Reducing computational cost of the multilevel Monte Carlo method by construction of suitable pathwise integrators

By

Viktor Reshniak

A dissertation submitted in partial fulfillment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Computational Science

Middle Tennessee State University

May 2017

Dissertation Committee:

Dr. Yuri Melnikov

Dr. Abdul Khaliq

Dr. Tibor Koritsanszky

ACKNOWLEDGEMENTS

This dissertation appeared as a result of my collaboration with researchers at Middle Tennessