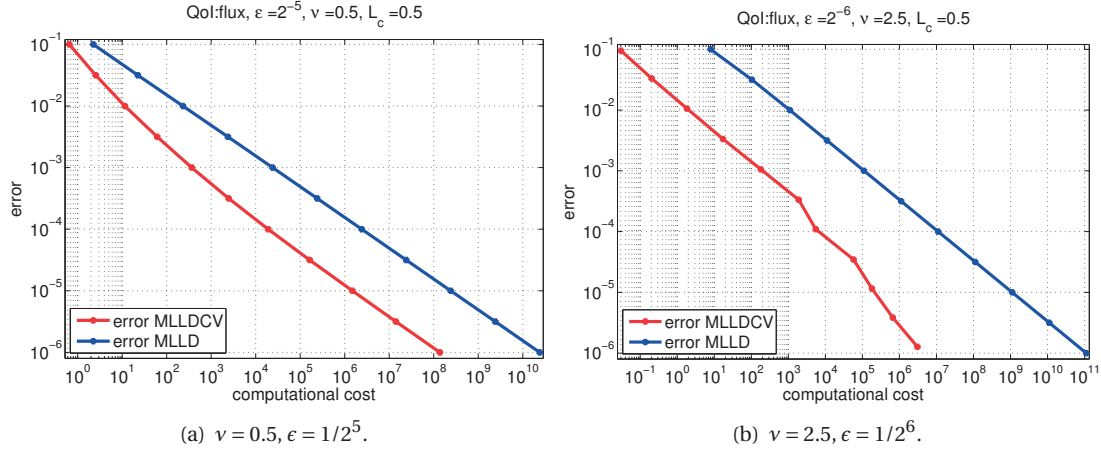


Figure 4.14 – Error vs Computational cost for  $\nu = 0.5$  (left) and  $\nu = 2.5$  (right). Error = Sparse Grid error + Statistical error. Computational cost estimated according to (4.18) and (4.19).

## 4.9 Conclusions

In this chapter we have presented a new control variate approach that combines the use of a sparse grid scheme with a Monte Carlo and a Multi Level Monte Carlo sampler. We then applied these MCCV and MLMCCV algorithms to solve an elliptic partial differential equation represented by the Darcy problem with log-normal permeability  $a = e^\gamma$  with  $\gamma$  having a covariance function belonging to the Matérn family 2.5. We consider as control variate the QoI computed starting from an auxiliary Darcy problem with smoothed input permeability. Since such a control variate is obtained starting from a smooth problem it is expected to be smooth also with respect to the stochastic parameters. The crucial assumption for this method to work is then to have at our disposal a good sparse grid quadrature scheme to compute the expected value of such control variate.

The proposed MLMCCV strategy considerably improves the performance of the standard MLMC method in terms of error versus computational cost, in both cases of rough and smooth coefficients, where we have always observed a gain with respect to the standard or level-dependent MLMC method.

The choice of the regularization parameter  $\varepsilon$  is rather delicate in the case of rough coefficients, as it should properly balance the variance reduction achievable in MLMC and the performance of the stochastic collocation.

Also the application of a level dependent truncation of the input random field has provided significant improvements with respect to the case in which a fixed truncation was considered; indeed, such choice improves both computational aspects, since it speeds up the samples generation since the FFT has to perform much less operations, and at the same time further improve the variance reduction with respect to the other cases.

# Appendix

## 4.A Proof of Lemma 4.3.1

We consider a fixed  $\omega \in \Omega$  and we will not specify the dependence on  $\omega$  in the proof. In order to prove Lemma 4.3.1 we will need the following three preliminary lemmas.

**Lemma 4.A.1.** *Let  $\gamma(x)$  be a deterministic function in  $C^\alpha(\mathbb{R}^d)$  and,  $\forall h \in \mathbb{R}^d$ , let us define  $D_{h,\beta}\gamma(x)$  as*

$$D_{h,\beta}\gamma(x) = \frac{\gamma(x+h) - \gamma(x)}{|h|^\beta}.$$

*Then,  $\forall 0 < \beta \leq \min(1, \alpha)$  it holds*

$$\|D_{h,\beta}\gamma\|_{C^{\alpha-\beta}(\mathbb{R}^d)} \leq (1 + 2\sqrt{d})\|\gamma\|_{C^\alpha(\mathbb{R}^d)}.$$

*and in particular*

$$|D_{h,\beta}\gamma|_{C^{\alpha-\beta}(\mathbb{R}^d)} \leq 2\sqrt{d}|\gamma|_{C^\alpha(\mathbb{R}^d)}.$$

*Proof.* Let us denote  $\alpha = A + s$ , with  $A \in \mathbb{N}$  and  $s \in (0, 1]$ .

• We first consider the case  $0 \leq \beta \leq \alpha \leq 1$  so that  $A = 0$  and  $s = \alpha$ . The norm we want to bound can be written as

$$\begin{aligned} \|D_{h,\beta}\gamma\|_{C^{\alpha-\beta}(\mathbb{R}^d)} &= \|D_{h,\beta}\gamma\|_{\mathcal{C}^0(\mathbb{R}^d)} + |D_{h,\beta}\gamma|_{C^{\alpha-\beta}(\mathbb{R}^d)} = \\ &= \underbrace{\|D_{h,\beta}\gamma\|_{\mathcal{C}^0(\mathbb{R}^d)}}_{(i)} + \underbrace{\sup_{x,t \in \mathbb{R}^d} \frac{|(D_{h,\beta}\gamma)(x+t) - (D_{h,\beta}\gamma)(x)|}{|t|^{\alpha-\beta}}}_{(ii)}. \end{aligned}$$

The first term can be bounded as

$$\begin{aligned} (i) &= \sup_{x \in \mathbb{R}^d} \frac{|\gamma(x+h) - \gamma(x)|}{|h|^\beta} \leq \sup_{x \in \mathbb{R}^d} \max \left\{ \sup_{|h| \geq 1} \frac{|\gamma(x+h) - \gamma(x)|}{|h|^\beta}, \sup_{|h| \leq 1} \frac{|\gamma(x+h) - \gamma(x)|}{|h|^\beta} \right\} \\ &\leq \sup_{x \in \mathbb{R}^d} \max \left\{ \sup_{|h| \geq 1} (|\gamma(x+h)| + |\gamma(x)|), \sup_{|h| \leq 1} \frac{|\gamma(x+h) - \gamma(x)|}{|h|^\alpha} |h|^{\alpha-\beta} \right\} \leq 2\|\gamma\|_{\mathcal{C}^0(\mathbb{R}^d)} + |\gamma|_{C^\alpha(\mathbb{R}^d)}. \end{aligned}$$

The second term can be bounded as

$$\begin{aligned}
 (ii) &= \max \left\{ \sup_{\substack{x, t \in \mathbb{R}^d \\ |t| \geq |h|}} \frac{|\gamma(x+t+h) - \gamma(x+t) - \gamma(x+h) + \gamma(x)|}{|h|^\beta |t|^{\alpha-\beta}}, \sup_{\substack{x, t \in \mathbb{R}^d \\ |t| \leq |h|}} \frac{|\gamma(x+h+t) - \gamma(x+h) - \gamma(x+t) + \gamma(x)|}{|h|^\beta |t|^{\alpha-\beta}} \right\} \\
 &\leq \max \left\{ \sup_{\substack{x, t \in \mathbb{R}^d \\ |t| \geq |h|}} \frac{|\gamma(x+t+h) - \gamma(x+t)| + |\gamma(x+h) - \gamma(x)|}{|h|^\beta |h|^{\alpha-\beta}}, \sup_{\substack{x, t \in \mathbb{R}^d \\ |t| \leq |h|}} \frac{|\gamma(x+t+h) - \gamma(x+h)| + |\gamma(x+t) - \gamma(x)|}{|t|^\beta |t|^{\alpha-\beta}} \right\} \\
 &\leq 2|\gamma|_{C^\alpha(\mathbb{R}^d)}
 \end{aligned}$$

Hence we get  $\|D_{h,\beta}\gamma\|_{C^{\alpha-\beta}(\mathbb{R}^d)} \leq 3\|\gamma\|_{C^\alpha(\mathbb{R}^d)}$  and  $|D_{h,\beta}\gamma|_{C^{\alpha-\beta}(\mathbb{R}^d)} \leq 2|\gamma|_{C^\alpha(\mathbb{R}^d)}$ .

• Let us consider now the case  $0 < \beta \leq 1 < \alpha$  so that  $\alpha = A + s$  with  $A \geq 1$ . The proof can be further divided in two parts:  $s > \beta$  and  $s < \beta$  since for  $s = \beta$  the result is obvious. We start with the case  $s > \beta$ . The norm that we want to bound can be written as

$$\begin{aligned}
 \|D_{h,\beta}\gamma\|_{C^{\alpha-\beta}(\mathbb{R}^d)} &= \|D_{h,\beta}\gamma\|_{\mathcal{C}^0(\mathbb{R}^d)} + \sum_{k=1}^A |D_{h,\beta}\gamma|_{\mathcal{C}^k(\mathbb{R}^d)} + |D_{h,\beta}\gamma|_{C^{\alpha-\beta}(\mathbb{R}^d)} \\
 &= \underbrace{\|D_{h,\beta}\gamma\|_{\mathcal{C}^0(\mathbb{R}^d)}}_{(i)} + \underbrace{\sum_{k=1}^A \max_{|i|_1=k} \|D^i(D_{h,\beta}\gamma)\|_{\mathcal{C}^0(\mathbb{R}^d)}}_{(ii)} + \underbrace{\max_{|i|_1=A} \sup_{x,t} \left\{ \frac{|D^i(D_{h,\beta}\gamma)(x+t) - D^i(D_{h,\beta}\gamma)(x)|}{|t|^{s-\beta}} \right\}}_{(iii)}.
 \end{aligned}$$

In what follows, we denote  $\xi_x^y$  a point of the segment  $\overline{xy}$ , i.e.  $\xi_x^y = \theta x + (1-\theta)y$  for some  $\theta \in [0, 1]$ .

The first term (i) can be bounded as

$$\begin{aligned}
 (i) &= \sup_{x \in \mathbb{R}^d} \frac{|\gamma(x+h) - \gamma(x)|}{|h|^\beta} \leq \sup_{x \in \mathbb{R}^d} \max \left\{ \sup_{|h| \geq 1} \frac{|\gamma(x+h) - \gamma(x)|}{|h|^\beta}, \sup_{|h| \leq 1} \frac{|\gamma(x+h) - \gamma(x)|}{|h|^\beta} \right\} \\
 &\leq \sup_{x \in \mathbb{R}^d} \max \left\{ \sup_{|h| \geq 1} (|\gamma(x+h)| + |\gamma(x)|), \sup_{|h| \leq 1} \frac{|\nabla \gamma(\xi_x^{x+h}) \cdot h|}{|h|^\beta} \right\} \leq 2\|\gamma\|_{\mathcal{C}^0(\mathbb{R}^d)} + \sqrt{d}|\gamma|_{\mathcal{C}^1(\mathbb{R}^d)}.
 \end{aligned}$$

Each term of (ii), for  $k = 1, \dots, A-1$ , can be bounded as

$$\begin{aligned}
 \max_{|i|_1=k} \left\{ \sup_{x \in \mathbb{R}^d} \frac{|D^i \gamma(x+h) - D^i \gamma(x)|}{|h|^\beta} \right\} &\leq \max_{|i|_1=k} \left\{ \sup_{x \in \mathbb{R}^d} \max \left\{ \sup_{|h| \geq 1} (|D^i \gamma(x+h)| + |D^i \gamma(x)|), \sup_{|h| \leq 1} \frac{|\nabla D^i \gamma(\xi_{i,x}^{x+h}) \cdot h|}{|h|^\beta} \right\} \right\} \\
 &\leq 2|\gamma|_{\mathcal{C}^k(\mathbb{R}^d)} + \sqrt{d}|\gamma|_{\mathcal{C}^{k+1}(\mathbb{R}^d)}.
 \end{aligned}$$

The last term of (ii) for  $k = A$ , analogously to what we did in the case  $0 < \beta \leq \alpha \leq 1$ , can be bounded as

$$|D_{h,\beta}\gamma|_{\mathcal{C}^A(\mathbb{R}^d)} \leq 2|\gamma|_{\mathcal{C}^A(\mathbb{R}^d)} + |\gamma|_{C^\alpha(\mathbb{R}^d)}.$$

Hence the term (ii) can be bounded as

$$(ii) \leq 2|\gamma|_{\mathcal{C}^1(\mathbb{R}^d)} + (2 + \sqrt{d}) \sum_{k=2}^A |\gamma|_{\mathcal{C}^k(\mathbb{R}^d)} + |\gamma|_{C^\alpha(\mathbb{R}^d)}.$$

The last term (iii) can be bounded as

$$(iii) \leq \max_{|i|_1=A} \left\{ \max \left\{ \sup_{\substack{x, t \in \mathbb{R}^d \\ |t| \geq |h|}} \frac{|D^i \gamma(x+t+h) - D^i \gamma(x+t)| + |D^i \gamma(x+h) - D^i \gamma(x)|}{|h|^\beta |t|^{s-\beta}}, \right. \right. \\ \left. \left. \sup_{\substack{x, t \in \mathbb{R}^d \\ |t| \leq |h|}} \frac{|D^i \gamma(x+t+h) - D^i \gamma(x+t)| + |D^i \gamma(x+h) - D^i \gamma(x)|}{|h|^\beta |t|^{s-\beta}} \right\} \right\} \leq 2|\gamma|_{C^\alpha(\mathbb{R}^d)}.$$

So the norm of  $\|D_{h,\beta} \gamma\|_{C^{\alpha-\beta}(\mathbb{R}^d)}$  can be bounded as

$$\|D_{h,\beta} \gamma\|_{C^{\alpha-\beta}(\mathbb{R}^d)} \leq (i) + (ii) + (iii) \leq 2\|\gamma\|_{C^0(\mathbb{R}^d)} + (2+\sqrt{d}) \sum_{k=1}^A |\gamma|_{C^k(\mathbb{R}^d)} + 3|\gamma|_{C^\alpha(\mathbb{R}^d)} \leq (2+\sqrt{d})\|\gamma\|_{C^\alpha(\mathbb{R}^d)}.$$

and  $|D_{h,\beta} \gamma|_{C^{\alpha-\beta}(\mathbb{R}^d)} \leq 2|\gamma|_{C^\alpha(\mathbb{R}^d)}$ .

• Let us consider now the case  $s < \beta$ . The quantity we want to bound becomes

$$\|D_{h,\beta} \gamma\|_{C^{\alpha-\beta}(\mathbb{R}^d)} = \|D_{h,\beta} \gamma\|_{C^0(\mathbb{R}^d)} + \sum_{k=1}^{A-1} |D_{h,\beta} \gamma|_{C^k(\mathbb{R}^d)} + |D_{h,\beta} \gamma|_{C^{\alpha-\beta}(\mathbb{R}^d)} \\ = \underbrace{\|D_{h,\beta} \gamma\|_{C^0(\mathbb{R}^d)}}_{(i)} + \underbrace{\sum_{k=1}^{A-1} \max_{|i|_1=k} \|D^i(D_{h,\beta} \gamma)\|_{C^0(\mathbb{R}^d)}}_{(ii)} + \underbrace{\max_{|i|_1=A-1} \sup_{x,t} \left\{ \frac{|D^i(D_{h,\beta} \gamma)(x+t) - D^i(D_{h,\beta} \gamma)(x)|}{|t|^{1+s-\beta}} \right\}}_{(iii)}.$$

We have already derived a bound for the terms (i) and (ii). The term (iii) can be bounded as follows:

$$(iii) = \max_{|i|_1=A-1} \max \left\{ \underbrace{\sup_{\substack{x, t \in \mathbb{R}^d \\ |t| \geq |h|}} \frac{|D^i(D_{h,\beta} \gamma)(x+t) - D^i(D_{h,\beta} \gamma)(x)|}{|t|^{1+s-\beta}}}_{I_i}, \underbrace{\sup_{\substack{x, t \in \mathbb{R}^d \\ |t| \leq |h|}} \frac{|D^i(D_{h,\beta} \gamma)(x+t) - D^i(D_{h,\beta} \gamma)(x)|}{|t|^{1+s-\beta}}}_{II_i} \right\}$$

By bounding separately the two terms we get

$$I_i = \sup_{\substack{x, t \in \mathbb{R}^d \\ |t| \geq |h|}} \frac{|D^i \gamma(x+t+h) - D^i \gamma(x+t) - D^i \gamma(x+h) + D^i \gamma(x)|}{|h|^\beta |t|^{1+s-\beta}} \\ = \sup_{\substack{x, t \in \mathbb{R}^d \\ |t| \geq |h|}} \frac{|(\nabla D^i \gamma(\xi_{i,x+t}^{x+t+h}) - \nabla D^i \gamma(\xi_{i,x}^{x+h})) \cdot h|}{|h|^\beta |t|^{1+s-\beta}} \\ \leq \sup_{\substack{x, t \in \mathbb{R}^d \\ |t| \geq |h|}} \frac{\max_{j=1,\dots,d} |\partial_{x_j}(D^i \gamma(\xi_{i,x+t}^{x+t+h})) - \partial_{x_j}(D^i \gamma(\xi_{i,x}^{x+h}))| \sqrt{d}|h| |\xi_{i,x+t}^{x+t+h} - \xi_{i,x}^{x+h}|^s}{|\xi_{i,x+t}^{x+t+h} - \xi_{i,x}^{x+h}|^s |h|^\beta |t|^{1+s-\beta}} \\ \leq \sqrt{d} |\gamma|_{C^\alpha(\mathbb{R}^d)} \sup_{\substack{x, t \in \mathbb{R}^d \\ |t| \geq |h|}} \frac{|h|^{1-\beta} (|h|^s + |t|^s)}{|t|^{1+s-\beta}} \leq 2\sqrt{d} |\gamma|_{C^\alpha(\mathbb{R}^d)}.$$

Similarly for the term  $II_i$ , we have:

$$\begin{aligned}
 II_i &= \sup_{\substack{x, t \in \mathbb{R}^d \\ |t| \leq |h|}} \frac{|D^i \gamma(x+t+h) - D^i \gamma(x+h) - D^i \gamma(x+t) + D^i \gamma(x)|}{|h|^\beta |t|^{1+s-\beta}} \\
 &= \sup_{\substack{x, t \in \mathbb{R}^d \\ |t| \leq |h|}} \frac{|(\nabla D^i \gamma(\xi_{i,x+h}^{x+t+h}) - \nabla D^i \gamma(\xi_{i,x}^{x+t})) \cdot t|}{|h|^\beta |t|^{1+s-\beta}} \\
 &\leq \sup_{\substack{x, t \in \mathbb{R}^d \\ |t| \leq |h|}} \frac{\max_{j=1, \dots, d} |\partial_{x_j}(D^i \gamma(\xi_{i,x+h}^{x+t+h}) - D^i \gamma(\xi_{i,x}^{x+t}))| \sqrt{d} |t| |\xi_{i,x+h}^{x+t+h} - \xi_{i,x}^{x+t}|^s}{|h|^\beta |t|^{1+s-\beta}} \\
 &\leq \sqrt{d} |\gamma|_{C^\alpha(\mathbb{R}^d)} \sup_{\substack{x, t \in \mathbb{R}^d \\ |t| \leq |h|}} \frac{|t|^{\beta-s} (|h|^s + |t|^s)}{|h|^\beta} \leq 2\sqrt{d} |\gamma|_{C^\alpha(\mathbb{R}^d)}.
 \end{aligned}$$

and then the term  $(iii)$  can be bounded as  $(iii) \leq 2\sqrt{d} |\gamma|_{C^\alpha(\mathbb{R}^d)}$ . Finally the norm of  $\|D_{h,\beta} \gamma\|_{C^{\alpha-\beta}(\mathbb{R}^d)}$  can be bounded as

$$\|D_{h,\beta} \gamma\|_{C^{\alpha-\beta}(\mathbb{R}^d)} \leq (i) + (ii) + (iii) \leq 2\|\gamma\|_{C^0(\mathbb{R}^d)} + (2 + \sqrt{d}) \sum_{k=1}^{A-1} |\gamma|_{C^k(\mathbb{R}^d)} + \sqrt{d} |\gamma|_{C^A(\mathbb{R}^d)} + (1 + 2\sqrt{d}) |\gamma|_{C^\alpha(\mathbb{R}^d)}.$$

By comparing this expression with the one obtained in the previous case it is possible to conclude that  $\forall h \in \mathbb{R}^d$ ,  $0 < \beta \leq \min\{\alpha, 1\}$  it holds

$$\|D_{h,\beta} \gamma\|_{C^{\alpha-\beta}(\mathbb{R}^d)} \leq (1 + 2\sqrt{d}) \|\gamma\|_{C^\alpha(\mathbb{R}^d)}, \quad |D_{h,\beta} \gamma|_{C^{\alpha-\beta}(\mathbb{R}^d)} \leq 2\sqrt{d} |\gamma|_{C^\alpha(\mathbb{R}^d)}.$$

□

**Lemma 4.A.2.** Let  $\kappa(x) \in C^\alpha(\mathbb{R}^d)$  be a deterministic function as in Lemma 4.A.1, and let  $\kappa^\epsilon(x)$  be a smoothed version of  $\kappa(x)$  defined as  $\kappa^\epsilon(x) = (\kappa * \phi_\epsilon)(x)$  with  $\phi_\epsilon(x)$  as in (4.2). It holds:

$$\|\kappa - \kappa^\epsilon\|_{C^0(\mathbb{R}^d)} \leq C(\alpha, d) |\kappa|_{C^{\min(\alpha, 2)}(\mathbb{R}^d)} \epsilon^{\min(\alpha, 2)},$$

where  $C(\alpha, d) = \frac{1}{\sqrt{2\pi}^d} \int_{\mathbb{R}^d} |y|^\alpha e^{-\frac{|y|^2}{2}} dy$ .

*Proof.* By definition we have

$$|(\kappa - \kappa^\epsilon)(x)| = \left| \int_{\mathbb{R}^d} (\kappa(x+y) - \kappa(x)) \phi_\epsilon(y) dy \right| \quad \forall x \in \mathbb{R}^d.$$

- Let us start with the case  $0 < \alpha \leq 1$ : if  $\kappa(x) \in C^\alpha(\mathbb{R}^d)$  then we obtain

$$|(\kappa - \kappa^\epsilon)(x)| \leq \int_{\mathbb{R}^d} \frac{|\kappa(x+y) - \kappa(x)|}{|y|^\alpha} |y|^\alpha \phi_\epsilon(y) dy \leq |\kappa|_{C^\alpha(\mathbb{R}^d)} \int_{\mathbb{R}^d} |y|^\alpha \phi_\epsilon(y) dy \leq C(\alpha, d) |\kappa|_{C^\alpha(\mathbb{R}^d)} \epsilon^\alpha.$$

- If  $1 < \alpha \leq 2$  we consider a Taylor expansion of  $\kappa(x+y)$  around  $x$  and set  $\alpha = 1 + s$  with  $s \in (0, 1]$ .



Since odd moments of a normal distribution vanish, we get:

$$\begin{aligned} |(\kappa - \kappa^\epsilon)(x)| &= \left| \int_{\mathbb{R}^d} \left( \nabla \kappa(x) \cdot y + (\nabla \kappa(\xi_x^{x+y}) - \nabla \kappa(x)) \cdot y \right) \phi_\epsilon(y) dy \right| \\ &\leq \int_{\mathbb{R}^d} \max_{|i|_1=1} \frac{|D^i \kappa(\xi_x^{x+y}) - D^i \kappa(x)|}{|\xi_x^{x+y} - x|^s} \sqrt{d} |y|^{1+s} \phi_\epsilon(y) dy \leq C(\alpha, d) \sqrt{d} |\kappa|_{C^\alpha(\mathbb{R}^d)} \epsilon^\alpha. \end{aligned}$$

• Finally by considering  $2 < \alpha$  and by expanding further the function  $\kappa$ , since the second moment of a normal distribution does not vanish, we get:

$$\begin{aligned} |(\kappa - \kappa^\epsilon)(x)| &= \left| \int_{\mathbb{R}^d} \left( \nabla \kappa(x) \cdot y + \sum_{j,k=1}^d \frac{\partial^2 \gamma}{\partial x_j \partial x_k}(\xi_x^{x+y}) y_j y_k \right) \phi_\epsilon(y) dy \right| \\ &\leq C(\alpha, d) |\kappa|_{\mathcal{C}^2(\mathbb{R}^d)} \epsilon^2. \end{aligned}$$

□

**Lemma 4.A.3.** Let  $\gamma(x) \in C^\alpha(\overline{D})$  and  $\gamma^\epsilon(x) \in C^\alpha(\overline{D})$  be two deterministic functions and let  $a(x) = e^{\gamma(x)}$  and  $a^\epsilon(x) = e^{\gamma^\epsilon(x)}$ . For any  $0 < \beta \leq \min(1, \alpha)$  it holds

$$\|a - a^\epsilon\|_{C^\beta(\overline{D})} \leq \|a\|_{\mathcal{C}^0(\overline{D})} \|1 + \frac{a^\epsilon}{a}\|_{\mathcal{C}^0(\overline{D})} (1 + |\gamma|_{C^\beta(\overline{D})}) \|\gamma - \gamma^\epsilon\|_{C^\beta(\overline{D})}.$$

*Proof.* We bound separately the terms coming from the definition of the  $C^\beta$  norm, namely  $\|a - a^\epsilon\|_{C^\beta(\overline{D})} = \|a - a^\epsilon\|_{\mathcal{C}^0(\overline{D})} + |a - a^\epsilon|_{C^\beta(\overline{D})}$ . For the first one we simply observe that

$$\|a - a^\epsilon\|_{\mathcal{C}^0(\overline{D})} \leq \|e^\gamma + e^{\gamma^\epsilon}\|_{\mathcal{C}^0(\overline{D})} \|\gamma - \gamma^\epsilon\|_{\mathcal{C}^0(\overline{D})}.$$

For the second term we start by considering the inequality

$$|a - a^\epsilon|_{C^\beta(\overline{D})} \leq \|e^\gamma\|_{\mathcal{C}^0(\overline{D})} \underbrace{|1 - e^{\gamma - \gamma^\epsilon}|}_{(i)} \underbrace{|e^\gamma|_{C^\beta(\overline{D})}}_{(ii)} \underbrace{\|1 - e^{\gamma - \gamma^\epsilon}\|_{\mathcal{C}^0(\overline{D})}}_{(iii)}.$$

The terms in the above equation can be bounded as follows:

$$\begin{aligned} (i) &\leq \|e^{\gamma^\epsilon - \gamma}\|_{\mathcal{C}^0(\overline{D})} \|\gamma^\epsilon - \gamma\|_{C^\beta(\overline{D})}, \\ (ii) &\leq \|e^\gamma\|_{\mathcal{C}^0(\overline{D})} |\gamma|_{C^\beta(\overline{D})}, \\ (iii) &\leq \|1 + e^{\gamma^\epsilon - \gamma}\|_{\mathcal{C}^0(\overline{D})} \|\gamma^\epsilon - \gamma\|_{\mathcal{C}^0(\overline{D})}. \end{aligned}$$

By putting everything together we obtain

$$\begin{aligned} \|a - a^\epsilon\|_{C^\beta(\overline{D})} &\leq \left( \|e^\gamma + e^{\gamma^\epsilon}\|_{\mathcal{C}^0(\overline{D})} + \|e^\gamma\|_{\mathcal{C}^0(\overline{D})} |\gamma|_{C^\beta(\overline{D})} \|1 + e^{\gamma^\epsilon - \gamma}\|_{\mathcal{C}^0(\overline{D})} \right) \|\gamma - \gamma^\epsilon\|_{\mathcal{C}^0(\overline{D})} \\ &\quad + \|e^\gamma\|_{\mathcal{C}^0(\overline{D})} \|e^{\gamma^\epsilon - \gamma}\|_{\mathcal{C}^0(\overline{D})} |\gamma^\epsilon - \gamma|_{C^\beta(\overline{D})} \\ &\leq \|e^\gamma\|_{\mathcal{C}^0(\overline{D})} \|1 + e^{\gamma^\epsilon - \gamma}\|_{\mathcal{C}^0(\overline{D})} (1 + |\gamma|_{C^\beta(\overline{D})}) \|\gamma^\epsilon - \gamma\|_{C^\beta(\overline{D})} \end{aligned}$$

which is the desired result. □

Thanks to these results we can prove Lemma 4.3.1.

*Proof.* (of Lemma 4.3.1.) From lemma 4.A.1 we have that  $\tilde{\gamma} \in C^\alpha(\mathbb{R}^d)$  implies  $D_{h,\beta}\tilde{\gamma} \in C^{\alpha-\beta}(\mathbb{R}^d)$   $\forall \beta \leq \min(\alpha, 1)$ . By using the definitions given in (4.6), thanks to Lemmas 4.A.2 and 4.A.1 we get

$$\begin{aligned} \|D_{h,\beta}\tilde{\gamma} - (D_{h,\beta}\tilde{\gamma})^\epsilon\|_{\mathcal{C}^0(\mathbb{R}^d)} &\lesssim |D_{h,\beta}\tilde{\gamma}|_{C^{\min(\alpha-\beta,2)}(\mathbb{R}^d)} \epsilon^{\min(\alpha-\beta,2)} \\ &\lesssim |\tilde{\gamma}|_{C^{\min(\alpha,2+\beta)}(\mathbb{R}^d)} \epsilon^{\min(\alpha-\beta,2)} \end{aligned}$$

Since  $D_{h,\beta}\gamma^\epsilon = (D_{h,\beta}\tilde{\gamma})^\epsilon$  and thanks to the fact that the previous estimate is valid uniformly in  $h$ , we can take the supremum of  $\|D_{h,\beta}\gamma - (D_{h,\beta}\gamma)^\epsilon\|_{\mathcal{C}^0(\mathbb{R}^d)}$  with respect to  $h$ . By doing this we get

$$|\tilde{\gamma} - \gamma^\epsilon|_{C^\beta(\mathbb{R}^d)} \leq \sup_{h \in \mathbb{R}^d} \|D_{h,\beta}\gamma - (D_{h,\beta}\gamma)^\epsilon\|_{\mathcal{C}^0(\mathbb{R}^d)} \lesssim |\tilde{\gamma}|_{C^{\min(\alpha,2+\beta)}(\mathbb{R}^d)} \epsilon^{\min(\alpha-\beta,2)}.$$

Now we get the desired result by observing that  $|\tilde{\gamma}|_{C^\alpha(\mathbb{R}^d)} \leq \|\varphi\|_{C^\alpha(\bar{D}_1)} \|\gamma\|_{C^\alpha(\bar{D}_1)}$ . In fact, since  $\varphi$  vanishes on  $D_1^c$ , we obtain

$$\begin{aligned} |\tilde{\gamma}|_{C^\alpha(\mathbb{R}^d)} &= \max \left\{ \sup_{\substack{x \in \bar{D}_1 \\ y \in \mathbb{R}^d}} \frac{|\gamma(x)(\varphi(x) - \varphi(y)) + \varphi(y)(\gamma(x) - \gamma(y))|}{|x - y|^\alpha}, \sup_{\substack{y \in \bar{D}_1 \\ x \in \mathbb{R}^d}} \frac{|\gamma(y)(\varphi(y) - \varphi(x)) + \varphi(x)(\gamma(y) - \gamma(x))|}{|x - y|^\alpha} \right\} \\ &\leq \max \left\{ \|\gamma\|_{\mathcal{C}^0(\bar{D}_1)} |\varphi|_{C^\alpha(\mathbb{R}^d)} + \sup_{x,y \in \bar{D}_1} \frac{|\varphi(y)(\gamma(x) - \gamma(y))|}{|x - y|^\alpha}, \|\gamma\|_{\mathcal{C}^0(\bar{D}_1)} |\varphi|_{C^\alpha(\mathbb{R}^d)} + \sup_{x,y \in \bar{D}_1} \frac{|\varphi(x)(\gamma(y) - \gamma(x))|}{|x - y|^\alpha} \right\} \\ &\leq \|\gamma\|_{\mathcal{C}^0(\bar{D}_1)} |\varphi|_{C^\alpha(\mathbb{R}^d)} + \|\varphi\|_{\mathcal{C}^0(\bar{D}_1)} |\gamma|_{C^\alpha(\bar{D}_1)} = \|\gamma\|_{\mathcal{C}^0(\bar{D}_1)} |\varphi|_{C^\alpha(\bar{D}_1)} + \|\varphi\|_{\mathcal{C}^0(\bar{D}_1)} |\gamma|_{C^\alpha(\bar{D}_1)} \leq \|\varphi\|_{C^\alpha(\bar{D}_1)} \|\gamma\|_{C^\alpha(\bar{D}_1)}. \end{aligned}$$

Hence, by considering the inequality given in Lemma 4.A.3:

$$\|a - a^\epsilon\|_{C^\beta(\bar{D})} \leq \|a\|_{\mathcal{C}^0(\bar{D})} \|1 + \frac{a^\epsilon}{a}\|_{\mathcal{C}^0(\bar{D})} (1 + |\log(a)|_{C^\beta(\bar{D})}) \|\gamma - \gamma^\epsilon\|_{C^\beta(\bar{D})} \quad \text{a.s. in } \Omega;$$

since in  $D$  it holds  $\gamma = \tilde{\gamma}$  we get  $\|\gamma - \gamma^\epsilon\|_{C^\beta(\bar{D})} = \|\tilde{\gamma} - \gamma^\epsilon\|_{C^\beta(\bar{D})} \leq \|\tilde{\gamma} - \gamma^\epsilon\|_{C^\beta(\mathbb{R}^d)}$  and we can conclude that

$$\begin{aligned} \|a - a^\epsilon\|_{C^\beta(\bar{D})} &\lesssim \|a\|_{\mathcal{C}^0(\bar{D})} \|1 + \frac{a^\epsilon}{a}\|_{\mathcal{C}^0(\bar{D})} (1 + |\gamma|_{C^\beta(\bar{D})}) \|\varphi\|_{C^{\min(\alpha,2+\beta)}(\bar{D}_1)} \|\gamma\|_{C^{\min(\alpha,2+\beta)}(\bar{D}_1)} \epsilon^{\min(\alpha-\beta,2)} \\ &\lesssim C_{4.3.1}(\omega, \alpha) \epsilon^{\min(\alpha-\beta,2)}. \end{aligned}$$

To prove the second bound concerning the  $\mathcal{C}^1$  norm when  $\alpha > 1$  we start again from the definition:

$$\|a - a^\epsilon\|_{\mathcal{C}^1(\bar{D})} = \|a - a^\epsilon\|_{\mathcal{C}^0(\bar{D})} + \max_{|i|_1=1} \|D^i(a - a^\epsilon)\|_{\mathcal{C}^0(\bar{D})}.$$

The first term, thanks to Lemma 4.A.2, can be bounded as

$$\|a - a^\epsilon\|_{\mathcal{C}^0(\bar{D})} \leq a_{\max} \|1 + e^{\gamma^\epsilon - \gamma}\|_{\mathcal{C}^0(\bar{D})} \|\gamma - \gamma^\epsilon\|_{\mathcal{C}^0(\bar{D})} \lesssim a_{\max} \|1 + e^{\gamma^\epsilon - \gamma}\|_{\mathcal{C}^0(\bar{D})} \|\varphi\|_{C^{\min(\alpha,2)}(\bar{D}_1)} \|\gamma\|_{C^{\min(\alpha,2)}(\bar{D}_1)} \epsilon^{\min(\alpha,2)}.$$

For the second term, since the derivatives and the convolution commute, we obtain

$$D^i \left( e^\gamma (1 - e^{\gamma^\epsilon - \gamma}) \right) \leq |e^\gamma D^i(\gamma)(1 - e^{\gamma^\epsilon - \gamma})| + |e^\gamma e^{\gamma^\epsilon - \gamma} D^i(\gamma^\epsilon - \gamma)| \leq e^\gamma \left( 1 + e^{\gamma^\epsilon - \gamma} \right) \left( |D^i(\gamma)| |\gamma^\epsilon - \gamma| + |D^i \gamma - (D^i \gamma)^\epsilon| \right)$$

therefore we get

$$\begin{aligned} \|D^i(a - a^\epsilon)\|_{\mathcal{C}^0(\bar{D})} &\leq a_{\max} \left\| 1 + e^{\gamma^\epsilon - \gamma} \right\|_{\mathcal{C}^0(\bar{D})} \left( \underbrace{\|D^i \gamma\|_{\mathcal{C}^0(\bar{D})}}_{\in C^\alpha(\bar{D})} \underbrace{\|\gamma - \gamma^\epsilon\|_{\mathcal{C}^0(\bar{D})}}_{\in C^{\alpha-1}(\bar{D})} + \underbrace{\|D^i \gamma - (D^i \gamma)^\epsilon\|_{\mathcal{C}^0(\bar{D})}}_{\in C^{\alpha-1}(\bar{D})} \right) \\ &\leq a_{\max} \left\| 1 + e^{\gamma^\epsilon - \gamma} \right\|_{\mathcal{C}^0(\bar{D})} \left( \|D^i \gamma\|_{\mathcal{C}^0(\bar{D})} \|\tilde{\gamma}\|_{C^{\min(\alpha,2)}(\mathbb{R}^d)} \epsilon^{\min(\alpha,2)} + \|D^i \tilde{\gamma}\|_{C^{\min(\alpha-1,2)}(\mathbb{R}^d)} \epsilon^{\min(\alpha-1,2)} \right) \end{aligned}$$

which implies

$$\max_{|i|_1=1} \|D^i(a - a^\epsilon)\|_{\mathcal{C}^0(\bar{D})} \lesssim a_{\max} \left\| 1 + e^{\gamma^\epsilon - \gamma} \right\|_{\mathcal{C}^0(\bar{D})} \left( \|\gamma\|_{\mathcal{C}^1(\bar{D})} \|\tilde{\gamma}\|_{C^{\min(\alpha,2)}(\mathbb{R}^d)} + \|\tilde{\gamma}\|_{C^{\min(\alpha,3)}(\mathbb{R}^d)} \right) \epsilon^{\min(\alpha-1,2)}.$$

Finally, by putting everything together, and by recalling that  $|\tilde{\gamma}|_{C^\beta(\mathbb{R}^d)} \leq \|\varphi\|_{C^\beta(\bar{D}_1)} \|\gamma\|_{C^\beta(\bar{D}_1)} \quad \forall \beta \in \mathbb{R}_+$ , we get the desired result:

$$\|a - a^\epsilon\|_{\mathcal{C}^1(\bar{D})} \lesssim a_{\max} \left\| 1 + e^{\gamma^\epsilon - \gamma} \right\|_{\mathcal{C}^0(\bar{D})} \left( 1 + \|\gamma\|_{\mathcal{C}^1(\bar{D})} \right) \|\gamma\|_{C^{\min(\alpha,3)}(\bar{D}_1)} \|\varphi\|_{C^{\min(\alpha,3)}(\bar{D}_1)} \epsilon^{\min(\alpha-1,2)}, \quad \text{a.s. in } \Omega.$$

□

## 4.B Optimal rates in Theorem 4.3

Here we present a sharper bound for the infimum  $\inf_{\substack{0 \leq \beta \leq \min(\alpha,1) \\ 0 < \eta + \beta \leq \alpha}} \frac{h^\beta \epsilon^{\min(\alpha-\beta-\eta,2)}}{(\alpha-\beta)^2 \sqrt{\eta}}$  than the one presented in Proposition 4.5.1 that can be obtained with very tedious calculations.

**Lemma 4.B.1.** *Let  $\epsilon \leq e^{-\frac{1}{2\alpha}}$ ; the following bounds hold:*

- $1/2 \leq \alpha \leq 1$ :

$$\inf_{\substack{0 \leq \beta < \alpha \\ 0 < \eta + \beta \leq \alpha}} \frac{h^\beta \epsilon^{\alpha-\beta-\eta}}{(\alpha-\beta)^2 \sqrt{\eta}} \lesssim \begin{cases} \begin{cases} h^\alpha \left| \log \frac{h}{\epsilon} \right|^2 |\log \epsilon|^{\frac{1}{2}}, & h \geq \epsilon^5, \\ h^\alpha |\log h|^{\frac{5}{2}}, & h \leq \epsilon^5 \end{cases}, & h \leq e^{-\frac{2}{\alpha}\epsilon}, \\ \epsilon^\alpha |\log \epsilon|^{\frac{1}{2}}, & h \geq e^{-\frac{2}{\alpha}\epsilon}. \end{cases}$$

- $1 < \alpha \leq 2$ :

$$\inf_{\substack{0 \leq \beta \leq 1 \\ 0 < \eta + \beta \leq \alpha}} \frac{h^\beta \epsilon^{\alpha-\beta-\eta}}{(\alpha-\beta)^2 \sqrt{\eta}} \lesssim \begin{cases} \begin{cases} h \epsilon^{\alpha-1} |\log h|^{\frac{1}{2}}, & h \leq \min(e^{-\frac{5}{2(\alpha-1)}}, e^{-\frac{2}{\alpha-1}\epsilon}), \\ h^\alpha |\log h|^{\frac{5}{2}}, & \min(e^{-\frac{5}{2(\alpha-1)}}, e^{-\frac{2}{\alpha-1}\epsilon}) \leq h \leq e^{-\frac{2}{\alpha-1}\epsilon}, \end{cases} \\ \begin{cases} \begin{cases} h^\alpha |\log h|^{\frac{5}{2}}, & h \leq \epsilon^5, \\ h^\alpha \left| \log \frac{h}{\epsilon} \right|^2 |\log \epsilon|^{\frac{1}{2}}, & h \geq \epsilon^5, \end{cases} & \text{if } e^{-\frac{2}{\alpha-1}\epsilon} \leq \epsilon^5, \\ \begin{cases} h^\alpha \left| \log \frac{h}{\epsilon} \right|^2 |\log \epsilon|^{\frac{1}{2}}, & \text{if } e^{-\frac{2}{\alpha-1}\epsilon} \geq \epsilon^5, \end{cases} & e^{-\frac{2}{\alpha-1}\epsilon} \leq h \leq e^{-\frac{2}{\alpha-1}\epsilon}, \end{cases} \\ \epsilon^\alpha |\log \epsilon|^{\frac{1}{2}}, & h \geq e^{-\frac{2}{\alpha}\epsilon}. \end{cases}$$

- $2 < \alpha \leq 3$ :

$$\inf_{\substack{0 \leq \beta \leq 1 \\ 0 < \eta + \beta \leq \alpha}} \frac{h^\beta \epsilon^{\alpha - \beta - \eta}}{(\alpha - \beta)^2 \sqrt{\eta}} \lesssim \begin{cases} h \epsilon^{\alpha-1} |\log h|^{\frac{1}{2}}, & h \leq \min(e^{-\frac{5}{2(\alpha-1)}}, e^{-\frac{2}{\alpha-1}} \epsilon), \\ h^\alpha |\log h|^{\frac{5}{2}}, & \min(e^{-\frac{5}{2(\alpha-1)}}, e^{-\frac{2}{\alpha-1}} \epsilon) \leq h \leq e^{-\frac{2}{\alpha-1}} \epsilon, \\ \begin{cases} h^\alpha |\log h|^{\frac{5}{2}}, & h \leq \epsilon^5, \\ h^\alpha \left| \log \frac{h}{\epsilon} \right|^2 |\log \epsilon|^{\frac{1}{2}}, & h \geq \epsilon^5, \end{cases} & \text{if } e^{-\frac{2}{\alpha-1}} \leq \epsilon^5, \\ & , e^{-\frac{2}{\alpha-1}} \epsilon \leq h \leq e^{-\frac{2}{\alpha}} \epsilon, \\ h^\alpha \left| \log \frac{h}{\epsilon} \right|^2 |\log \epsilon|^{\frac{1}{2}}, & \text{if } e^{-\frac{2}{\alpha-1}} \geq \epsilon^5, \\ \begin{cases} \epsilon^\alpha |\log \epsilon|^{\frac{1}{2}}, & \text{if } \epsilon \geq e^{-\frac{1}{2(\alpha-2)}}, \\ h^{\alpha-2} \epsilon^2 |\log \epsilon|^{\frac{1}{2}}, & \text{if } \epsilon \leq e^{-\frac{1}{2(\alpha-2)}}, \end{cases} & h \geq e^{-\frac{2}{\alpha}} \epsilon. \end{cases}$$

- $\alpha > 3$ :

$$\inf_{\substack{0 \leq \beta \leq 1 \\ 0 < \eta + \beta \leq \alpha}} \frac{h^\beta \epsilon^{\alpha - \beta - \eta}}{(\alpha - \beta)^2 \sqrt{\eta}} \lesssim h \epsilon^2.$$

## 5 Extension to Transport Problem

### 5.1 Introduction

In this chapter we extend the methodology developed so far to deal also with transport problems that may occur when studying flows in porous media. Here the main focus lies on describing the behavior of a contaminant driven by the Darcy velocity field and subject to molecular diffusion that takes place in the medium at the porous level. The motion of such pollutant particles can be modeled through a stochastic differential equation (SDE) with a Darcy velocity as drift term and a suitable “small” diffusion coefficient to take into account the molecular diffusion. In hydrology, instead than the molecular diffusion, people usually consider the dispersion occurring at a microscopic level which comes from the changes of the velocity due to the pore structure of the medium (see for instance [9, 10]). However, since at a macroscopic level such changes are actually negligible, or at least not dominant, with respect to the ones induced by the permeability heterogeneities, we will consider the simpler model of an SDE with a small diffusion coefficient. Unlike commonly used approaches which are based on ensembles of pathwise solutions of the SDE model in order to compute statistical properties related to a contaminant’s evolution, we will consider a PDE approach instead, which is related to the underlying SDE model through the famous Feynman-Kac representation formula. The reasons for resorting to a PDE approach are twofold. Firstly, it stems from the fact that lack of regularity of the Darcy velocity renders a corresponding pathwise SDE approach unfeasible (questionable at best) as the existence of a unique pathwise solution may not be guaranteed; secondly, as we are mainly interested in problems posed on two dimensional domains, many reliable appropriate techniques are available, so that quantities, such as first exit times or probabilities, are easily computed using a PDE approach.

In this kind of applications people are often interested in understanding and properly modeling how a specific solute (e.g. a pollutant) moves under the influence of the Darcy velocity: one way of dealing with such an issue is to consider an initial concentration of a solute and look at its corresponding spread and macro-dispersion which are computed starting from the solution of a time dependent advection diffusion equation (see for instance [51]); here we consider a different approach and focus on the computation of one of the most practically relevant quantity of interest in this sort of application, the so called capture zone, which is the collection of points starting from which a fluid particle reaches a certain boundary in the physical domain in a given (finite or infinite) time horizon with a prescribed probability. The delineation of these capture zones and, more generally, tracking particle problems are relevant in hydrology applications and have been addressed by several authors in the literature [67, 37, 57, 58]. Here we consider this problem in a

more complex framework presenting two sources of randomness: one coming from the uncertain Darcy velocity, i.e. from the uncertainty on the permeability field, and the other being given by the Brownian motion, modeling the diffusive effects.

For a computational treatment of statistical properties related to these transport problems, we will use the methods developed in the previous chapters, namely sparse grid as well as MLMC and MLMCCV schemes. In particular we will focus on two quantities of interest, namely: (i) the expected arrival time of a particle released in a generic position  $\mathbf{x} \in D$  to a specific boundary and (ii) the probability of the event that such arrival times is smaller than a prescribed time horizon  $T$  (capture zone at time  $T$  at a given probability level). Observe that such quantities of interest are not random variables anymore but random fields as they depend on the starting position  $\mathbf{x} \in D$ ; moreover this random field depends on two sources of randomness corresponding to the Darcy velocity and the Brownian motion, respectively, which we emphasize by using the notation  $Q = Q(\mathbf{x}, \omega, \omega')$ , where  $\omega \in \Omega$  and  $\omega' \in \Omega'$  denote the different random events; more precisely  $\omega$  refers to the random Darcy velocity and  $\omega'$  to the Brownian motion. We assume these different sources of stochasticity to be independent and write the expectation of  $Q$  as  $\mathbb{E}^{\omega, \omega'}[Q] = \mathbb{E}^{\omega}[\mathbb{E}^{\omega'}[Q|\omega]]$  so that we first get rid of the dependence on  $\omega'$  by computing the internal expectation and then the one with respect to  $\omega$ . In this “two step” strategy we will use a PDE approach to take care of the internal expected value which requires the solution of a parabolic (or elliptic) PDE with a stochastic (dominant) transport given by the Darcy velocity and a (small) uniform diffusion term. Then, in order to compute the expected value with respect to  $\omega$ , we will use one of the methods developed in the previous chapters: the particular choice is therefore deeply linked to the smoothness of the mapping  $\mathbf{x} \rightarrow \mathbb{E}^{\omega'}[Q(\mathbf{x}, \omega, \cdot)|\omega]$ . Consequently, the regularity of this mapping dictates whether or not a sparse grid approximation technique is preferable over a MC based one.

To study the delineation of capture zones we also propose a streamline approach in a particularly relevant study case by coupling a sparse grid approximation of the stream-function and a MC sampler whose computation is particularly cheap.

The structure of the chapter is the following: in the first section we go through the specifics of the model. In Section 5.3 we study the case of an undisturbed flow and look at arrival times to the outlet by focusing on both modeling issues and numerical aspects. In Section 5.4 we model the case in which a pumping well extracts fluid from the ground: in this case we consider the arrival times to the well in order to compute the capture zones; notice that, for a fixed realization of the Darcy velocity, the arrival times are not uniformly bounded random variables, and can take also infinite values; we provide numerical results and introduce a stream-formulation approach to better address the case in which we look at infinite time horizons. In Section 5.5 we draw some conclusions.

## 5.2 Problem setting

In this section we introduce the general set up that will be used in the chapter. As anticipated in the introduction we aim at modeling the motion of fluid, or contaminant, particles driven by the Darcy velocity  $\mathbf{u}(\mathbf{x}, \omega) = -a(\mathbf{x}, \omega) \nabla p(\mathbf{x}, \omega)$ . To this end we first write the stochastic differential equation (SDE) that describes the trajectory of a single particle in our physical domain:

$$X_{\mathbf{x},t}^{\sigma'}(s, \omega) = \mathbf{x} + \int_t^s \mathbf{u}(X_{\mathbf{x},t}^{\sigma'}(r, \omega), \omega) dr + \sqrt{2}\sigma' \int_t^s dW(r, \omega'), \quad (5.1)$$

where  $X_{\mathbf{x},t}^{\sigma'}(s, \omega)$  denotes the position at time  $s$  of the particle that starts from the position  $\mathbf{x}$  at time  $t$ ,  $\sigma'$  represents the strength of molecular diffusion, modeled as a standard Brownian motion  $W(s, \omega')$  which we assume to be independent of  $\mathbf{u}(\mathbf{x}, \omega)$ , and  $\omega = (\omega, \omega')$ . The solution  $X_{\mathbf{x},t}^{\sigma'}(s, \omega)$  depends on two independent random elementary events,  $\omega \in \Omega$  and  $\omega' \in \Omega'$  that model the uncertainty related to the permeability of the subsurface and the one related to the diffusion occurring at the pore level, respectively. Consistently, the resulting sampling space will be denoted by  $\Omega = \Omega \times \Omega'$ . Observe that the SDE (5.1) holds in an *almost sure* sense with respect to  $\Omega'$ ; more precisely we have (see [65]):

- if, for a given Brownian motion  $W(s, \omega')$ , the trajectory  $X_{\mathbf{x},t}^{\sigma'}(s, \omega)$  is an adapted process with respect the natural filtration generated by  $W(s, \omega')$  solving (5.1), we say that  $X_{\mathbf{x},t}^{\sigma'}(s, \omega)$  is a weak solution of (5.1);
- if the request above holds for any Brownian motion, i.e.  $\mathbb{P}^{\omega'}$ -a.s.  $\omega' \in \Omega'$ , then we say that  $X_{\mathbf{x},t}^{\sigma'}(s, \omega)$  is a strong solution of (5.1).

We also recall the following definitions (see [20]):

**Definition 5.2.1.** *There is uniqueness in law if, given two solutions of problem (5.1)  $X_{\mathbf{x},t}^{\sigma',i}(s, \omega)$ ,  $i = 1, 2$ , defined on different filtered probability spaces, the two processes have the same law.*

**Definition 5.2.2.** *There is pathwise uniqueness if, given two solutions of problem (5.1)  $X_{\mathbf{x},t}^{\sigma',i}(s, \omega)$ ,  $i = 1, 2$ , defined on the same filtered probability space, it holds that  $\mathbb{P}^{\omega'}(X_{\mathbf{x},t}^{\sigma',1}(s, \omega) = X_{\mathbf{x},t}^{\sigma',2}(s, \omega) \quad \forall s \geq t | \omega) = 1$ .*

Standard SDE results guarantee existence and pathwise uniqueness of a strong solution of (5.1) for drift terms that are at least Lipschitz continuous. On the other hand, in [18, Theorem 3.1] the authors proved that, by assuming  $a \in C^\alpha(D)$ ,  $f \in L^q(D)$  for some  $q$  such that  $1 - d/q > \alpha$  and  $D$  a bounded  $\mathcal{C}^2$  domain, then the solution  $p$  of the Darcy problem with fully homogeneous Dirichlet conditions is  $C^{\alpha+1}(D)$  continuous; consequently the corresponding velocity  $\mathbf{u}(\cdot, \omega) \in \{C^\alpha(D)\}^2$ ,  $\mathbb{P}^\omega$ -a.s.  $\omega \in \Omega$  with suitable  $L_{\mathbb{P}^\omega}^q$  integrability conditions; hence, assuming to deal with  $\{C^\alpha(D)\}^2$  velocity fields, the SDE (5.1) is guaranteed to have a pathwise unique strong solution only when  $\alpha \geq 1$ . Conversely, we do not have a sufficient condition at our disposal which guarantees pathwise uniqueness of strong solutions when  $a \in C^\alpha(D)$  with  $\alpha < 1$ . In such case we resort to the notion of weak solution and refer to a sufficient condition given in [73] which states:

**Lemma 5.2.1.** *Suppose that the transport field  $\mathbf{u}(\cdot, \omega)$  in (5.1) is measurable and bounded for a.e.  $\omega \in \Omega$ ; then there exists a weak solution of (5.1) for which uniqueness in law holds.*

*Proof.* See [73, Theorem 5.6] □

An overview of other sufficient conditions that guarantee weak existence and uniqueness in law can be found, for instance, in [20].

Our final goal is the computation of quantities of interest related to the SDE (5.1); to achieve this goal in what follows we first need to introduce the first passage time (FPT) to the domain  $D$ , defined as

$$\tau_{\mathbf{x},t}^{\sigma'}(\omega) = \inf\{s > t : X_{\mathbf{x},t}^{\sigma'}(s, \omega) \notin \bar{D}\}. \quad (5.2)$$

Here  $\tau_{\mathbf{x},t}^{\sigma'}(\omega)$  represents the first time at which a particle starting from  $\mathbf{x} \in D$  at time  $t$  would exit the domain  $D$ .

So far we looked at the problem of the transport of particles only from an SDE perspective; the following step will be to consider a Feynman-Kac representation formula to link a suitable probabilistic formula to the solution of a parabolic PDE. Observe that, since the Darcy velocity is given only in our computational domain  $D$ , we can only use models and equations that require evaluations in  $D$ , only. Consequently, we need to make additional, yet reasonable, assumptions on the behavior of the particles when they reach the boundary  $\partial D$ : in particular, we will consider either absorbing or reflecting conditions on the boundary  $\partial D$  and in the following, case by case, we will detail the choice we made depending on the problem under consideration.

In light of these considerations, from a practical point of view, instead of solving the SDE (5.1) for many starting points  $\mathbf{x} \in D$  we will consider a PDE approach. To this end, we look at the infinitesimal generator

$$\mathcal{L}_{\sigma'} = \sum_{i=1}^d u_i(\mathbf{x}, \omega) \frac{\partial}{\partial x_i} + (\sigma')^2 \frac{\partial^2}{\partial x_i^2} \quad (5.3)$$

associated with the SDE (5.1) and we consider the following time dependent backward PDE

$$\begin{cases} \frac{\partial \theta_T^{\sigma'}(\mathbf{x}, t, \omega)}{\partial t} + \mathcal{L}_{\sigma'} \theta_T^{\sigma'}(\mathbf{x}, t, \omega) = 0 & \text{for } \mathbf{x} \in D, t \in [0, T), \\ \theta_T^{\sigma'}(\mathbf{x}, t, \omega) = g(\mathbf{x}, t) & \text{for } \mathbf{x} \in \Gamma_D, t \in [0, T), \\ \partial_{\mathbf{n}} \theta_T^{\sigma'}(\mathbf{x}, t, \omega) = 0 & \text{for } \mathbf{x} \in \Gamma_N, t \in [0, T), \\ \theta_T^{\sigma'}(\mathbf{x}, T, \omega) = h(\mathbf{x}) & \text{for } \mathbf{x} \in D. \end{cases} \quad \text{a.s. in } \Omega \quad (5.4)$$

Here, as usual,  $\Gamma_D$  and  $\Gamma_N$  denote the Dirichlet and Neumann boundaries,  $g$  represents the Dirichlet datum and  $h$  the final condition at time  $t = T$ . As we will see later, the solution  $\theta_T^{\sigma'}(\mathbf{x}, t, \omega)$  of the PDE (5.4), thanks to a Feynman-Kac type formula, can be expressed as an expected value with respect to  $\omega'$  involving the process  $X_{\mathbf{x},t}^{\sigma'}$ . Moreover, here the Dirichlet and Neumann conditions represent, from an SDE point of view, absorbing and reflecting conditions, respectively: this means that the solution of the parabolic PDE (5.4) is related to a process  $X_{\mathbf{x},t}^{\sigma'}$  which is allowed to leave the domain only through the Dirichlet boundary, while it is reflected towards the interior of the domain when it hits the Neumann boundary. To introduce a representation formula we have to consider the FPT to the domain defined in (5.2). Observe that, as a consequence of the absorbing and reflecting conditions imposed on the process  $X_{\mathbf{x},t}^{\sigma'}(s, \omega)$ , the FPT actually represents the first time at which a particle starting from  $\mathbf{x}$  at time  $t$  would exit the domain through the Dirichlet boundary. Observe also that, in order to make the definition of the trajectory  $X_{\mathbf{x},t}^{\sigma'}(s, \omega)$  consistent with the reflecting conditions imposed on the Neumann boundary, we have to slightly modify the SDE (5.1) by considering an additional term that gets activated when a particle hits the Neumann boundary. To make the presentation clearer we will consider separately the case in which no Neumann reflecting conditions are present, and the one in which mixed boundary conditions are imposed.

Since in this section we consider a particular realization of the Darcy velocity  $\mathbf{u}(\mathbf{x}, \omega)$ , for the sake of notation we will often omit the dependence on  $\omega$ , when this does not create any ambiguity.



### 5.2.1 Pure Dirichlet case: $\Gamma_D = \partial D$

To establish a link between the solution of the SDE  $X_{\mathbf{x},t}^{\sigma'}$  defined in (5.1) and the one of the PDE,  $\theta_T^{\sigma'}$ , in (5.4), we will consider the process  $X_{\mathbf{x},t}^{\sigma'}$  absorbed (or stopped) as soon as it hits the boundary  $\partial D$  at time  $\tau_{\mathbf{x},t}^{\sigma'}(\omega)$  and use the Feynman-Kac representation formula (see e.g. [65]) which states:

$$\theta_T^{\sigma'}(\mathbf{x}, t, \omega) = \mathbb{E}^{\omega'}[g(X_{\mathbf{x},t}^{\sigma'}(\tau_{\mathbf{x},t}^{\sigma'}(\omega), \omega), \tau_{\mathbf{x},t}^{\sigma'}(\omega)) \mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) \leq T\}} + h(X_{\mathbf{x},t}^{\sigma'}(T, \omega)) \mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) > T\}} | \omega].$$

Notice that here, since we deal with two sources of randomness, we first fix the Darcy velocity and then average over all Brownian motions: hence, the use of a conditional expectation with respect to  $\omega$ . However, to properly use such representation formula, we first need to ensure the existence of the solution of the SDE (5.1) and that of the PDE (5.4) in some suitable sense. Since we want to deal with a wide range of permeability fields, i.e. resulting from a Matérn covariance function (2.5) or (2.6) with parameter  $\nu \geq 0.5$ , we will use the concept of viscosity solution introduced by P. L. Lions and co-workers a few decades ago [6]. Suppose for simplicity that the domain  $D$  is bounded, connected, and with  $\mathcal{C}^2$  boundary; we have the following definitions:

**Definition 5.2.3.** *Let  $\omega$  be fixed and  $\theta : D \times [0, T] \rightarrow \mathbb{R}$  be a lower semicontinuous function. Then  $\theta$  is also a viscosity sub-solution of (5.4) if  $\forall \psi \in \mathcal{C}^\infty(D \otimes [0, T])$  it holds: if  $\theta(\mathbf{x}, t) - \psi(\mathbf{x}, t)$  has a local maximum at  $(\hat{\mathbf{x}}, \hat{t})$ , then we have*

$$\begin{aligned} \theta(\mathbf{x}, T) &\leq h(\mathbf{x}, T), \\ \frac{\partial \psi}{\partial t}(\hat{\mathbf{x}}, \hat{t}) + \mathcal{L}_{\sigma'} \psi(\hat{\mathbf{x}}, \hat{t}) &\leq 0. \end{aligned}$$

**Definition 5.2.4.** *Let  $\omega$  be fixed and  $\theta : D \times [0, T] \rightarrow \mathbb{R}$  be an upper semicontinuous function. Then  $\theta$  is also a viscosity super-solution of (5.4) if  $\forall \psi \in \mathcal{C}^\infty(D \otimes [0, T])$  it holds: if  $\theta(\mathbf{x}, t) - \psi(\mathbf{x}, t)$  has a local minimum at  $(\hat{\mathbf{x}}, \hat{t})$ , then we have*

$$\begin{aligned} \theta(\mathbf{x}, T) &\geq h(\mathbf{x}, T), \\ \frac{\partial \psi}{\partial t}(\hat{\mathbf{x}}, \hat{t}) + \mathcal{L}_{\sigma'} \psi(\hat{\mathbf{x}}, \hat{t}) &\geq 0 \end{aligned}$$

**Definition 5.2.5.** *Let  $\omega$  be fixed. The function  $\theta_T^{\sigma'}(\hat{\mathbf{x}}, t, \omega) : D \times [0, T] \rightarrow \mathbb{R}$  is a viscosity solution if it is a viscosity sub-solution and a viscosity super-solution at the same time.*

Viscosity solutions of (5.4) can be found under weaker assumptions on the transport field, in particular it is possible to recover the well posedness of PDE (5.4) also for non differentiable velocities. Moreover it is possible to link such a result to the solution of the SDE (5.1) also when the latter is not defined in a strong sense and pathwise uniqueness is not guaranteed. The following result, given in [65, Theorem 3.46]), holds:

**Lemma 5.2.2.** *Let the set  $\{(\mathbf{x}, t) \in \Gamma_D \times [0, T] : \mathbb{P}^{\omega'}(\tau_{\mathbf{x},t}^{\sigma'}(\omega) > t | \omega) = 0\}$  be closed for any  $\omega \in \Omega$  and let the boundary and final datum be such that  $g(\mathbf{x}, T) = h(\mathbf{x})$  on  $\Gamma_D$ . Furthermore assume  $g \in \mathcal{C}(\Gamma_D \times [0, T])$ ,  $h \in \mathcal{C}(\Gamma_D)$  and  $\mathbf{u} \in \{C^\alpha(D)\}^2$  with  $0 < \alpha$ . Then equation (5.4) admits a unique viscosity solution such that, for a.e.  $\omega \in \Omega$ ,  $\theta_T^{\sigma'}(\cdot, \cdot, \omega) \in \mathcal{C}(\Gamma_D) \times \mathcal{C}([0, T])$ . Moreover the following representation formula holds:*

$$\theta_T^{\sigma'}(\mathbf{x}, t, \omega) = \mathbb{E}^{\omega'}[g(X_{\mathbf{x},t}^{\sigma'}(\tau_{\mathbf{x},t}^{\sigma'}(\omega), \omega), \tau_{\mathbf{x},t}^{\sigma'}(\omega)) \mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) \leq T\}} + h(X_{\mathbf{x},t}^{\sigma'}(T, \omega)) \mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) > T\}} | \omega]. \quad (5.5)$$

*Proof.* See [65]. □

Under the assumptions of Lemma 5.2.2 equation (5.4) allows us to efficiently compute statistical properties (e.g. expectations) related to the FPT  $\tau_{\mathbf{x},t}^{\sigma'}(\omega)$  defined through the stochastic process  $X_{\mathbf{x},t}^{\sigma'}$ . In fact, we only have to properly choose the final and the boundary data  $h$  and  $g$ . It is important to observe that the result stated in Lemma 5.2.2 does not require any Lipschitz condition on the transport field  $\mathbf{u}$  and, at the same time, ensures existence and uniqueness of a viscosity solution of the PDE, even if pathwise uniqueness does not hold. Since  $\theta_T^{\sigma'}(\mathbf{x}, t, \omega)$  is defined through an expected value, in this case it is sufficient to have uniqueness in law (as guaranteed by Lemma 5.2.1) of the process  $X_{\mathbf{x},t}^{\sigma'}$  to properly define such expectation since the probability space  $(\Omega', \mathcal{F}', \mathbb{P}')$  and the Brownian motion  $W(s, \omega')$  do not play a crucial role.

In the following we will mainly focus on two choices of  $g$  and  $h$ , which allow us to link the solution of the parabolic PDE (5.4),  $\theta_T^{\sigma'}(\mathbf{x}, t, \omega)$ , to quantities of interest both related to the FPT through the Feynman-Kac formula (5.5). Specifically we consider:

- **Mean** of  $\min(\tau_{\mathbf{x},t}^{\sigma'}, T)$ :  $g(\mathbf{x}, t) = t$ ,  $h(\mathbf{x}) = T$ . This choice leads to the expectation of the minimum between the FPT of a fluid particle starting from  $\mathbf{x}$  at time  $t \leq T$  and the considered time horizon  $T$ :

$$\theta_T^{\sigma'}(\mathbf{x}, t, \omega) = \mathbb{E}^{\omega'}[\tau_{\mathbf{x},t}^{\sigma'}(\omega) \mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) \leq T\}} + T \mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) > T\}} | \omega] = \mathbb{E}^{\omega'}[\min\{\tau_{\mathbf{x},t}^{\sigma'}(\omega), T\} | \omega]. \quad (5.6)$$

- **Exit probability**:  $g(\mathbf{x}, t) = 1$ ,  $h(\mathbf{x}) = 0$ . This choice leads to the exit-probability, i.e. the probability that a particle starting from  $\mathbf{x}$  at time  $t \leq T$ , leaves the domain within the finite time horizon  $T$ :

$$\theta_T^{\sigma'}(\mathbf{x}, t, \omega) = \mathbb{E}^{\omega'}[\mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) \leq T\}} | \omega] = \mathbb{P}^{\omega'}(\tau_{\mathbf{x},t}^{\sigma'}(\omega) \leq T | \omega). \quad (5.7)$$

### Mean first passage time

Here we further detail a strategy that can be followed to compute the mean first passage time through the boundary  $\partial D$ . Notice first that the solution  $\theta_T^{\sigma'}(\mathbf{x}, t, \omega)$  to (5.6) can directly be derived from Lemma 5.2.2. However this solution represents the expectation  $\mathbb{E}^{\omega'}[\min\{\tau_{\mathbf{x},t}^{\sigma'}, T\} | \omega]$  while we would like to consider  $\mathbb{E}^{\omega'}[\tau_{\mathbf{x},t}^{\sigma'} | \omega]$ . Since it is reasonable to assume  $\tau_{\mathbf{x},t}^{\sigma'}(\omega)$  to be a bounded random variable, it is possible to take  $T$  sufficiently large to obtain a formula for the mean FTP.

An alternative approach to directly compute the mean of the FPT is to consider a suitable elliptic PDE. In fact, by eliminating the dependence on  $t$ , we define the FPT as  $\tau_{\mathbf{x}}^{\sigma'} = \tau_{\mathbf{x},0}^{\sigma'}$ . Notice that  $\tau_{\mathbf{x}}^{\sigma'}(\omega)$  represents the time that a particle needs to reach the boundary starting from  $\mathbf{x}$ , while  $\tau_{\mathbf{x},t}^{\sigma'}(\omega)$  is the time at which the particle starting from  $\mathbf{x}$  at time  $t$  reaches the boundary. Starting from this FPT it is possible to link the expected value of  $\tau_{\mathbf{x}}^{\sigma'}$  with respect to all Brownian motion, which will be denoted as  $\bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega) = \mathbb{E}^{\omega'}[\tau_{\mathbf{x}}^{\sigma'}(\omega) | \omega]$ , to the solution of an elliptic PDE; therefore  $\bar{\tau}_{\mathbf{x}}^{\sigma'}$  satisfies the following problem with homogeneous Dirichlet conditions:

$$\begin{cases} \mathcal{L}_{\sigma'} \bar{\tau}_{\mathbf{x}}^{\sigma'} = -1 & \text{in } D \times [0, T), \\ \bar{\tau}_{\mathbf{x}}^{\sigma'} = 0 & \text{on } \Gamma_D \times [0, T). \end{cases} \quad (5.8)$$

Intuitively, one can obtain this result by choosing, as in (5.6),  $g = t$  and  $h = T$ , and by looking at the problem solved by  $\eta_T^{\sigma'}(\mathbf{x}, t, \omega) = \theta_T^{\sigma'}(\mathbf{x}, t, \omega) - t$  with  $\theta_T^{\sigma'}(\mathbf{x}, t, \omega)$  again defined as in (5.6). Then, by taking the limit for  $T \rightarrow \infty$  and recalling that the Darcy velocity does not depend on time, there holds  $\bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega) = \lim_{T \rightarrow \infty} \eta_T^{\sigma'}(\mathbf{x}, 0, \omega)$ .

When we are interested in the computation of the mean first passage time, equation (5.8) is much more appealing than equation (5.4) from a computational point of view since the time-dependence is dropped. Equation (5.4), provided to take  $T$  sufficiently large, produces a time-dependent solution that, evaluated at time  $t = 0$ , reproduces the solution  $\bar{\tau}_{\mathbf{x}}^{\sigma'}$  of (5.8). For this elliptic equation, similarly to the parabolic PDE case, a well posedness result holds and again the solution  $\bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega)$  can be obtained thanks to a Feynman-Kac representation formula. The following result holds, given in [65, Theorem 3.49]), holds:

**Lemma 5.2.3.** *Let the set  $\{(\mathbf{x}, t) \in \Gamma_D : \mathbb{P}^{\omega'}(\tau_{\mathbf{x}}^{\sigma'}(\omega) > t | \omega) = 0\}$  be closed for any  $\omega \in \Omega$ . Moreover assume  $g \in \mathcal{C}(\Gamma_D)$  and  $\mathbf{u} \in [C^\alpha(D)]^2$  with  $0 < \alpha$ . Then the PDE (5.8) admits a unique viscosity solution such that, for a.e.  $\omega \in \Omega$ ,  $\theta^{\sigma'}(\mathbf{x}, \omega) \in \mathcal{C}(\bar{D})$ . Furthermore the following representation formula holds:*

$$\bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega) = \mathbb{E}^{\omega'}[\tau_{\mathbf{x}}^{\sigma'}(\omega) | \omega]. \quad (5.9)$$

*Proof.* See [65]. □

Again, viscosity solutions can be also defined for a non-differentiable Darcy velocity  $\mathbf{u}$ . Analogously to the previous case, the uniqueness of the time independent viscosity solution  $\bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega)$  is related to the uniqueness in law of the trajectories of the PDE (5.1).

### Exit probability

Here we further detail the case in which we might be interested in computing the probability of a particle starting from  $\mathbf{x}$  exiting the domain in a prescribed time horizon  $T$ . In this case we look at the solution  $\theta_T$  defined in (5.7) evaluated at time  $t = 0$ . Notice that in this case equation (5.7) can not be derived directly using Lemma 5.2.2 as the compatibility assumption is not met. However the equality (5.7) turns out to be valid nonetheless.

**Lemma 5.2.4.** *Consider the problem (5.4), with the choice  $g = 1$ ,  $h = 0$ . Even if the compatibility hypothesis  $g(\mathbf{x}, T) = h(\mathbf{x})$ ,  $\mathbf{x} \in \Gamma_D$ , in Lemma 5.2.2 is not satisfied, equation (5.7) still holds.*

*Proof.* Denote the support of  $h : D \rightarrow \mathbb{R}$  as  $\text{supp}(h) = \{\mathbf{x} \in D : h(\mathbf{x}) \neq 0\}$  and consider a sequence of non-negative functions  $\{h_n(\mathbf{x})\}_{n \in \mathbb{N}}$  such that:

- $h_n(\mathbf{x}) = 1 \ \forall \mathbf{x} \in \Gamma_D$ ,
- $h_n \in \mathcal{C}(\bar{D})$  and  $1 \geq h_1(\mathbf{x}) \geq h_2(\mathbf{x}) \geq \dots \geq h_n(\mathbf{x}) \dots$ ,
- $\text{supp}(h_n) \subset \{\mathbf{x} : \text{dist}(\mathbf{x}, \Gamma_D) \leq 1/n\}$ .
- denote by  $\theta_{T,n}^{\sigma'}(\mathbf{x}, t, \omega)$  the solution of (5.4) corresponding to the choice  $g = 1$ ,  $h = h_n$ , for which the representation formula (5.5) holds:

$$\theta_{T,n}^{\sigma'}(\mathbf{x}, t, \omega) = \mathbb{P}^{\omega'}(\tau_{\mathbf{x},t}^{\sigma'}(\omega) \leq T | \omega) + \mathbb{E}^{\omega'}[h_n(X_{\mathbf{x},t}^{\sigma'}(T)) \mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) > T\}} | \omega]. \quad (5.10)$$

We now have to verify that  $\lim_{n \rightarrow \infty} \mathbb{E}^{\omega'} [h_n(X_{\mathbf{x},t}^{\sigma'}(T)) \mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) > T\}} | \omega] = 0$ . We begin by observing that

$$\mathbb{E}^{\omega'} [h_n(X_{\mathbf{x},t}^{\sigma'}(\tau_{\mathbf{x},t}^{\sigma'}(\omega))) \mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) > T\}} | \omega] < \mathbb{E}^{\omega'} [\mathbf{1}_{\{X_{\mathbf{x},t}^{\sigma'}(T) \in \text{supp}(h_n)\}} \mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) > T\}} | \omega].$$

Hence it is possible to bound the right hand side of this inequality by  $\mathbb{P}^{\omega'} \{X_{\mathbf{x},t}^{\sigma'}(T) \in \text{supp}(h_1)\} < \infty$  and use the dominate convergence theorem to obtain

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}^{\omega'} [h_n(X_{\mathbf{x},t}^{\sigma'}(T)) \mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) > T\}} | \omega] &= \mathbb{E}^{\omega'} [\lim_{n \rightarrow \infty} h_n(X_{\mathbf{x},t}^{\sigma'}(T)) \mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) > T\}} | \omega] \\ &= \mathbb{E}^{\omega'} [\mathbf{1}_{\{X_{\mathbf{x},t}^{\sigma'}(T) \in \Gamma_D\}} \mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) > T\}} | \omega] = 0, \end{aligned}$$

since either  $\tau_{\mathbf{x},t}^{\sigma'}(\omega) > T$  and  $X_{\mathbf{x},t}^{\sigma'}(T) \notin \Gamma_D$ , or  $\tau_{\mathbf{x},t}^{\sigma'}(\omega) \leq T$ , so that  $\mathbf{1}_{\{\tau_{\mathbf{x},t}^{\sigma'}(\omega) > T\}} = 0$ . In both cases we have that the limit for  $n \rightarrow \infty$  tends to 0 and equation (5.7) holds. Indeed observe that this argument shows a  $L^1(D \times [0, T])$  convergence of  $\theta_{T,n}^{\sigma'}(\mathbf{x}, t, \omega)$  to  $\theta_T^{\sigma'}(\mathbf{x}, t, \omega)$  as indicated in (5.7). Moreover, observe that the solution  $\theta_{T,n}^{\sigma'}$  of the PDE with final datum  $h_n$ , when  $n \rightarrow \infty$ , converges in  $L^2([0, T], H^1(D))$ .  $\square$

### Vanishing diffusion limit

In many practical situations the values of  $\sigma'$  are typically very small. Therefore we will also investigate the case of a vanishing diffusion coefficient  $\sigma' = 0$ . In this case, of course, since we are changing the nature of the mathematical model we will have to properly modify the boundary conditions to deal with the resulting hyperbolic equations. It will then be crucial to be able to identify, possibly uniformly with respect to all possible Darcy velocities, the inflow boundary from which the characteristics start. From the pathwise SDE point of view, by denoting with  $X_{\mathbf{x},t} = X_{\mathbf{x},t}^0$  the solution of (5.1) in the vanishing diffusion case, it is possible to show that  $X_{\mathbf{x},t}^{\sigma'}$  converges to  $X_{\mathbf{x},t}$  when  $\mathbf{u} \in \{C^1(D)\}^2$ ; this can be done by applying Gronwall's Lemma to the equation

$$\begin{aligned} |X_{\mathbf{x},t}^{\sigma'}(s) - X_{\mathbf{x},t}(s)| &= \left| \int_t^s \mathbf{u}(X_{\mathbf{x},t}^{\sigma'}(r)) - \mathbf{u}(X_{\mathbf{x},t}(r)) dr + \sigma' W(s) \right| \\ &\leq \int_t^s \frac{|\mathbf{u}(X_{\mathbf{x},t}^{\sigma'}(r)) - \mathbf{u}(X_{\mathbf{x},t}(r))|}{|X_{\mathbf{x},t}^{\sigma'}(s) - X_{\mathbf{x},t}(s)|} |X_{\mathbf{x},t}^{\sigma'}(s) - X_{\mathbf{x},t}(s)| dr + |\sigma' W(s)| \\ &\leq \|\mathbf{u}\|_{\{C^1(D)\}^2} \int_t^s |X_{\mathbf{x},t}^{\sigma'}(s) - X_{\mathbf{x},t}(s)| dr + |\sigma' W(s)| \\ &\leq |\sigma' W(s)| \exp\{\|\mathbf{u}\|_{\{C^1(D)\}^2} (s - t)\}, \end{aligned}$$

upon noticing that the right hand side tends to zero when  $\sigma' \rightarrow 0$ . A similar result at the SDE level can not be recovered when dealing with rough velocities  $\mathbf{u} \in \{C^\alpha(D)\}^2$ ,  $\alpha < 1$ . On the other hand, if we consider the PDEs (5.4) and (5.8), the viscosity solutions are still well defined even in the case of vanishing diffusion. This can be seen, for instance, by applying again the dominated convergence theorem analogously to what was previously done in Lemma 5.2.4. However, as a consequence of the dominated convergence it is possible only to prove  $L^1(D \times [0, T])$  convergence. As the vanishing diffusion limits  $\theta_T(\mathbf{x}, \omega) = \theta_T^0(\mathbf{x}, \omega)$  and  $\bar{\tau}_{\mathbf{x}}(\omega) = \bar{\tau}_{\mathbf{x}}^0(\omega)$  could be discontinuous, it is not possible to prove a convergence result with respect to the  $\mathcal{C}^0$  norm.

### 5.2.2 Mixed case: $|\Gamma_N| > 0$

In this subsection we briefly address an extension of the previous results to the case of mixed (Dirichlet-Neumann) boundary conditions as in problem (5.4). Notice that in order to make the definition of the trajectory  $X_{\mathbf{x},t}^{\sigma'}(s, \omega)$  consistent with the reflecting conditions imposed on the Neumann boundary, we have to slightly modify the SDE (5.1) by considering an additional term that will be activated whenever a particle hits the Neumann boundary. To do so, we first need the following definition.

**Definition 5.2.6.** *The total variation of a function  $f : [a, b] \rightarrow \mathbb{R}$ ,  $a \leq b \in \mathbb{R}$ , which we denote by  $V_b^a f$ , is defined as*

$$V_b^a f = \sup_{n \in \mathbb{N}^+} \left( \sum_{i=0}^n |f(x_{i+1}) - f(x_i)| \right), \text{ with } x_0 = a \leq \dots \leq x_n = b.$$

We denote by  $BV([a, b], \mathbb{R})$  the corresponding space of functions with bounded total variation.

Now, suppose that we stop the process  $X_{\mathbf{x},t}^{\sigma'}(s, \omega)$  as soon as it hits the Dirichlet boundary. This process evolves according to the following reflected SDE (see [65]),

$$\begin{cases} X_{\mathbf{x},t}^{\sigma'}(s, \omega) + K_{\mathbf{x},t}(s, \omega) = \mathbf{x} + \int_t^s \mathbf{u}(X_{\mathbf{x},t}^{\sigma'}(r, \omega), \omega) dr + \int_t^s \sqrt{2}\sigma' dW(r, \omega'), \\ K_{\mathbf{x},t}(s, \omega) = \int_t^s \mathbf{n}(X_{\mathbf{x},t}^{\sigma'}(r, \omega)) \mathbf{1}_{\{X_{\mathbf{x},t}^{\sigma'}(s, \omega) \in \Gamma_N\}} dV_t^s |K_{\mathbf{x},t}(r, \omega)|, \end{cases} \quad (5.11)$$

until it is stopped. Here  $\mathbf{n}$  is the outward normal to the domain and  $K_{\mathbf{x},t}$  represents the process that ensures that the process  $X_{\mathbf{x},t}^{\sigma'}$  is reflected into  $D$  whenever it hits the Neumann boundary. The FPT  $\tau_{\mathbf{x}}^{\sigma'}(\omega)$  of the reflected process is then defined as in (5.2) but using the reflected trajectories (5.11). It is possible to use the same Keynman-Kac representation formula (5.5) to compute both the mean FPT and the exit probability from the domain. This is true since we are dealing with homogeneous boundary conditions. A more general formula dealing also with non homogeneous Neumann conditions can be found in [65].

In the next section we apply, and properly adapt, these ideas to the cases of an undisturbed flow from left to right and to a case of a flow induced by a pumping well that extracts fluid from the subsurface.

## 5.3 Case 1: an undisturbed flow from left to right

In this section we again consider the Darcy velocity obtained starting from equation (4.27), modeling an undisturbed flow from left to right on the square domain  $D = (0, 1)^2$ . Starting from such a velocity field, we now consider the arrival times to the outlet. Specifically, we define the arrival time as the time needed by the particle to pass the outlet  $\Gamma_{out} = \{\mathbf{x} \in \bar{D} \text{ s.t. } x_1 = 1\}$ . To model this situation at the SDE level we use a reflecting condition on the top and bottom boundaries  $\{\mathbf{x} \in \bar{D} \text{ s.t. } x_2 \in \{0, 1\}\}$  and also on the inlet  $\Gamma_{in} = \{\mathbf{x} \in \bar{D} \text{ s.t. } x_1 = 0\}$ , according to problem (5.11). Observe that the underlying velocity is tangential to the domain on the top and bottom boundaries. As the main transport goes from left to right it is reasonable to assume that for any particle leaving the domain from these two boundaries there will be another entering the domain with an opposite vertical velocity. Moreover, notice that in the vanishing diffusion limit, at least for smooth velocity fields  $\mathbf{u} \in \{C^1(D)\}^2$ , for which the pathwise uniqueness of trajectories is guaranteed, there will not be any particle starting from the interior of  $D$  that will touch those boundaries, nor will there be particles entering the domain

through these boundaries from the outside. In other words, the reflection never occurs. At the same time notice that the flow enters the domain from the inlet; in practice the flow enters the domain from here and continues towards the right; since the diffusion  $\sigma'$  we consider is quite small it is very unlikely (although not impossible) that a trajectory starting from a point  $\mathbf{x}$  at a certain distance to the inlet could go back and hit the inlet; even in the case this happens, it is reasonable to assume that the particle would be pushed back into the domain. This situation can be reasonably modeled via reflecting conditions since, especially for small diffusion, it is quite unlikely that particles would try to exit the domain from one of these three reflecting boundaries. Moreover such a modeling assumption would influence the solution only in the proximity of such boundaries. The only portion of  $\partial D$  from which particles can exit the domain is  $\Gamma_{out}$ . We define therefore our FPT again as in (5.2), by considering now a process  $X_{\mathbf{x},t}^\sigma$  that is reflected as soon as hits the top, bottom and left boundary. Specifically we will be interested in the computation of the time independent FPT  $\tau_{\mathbf{x}}^{\sigma'}(\omega) = \tau_{\mathbf{x},0}^{\sigma'}$  defined according to (5.2). As motivated in the previous section, we will tackle the problem using a PDE approach. In fact, by imposing Dirichlet conditions on  $\Gamma_{out}$  and homogeneous Neumann conditions on the other three edges of the square, we look for a suitable PDE whose solution represents our QoI.

According to what has been stated in the previous section, it is possible to compute the expected value of the FPT  $\tau_{\mathbf{x},t}^{\sigma'}(\omega)$  by solving the corresponding parabolic PDE. Since in this case the expected value of  $\tau_{\mathbf{x}}^{\sigma'}(\omega)$  is bounded as we have an underlying transport from left to right, it is computationally advantageous to consider the elliptic PDE instead of the parabolic one:

$$\begin{cases} \mathcal{L}_{\sigma'} \bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega) = -1 & \text{in } D, \\ \bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega) = 0 & \text{on } \Gamma_{out}, \\ \nabla \bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega) \cdot \mathbf{n} = 0 & \text{on } \partial D \setminus \Gamma_{out}. \end{cases} \quad (5.12)$$

If, on the other hand, we are interested in computing the probability of  $\tau_{\mathbf{x}}^{\sigma'}(\omega)$  being smaller than a prescribed  $T$  then we consider the parabolic PDE

$$\begin{cases} \frac{\partial \theta_T^{\sigma'}(\mathbf{x}, t, \omega)}{\partial t} + \mathcal{L}_{\sigma'} \theta_T^{\sigma'}(\mathbf{x}, t, \omega) = 0 & \text{in } D \times [0, T), \\ \theta_T^{\sigma'}(\mathbf{x}, t, \omega) = 1 & \text{on } \Gamma_{out} \times [0, T), \\ \partial_{\mathbf{n}} \theta_T^{\sigma'}(\mathbf{x}, t, \omega) = 0 & \text{on } \partial D \setminus \Gamma_{out} \times [0, T), \\ \theta_T^{\sigma'}(\mathbf{x}, T, \omega) = 0 & \text{on } D. \end{cases} \quad (5.13)$$

At this point we have to deal with quantities of interest which are random fields indexed with respect to space and time. As usual we consider a truncation of the input permeability field  $a$  used to compute the Darcy velocity, by parametrizing the random event  $\omega$  with  $\mathbf{y}$  (see Chapter 3). Therefore also  $\theta_T^{\sigma'}(\mathbf{x}, t, \omega)$  and  $\bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega)$  are evaluated as  $\theta_T^{\sigma'}(\mathbf{x}, t, \mathbf{y})$  and  $\bar{\tau}_{\mathbf{x}}^{\sigma'}(\mathbf{y})$  respectively.

Now the idea is to adapt the methodology of sparse grid approximations and MCCV methods to this class of problems (either (5.12) or (5.13)).

### 5.3.1 Spatial Discretization

Since we consider the case of small molecular diffusion  $\sigma'$ , equation (5.13) is typically transport dominated. Moreover when dealing with exit probabilities, the discontinuous condition on the parabolic boundary  $\Gamma_{out} \times [0, T) \cup D \times \{t = T\}$  induces a propagation of sharp fronts (in the vanishing diffusion limit the solution has a propagating discontinuity). We therefore use a discontinuous Galerkin (DG) formulation for its good stability properties (see [14, 66, 50]). Here we briefly introduce the DG weak formulation we used to approximate the PDE (5.13) or (5.12). First we introduce jumps and averages of scalar and vector valued functions across edges of the considered regular triangulation  $\mathcal{T}_h$ . Let  $e$  be an interior edge between the triangles  $T_1$  and  $T_2$ ,  $\mathcal{E}^i$  the set of all interior edges,  $\mathcal{E}^D$  the set of all Dirichlet edges and  $\mathcal{E}^N$  the set of all Neumann edges; moreover let  $\mathbf{n}_1$  and  $\mathbf{n}_2$  be the two normal vectors on  $e$  pointing exterior to  $T_1$  and  $T_2$  respectively. For a scalar piecewise linear function  $v$  on  $\mathcal{T}_h$ , such that  $v_i = v|_{T_i}$ , we define

$$\{v\} = \frac{1}{2}(v_1 + v_2), \quad [v] = v_1 \mathbf{n}_1 + v_2 \mathbf{n}_2 \quad \text{on } e \in \mathcal{E}^i. \quad (5.14)$$

Analogously for a vector valued function we define

$$\{\mathbf{v}\} = \frac{1}{2}(\mathbf{v}_1 + \mathbf{v}_2), \quad [\mathbf{v}] = \mathbf{v}_1 \cdot \mathbf{n}_1 + \mathbf{v}_2 \cdot \mathbf{n}_2 \quad \text{on } e \in \mathcal{E}^i. \quad (5.15)$$

If  $e \in \mathcal{E}^D \cup \mathcal{E}^N$  then  $\{v\} = v$ ,  $[v] = v\mathbf{n}$ ,  $\{\mathbf{v}\} = \mathbf{v}$  and  $[\mathbf{v}] = \mathbf{v} \cdot \mathbf{n}$  where there is no ambiguity of notation. With these definitions at hand we can define the bilinear forms and the functional that will be involved in the weak formulation: we have

$$a_h(w, v) = \sum_{T \in \mathcal{T}_h} \int_T (\sigma')^2 \nabla w \cdot \nabla v - w(\mathbf{u} \cdot \nabla v) d\mathbf{x}, \quad (5.16)$$

$$b_h(w, v) = \sum_{e \in \mathcal{E}^i \cup \mathcal{E}^D} \int_e (\sigma')^2 \{\nabla w\} \cdot [v] + \eta(\sigma')^2 \{\nabla v\} \cdot [w] + \frac{\beta}{|e|} [w][v] d\gamma - \sum_{e \in \mathcal{E}^i \cup \mathcal{E}^N} \int_e \{w\mathbf{u}\} \cdot [v] d\gamma, \quad (5.17)$$

$$\ell_h(v) = \sum_{e \in \mathcal{E}^D} \int_e g\mathbf{u} \cdot \mathbf{n} v d\gamma - \eta(\sigma')^2 \nabla v \cdot \mathbf{n} g + \frac{\beta}{|e|} v g d\gamma. \quad (5.18)$$

Here  $\beta > 0$  is a parameter penalizing jumps of the solution across the edges while  $\eta$  can be chosen equal to 0, 1 or  $-1$  depending on the desired treatment of the diffusion term. The spatial standard DG weak formulation of equation (5.4) therefore is: find  $\theta_{T,h}^{\sigma'}(\mathbf{x}, t, \mathbf{y}) \in V_h$  such that

$$\sum_{T \in \mathcal{T}_h} \int_T -\frac{\partial \theta_{T,h}^{\sigma'}(\mathbf{x}, t, \mathbf{y})}{\partial t} v d\mathbf{x} + a_h(\theta_{T,h}^{\sigma'}(\mathbf{x}, t, \mathbf{y}), v) + b_h(\theta_{T,h}^{\sigma'}(\mathbf{x}, t, \mathbf{y}), v) = \ell(v), \quad \forall v \in V_h, \quad a.e \text{ in } \Omega, \quad (5.19)$$

where  $V_h = \{v \in L^2(D) : v|_T \in \mathbb{P}^1(T), T \in \mathcal{T}_h\}$ . In the numerical results that we will present later, instead of considering the simple average  $\{\mathbf{u}w\}$  we actually considered the upwind value of  $\mathbf{u}w$ , namely

$$\{\mathbf{u}w\}^{up} = \begin{cases} \mathbf{u}w^1 & \text{if } \mathbf{u} \cdot \mathbf{n}_1 > 0, \\ \mathbf{u}w^2 & \text{if } \mathbf{u} \cdot \mathbf{n}_1 < 0, \\ \mathbf{u}\{w\} & \text{if } \mathbf{u} \cdot \mathbf{n}_1 = 0, \end{cases}$$



which leads to the DG upwind (DGUP) weak formulation of equation (5.4): find  $\theta_{T,h}^{\sigma'}(\mathbf{x}, t, \mathbf{y}) \in V_h$  such that

$$\sum_{T \in \mathcal{T}_h} \int_T -\frac{\partial \theta_{T,h}^{\sigma'}(\mathbf{x}, t, \mathbf{y})}{\partial t} v d\mathbf{x} + a_h(\theta_{T,h}^{\sigma'}(\mathbf{x}, t, \mathbf{y}), v) + b_h^{up}(\theta_{T,h}^{\sigma'}(\mathbf{x}, t, \mathbf{y}), v) = \ell(v), \quad \forall v \in V_h, \quad a.e \text{ in } \Omega \quad (5.20)$$

where  $b_h^{up}(w, v) = b_h(w, v) + \sum_{e \in \mathcal{E}^i} \int_e \frac{|\mathbf{u} \cdot \mathbf{n}|}{2} [w][v] d\gamma$ . This can be seen as a stabilized version of the original DG scheme. Concerning the convergence of such DGUP scheme, since we would like to have a uniform estimate with respect to  $\sigma'$ , we refer to the result given in [14] for pure transport equations: there the authors proved as, for a fixed transport  $\mathbf{u}$ , the error between the exact and the DGUP solution verifies the following estimate:

$$\|\theta_{T,h}^{\sigma'}(\cdot, t, \mathbf{y}) - \theta_T^{\sigma'}(\cdot, t, \mathbf{y})\|_{DG(D)} \leq C(t) h^{\frac{3}{2}}$$

where  $\|\cdot\|_{DG(D)} = \left( \|\cdot\|_{L^2(D)} + \sum_{e \in \mathcal{E}_h} \int_e \frac{|\mathbf{u} \cdot \mathbf{n}|}{2} [\cdot][\cdot] d\gamma \right)^{\frac{1}{2}}$ . Notice however that such an estimate has been proved only for a  $\mathcal{C}^1(\bar{D})$  transport field and under suitable regularity assumptions on  $\theta_{T,h}^{\sigma'}(\cdot, t, \mathbf{y})$  (see [14]).

To conclude this section, to actually use this DGUP scheme we perform an implicit treatment of the time derivative and solve for any time step a spatial problem as in (5.20).

### 5.3.2 Approximation of the mean FPT through MLMCCV

So far we introduced a parabolic PDE to approximate the probability of the FPT being smaller than  $T$  and an elliptic PDE to compute the expected value of the FPT. We are still left with the choice of a numerical method for the (external) expectation with respect to the random Darcy velocity (hence the random log-permeability). Let us focus on the expected arrival time  $\bar{\tau}_{\mathbf{x}}^{\sigma'}(\mathbf{y})$ . To approximate this function, it is again possible to use a sparse grid scheme or a MC-based one. It is hence crucial to understand whether or not this quantity is smooth with respect to  $\mathbf{y}$ . It turns out that, a smooth velocity  $\mathbf{u}$  generates smooth arrival times, at least for a sufficiently large time horizon. In fact, for the parabolic problem (5.13), we obtain non smooth solutions for sufficiently small time horizon since, due to the incompatibility between the boundary condition  $g = t$  and the final datum  $T$ , we have a non smooth front evolving towards the left. However, as soon as we consider sufficiently large time horizon (hence also for the elliptic problem (5.12)), such front exits the domain from the left boundary  $\Gamma_{in}$  and the regularity of the PDE solution depends only on the regularity of  $\mathbf{u}$ . This can be seen by taking derivatives of the parabolic PDE with respect to the stochastic parameters. Indeed:

$$\frac{\partial(\partial_{y_k} \theta_T^{\sigma'}(\mathbf{x}, t, \mathbf{y}))}{\partial t} + \mathbf{u} \cdot \nabla(\partial_{y_k} \theta_T^{\sigma'}(\mathbf{x}, t, \mathbf{y})) + (\sigma')^2 \Delta(\partial_{y_k} \theta_T^{\sigma'}(\mathbf{x}, t, \mathbf{y})) = -\partial_{y_k} \mathbf{u} \cdot \nabla \theta_T^{\sigma'}(\mathbf{x}, t, \mathbf{y}),$$

so that, by multiplying the equation by  $\partial_{y_k} \theta_T^{\sigma'}(\mathbf{x}, t, \mathbf{y})$  and integrating over  $D$ , we get

$$\frac{1}{2} \frac{d \|\partial_{y_k} \theta_T^{\sigma'}(\mathbf{x}, t, \mathbf{y})\|_{L^2(D)}^2}{dt} - (\sigma')^2 \|\nabla(\partial_{y_k} \theta_T^{\sigma'}(\mathbf{x}, t, \mathbf{y}))\|_{L^2(D)}^2 = - \int_D \partial_{y_k} \mathbf{u} \cdot \nabla \theta_T^{\sigma'}(\mathbf{x}, t, \mathbf{y}) \partial_{y_k} \theta_T^{\sigma'}(\mathbf{x}, t, \mathbf{y}) d\mathbf{x}.$$



### 5.3. Case 1: an undisturbed flow from left to right

Now, integrating with respect to the time from  $t$  to  $T$  we obtain

$$\begin{aligned} \frac{1}{2} \|\partial_{y_k} \theta_T^{\sigma'}(\mathbf{x}, t, \mathbf{y})\|_{L^2(D)}^2 + \int_t^T (\sigma')^2 \|\nabla(\partial_{y_k} \theta_T^{\sigma'}(\mathbf{x}, s, \mathbf{y}))\|_{L^2(D)}^2 ds \leq \frac{1}{2} \|\partial_{y_k} \theta_T^{\sigma'}(\mathbf{x}, T, \mathbf{y})\|_{L^2(D)}^2 \\ + \int_t^T \|\partial_{y_k} \mathbf{u}\|_{L^\infty(\bar{D})} \|\nabla \theta_T^{\sigma'}(\mathbf{x}, s, \mathbf{y})\|_{L^2(D)} \|\partial_{y_k} \theta_T^{\sigma'}(\mathbf{x}, ts, \mathbf{y})\|_{L^2(D)} ds. \end{aligned}$$

We observe that, in order to obtain a bound on the derivative of the solution with respect to one stochastic parameter, we need extra spatial regularity of the solution of the PDE, i.e.  $\nabla \theta_T^{\sigma'}(\mathbf{x}, s, \mathbf{y}) \in L^2(D)$ , and at least one derivative with respect to  $y_k$ , i.e.  $\|\partial_{y_k} \mathbf{u}\|_{L^\infty(\bar{D})}$ , so that a regularity result for  $\partial_{y_k} \theta_T^{\sigma'}(\mathbf{x}, t, \mathbf{y})$  can be recovered by using the Gronwall Lemma. Notice that this argument also applies

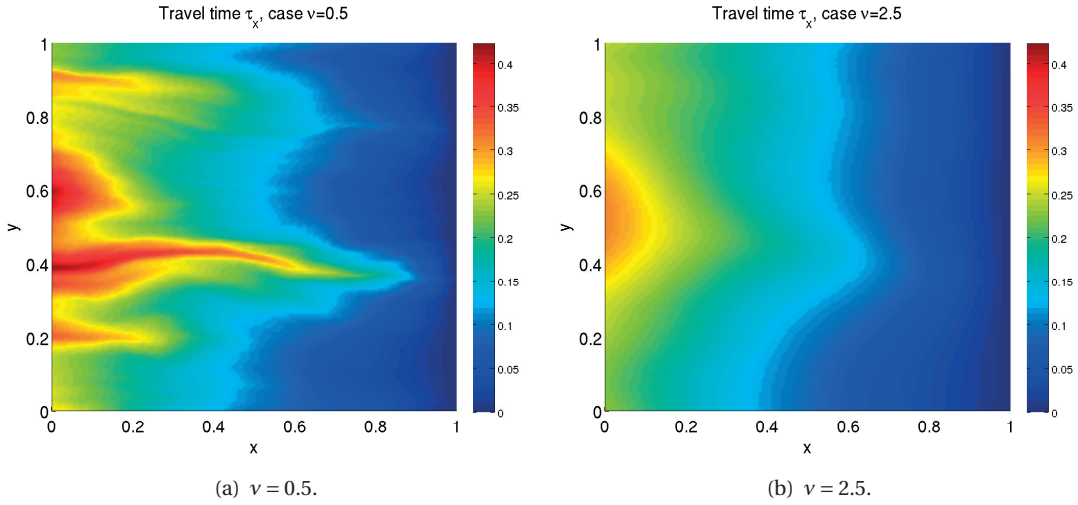


Figure 5.1 – Two realizations of the mean FPT through the outlet in the vanishing diffusion case for  $\nu = 0.5$  (left) and  $\nu = 2.5$  (right). Observe that the random event  $\mathbf{y}$  used to generate the realizations is the same.

to the parabolic PDE in the case of vanishing diffusion, of course by losing the control on the  $H^1(D)$  norm of the solution, since the second order PDE degenerates to a first order one. Figure 5.1 shows two realizations of  $\bar{\tau}_{\mathbf{x}}(\mathbf{y})$  obtained starting with a rough ( $C^{1/2}(\bar{D})$ , on the left) and a smooth ( $C^{5/2}(\bar{D})$ , on the right) velocity field. As predicted, also in the vanishing diffusion limit, the smoothness of the arrival times reflects the one of the transport field. These considerations make the problems (both the parabolic and the elliptic) suitable for a sparse grid approximation of the FPT; however, since in the former case we have to consider a large time horizon  $T$ , we rather prefer to focus on the elliptic problem and look at the sparse grid approximation of the arrival time  $\bar{\tau}_{\mathbf{x}}^{\sigma', \epsilon}(\mathbf{y})$  whenever the input Darcy velocity is smooth, i.e. when the input permeability  $a$  is smooth. On the other hand, when dealing with rough permeabilities, we can introduce an auxiliary problem with a regularized velocity  $\mathbf{u}^\epsilon$  and then consider the corresponding smoothed arrival time  $\bar{\tau}_{\mathbf{x}}^{\sigma', \epsilon}(\mathbf{y})$ , satisfying

$$\begin{cases} \mathbf{u}^\epsilon \nabla \bar{\tau}_{\mathbf{x}}^{\sigma', \epsilon}(\mathbf{y}) + (\sigma')^2 \Delta \bar{\tau}_{\mathbf{x}}^{\sigma', \epsilon}(\mathbf{y}) = -1 & \text{in } D, \\ \bar{\tau}_{\mathbf{x}}^{\sigma', \epsilon}(\mathbf{y}) = 0 & \text{on } \Gamma_{out}, \\ \nabla \bar{\tau}_{\mathbf{x}}^{\sigma', \epsilon}(\mathbf{y}) \cdot \mathbf{n} = 0 & \text{on } \partial D \setminus \Gamma_{out}. \end{cases} \quad (5.21)$$

The expansion (in  $\mathbf{y}$ ) of the mean arrival time can be effectively approximated by an adaptive sparse

grid. Notice that, since the QoI is now a random field and not a random variable anymore, in order to compute the profit associated to each multi-index we have to use a spatial norm to quantify the difference between the current sparse grid approximation and the one obtained by adding the current multi-index to the current set  $\mathbf{I}$ . We decided to use an  $L^1(D)$  norm to tackle this issue.

We can now introduce a MC-type scheme with a control variate variance reduction technique by defining

$$\bar{\tau}_{\mathbf{x}}^{\sigma', CV}(\mathbf{y}) = \bar{\tau}_{\mathbf{x}}^{\sigma'}(\mathbf{y}) - \bar{\tau}_{\mathbf{x}}^{\sigma', \epsilon}(\mathbf{y}) + \mathbb{E}[\bar{\tau}_{\mathbf{x}}^{\sigma', \epsilon}(\cdot)]; \quad (5.22)$$

By following the strategy proposed in Chapter 4, we define the MLMCCV estimator as

$$\hat{\tau}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV}(\mathbf{x}) = \sum_{\ell=0}^L \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} \left( \bar{\tau}_{\ell}^{\sigma', CV}(\mathbf{x}, \mathbf{y}_{i, \ell}) - \bar{\tau}_{\ell-1}^{\sigma', CV}(\mathbf{x}, \mathbf{y}_{i, \ell}) \right) + \mathcal{Q}_W[\bar{\tau}_L^{\sigma', \epsilon}(\mathbf{x}, \cdot)], \quad \bar{\tau}_{-1}^{\sigma', CV}(\mathbf{x}, \mathbf{y}_{i, \ell}) = 0, \quad (5.23)$$

where, as usual,  $\mathbf{y}_{i, \ell}$  for  $\ell = 0, \dots, L$  and  $i = 1, \dots, M_\ell$ , are independent identically distributed standard Gaussian random variables, and  $\bar{\tau}_{\ell}^{\sigma', CV}$  denotes the QoI evaluated on level  $\ell$  with mesh size  $h_\ell$ . Analogously we obtain a mean square error given by the sum of the statistical, the sparse grid and the finite element error.

**Lemma 5.3.1.** *The pointwise mean square error of the estimator (5.23) can be bounded by*

$$\begin{aligned} e(\hat{\tau}_{\{h_\ell\}, \{M_\ell\}, W}^{MLMCCV}(\mathbf{x}))^2 &\leq \sum_{\ell=0}^L \frac{\mathbb{V}ar(\bar{\tau}_{\ell}^{\sigma', CV}(\mathbf{x}, \mathbf{y}_{i, \ell}) - \bar{\tau}_{\ell-1}^{\sigma', CV}(\mathbf{x}, \mathbf{y}_{i, \ell}))}{M_\ell} + 2 \left( \mathbb{E}[\bar{\tau}_L^{\sigma', \epsilon}(\mathbf{x}, \cdot)] - \mathcal{Q}_W[\bar{\tau}_L^{\sigma', \epsilon}(\mathbf{x}, \cdot)] \right)^2 \\ &\quad + 2 \mathbb{E}[\bar{\tau}_L^{\sigma'}(\mathbf{x}, \cdot) - \bar{\tau}^{\sigma'}(\mathbf{x}, \cdot)]^2. \end{aligned} \quad (5.24)$$

### 5.3.3 Delineation of capture zones for small temporal horizon by stochastic collocation on mean FPT

Let us focus now on the computation of the probability of the arrival time being smaller than a prescribed time  $T$ . In particular we want to delineate the so called capture zones, i.e. the zone starting from which a particle will reach the boundary  $\Gamma_{out}$  with a certain probability in a prescribed time horizon  $T$ , that is:

$$\Sigma_{T, \alpha} = \{\mathbf{x} : \mathbb{P}^\omega(\tau_{\mathbf{x}, 0}^{\sigma'}(\omega) \leq T) \geq \alpha\}, \quad 0 < \alpha < 1. \quad (5.25)$$

The most natural approach to compute such a capture zone (5.25) is to solve the parabolic equation (5.13) for each realization of the Darcy velocity and then approximate the expected value with respect to  $\omega$  with a MC or MLMC scheme. This is the most straightforward strategy even if it is the most costly one, as it entails the solution of a time dependent PDE for each realization of the Darcy velocity. Moreover, the solution of this PDE is not a smooth function of the ‘‘Darcy’’ randomness, at least in the vanishing diffusion limit. This is due to the incompatible boundary data ( $\theta_T^{\sigma'}(\mathbf{x}, t) = 1$  on  $\Gamma_{out}$  and final datum  $\theta_T^{\sigma'}(\mathbf{x}, T) = 0$  on  $D$ ) which induce a discontinuity in the limit case  $\sigma' = 0$  propagating at ‘‘random’’ speed.

In the case of non vanishing diffusion  $\sigma' > 0$ , the solution depends analytically on the random

### 5.3. Case 1: an undisturbed flow from left to right

variables  $\mathbf{y}$  parametrizing the Darcy flow (see [62]), however with very large growth of derivatives (due to the sharp fronts), and, equivalently, very small size of the analyticity region. Hence, sparse grid approximation schemes are not suited to tackle this problem. For the same reason, also the control variate approach does not bring any significant improvement with respect to the standard MLMC scheme. In Figure 5.2 we can see how the solution introduced in (5.7) evolves in time,

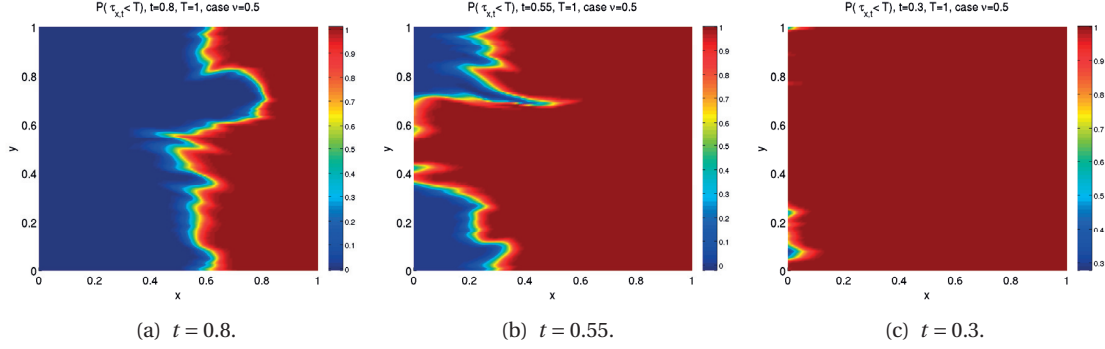


Figure 5.2 –  $\mathbb{P}^{\omega'}(\tau_{\mathbf{x},t}^{\sigma'} \leq T | \omega)$  for different values of  $t$ .  $\sigma = 1$ ,  $p_0 = 2$ ,  $\sigma' = 0$ ,  $\nu = 0.5$ ,  $L_c = 0.5$ .

basically propagating backwards according to the Darcy velocity until the propagating front covers the whole domain. This fact suggests that it does not make sense to consider an infinite time horizon  $T$  since all the particles will reach, sooner or later, the outlet in both cases with or without molecular diffusion. Consistently, also the probability front will reach the inlet so that a uniform unit probability will be obtained for sufficiently large time horizons.

In the case of vanishing diffusion it is preferable to use again an approach based on arrival times, which allows the use of sparse grid approximations. As we will see later on in the numerical results, the sparse grid is actually very effective in approximating  $\bar{\tau}_{\mathbf{x}}(\mathbf{y})$  when dealing with a smooth velocity field (respectively  $\bar{\tau}_{\mathbf{x}}^{\varepsilon}(\mathbf{y})$  when starting from rough velocities and considering smoothed versions to build a control variate for MC simulations). In this case the probability of the arrival time being smaller than  $T$  can be equivalently written as

$$\mathbb{E}^{\omega}[\mathbf{1}_{\{\tau_{\mathbf{x}}(\cdot) \leq T\}}] = \mathbb{E}^{\omega}[\mathbf{1}_{\{\bar{\tau}_{\mathbf{x}}(\cdot) \leq T\}}].$$

Therefore, for smooth Darcy velocities,  $\bar{\tau}_{\mathbf{x}}$  is a smooth function of  $\mathbf{y}$  and can be effectively approximated by a sparse grid scheme. Hence we propose the following approximation

$$\mathbb{E}^{\omega}[\mathbf{1}_{\{\bar{\tau}_{\mathbf{x}}(\cdot) \leq T^*\}}] \approx \mathbb{E}^{\omega}[\mathbf{1}_{\{\mathcal{I}_1[\bar{\tau}_{\mathbf{x}}](\cdot) \leq T^*\}}] \approx \frac{1}{M} \sum_{i=1}^M \mathbf{1}_{\{\mathcal{I}_1[\bar{\tau}_{\mathbf{x}}](\mathbf{y}_i) < T^*\}} \quad (5.26)$$

where  $\mathcal{I}_1$  denotes the sparse grid interpolant defined in (3.4). In practice, since there is no diffusion, the solution of the PDE (5.8) is not an expectation anymore (the corresponding ODE (5.1) has deterministic trajectories); therefore, the probability over all possible Brownian motions and Darcy velocities becomes an expected value only with respect to  $\omega$ . At this point, since the sparse grid is able to accurately approximate the (deterministic) arrival times, we can replace  $\bar{\tau}_{\mathbf{x}}$  with its sparse grid interpolant by committing a small error. Since in the previous chapter we showed that the sparse grid is well suited not only to perform quadrature but also interpolation, we now use a MC scheme to estimate  $\mathbb{E}^{\omega}[\mathbf{1}_{\{\mathcal{I}_1[\bar{\tau}_{\mathbf{x}}](\cdot) \leq T\}}]$ , i.e. we sample from the sparse grid interpolant; the idea is to operate as follows:

1. compute the sparse grid interpolant of  $\bar{\tau}_{\mathbf{x}}$ ;
2. given a new sample  $\mathbf{y}_i$ , compute the new sample of the FPT by sparse grid interpolation, i.e. by evaluating the sparse grid interpolant  $\mathcal{S}_1[\bar{\tau}_{\mathbf{x}}]$  in  $\mathbf{y}_i$ ;
3. iterate the procedure to compute a MC or MLMC estimator of the mean of the indicator function.

It is clear that a MC scheme does not present good convergence properties in terms of asymptotic rates but, since the cost of obtaining a sample in this case is significantly smaller than the one we needed to solve a PDE, such strategy outperforms the others in terms of computational cost. Finally the capture zone (5.25) is computed by looking at the level-set of the function  $\mathbb{E}^\omega[\mathbf{1}_{\{\mathcal{S}_1[\bar{\tau}_{\mathbf{x}}](\cdot) \leq T\}}]$ .

Observe that the estimate in (5.26) can be useful also when dealing with positive molecular diffusion  $\sigma'$ . In this case, by using again the solution of the elliptic PDE that gives us  $\bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega) = \mathbb{E}^{\omega'}[\tau_{\mathbf{x}}^{\sigma'}(\omega)|\omega]$ , thanks to [1, Theorem 2.4] we obtain

$$\begin{aligned} |\mathbb{E}^\omega[\mathbb{E}^{\omega'}[\mathbf{1}_{\{\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T\}}|\omega]] - \mathbb{E}^\omega[\mathbf{1}_{\{\bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega) \leq T\}}|\omega]]| &= |\mathbb{E}^\omega[\mathbb{E}^{\omega'}[\mathbf{1}_{\{\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T\}} - \mathbf{1}_{\{\bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega) \leq T\}}|\omega]]| \\ &\leq C \mathbb{E}^\omega[\|\tau_{\mathbf{x},0}^{\sigma'}(\omega) - \bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega)\|_{L^2(\Omega')}^{\frac{2}{3}}]. \end{aligned} \quad (5.27)$$

Such an estimate shows the error one would commit by approximating the correct probability  $\mathbb{P}^\omega(\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T) = \mathbb{E}^\omega[\mathbb{E}^{\omega'}[\mathbf{1}_{\{\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T\}}|\omega]]$  with  $\mathbb{E}^\omega[\mathbb{E}^{\omega'}[\mathbf{1}_{\{\bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega) \leq T\}}|\omega]]$ . This error depends on the variance (with respect to all Brownian motions) of the arrival time, i.e.  $\|\tau_{\mathbf{x},0}^{\sigma'}(\omega) - \bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega)\|_{L^2(\Omega')}$ . Observe also that the constant  $C$  in the above bound depends on  $\sigma'$  and on the  $L^\infty$  norm of the probability density function of  $\tau_{\mathbf{x},0}^{\sigma'}(\cdot, \omega)$  (conditioned to the event  $\omega$ ).

### 5.3.4 Numerical results

In this subsection we present the numerical results concerning the case of an undisturbed flow from left to right. As previously mentioned, when starting from a smooth velocity field it is possible to use a sparse grid approximation to compute the expected value of  $\bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega)$ . In this case the computational cost to assemble the sparse grid is not so expensive since we have to solve an elliptic problem for each node of the sparse grid (similar to the case of a standard Darcy problem). We first present the convergence plot of the sparse grid approximation error; we consider a “deltaint”-based adaptive sparse grid with non nested knots; by denoting our QoI  $Q(\mathbf{x}, \omega) = \bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega)$  we measure the sparse grid error as

$$\|\mathbb{E}[Q(\mathbf{x}, \omega)] - \mathcal{Q}_W[Q(\mathbf{x}, \omega)]\|_{L^2(D)} \approx \|\mathcal{Q}_{W*}[Q(\mathbf{x}, \omega)] - \mathcal{Q}_W[Q(\mathbf{x}, \omega)]\|_{L^2(D)}$$

being  $\mathcal{Q}_{W*}[Q(\mathbf{x}, \omega)]$  our reference quadrature, computed with a sparse grid built over 3593 Hermite points corresponding to a work  $W* = 4017$ . We recall that in this case during the sparse grid construction, to deal with random fields, the profits  $P(\mathbf{i})$  introduced in (3.9) through (3.17) have been replaced with

$$\Delta E(\mathbf{i}) = \|\mathbb{E}[\Delta^{\mathbf{m}(\mathbf{i})}[Q]]\|_{L^1(D)} = \|\mathcal{Q}_{\mathbf{j}}^{\mathbf{m}}[Q] - \mathcal{Q}_{\mathbf{i}}^{\mathbf{m}}[Q]\|_{L^1(D)}.$$

In Figure 5.3 we can observe how the sparse grid convergence behaves well, significantly improving the performance of a standard MC scheme when dealing with smooth arrival times.

### 5.3. Case 1: an undisturbed flow from left to right

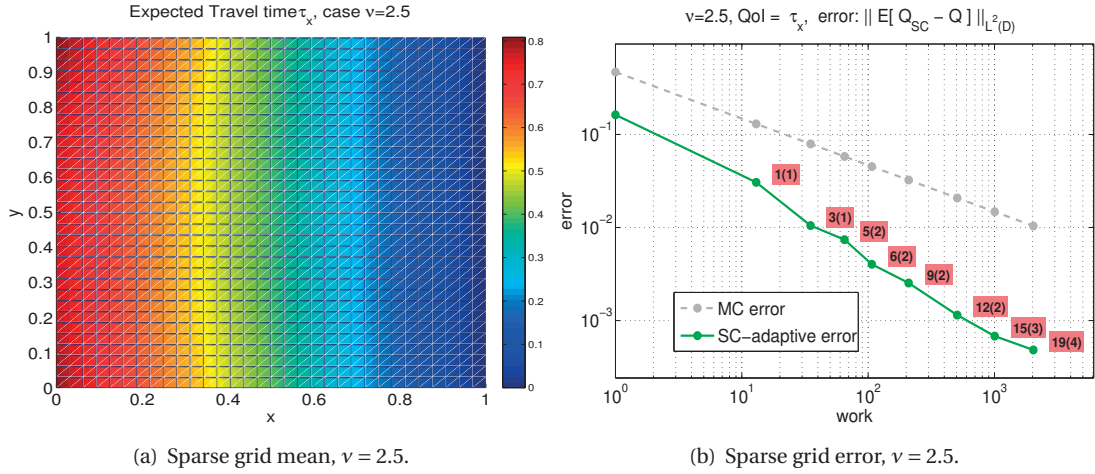


Figure 5.3 – Case  $v = 2.5$ : on the left the sparse grid approximation of  $\mathbb{E}^\omega[Q(\mathbf{x}, \cdot)]$ ; on the right the sparse grid error committed on such quantity in the  $L^2$  spatial norm versus the work evaluated as number of solves needed to build such approximation.  $\sigma' = 0.01$ ,  $L_C = 0.5$ ,  $p_0 = 1$ .

As previously anticipated, the smoothness of  $Q(\mathbf{x}, \mathbf{y})$  in  $\mathbf{y}$  is strictly related to the smoothness of the transport  $\mathbf{u}$ , and then to the smoothness of the realizations of the permeability field, i.e. we have smooth realizations of  $Q$  (respectively rough) when starting from a smooth (rough) transport  $\mathbf{u}$ . It is therefore possible to apply a MLMCCV strategy to this problem. In Figure 5.4 and 5.5 we show the variance reduction obtained by considering a MLMCCV estimator with respect to the one obtained with a standard MLMC scheme. In Figure 5.4 we consider a rough input permeability that generates rough arrival times. What happens is consistent with the considerations done in the case of the Darcy problem. In fact, in this case the variance reduction seems to concern only the coarsest levels; on the other hand from a certain level on, depending on  $\epsilon$ , the variance of the MLMCCV and MLMC estimators become very close. Figure 5.4-(c) might suggest that the variance reduction actually appears on all levels; however, since the two curves (the MLMC and the MLMCCV ones) are getting closer and closer, we claim that for finer levels they should overlap, or at least behave as in Figure 5.4-(a,b). In Figure 5.5 we report the results obtained starting from a smooth input permeability

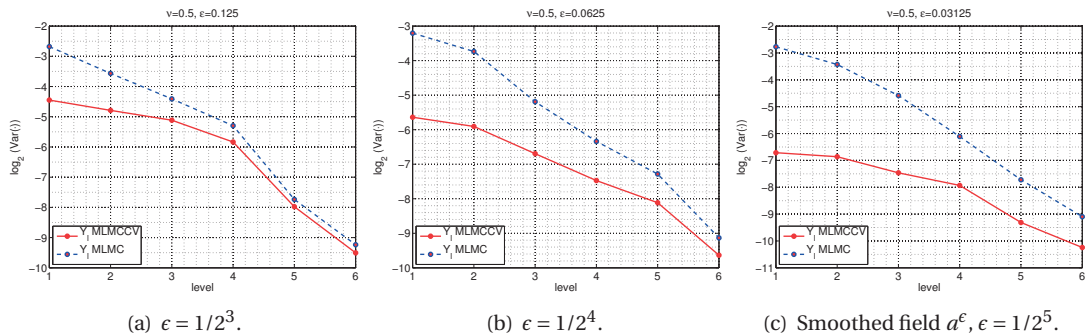


Figure 5.4 – Variance of  $Q_\ell$  and  $Y_\ell(\mathbf{x}, \omega) = Q_\ell(\mathbf{x}, \omega) - Q_{\ell-1}(\mathbf{x}, \omega)$  on each level, for  $v = 0.5$  and different values of  $\epsilon$ .  $\sigma' = 0.01$ ,  $L_C = 0.5$ ,  $p_0 = 1$ .

that generates smooth arrival times. Also in this case we obtain similar results to the ones presented in the Darcy problem: the variance reduction with respect to the MLMC case appears on all levels

and gets significantly larger when considering smaller and smaller values of  $\epsilon$ . In Figure 5.6 and

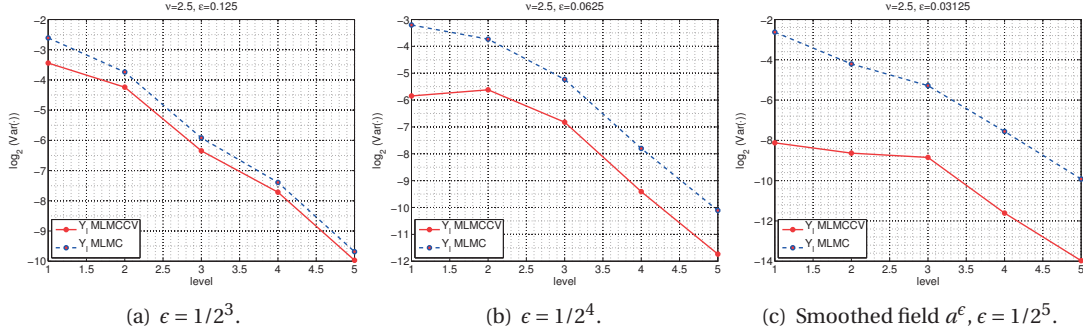


Figure 5.5 – Variance of  $Q_\ell$  and  $Y_\ell(\mathbf{x}, \omega) = Q_\ell(\mathbf{x}, \omega) - Q_{\ell-1}(\mathbf{x}, \omega)$  on each level, for  $\nu = 2.5$  and different values of  $\epsilon$ .  $\sigma' = 0.01$ ,  $L_C = 0.5$ ,  $p_0 = 1$ .

5.7 we report the corresponding variance reduction obtained in the case of vanishing diffusion. It

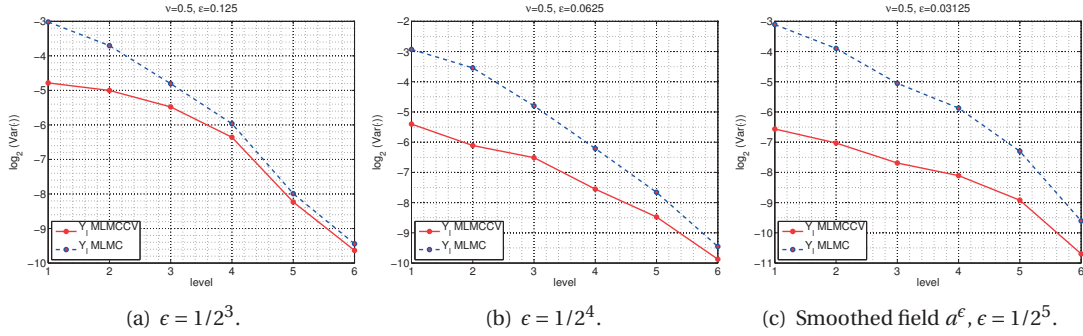


Figure 5.6 – Variance of  $Q_\ell$  and  $Y_\ell(\mathbf{x}, \omega) = Q_\ell(\mathbf{x}, \omega) - Q_{\ell-1}(\mathbf{x}, \omega)$  on each level, for  $\nu = 0.5$  and different values of  $\epsilon$ .  $\sigma' = 0$ ,  $L_C = 0.5$ ,  $p_0 = 1$ .

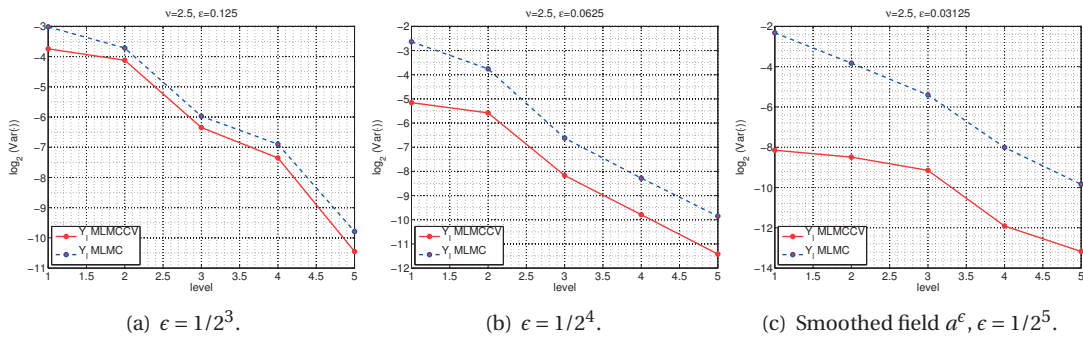


Figure 5.7 – Variance of  $Q_\ell$  and  $Y_\ell(\mathbf{x}, \omega) = Q_\ell(\mathbf{x}, \omega) - Q_{\ell-1}(\mathbf{x}, \omega)$  on each level, for  $\nu = 2.5$  and different values of  $\epsilon$ .  $\sigma' = 0$ ,  $L_C = 0.5$ ,  $p_0 = 1$ .

is possible to observe how, also in this case, the variance reduction obtained are comparable with the ones obtained with positive diffusion. The MLMCCV approach is then well suited to solve the problem of approximating the expected value of the arrival time to the outlet in all cases.

To approximate the capture zones (5.25) we have to simulate the probability of the arrival time



#### 5.4. Case 2: a flow induced by the presence of an extracting well

being smaller than  $T$ , i.e. the quantity  $\mathbb{P}(\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T)$ . Observe that for each realization of the Darcy velocity (even the smooth ones) the corresponding probability  $\mathbb{E}^{\omega'}[\mathbf{1}_{\{\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T\}} | \omega]$  presents a sharp front for small values of diffusion  $\sigma'$ . Therefore we could think to use a standard MLMC scheme to estimate the quantity  $\mathbb{P}(\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T)$  by denoting out QoI as  $Q(\mathbf{x}, \omega) = \mathbb{E}^{\omega'}[\mathbf{1}_{\{\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T\}} | \omega]$  and by considering the corresponding MLMC estimator according to (4.9). In Figure 5.8 we show the mean and the variance of  $Q$  computed with a MLMC scheme. As expected the probability of ending up

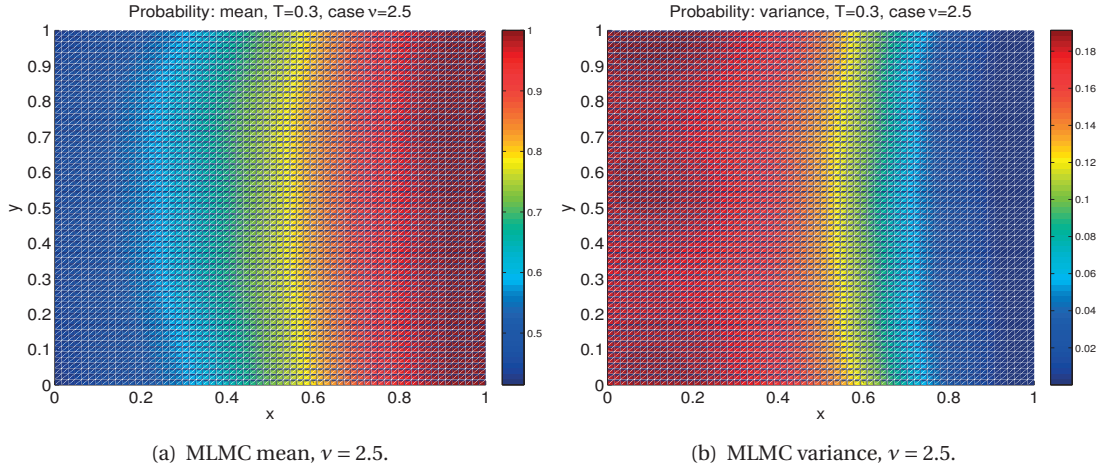


Figure 5.8 – Case  $\nu = 2.5$ : MLMC approximation of mean (left) and variance (right) of  $Q(\mathbf{x}, \omega) = \mathbb{E}^{\omega'}[\mathbf{1}_{\{\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T\}}]$ .  $\sigma' = 0$ ,  $L_C = 0.5$ ,  $p_0 = 3$ ,  $T = 0.3$ ,  $L = 5$ .

in the outlet in  $T$  time units decreases with the distance from  $\Gamma_{out}$  since the transport is mainly in the horizontal direction from left to right; by contrast the variance increases with the distance from  $\Gamma_{out}$  since these probability fronts, as Figure 5.2 suggests, reach different zones of the domain depending on the realization of the Darcy velocity. We do not report here the results concerning the variance reduction of the MLMC scheme applied to the solution of the parabolic PDE since they will be presented among the numerical results of the next section.

We outline again that it is possible to reuse all this methodology developed to effectively estimate the mean arrival time. In this case for each realization of the Darcy velocity it suffices to solve an elliptic PDE in order to obtain the quantity  $\bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega)$ ; since this arrival time has smooth realizations we have regularity also with respect to the stochastic parameters and we can successfully apply a sparse grid scheme to approximate the mean of  $\bar{\tau}_{\mathbf{x}}^{\sigma'}(\mathbf{y})$ ; then, according to what stated in the previous subsection, we can apply a convenient MC sampler to estimate  $\mathbb{E}^{\omega}[\tau_{\mathbf{x},0}^{\sigma'}(\omega)] = \mathbb{E}^{\omega}[\bar{\tau}_{\mathbf{x}}^{\sigma'}(\omega)]$  taking samples from the sparse grid interpolant  $\mathcal{I}_1[\bar{\tau}_{\mathbf{x}}^{\sigma'}]$ .

#### 5.4 Case 2: a flow induced by the presence of an extracting well

In this section we focus on a different scenario, more interesting from the application's point of view. We consider in fact an aquifer in which a well extracts drinkable water at a constant rate  $q$ . To model the presence of the pumping well located at  $\mathbf{x}_0$  and compute the corresponding flow we could either consider an external forcing term  $f(\mathbf{x}) = -q\delta_{\mathbf{x}_0}(\mathbf{x})$ , being  $\delta_{\mathbf{x}_0}$  the Dirac function centered in  $\mathbf{x}_0$ , and  $q$  the constant extraction rate, or consider a small hole around  $\mathbf{x}_0$  having the size of the physical well and Neumann conditions to enforce the total flux around the hole to be equal to  $q$ . In this case

we consider for simplicity a circular well of radius  $r \ll 1$ , and assume the flow in proximity of the well to be radial. By adding this “well effect” to the previous undisturbed flow case, we end up with the solution of the following Darcy problem

$$\begin{cases} -\operatorname{div}(a(\mathbf{x}, \omega) \nabla p(\mathbf{x}, \omega)) = 0, & \mathbf{x} \in D \setminus \bar{B}(\mathbf{x}_0, r), \\ p(\mathbf{x}, \omega) = p_0(1 - x_1) & \mathbf{x} \in x_1 = \{0, 1\}, \\ a(\mathbf{x}, \omega) \partial_{\mathbf{n}} p(\mathbf{x}, \omega) = 0 & \mathbf{x} \in x_2 = \{0, 1\}, \\ a(\mathbf{x}, \omega) \partial_{\mathbf{n}} p(\mathbf{x}, \omega) = \frac{q}{2\pi r} & \mathbf{x} \in \partial B(\mathbf{x}_0, r) \end{cases} \quad (5.28)$$

where  $B(\mathbf{x}_0, r)$  denotes the open ball of radius  $r$  and center  $\mathbf{x}_0$ . For  $r \rightarrow 0$ , it is then reasonable from a modeling point of view to consider the similar problem

$$\begin{cases} -\operatorname{div}(a(\mathbf{x}, \omega) \nabla p(\mathbf{x}, \omega)) = -q\delta_{\mathbf{x}_0}(\mathbf{x}), & \mathbf{x} \in D, \\ p(\mathbf{x}, \omega) = p_0(1 - x_1) & \mathbf{x} \in x_1 = \{0, 1\}, \\ a(\mathbf{x}, \omega) \partial_{\mathbf{n}} p(\mathbf{x}, \omega) = 0 & \mathbf{x} \in x_2 = \{0, 1\}. \end{cases} \quad (5.29)$$

Depending on how large the size of the physical well is with respect to the size of the domain one can use either equation (5.28) or (5.29) to model the flow. Notice that this problem, because of the  $\delta_{\mathbf{x}_0}(\mathbf{x})$  as right hand side, is not well posed in  $H^1(D)$ ; however, it can be rewritten in terms of  $\tilde{p}(\mathbf{x}, \omega) = p(\mathbf{x}, \omega) + G(\mathbf{x}, \omega)$ , being  $G(\mathbf{x}, \omega)$  the Green's function that solves  $-\operatorname{div}(a(\mathbf{x}_0, \omega) \nabla G(\mathbf{x}, \omega)) = -q\delta_{\mathbf{x}_0}(\mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^2$ , i.e.  $G(\mathbf{x}, \omega) = -\frac{q}{2\pi a(\mathbf{x}_0, \omega)} \log(|\mathbf{x} - \mathbf{x}_0|)$ . With this change of variable we obtain a problem for  $\tilde{p}$  which reads  $-\operatorname{div}(a(\mathbf{x}, \omega) \nabla \tilde{p}(\mathbf{x}, \omega)) = -\operatorname{div}((a(\mathbf{x}, \omega) - a(\mathbf{x}_0, \omega)) \nabla G(\mathbf{x}, \omega))$ ,  $\mathbf{x} \in D$  with proper boundary conditions. Now, since  $a \in C^\alpha(\bar{D})$ , the term  $a(\mathbf{x}, \omega) - a(\mathbf{x}_0, \omega)$  absorbs part of the singularity of  $\nabla G(\mathbf{x}, \omega)$  in  $\mathbf{x}_0$  making possible to recover the well posedness in  $H^1(D)$  for  $\tilde{p}$ . However, since we are interested in computing arrival times to the well, we need to find a way to properly define our arrival time. Observe that, depending on how large is the extraction rate  $q$ , some of the particles coming from the left part of the domain will be attracted by the well while others will continue their motion up to the right boundary  $\Gamma_{out}$ . To properly define the arrival time  $\tau_{\mathbf{x}, t}^{\sigma'}(\omega)$  to the well in the bounded domain  $D$ , i.e. when the Darcy velocity is given only in  $D$ , we chose to reflect the process  $X_{\mathbf{x}, t}^\sigma$  whenever it hits the external boundary  $\partial D$  and to absorb it as soon as it enters the internal ball  $B(\mathbf{x}_0, r)$ . Moreover we assume the domain to be large enough so that the perturbation caused by the extraction well does not influence the outlet  $\Gamma_{out}$ , i.e.  $u_1(\mathbf{x}, \omega) > 0$  for any  $\mathbf{x} \in \Gamma_{out}$ .

If we work with equation (5.28) we can define our FPT as

$$\tau_{\mathbf{x}, t}^{\sigma'}(\omega) = \inf(s \geq t : X_{\mathbf{x}, t}^\sigma(s) \in B(\mathbf{x}_0, r)). \quad (5.30)$$

Observe that, with this definition, the particles that will reach  $\Gamma_{out}$  will be reflected until the diffusion pushes them back to the well. Observe also that, in the limit of vanishing diffusion, for some starting points  $\mathbf{x} \in D$  the arrival time could be actually  $\infty$  since, for each realization of the transport, some particles might reach the outlet  $\Gamma_{out}$  instead of the well and, in the absence of diffusion, they can not diffuse back to the well.

On the other hand, when working with equations (5.29), to properly define the arrival time to the well we have to consider some bounded region surrounding the well small enough to assume that once a particle reaches such region it will reach the well with unit probability in a negligible time.



#### 5.4. Case 2: a flow induced by the presence of an extracting well

Observe that in this case the Darcy velocity presents a singularity in  $\mathbf{x} = \mathbf{x}_0$ , since the solution  $p$  of the Darcy problem (5.29) behaves locally around the well as the Green's function of the Laplace operator in a neighborhood of such point. Also in this case, then, we consider a small ball surrounding the location of the well and again we define the arrival time through equation (5.30), where now  $r$  does not represent the physical size of the well but it is rather an artificial barrier.

In any case we end up with the solution of a time-dependent PDE:

$$\begin{cases} \frac{\partial \theta_T^{\sigma'}(\mathbf{x}, t, \omega)}{\partial t} + \mathcal{L}_{\sigma'} \theta_T^{\sigma'}(\mathbf{x}, t, \omega) = 0 & \text{in } (D \setminus \bar{B}(\mathbf{x}_0, r)) \times [0, T), \\ \theta_T^{\sigma'}(\mathbf{x}, t, \omega) = g(\mathbf{x}, t) & \text{on } \partial B(\mathbf{x}_0, r) \times [0, T), \\ \partial_{\mathbf{n}} \theta_T^{\sigma'}(\mathbf{x}, t, \omega) = 0 & \text{on } \partial D \times [0, T), \\ \theta_T^{\sigma'}(\mathbf{x}, T, \omega) = h(\mathbf{x}) & \text{on } D. \end{cases} \quad (5.31)$$

As in the previous case of undisturbed flow, proper choices of the boundary and final conditions allow to link the solution of this time dependent PDE to the expectation of some quantities through the Feynman-Kac formula (5.5). Again the goals will be the computation of quantities related to the FPT as  $\mathbb{E}^\omega[\min(\tau_{\mathbf{x}_0}^{\sigma'}(\omega), T)]$  and of the probability to end up in the well in a time smaller than  $T$ , i.e.  $\mathbb{E}^\omega[\mathbf{1}_{\{\tau_{\mathbf{x}_0}^{\sigma'}(\omega) \leq T\}}] = \mathbb{P}^\omega(\tau_{\mathbf{x}_0}^{\sigma'}(\omega) \leq T)$ .

##### 5.4.1 A MLMC approach to approximate expected arrival times

Let us focus on the approximation of the mean of the arrival time  $\tau_{\mathbf{x},t}^{\sigma'}(\omega)$ , i.e.  $\mathbb{E}^\omega[\tau_{\mathbf{x},t}^{\sigma'}(\omega)]$  and consider again the PDE (5.31). By putting Dirichlet conditions on  $\partial B(\mathbf{x}_0, r)$  we are absorbing the process once it hits this boundary. On the other hand, we reflect it when it hits the other external boundaries. The more remarkable difference with respect to the undisturbed case is that now, in the vanishing diffusion case, not all the particles will eventually reach the actual outflow which is given by the Dirichlet boundary  $\partial B(\mathbf{x}_0, r)$ . Moreover, as previously anticipated, in the vanishing diffusion limit the arrival time defined in (5.30) can be infinite for some points  $\mathbf{x} \in D$ , while in the case  $\sigma' > 0$ , even if it is finite for any realization of the Brownian motion, it might not be uniformly bounded from above with respect to the random event  $\omega$  modeling the Darcy velocities. For practical purposes we are interested in approximating such arrival times only for those points belonging to trajectories that end up in the well under the transport of the Darcy velocity; since, according to our reflected SDE, also trajectories that hit first the outlet  $\Gamma_{out}$  reach eventually the well, even if this might induce extremely large associated arrival times, we rather prefer to keep bounded from above our quantity of interest by considering instead the more convenient quantity  $\mathbb{E}^\omega[\min(\tau_{\mathbf{x},t}^{\sigma'}(\omega), T)]$ .

By choosing again  $g = t$  and  $h = T$  in (5.31), we start from a solution constant at time  $t = T$ ; then, by moving back in time, the datum  $g = t$  on  $\partial B(0, r)$  starts propagating backward in time towards left driven by the Darcy velocity and at the same time under the effect of the diffusion  $(\sigma')^2$ .

**Remark 5.4.1.** A similar (yet different) way to model this situation in the general case  $\sigma' > 0$  was to put a Dirichlet absorbing condition also on  $\Gamma_{out}$  with corresponding Dirichlet datum equal to  $T$ ; in this case such datum, when retropropagating, would not generate any propagating front since it would match the final condition  $h = T$ . Such choice leads to the same solution of the parabolic PDE, i.e.  $\mathbb{E}^\omega[\min(\tau_{\mathbf{x}_0}^{\sigma'}(\omega), T)]$ , being now  $\tau_{\mathbf{x}_0}^{\sigma'}$  the time needed to reach either the internal ball  $B(\mathbf{x}_0, r)$  or the boundary  $\Gamma_{out}$ . Nevertheless, because of the two different Dirichlet boundary conditions, namely  $g = t$  on  $\partial B(\mathbf{x}_0, r)$  and  $g = T$  on  $\Gamma_{out}$ , it is not possible to write an equation for the mean FPT as in the

undisturbed case while, by putting Neumann conditions on all the external boundaries, it is possible to write such equation for the mean FPT and numerically verify how the expected arrival time is not bounded for any starting position  $\mathbf{x}$ .

Since close to the well the velocity is practically radial, as it is mainly driven by the presence of the pumping well, the solution of the PDE presents a front evolving first radially; then, at a small distance to the well, the Darcy's flow, which is left-to-right, becomes dominant and the front starts moving toward left, as suggested by Figure 5.1-(a). The solution of the parabolic PDE with this choice of boundary conditions (Neumann everywhere on  $\partial D$ ) has the following interpretation (as in (5.6))

$$\theta_T^{\sigma'}(\mathbf{x}, t, \omega) = \mathbb{E}^{\omega'}[\min\{\tau_{\mathbf{x},t}^{\sigma'}(\omega), T\}|\omega].$$

Observe again that the minimum in this formula plays a crucial role as the arrival time could actually be infinite for some points  $\mathbf{x}$ . Our final goal is to compute the expected value  $\mathbb{E}^{\omega}[\min(\tau_{\mathbf{x},0}^{\sigma}(\omega), T)]$ . This can be achieved by using conditional expectation  $\mathbb{E}^{\omega}[\min(\tau_{\mathbf{x},0}^{\sigma}(\omega), T)] = \mathbb{E}^{\omega}[\mathbb{E}^{\omega'}[\min(\tau_{\mathbf{x},0}^{\sigma}(\omega), T)|\omega]]$ . We have therefore to approximate the mapping  $\omega \rightarrow \mathbb{E}^{\omega'}[\min\{\tau_{\mathbf{x},t}^{\sigma'}(\omega, \omega'), T\}|\omega]$ . Unfortunately, we do not expect the quantity  $\mathbb{E}^{\omega'}[\min\{\tau_{\mathbf{x},0}^{\sigma'}(\omega, \omega'), T\}|\omega]$  to be smooth, as a function of the spatial variable  $\mathbf{x}$ , because of the presence of the minimum. This fact, according to the computations carried out in Subsection 5.3.2, implies that the function  $\mathbf{y} \rightarrow \mathbb{E}^{\omega'}[\min\{\tau_{\mathbf{x},0}^{\sigma'}(\mathbf{y}, \omega'), T\}|\mathbf{y}]$  is differentiable only in the case of non vanishing diffusion for any  $T$ , although with a  $\mathcal{C}^1$  norm which explodes as  $\sigma'$  tends to 0.

**Remark 5.4.2.** Observe that also in the undisturbed case the mapping  $\mathbf{x} \rightarrow \mathbb{E}^{\omega'}[\min\{\tau_{\mathbf{x},t}^{\sigma'}(\omega, \omega'), T\}|\omega]$  might not be smooth; this is the case for instance when considering small values of  $T$  for which the Dirichlet datum on  $\Gamma_{out}$  is not retropropagated far enough to reach the opposite boundary  $\Gamma_{in}$ . However in that case, since the vanishing diffusion characteristics coming from the Dirichlet boundary (with datum  $g = t$ ) cover all the domain, it is possible to choose  $T$  sufficiently large to get rid of the minimum function in (5.6) and obtain a smooth mapping. This can not be done in the case of the extraction well since, also for infinite time horizon  $T$ , in order to properly model the arrival time to the well, we can put on  $\Gamma_{out}$  either a homogeneous reflecting Neumann condition or a Dirichlet absorbing condition  $g = T$ , both leading to realizations characterized by the presence of sharp fronts.

Such considerations suggest that, in this case, the use of a sparse grid approximation of the map  $\mathbf{y} \rightarrow \mathbb{E}^{\omega'}[\min\{\tau_{\mathbf{x},t}^{\sigma'}(\mathbf{y}, \omega'), T\}|\mathbf{y}]$  is not recommended. Also the proposed variance reduction approach, that strongly depends on the possibility of using a sparse grid quadrature to successfully approximate the mean of the control variate, is not well suited. In Figure 5.1 we show two realizations of  $\mathbf{y} \rightarrow \mathbb{E}^{\omega'}[\min\{\tau_{\mathbf{x},t}^{\sigma'}(\mathbf{y}, \omega'), T\}|\mathbf{y}]$ , obtained starting from the same Darcy velocity and different molecular diffusions  $\sigma'$ . Looking at these plots it is clear that, although they have been generated with  $T = 0.5$ , these arrival times will not be smooth with respect to  $\mathbf{x}$  even if we considered a very large time horizon  $T$  since they present sharp fronts, especially the plot on the left, that separate the domain in two regions corresponding to points starting from which a particle will or will not eventually reach the well.

Denote our QoI as  $Q^{\sigma'}(x, 0, \mathbf{y}) = \mathbb{E}^{\omega'}[\min\{\tau_{\mathbf{x},0}^{\sigma'}(\mathbf{y}, \omega'), T\}|\mathbf{y}]$ . We propose in this case to use a MLMC strategy without any control variate technique to approximate the mean of  $Q$  as detailed in Chapter 4 in order to estimate the minimum between the arrival time and  $T$ .

Since we want to solve the parabolic PDE (5.31) also for very small diffusion, in order to avoid numerical oscillations on coarse grids, we consider a MLMC estimator with a level-dependent

#### 5.4. Case 2: a flow induced by the presence of an extracting well

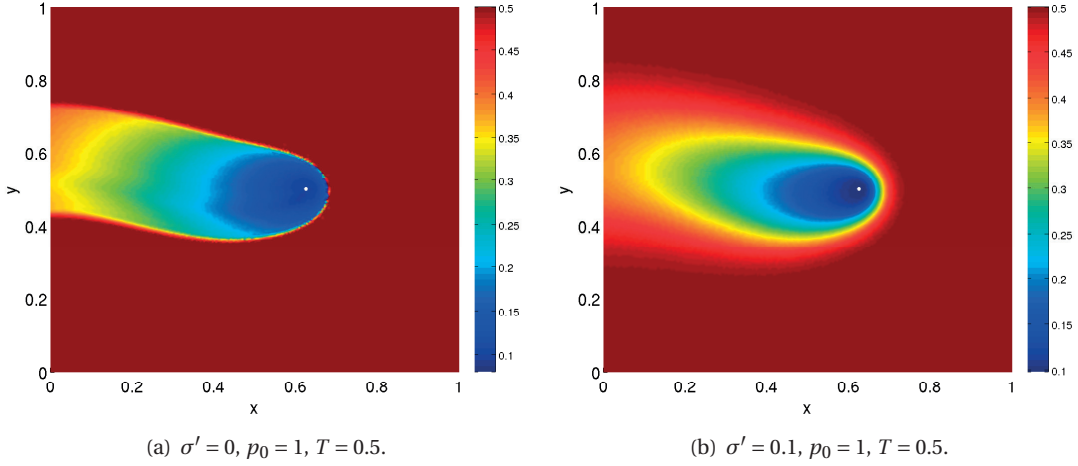


Figure 5.1 – Realizations of  $\mathbb{E}^{\omega'}[\min\{\tau_{\mathbf{x},t}^{\sigma'}(\mathbf{y}, \omega'), T\}|\mathbf{y}]$  for  $\nu = 2.5$  in the vanishing diffusion case (left) and for a relatively large diffusion (right). Observe that the random event  $\mathbf{y}$  used to generate the Darcy velocity is the same.

diffusion coefficient  $\sigma'$ . More precisely, we denote  $Q_\ell^{\sigma'}(\mathbf{y}_{\ell,i})$  the solution of the parabolic PDE (5.31) obtained for the  $i$ -th realization of the Darcy velocity, mesh size  $h_\ell$ , time step  $t_\ell \sim \sqrt{h_\ell}$  and level dependent diffusion  $\sigma'_\ell = \max(\sigma', c\sqrt{h_\ell})$ , with  $c > 0$  properly chosen. With this definition, we can build the MLMC estimator of  $Q^{\sigma'}$  as

$$\widehat{Q^{\sigma'}}_{\{h_\ell\}, \{M_\ell\}, \{\sigma'_\ell\}}^{MLMC}(\mathbf{x}) = \sum_{\ell=0}^L \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} \left( Q_{h_\ell}^{\sigma'_\ell}(\mathbf{x}, \mathbf{y}_{\ell,i}) - Q_{h_{\ell-1}}^{\sigma'_{\ell-1}}(\mathbf{x}, \mathbf{y}_{\ell,i}) \right), \text{ where } Q_{h_{-1}}^{\sigma'_{-1}} = 0, \quad (5.32)$$

where  $\mathbf{y}_{\ell,i}$  are independent identically distributed draws from  $\rho(\mathbf{y})$ ; the mean square error associated to this estimator, by omitting the dependence with respect to  $\mathbf{x}$ , is

$$e(\widehat{Q^{\sigma'}}_{\{h_\ell\}, \{M_\ell\}, \{\sigma'_\ell\}}^{MLMC})^2 = \mathbb{E}[(\widehat{Q^{\sigma'}}_{\{h_\ell\}, \{M_\ell\}, \{\sigma'_\ell\}}^{MLMC} - \mathbb{E}[Q^{\sigma'}])^2] = \sum_{\ell=0}^L \frac{\text{Var}(Q_{h_\ell}^{\sigma'_\ell}(\mathbf{y}_\ell) - Q_{h_{\ell-1}}^{\sigma'_{\ell-1}}(\mathbf{y}_\ell))}{M_\ell} + (\mathbb{E}[Q_{h_L}^{\sigma'_L} - Q^{\sigma'}])^2. \quad (5.33)$$

##### 5.4.2 Delineation of capture zones for small temporal horizon

Let us consider now the problem of the delineation of the so called capture zone, i.e. the zone starting from which a particle reaches the pumping well with at least a certain probability in a prescribed time horizon, namely

$$\Sigma_{T,\alpha} = \{\mathbf{x} : \mathbb{P}^{\omega'}(\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T) \geq \alpha\}, \quad (5.34)$$

with  $\alpha$  a positive (small) real number. To compute such capture zones (5.34) we start again from the parabolic PDE (5.31). By choosing  $g = 1$  and  $h = 0$  we obtain  $\theta_T^{\sigma'}(\mathbf{x}, t, \omega) = \mathbb{P}^{\omega'}(\tau_{\mathbf{x},t}^{\sigma'}(\omega) \leq T|\omega)$ . Again the mapping  $\mathbf{x} \rightarrow \mathbb{P}^{\omega'}(\tau_{\mathbf{x},t}^{\sigma'}(\omega) \leq T|\omega)$  is not smooth for the vanishing diffusion case, while in the diffusive case it presents degenerating derivatives (with respect to  $\mathbf{x}$ ) when  $\sigma'$  goes to 0. For small diffusion we have a sharp front evolving (see realizations in Figure 5.2); therefore the use of a sparse grid scheme to exploit the dependence  $\mathbf{y} \rightarrow \mathbb{P}^{\omega'}(\tau_{\mathbf{x},t}^{\sigma'}(\mathbf{y}) \leq T|\mathbf{y})$  is again not recommended, since the

corresponding analyticity region has a very small size.

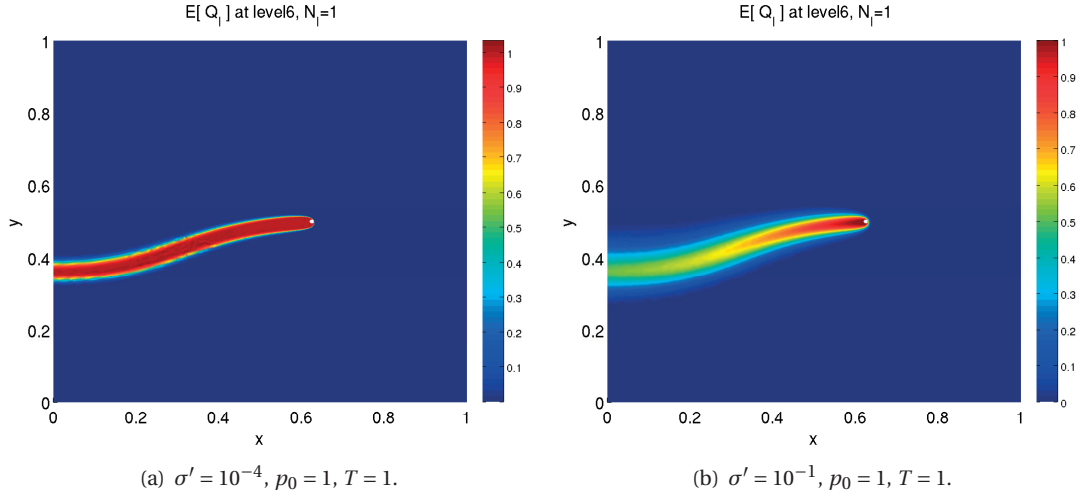


Figure 5.2 – Realization of  $\mathbb{P}^{\omega'}(\tau_{\mathbf{x},t}^{\sigma'}(\mathbf{y}) \leq T | \mathbf{y})$  for  $\nu = 2.5$  in the vanishing diffusion case (left) and for a relatively large diffusion (right). Observe that the random event  $\mathbf{y}$  used to generate the realizations is the same.

Again a MLMC scheme seems to be a more appropriate choice. Denoting our QoI as  $Q^{\sigma'}(x, \mathbf{y}) = \mathbb{P}^{\omega'}(\tau_{\mathbf{x},0}^{\sigma'}(\mathbf{y}) \leq T | \mathbf{y})$ , we propose a MLMC strategy on  $Q^{\sigma'}$  as detailed in the previous subsection by taking again samples  $Q_{\ell}^{\sigma'}(\mathbf{y}_{\ell,i})$  defined through the solution of the PDE at level  $\ell$  and using equation (5.33). Notice that in this case it would make sense to look at the limit for  $T \rightarrow \infty$  since the probability does not blow up. On the other hand, as we will see in the next section, for the infinite time horizon case another strategy will be proposed to avoid the solution of a time dependent PDE over a large time horizon.

### 5.4.3 Delineation of capture zones for infinite temporal horizon: a streamline approach

In the previous subsection we dealt with the case of a small temporal horizon  $T$ . Here we present a methodology that allows us to compute the capture zones (5.34) in the limit when  $T \rightarrow \infty$  without solving any time dependent PDE. To this end, for the sake of simplicity, we consider the formulation (5.29) of the Darcy problem. Let us focus on the case of vanishing diffusion. As we mentioned before, viscosity solutions for the PDE (5.31) can be found also for pure transport equation, so that actually it holds  $\theta_T^0(x, 0, \omega) = \mathbf{1}_{\{\tau_{\mathbf{x},0}^0(\omega) \leq T\}}$ .

Here we want to approximate the indicator function for  $T = \infty$  by introducing a streamline approach. First notice that, in the vanishing diffusion case, there holds  $\mathbb{P}^{\omega}(\tau_{\mathbf{x},0}^0(\omega) \leq \infty) = \mathbb{E}^{\omega}[\mathbf{1}_{\{\tau_{\mathbf{x},0}^0(\omega) \leq \infty\}}]$ . Our idea is to approximate such expected value through a MC scheme by properly sampling the indicator function  $\mathbf{1}_{\{\tau_{\mathbf{x},0}^0(\omega) \leq \infty\}}$  from a suitable sparse grid interpolant. Since as previously mentioned arrival times and probabilities are not suited to be approximated by sparse grids, we introduce now a streamline approach to delineate the capture zones introduced in (5.34). We start by decomposing the velocity field in its regular and singular part, namely

$$\mathbf{u}(\mathbf{x}, \omega) = \mathbf{u}_R(\mathbf{x}, \omega) + \mathbf{u}_S(\mathbf{x}), \quad (5.35)$$

#### 5.4. Case 2: a flow induced by the presence of an extracting well

with  $\mathbf{u}_R$  divergence free in  $D$  and  $\mathbf{u}_S(\mathbf{x})$  is the gradient of the two dimensional Green's function  $\phi = \log(|\mathbf{x} - \mathbf{x}_0|)/(2\pi)$  solution of  $\Delta\phi = -q\delta_{\mathbf{x}_0}(\mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^2$ ; at the same time  $\mathbf{u}_S(\mathbf{x})$  can be seen as the curl of a suitable stream function  $\Psi_S$ , namely  $\mathbf{u}_S(\mathbf{x}) = \nabla \times \Psi_S(\mathbf{x})$ , with  $\Psi_S(\mathbf{x}) = -\frac{q\theta(\mathbf{x}, \mathbf{x}_0)}{2\pi}$ ; here  $\theta(\mathbf{x}, \mathbf{x}_0)$  is the counterclockwise angle that the segment  $\overline{\mathbf{x}\mathbf{x}_0}$  forms with respect to the horizontal axis. Notice that by construction the singular part of the transport does not depend on  $\omega$ . Observe also that since the singular Darcy velocity  $\mathbf{u}_S$  is defined on a not simply connected domain, i.e.  $D \setminus \{\mathbf{x} = \mathbf{x}_0\}$ , we have to introduce an artificial cut in  $D$  in order to deal with a simply connected domain  $\tilde{D}$  and give sense to the previous definition of singular stream function  $\Psi_S(\mathbf{x})$ ; we operate this cut by considering a domain  $\tilde{D} = D \setminus \{\mathbf{x} : x_2 = x_{0,2} \text{ and } x_1 \leq x_{0,1}\}$ . Such stream function  $\Psi_S(\mathbf{x}, \omega)$ , associated to the singular part of the transport, is discontinuous along the cut introduced in  $\tilde{D}$ , with continuous derivatives along the cut. On the other hand, since the regular part of the velocity  $\mathbf{u}_R$  is divergence free, there exists an associated stream function  $\Psi_R$  such that  $\mathbf{u}_R(\mathbf{x}, \omega) = \nabla \times \Psi_R(\mathbf{x}, \omega)$ . Notice that in this case, since  $\mathbf{u}_R$  is well defined on  $D$ , we do not need to introduce any artificial cut to properly define it.

At this point, starting from  $\nabla \times \mathbf{u}_R = \nabla \times (\mathbf{u} - \mathbf{u}_S)$ , it is possible to derive a Laplace equation for  $\Psi_R$ ; concerning the boundary conditions, we fix the stream function  $\Psi_R$  at an arbitrary value on the top boundary of the domain  $D$ , and then recover the natural Neumann condition on the other boundaries by observing that  $\nabla \Psi_R \cdot \mathbf{n} = \mathbf{u}_R \cdot \mathbf{t}$ , where  $\mathbf{t}$  is the tangent vector to the boundary. We have

$$\begin{cases} -\Delta \Psi_R(\mathbf{x}, \omega) = \frac{\partial u_2(\mathbf{x}, \omega)}{\partial x_1} - \frac{\partial u_1(\mathbf{x}, \omega)}{\partial x_2}, & \mathbf{x} \in D, \\ \Psi_R(\mathbf{x}, \omega) = 0 & \mathbf{x} \in \{x_2 = 1\}, \\ \partial_{\mathbf{n}} \Psi_R(\mathbf{x}, \omega) = (\mathbf{u}(\mathbf{x}, \omega) - \mathbf{u}_S(\mathbf{x})) \cdot \mathbf{t} & \mathbf{x} \in x_1 = \{0, 1\} \cup x_2 = 0. \end{cases} \quad (5.36)$$

With this strategy, it turns out that  $\Psi(\mathbf{x}, \omega) = \Psi_R(\mathbf{x}, \omega) + \Psi_S(\mathbf{x}) : \tilde{D} \rightarrow \mathbb{R}$  is an actual stream function of the transport  $\mathbf{u}$ ; then, the idea is to identify the limit trajectory that passes through the saddle point of  $\Psi$ , located to the right of the well, and use such trajectory to delineate the zone in which all the trajectories that sooner or later will end up in the well are concentrated.

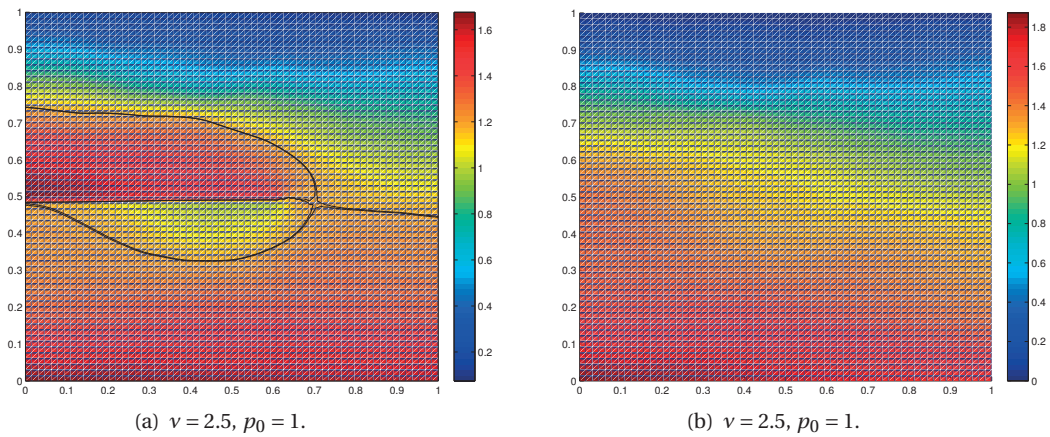


Figure 5.3 – Realizations of  $\Psi(\mathbf{x}, \omega)$  (left) and  $\Psi_R(\mathbf{x}, \omega)$  (right) corresponding to the random event  $\omega_i$ .

In Figure 5.3-(a) we give an illustration of what we just stated; the trajectories outlined are the ones that are located immediately close to the limiting one; we can observe how the total stream function is discontinuous along the cut introduced in  $\tilde{D}$  corresponding to the choice  $\theta(\mathbf{x}, \mathbf{x}_0) = \pi$



since in our notation  $\theta \in (-\pi, \pi)$ . On the other hand in Figure 5.3-(b) it is possible to see that the corresponding realization of  $\Psi_R$  is smooth with respect to the spatial coordinate  $\mathbf{x}$ : this implies that also the mapping  $\mathbf{y} \rightarrow \Psi_R(\mathbf{x}, \mathbf{y})$  is smooth and, in light of what previously stated, it can be successfully approximated by a sparse grid approximation. Notice however that, for every knot of the sparse grid we have now to solve a Laplace problem with non homogeneous boundary conditions; in view of a MC estimation of  $\mathbb{E}^\omega[\mathbf{1}_{\{\tau_{\mathbf{x},0}^0(\omega) \leq \infty\}}]$  we will need to interpolate the sparse grid approximation of  $\Psi_R$  in many points  $\mathbf{y}_i \in \Gamma$ ; this is done with the following procedure, i.e. by assembling the sparse grid interpolant  $\mathcal{S}_1[\Psi_R](\mathbf{x}, \mathbf{y})$  and by operating as follows:

1. given a new sample  $\mathbf{y}_i$  in the MC estimator we generate a new sample of  $\Psi_R$ , namely  $\Psi_R^i = \mathcal{S}_1[\Psi_R](\mathbf{y}_i)$ , through sparse grid interpolation;
2. thanks to  $\Psi(\mathbf{x}, \mathbf{y}_i) = \Psi_R(\mathbf{x}, \mathbf{y}_i) + \Psi_S(\mathbf{x})$  we compute the sample  $\Psi^i$  of the total stream function;
3. we find numerically the saddle point  $\mathbf{x}^*$  of  $\Psi^i$  and, moving from it, we delineate the two limit trajectories;
4. we obtain a realization of the indicator function  $\mathbf{1}_{\{\tau_{\mathbf{x},0}^0(\omega_i) \leq \infty\}}$  by assigning 1 to the points between the two limit trajectories and 0 to the others: we denote such approximation as  $\mathbf{1}_{\{A(\Psi^i)\}}(\mathbf{x})$ , being  $A(\Psi)$  the region included between the two limit trajectories associated to the stream function  $\Psi$ .

By repeating many times this procedure, it is possible to efficiently compute a MC estimator of  $\mathbb{E}^\omega[\mathbf{1}_{\{\tau_{\mathbf{x},0}^0(\omega) \leq \infty\}}]$ , namely

$$\hat{Q}_{h,M}^{MC}(\mathbf{x}) = \frac{1}{M} \sum_{i=1}^M \mathbf{1}_{\{A(\Psi_h^i)\}}(\mathbf{x}). \quad (5.37)$$

In practice, the sparse grid interpolant of  $\Psi_R$  is built starting from a piecewise linear finite element approximation, defined on a structured mesh, of problem (5.36). Hence, in step 3, each interpolated stream function  $\Psi^i$ , obtained as sum of a standard linear finite element function ( $\Psi_{R,h}^i$ ) and a discontinuous finite element function ( $\Psi_{S,h}$ ), will be a piecewise linear function (discontinuous along the cut) and the numerical computation of the saddle point  $\mathbf{x}^*$  is done as follows:

1. denote  $e_i$ ,  $i = 1, \dots, 6$ , the edges which share the vertex  $\mathbf{x}^{*1}$  and  $\mathbf{u}_{i,j}$ ,  $j = 1, 2$ , the velocities evaluated on the two adjacent triangles that share the edge  $e_i$ ; moreover denote  $\mathbf{t}_{e_i}$  the vector aligned with the edge  $e_i$ .
2. suppose that the edges are ordered in such a way to turn in a clockwise sense around  $\mathbf{x}^*$  and compute for any  $i$  the quantity  $r(i) = (\mathbf{u}_{i,1} + \mathbf{u}_{i,2}) \cdot \mathbf{t}_{e_i}$ ;
3. count the changes of sign of  $r$  when passing from  $r(1)$  to  $r(2)$ , till arriving to the change of sign from  $r(6)$  to  $r(1)$ ;
4. if the number of sign changes is 4 then we have found the saddle point  $\mathbf{x}^*$ , otherwise we move to another vertex.

---

<sup>1</sup>Observe that the numerical saddle point has to coincide with a vertex of the triangulation.

#### 5.4. Case 2: a flow induced by the presence of an extracting well

Observe that the number of changes has to be 4 in order to obtain a saddle point structure around the vertex  $\mathbf{x}^*$ ; otherwise if the number of changes is 0 then we are facing a maximum (or minimum) while if it is equal to 2 then we are facing a common point with non horizontal tangent plane. After an initial (small) cost needed to build the sparse grid interpolant  $\mathcal{S}_1[\Psi_R](\mathbf{x}, \mathbf{y})$ , to generate a sample we just need to use the sparse grid to perform an interpolation on a given value  $\mathbf{y}_i$ ; observe that in Chapter 3 the sparse grid approximation has been shown to be suited to perform both quadrature and interpolation.

So far we have considered a vanishing diffusion case; in general we could be interested in considering also cases with (small) positive diffusion. Unfortunately a similar approach can not be straightforwardly extended to cover also this case. We propose to use the same estimator (5.37) (built for the vanishing diffusion case) also in the case of positive diffusion  $\sigma' > 0$  and estimate the error committed by such approximation. By using again [1, Theorem 2.4] it holds

$$|\mathbb{P}^\omega(\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T) - \mathbb{P}^\omega(\tau_{\mathbf{x},0}^0(\omega) \leq T)| = |\mathbb{E}^\omega[\mathbb{E}^{\omega'}[\mathbf{1}_{\{\tau_{\mathbf{x},0}^{\sigma'}(\omega, \omega') \leq \infty\}} - \mathbf{1}_{\{\tau_{\mathbf{x},0}^0(\omega) \leq \infty\}} | \omega]]| \leq \mathbb{E}^\omega[\|\tau_{\mathbf{x},0}^{\sigma'}(\omega, \omega') - \tau_{\mathbf{x},0}^0(\omega)\|_{L^p(\Omega)}^{\frac{p}{1+p}}]. \quad (5.38)$$

Notice that now we are not comparing the results obtained with positive diffusion with the ones obtained by considering  $\bar{\tau}_{\mathbf{x}}^{\sigma'}$  as in Subsection 5.3.3. Nevertheless, also this quantity is expected to be negligible when considering small diffusion  $\sigma'$ .

##### 5.4.4 Numerical results

We start first by presenting some results obtained by using a standard MLMC scheme to approximate the QoI  $Q^{\sigma'}(\mathbf{x}, \omega)$  representing the probability with respect to all Brownian motion of ending up in the well within a time  $T$  for a given realization of the Darcy velocity. This QoI is obtained as solution of the parabolic PDE (5.31), i.e.  $Q^{\sigma'}(\mathbf{x}, \omega) = \mathbb{P}^{\omega'}(\tau_{\mathbf{x},0}^{\sigma'}(\omega, \cdot) \leq T | \omega)$ . In Figure 5.4 we report a MLMC approximation of  $\mathbb{P}^\omega(\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T)$  to give an idea of the quantity that we want to estimate. By

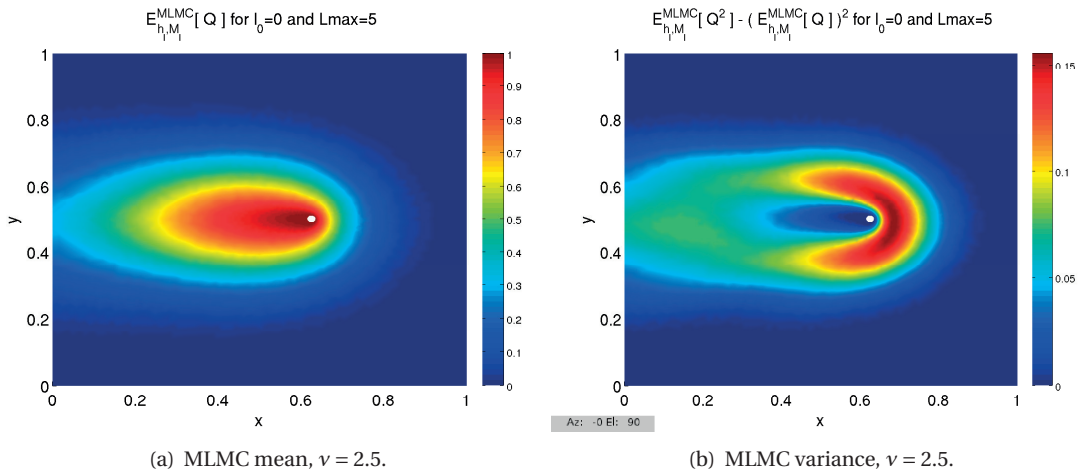


Figure 5.4 – Case  $v = 2.5$ : MLMC approximation of mean (left) and variance (right) of  $Q(\mathbf{x}, \omega) = \mathbb{E}^{\omega'}[\mathbf{1}_{\{\tau_{\mathbf{x},0}^{\sigma'}(\omega) \leq T\}}]$ .  $\sigma' = 0.01$ ,  $L_C = 0.5$ ,  $p_0 = 1$ ,  $q = 0.5$ ,  $r = 0.01$ ,  $T = 0.5$ .

looking at the MLMC mean, as expected, the probability of ending up in the well in at most  $T = 0.5$

time units decreases as we move away from the well in the direction of the underlying flow (from left to right). From this solution, the capture zones can be computed simply by taking level sets of this function. The plot displaying the variance is interesting as it shows that the zones affected by the largest variability are the ones immediately behind the well, with respect to the direction of the flow. On the other hand there is basically no variability for the particles that start immediately on the left of the well (as they fall with probability 1 into the well in short time) and for the ones that are outside the capture zone (as they never reach the well). Here we report only the results obtained in the case  $\nu = 2.5$  since the other concerning the rough case  $\nu = 0.5$  are analogous.

In Figure 5.5 the variance reduction obtained with a MLMC scheme with respect to a standard MC one is shown. We show the results obtaining by using the proposed level dependent diffusion approach and compare them with the ones obtained with a standard MLMC scheme with fixed diffusion. Notice that the level diffusion is such that, on the finest level  $L$ , it holds  $\sigma'_L = \sigma'$ ; the

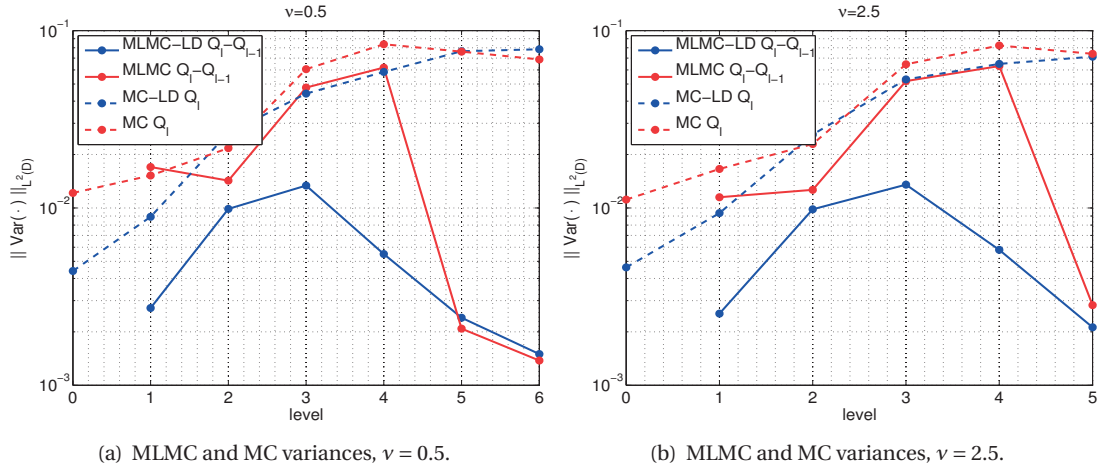


Figure 5.5 – MLMC variance reduction (continuous line) versus MC variance (dashed line) in the case of level dependent diffusion (in blue, label LD) and fixed diffusion (in red) for  $\nu = 0.5$  (right) and  $\nu = 2.5$  (left). The variances have been computed with respect the  $L^2$  spatial norm.  $Q(\mathbf{x}, \omega) = \mathbb{E}^{\omega'}[\mathbf{1}_{\{\tau_{\mathbf{x},t}^{\omega'}(\omega) \leq T\}}]$ .  $\sigma' = 0.01$ ,  $L_C = 0.5$ ,  $p_0 = 1$ ,  $q = 0.5$ ,  $r = 0.01$ .

comparison shows how such level dependent diffusion approach actually improves the variance reduction achievable with a MLMC scheme. Although these results seem promising, in the sense that we obtain a good variance reduction with respect to the standard MC case, it is important to remark how the computational cost needed to achieve practical tolerances could actually be very high, since here we are dealing with time dependent PDEs.

In the following we report the results obtained with the streamline approach introduced in the previous section in the vanishing diffusion case. In Figure 5.6 we show the convergence of the error committed by approximating the regular stream function  $\Psi_R(\mathbf{x}, \omega)$  with a sparse grid. As in the previous section we consider a “deltaint”-based adaptive sparse grid with non nested knots; we denote our QoI as  $Q(\mathbf{x}, \omega) = \Psi_R(\mathbf{x}, \omega)$  and again we measure the sparse grid error as

$$\|E[Q(\mathbf{x}, \omega)] - \mathcal{Q}_W[Q(\mathbf{x}, \omega)]\|_{L^2(D)} \approx \|\mathcal{Q}_{W*}[Q(\mathbf{x}, \omega)] - \mathcal{Q}_W[Q(\mathbf{x}, \omega)]\|_{L^2(D)}$$

being  $\mathcal{Q}_{W*}[Q(\mathbf{x}, \omega)]$  the reference quadrature, computed with a sparse grid built over 937 Hermite points corresponding to a work  $W^* = 1029$ . The obtained convergence of the sparse grid quadrature



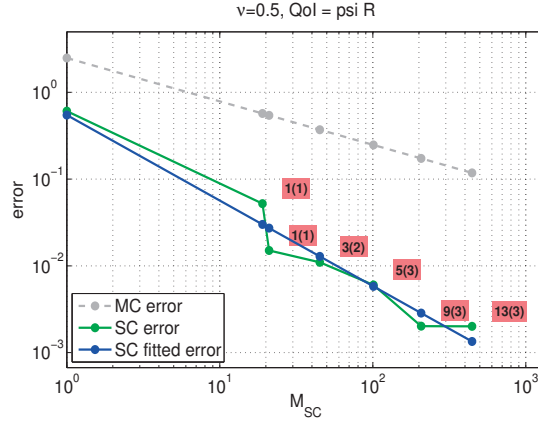
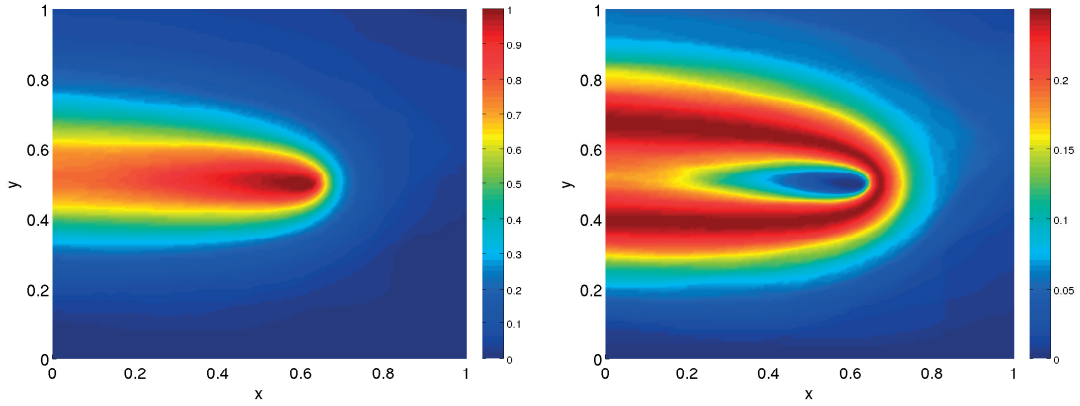


Figure 5.6 – Case  $\nu = 2.5$ : sparse grid approximation of the stream function  $\Psi_R(\mathbf{x}, \omega)$ .  $L_C = 0.5$ ,  $p_0 = 1$ .

of  $\Psi_R$  is again satisfactory, especially if compared with the corresponding MC error. Once we have the regular stream function we can then use the sparse grid interpolation to generate other possible stream functions  $\Psi$  and then build a MC estimator of the probability in (5.34). A realization of such MC estimator obtained with  $M = 1000$  samples is shown in Figure 5.7 in which we report the probability of ending up in the well on an infinite time horizon and the corresponding variability, i.e. a MC estimation of  $\mathbb{E}^\omega [\mathbf{1}_{\{\tau_{\mathbf{x},0}(\omega) \leq \infty\}}^2] - \mathbb{E}^\omega [\mathbf{1}_{\{\tau_{\mathbf{x},0}(\omega) \leq \infty\}}]^2$ .



(a) MC mean obtained with the streamline approach,  $\nu = 2.5$ . (b) MC variance obtained with the streamline approach,  $\nu = 2.5$ .

Figure 5.7 – Case  $\nu = 2.5$ : MC approximation of mean (left) and variance (right) of  $Q(\mathbf{x}, \omega) = \mathbf{1}_{\{\tau_{\mathbf{x},0}(\omega) \leq \infty\}}$ .  $L_C = 0.5$ ,  $p_0 = 2$ .

## 5.5 Conclusions

In this Chapter we focused on problems related to the transport of particles in aquifers. In particular we considered a model that allowed us to take into account also the molecular diffusion that could occur at the pore level and developed a methodology to fully address the problem. The main goals of the analysis were the computations of arrival times to specific boundaries of the domain, and the delineation of the so called capture zones. Following this goal we gave an overview of several

possible choices that could be made from both, a modeling and a numerical point of view. We addressed several scenarios and possible numerical approaches outlining their advantages and limitations.

In particular we first consider an undisturbed flow; in this case we showed that a MLMCCV scheme is able to effectively approximate the expected FPT and proposed a strategy to adapt such result also to compute the probability of the FPT being smaller than a prescribed time horizon: we are still left with a rigorous error analysis to keep under control the error term (5.27).

Then we moved to a more involved case in which the underlying from-left-to-right flow is combined with the presence of an extracting well in the domain. In this case, unfortunately, the sparse grid can not be applied to effectively approximate either the expected FPT or the related probabilities. The best strategy we could propose to tackle the corresponding parabolic PDEs (for both the expected FPT and the exit probability) was a MLMC scheme which uses level dependent diffusion coefficients. To tackle the problem of the delineation of capture zones in the case of infinite time horizon we proposed a streamline approach: by considering the vanishing diffusion case, here we perform an effective sparse grid approximation of the stream function related to the Darcy velocity, and by tracking suitable trajectories, we delineate the corresponding capture zone. Then, via sparse grid interpolation of the stream function, we build a convenient MC estimator; also in this case a rigorous error analysis to keep under control the error term (5.38) will be needed.

## 6 Conclusions and future work

In this Thesis we have studied linear PDEs with random coefficients to address specific problems arising in hydrology related to groundwater flows taking place in a heterogeneous saturated porous medium. We have extensively studied the Darcy boundary value problem, consisting in an elliptic PDE having the pressure as unknown and the permeability of the medium as input diffusion coefficient. Such quantity has been modeled as a lognormally distributed random field  $a(\mathbf{x}, \omega) = e^{\gamma(\mathbf{x}, \omega)}$ , being  $\gamma(\mathbf{x}, \omega)$  a Gaussian random field with given Matérn-type covariance structure, which, in turn, generates fields having a wide range of possible smoothness.

The goal of our analysis has been to develop efficient numerical schemes able to compute statistics of quantities of interest related to the solution  $p$ , representing the pressure, of the Darcy problem. We have considered two main approaches, a deterministic and a sampling one: the first is based on a sparse grid stochastic collocation approximation of the quantity of interest while the second is based on Monte-Carlo type schemes that use a suitable variance reduction technique.

Concerning the sparse grid approximation of the Darcy problem we have focused on algorithms able to exploit the anisotropy of the quantity of interest with respect to the stochastic parameters and considered two approaches: a quasi-optimal one, extending results already present in literature also to the case of non-nested collocation points, and an adaptive one, in which the construction of the underlying sparse grid is driven by different (computable) estimates of the hierarchical surplus operators  $\Delta^{\mathbf{m}^{(i)}}$ . The two strategies have then been implemented in a dimension adaptive fashion, namely starting with  $N_b$  random variables as buffer and then progressively adding variables on the fly whenever a new direction is activated. It is important to remark that this implementation is able to work formally with an infinite (countable) number of random variables, since it does not require any a priori truncation of the input permeability. Such strategies have been verified to be very effective to solve the Darcy problem when considering a smooth random permeability as input; our numerical results show a much better performance of the adaptive sparse grid algorithms than standard MC techniques. At the same time it is important to notice that the observed convergence rates obtained with our numerical experiments are not so distant from the theoretical ones available in literature for elliptic problems parametrized with uniform random variables. On the other hand, such advanced adaptive sparse grid approximations have shown deteriorating performance when applied to either shortly correlated random fields ( $L_c \ll 1$ ) or rough input random fields. In the case of smooth but shortly correlated input random fields we expect that the fast convergence rate is achieved only asymptotically when the sparse grid is rich enough to resolve the fine scale structures of the solution. In the case of longly correlated but rough random fields, the convergence rate is

negatively affected by the lack of smoothness and might not be competitive with a MC approach. The reason comes from the slow decay of the eigenvalues of the covariance operator which forces the adaptive sparse algorithm to include too many random variables even for large tolerances, thus feeling a strong curse of dimensionality.

However, the analysis of such sparse grid approximations is still incomplete and the next crucial step would be to complete a rigorous convergence analysis of the dimension adaptive algorithms in the case of PDE with log-normal coefficients (partial results are only available so far for sequences of uniform random variables and uniformly bounded and positive diffusion coefficients.)

We also considered sampling schemes in order to deal with Darcy problems characterized by a limited spatial regularity. In particular we started from the popular Multi Level Monte Carlo method and proposed an improved version that uses a suitable control variate to reduce the variance of the corresponding MLMCCV estimator. Among all possible choices, we considered as a control variate the quantity of interest  $Q^\epsilon$  obtained starting from the solution of a smoothed auxiliary Darcy problem, in which we have replaced the original permeability field  $a$  with a smoothed version  $a^\epsilon$ . The key idea was to consider a smooth control variate  $Q^\epsilon$  for which a sparse grid quadrature approximation may still be efficiently computed. Hence the necessity of carefully choosing the smoothing parameter  $\epsilon$  to achieve both a significant variance reduction and, at the same time, good sparse grid approximation properties for the mean of the control variate. We performed a rigorous analysis of the statistical error associated the MLMCCV estimator and provided explicit bounds indicating the combined effect of the discretization parameter  $h_\ell$  on the different levels and of the smoothing parameter  $\epsilon$ . Finally we showed how, asymptotically, the complexity of the MLMCCV scheme improves the one of the standard MLMC one, and provided several numerical results covering different scenarios to show the effectiveness of the method with respect to a standard MLMC approach, in terms of complexity against tolerance achieved on the root MSE. An important remark is that, for the moment, we just tuned the parameter  $\epsilon$  ad hoc. It would be then interesting to either include the selection of  $\epsilon$  in the optimization procedure used to compute optimal sample sizes and number of sparse grid nodes, or considering a level dependent choice of  $\epsilon$ .

We then focused on another relevant problem in groundwater flows, i.e. the computation of statistics related to arrival times defined through particle trajectories subject to (small) molecular diffusion and driven by the stochastic Darcy velocity. In particular we addressed again the reference scenario of a flow induced by a pressure drop between two opposite boundaries of the domain (left-to-right flow) and a scenario in which we have modeled also the presence of an extracting well into the physical domain. In the first case we looked at arrival times to the outlet of the domain while in the second case we looked at arrival times to the well. We extended, whenever possible, the methodologies developed for the Darcy flow to conveniently compute expectations and probabilities of arrival times, which have been both properly linked to the solution of suitable parabolic (or sometimes elliptic) PDEs via the famous Feynman-Kac formula.

In the case of the left-to-right flow, we showed how the extension of the MLMCCV scheme was particularly effective for approximating expected arrival times to the outlet; notice that in this case, for each realization of the Darcy velocity, we had to solve a simple elliptic equation. We were not able to directly extend this argument also to approximate probabilities related to the arrival times; however, in this case, we proposed to still use the sparse grid interpolant of the arrival time to conveniently generate samples for a MC-type estimator; this approximation contains a bias proportional to the variance (with respect to all Brownian motion) of the arrival time, cf. (5.27), which is small for small diffusions  $\sigma'$ .

---

On the other hand, when considering the scenario with the extraction well, since in all cases the realizations of our quantities of interest were always presenting sharp evolving fronts, the application of sparse grid and MLMCCV strategies was unfeasible; we therefore resorted to a standard MLMC scheme, however, with level dependent diffusion to improve the stability of the numerical DG scheme on coarse levels, and compared the results obtained in terms of variance reduction with a standard MLMC. Finally, we proposed also a streamline approach to tackle the problem of the delineation of capture zones in the case of vanishing diffusion and infinite time horizon. Here we used again a suitable sparse grid interpolant to efficiently obtain realizations of the indicator function representing the points starting from which a particle either reaches the well or not.

In the last chapter we presented different scenarios and shown numerical results; yet, there are still several open issues that will have to be properly addressed. One of the most important is certainly a more rigorous investigation of the theoretical aspects and convergence properties of the schemes so far available only for elliptic problems. We also proposed to use convenient MC samplers to estimate the exit-probability in different scenarios: a rigorous analysis able to keep track of the different sources of errors affecting this schemes should be performed. In this case we should quantify and bound the interpolation error affecting the sampling as well as other error terms discussed in Chapter 5, cf. (5.27), (5.38).



# Bibliography

- [1] R. Avikainen. On irregular functionals of sdes and the euler scheme. *Finance and Stochastics*, 13(3):381–401, 2009.
- [2] I. Babuška, R. Tempone, and G. E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numer. Anal.*, 42(2):800–825, 2004.
- [3] 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(3):1005–1034, 2007.
- [4] I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Review*, 52(2):317–355, June 2010.
- [5] J. Bäck, F. Nobile, L. Tamellini, and R. Tempone. Stochastic spectral Galerkin and collocation methods for PDEs with random coefficients: a numerical comparison. In J. Hesthaven and E. Ronquist, editors, *Spectral and High Order Methods for Partial Differential Equations*, volume 76 of *Lecture Notes in Computational Science and Engineering*, pages 43–62. Springer, 2011. Selected papers from the ICOSAHOM '09 conference, June 22–26, Trondheim, Norway.
- [6] M. Bardi, M. G. Crandall, L. C. Evans, H. M. Soner, and P. E. Souganidis. *Viscosity Solutions and Applications*, volume 1660 of *C.I.M.E. Foundation Subseries*. Springer-Verlag Berlin Heidelberg, 1997.
- [7] A. Barth, C. Schwab, and N. Zollinger. Multi-level monte carlo finite element method for elliptic pdes with stochastic coefficients. *Numer. Math.*, 119(1):123–161, 2011.
- [8] V. Barthelmann, E. Novak, and K. Ritter. High dimensional polynomial interpolation on sparse grids. *Adv. Comput. Math.*, 12(4):273–288, 2000.
- [9] J. Bear and A. H.-D. Cheng. *Modeling groundwater flow and contaminant transport*. Theory and applications of transport in porous media. Springer, Dordrecht, Heidelberg, London, 2010.
- [10] A. Beaudoin, J. de Dreuzy, and J. Erhel. Numerical monte carlo analysis of the influence of pore-scale dispersion on macrodispersion in 2-d heterogeneous porous media. *Water Resources Research*, 46(12):n/a–n/a, 2010. W12537.
- [11] J. Beck, F. Nobile, L. Tamellini, and R. Tempone. On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods. *Mathematical Models and Methods in Applied Sciences*, 22(09), 2012.
- [12] J. Beck, F. Nobile, L. Tamellini, and R. Tempone. A Quasi-optimal Sparse Grids Procedure for Groundwater Flows. In M. Azaiez, H. El Fekih, and J. S. Hesthaven, editors, *Spectral and High*



*Order Methods for Partial Differential Equations - ICOSAHOM 2012*, volume 95 of *Lecture Notes in Computational Science and Engineering*, pages 1–16. Springer, 2014. Selected papers from the ICOSAHOM '12 conference.

- [13] J. Beck, F. Nobile, L. Tamellini, and R. Tempone. A quasi-optimal sparse grids procedure for groundwater flows. In M. Azaïez, H. El Fekih, and J. S. Hesthaven, editors, *Spectral and High Order Methods for Partial Differential Equations - ICOSAHOM 2012*, volume 95 of *Lecture Notes in Computational Science and Engineering*, pages 1–16. Springer International Publishing, 2014.
- [14] F. Brezzi, L. D. Marini, and E. Süli. Discontinuous galerkin methods for first-order hyperbolic problems. *Mathematical Models and Methods in Applied Sciences*, 14(12):1893–1903, 2004.
- [15] H. Bungartz and M. Griebel. Sparse grids. *Acta Numer.*, 13:147–269, 2004.
- [16] J. Charrier. *Analyse numérique d'équations aux dérivées partielles à coefficients aléatoires, applications à l'hydrologéologie*. PhD thesis, ENS Cachan-Bretagne, 2011.
- [17] J. Charrier. Strong and weak error estimates for elliptic partial differential equations with random coefficients. *SIAM J. Numer. Anal.*, 50(1), 2012.
- [18] J. Charrier and A. Debussche. Weak truncation error estimates for elliptic pdes with lognormal coefficients. *Stochastic Partial Differential Equations: Analysis and Computations*, 1(1):63–93, 2013.
- [19] J. Charrier, R. Scheichl, and A. Teckentrup. Finite element error analysis of elliptic pdes with random coefficients and its application to multilevel monte carlo methods. *SIAM Journal on Numerical Analysis*, 51(1):322–352, 2013.
- [20] A. Cherny and H.-J. Engelbert. 1. stochastic differential equations. In *Singular Stochastic Differential Equations*, volume 1858 of *Lecture Notes in Mathematics*, pages 5–25. Springer Berlin Heidelberg, 2005.
- [21] J.-P. Chilès and P. Delfiner. *Geostatistics: Modeling spatial uncertainty, second edition*. John Wiley & Sons, Inc., 2012.
- [22] A. Chkifa, A. Cohen, and C. Schwab. High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs. *Foundations of Computational Mathematics*, pages 1–33, 2013.
- [23] A. Chkifa, A. Cohen, and C. Schwab. High-dimensional adaptive sparse polynomial interpolation and applications to parametric pdes. *Journ. Found. Comp. Math.*, 14(4):601–633, 2013.
- [24] A. Chkifa, A. Cohen, and C. Schwab. Breaking the curse of dimensionality in sparse polynomial approximation of parametric pdes. *Journ. Math. Pures et Appliquees*, 103(2):400–428, 2015.
- [25] K. Cliffe, M. Giles, R. Scheichl, and A. Teckentrup. Multilevel monte carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science*, 14(1):3–15, 2011.
- [26] A. Cohen and R. Devore. Approximation of high-dimensional parametric pdes. *Acta Numerica*, 24:1–159, 5 2015.

- 
- [27] A. Cohen, R. DeVore, and C. Schwab. Analytic regularity and polynomial approximation of parametric and stochastic pdes. *Analysis and Applications*, 09(01):11–47, 2011.
- [28] N. Collier, A. Haji-Ali, F. Nobile, and R. Tempone. A continuation multilevel monte carlo algorithm. *BIT Numerical Mathematics*, 55(2):399–432, 2015.
- [29] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and its Applications (No. 45). Cambridge University Press, 1992.
- [30] B. A. Davey and H. A. Priestley. *Introduction to lattices and order*. Cambridge University Press, New York, second edition, 2002.
- [31] C. R. Dietrich and G. N. Newsam. Fast and exact simulation of stationary gaussian processes through circulant embedding of the covariance matrix. *SIAM J. Sci. Comput.*, 18(4):1088–1107, July 1997.
- [32] P. Diggle and P. J. Ribeiro. *Model-based geostatistics*. Springer, 2007.
- [33] M. S. Eldred and J. Burkardt. Comparison of non-intrusive polynomial chaos and stochastic collocation methods for uncertainty quantification. American Institute of Aeronautics and Astronautics Paper 2009-0976.
- [34] H. C. Elman, C. W. Miller, E. T. Phipps, and R. S. Tuminaro. Assessment of Collocation and Galerkin approaches to linear diffusion equations with random data. *International Journal for Uncertainty Quantification*, 1(1):19–33, 2011.
- [35] O. Ernst and B. Sprungk. Stochastic collocation for elliptic PDEs with random data: The lognormal case. In J. Garcke and D. Pflüger, editors, *Sparse Grids and Applications - Munich 2012*, volume 97 of *Lecture Notes in Computational Science and Engineering*, pages 29–53. Springer International Publishing, 2014.
- [36] J. Foo, X. Wan, and G. Karniadakis. The multi-element probabilistic collocation method (ME-PCM): Error analysis and applications. *Journal of Computational Physics*, 227(22):9572–9595, 2008.
- [37] S. Franzetti and A. Guadagnini. Probabilistic estimation of well catchments in heterogeneous aquifers. *Journal of Hydrology*, 174(1-2):149–171, 1996.
- [38] J. Galvis and M. Sarkis. Approximating infinity-dimensional stochastic Darcy’s equations without uniform ellipticity. *SIAM J. Numer. Anal.*, 47(5):3624–3651, 2009.
- [39] A. Genz and B. D. Keister. Fully symmetric interpolatory rules for multiple integrals over infinite regions with Gaussian weight. *J. Comput. Appl. Math.*, 71(2):299–309, 1996.
- [40] T. Gerstner and M. Griebel. Dimension-adaptive tensor-product quadrature. *Computing*, 71(1):65–87, 2003.
- [41] M. B. Giles. Multi-level monte carlo path simulation. *OPERATIONS RESEARCH*, 56(3):607–617, 2008.
- [42] C. Gittelson. An adaptive stochastic galerkin method for random elliptic operators. *Mathematics of Computation*, 82:1515–1541, 2013.

- [43] C. J. Gittelsohn. Stochastic Galerkin discretization of the log-normal isotropic diffusion problem. *Math. Models Methods Appl. Sci.*, 20(2):237–263, 2010.
- [44] I. Graham, F. Kuo, D. Nuyens, R. Scheichl, and I. Sloan. Quasi-monte carlo methods for elliptic pdes with random coefficients and applications. *Journal of Computational Physics*, 230(10):3668–3694, May 2011.
- [45] I. G. Graham, F. Y. Kuo, J. A. Nichols, R. Scheichl, C. Schwab, and I. H. Sloan. Quasi-monte carlo finite element methods for elliptic pdes with lognormal random coefficients. *Numerische Mathematik*, 2014.
- [46] I. G. Graham, F. Y. Kuo, J. A. Nichols, R. Scheichl, C. Schwab, and I. H. Sloan. Quasi-monte carlo finite element methods for elliptic pdes with lognormal random coefficients. *Numerische Mathematik*, 131(2):329–368, 2015.
- [47] P. Grisvard. *Elliptic problems in nonsmooth domains*. Classics in applied mathematics. SIAM, Philadelphia, Pa., 2011. SIAM : Society for industrial and applied mathematics.
- [48] A. Haji-Ali, F. Nobile, and R. Tempone. Multi Index Monte Carlo: when sparsity meets sampling. arXiv arXiv:1405.3757, e-print, 2014.
- [49] H. Harbrecht, M. Peters, and M. Siebenmorgen. Multilevel accelerated quadrature for PDEs with log-normal distributed random coefficient. preprint 2013-18, Universität Basel, 2013.
- [50] O. Havle, V. Dolejší, and M. Feistauer. Discontinuous galerkin method for nonlinear convection-diffusion problems with mixed dirichlet-neumann boundary conditions. *Applications of Mathematics*, 55(5):353–372, 2010.
- [51] C. J. Numerical analysis of the advection-diffusion of a solute in porous media with uncertainty. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):650–685, 2015.
- [52] J. D. Jakeman, R. Archibald, and D. Xiu. Characterization of discontinuities in high-dimensional stochastic problems on adaptive sparse grids. *J. Comput. Phys.*, 230(10), 2011.
- [53] A. Klimke. *Uncertainty modeling using fuzzy arithmetic and sparse grids*. PhD thesis, Universität Stuttgart, Shaker Verlag, Aachen, 2006.
- [54] A. Lang and J. Potthoff. Fast simulation of gaussian random fields. *Monte Carlo Methods and Applications*, 17(3):195–214, 2011.
- [55] G. J. Lord, C. E. Powell, and T. Shardlow. *An Introduction to Computational Stochastic PDEs*. Cambridge University Press, 2014. Cambridge Books Online.
- [56] S. Martello and P. Toth. *Knapsack problems: algorithms and computer implementations*. Wiley-Interscience series in discrete mathematics and optimization. J. Wiley & Sons, 1990.
- [57] F. Müller, P. Jenny, and D. Meyer. Probabilistic collocation and lagrangian sampling for advective tracer transport in randomly heterogeneous porous media. *Advances in Water Resources*, 34(12):1527–1538, 2011.
- [58] F. Müller, P. Jenny, and D. Meyer. Solver-based vs. grid-based multilevel monte carlo for two phase flow and transport in random heterogeneous porous media. *Journal of Computational Physics*, 268:39 – 50, 2014.

- 
- [59] A. Narayan and J. D. Jakeman. Adaptive Leja Sparse Grid Constructions for Stochastic Collocation and High-Dimensional Approximation. *SIAM Journal on Scientific Computing*, 36(6):2952–A2983, 2014.
  - [60] F. Nobile, L. Tamellini, and R. Tempone. Convergence of quasi-optimal sparse grids approximation of Hilbert-valued functions: application to random elliptic PDEs. Mathicse report 12/2014, EPFL, 2014. Submitted.
  - [61] F. Nobile, L. Tamellini, and R. Tempone. Comparison of Clenshaw–Curtis and Leja quasi-optimal sparse grids for the approximation of random PDEs. In *Spectral and High Order Methods for Partial Differential Equations - ICOSAHOM '14*, volume 105 of *Lecture Notes in Computational Science and Engineering*. Springer, 2016. To appear. Also available as MATHICSE report 41/2014.
  - [62] 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(6-7):979–1006, 2009.
  - [63] F. Nobile, R. Tempone, and C. Webster. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2411–2442, 2008.
  - [64] F. Nobile, R. Tempone, and C. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2309–2345, 2008.
  - [65] E. Pardoux and A. Țăbașanu. *Stochastic differential equations, backward SDEs, partial differential equations*. Stochastic Modelling and Applied Probability. Springer International Publishing, 2014.
  - [66] A. M. Quarteroni and A. Valli. *Numerical Approximation of Partial Differential Equations*. Springer Publishing Company, Incorporated, 1st ed. 1994. 2nd printing edition, 2008.
  - [67] M. Riva, L. Guadagnini, A. Guadagnini, T. Ptak, and E. Martac. Probabilistic study of well capture zones distribution at the lauswiesen field site. *Journal of contaminant hydrology*, 88(1-2):92–118, 2006.
  - [68] C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2004.
  - [69] C. Schillings and C. Schwab. Sparse, adaptive Smolyak quadratures for Bayesian inverse problems. *Inverse Problems*, 29(6), 2013.
  - [70] C. Schwab and R. A. Todor. Karhunen–loève approximation of random fields by generalized fast multipole methods. *Journal of Computational Physics*, 217(1):100 – 122, 2006.
  - [71] S. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Dokl. Akad. Nauk SSSR*, 4:240–243, 1963.
  - [72] M. L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York, 1999.
  - [73] D. W. Stroock and S. R. S. Varadhan. Diffusion processes with continuous coefficients, I. *Communications on Pure and Applied Mathematics*, 22:345–400, 1969.

- [74] A. Teckentrup, R. Scheichl, M. Giles, and E. Ullmann. Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients. *Numerische Mathematik*, 125(3):569–600, 2013.
- [75] F. Tesei and F. Nobile. A Multi Level Monte Carlo method with control variate for elliptic PDEs with log-normal coefficients. *Stochastic Partial Differential Equations: Analysis and Computations*, pages 1–47, 2015.
- [76] G. Wasilkowski and H. Wozniakowski. Explicit cost bounds of algorithms for multivariate tensor product problems. *Journal of Complexity*, 11(1):1 – 56, 1995.
- [77] D. Xiu and G. Karniadakis. Modeling uncertainty in flow simulations via generalized polynomial chaos. *J. Comput. Phys.*, 187(1):137–167, May 2003.
- [78] D. Zhang. In D. Zhang, editor, *Stochastic Methods for Flow in Porous Media*. Academic Press, San Diego, 2002.
- [79] G. Zhang, D. Lu, M. Ye, M. Gunzburger, and C. Webster. An adaptive sparse-grid high-order stochastic collocation method for bayesian inference in groundwater reactive transport modeling. *Water Resources Research*, 49(10):6871–6892, 2013.

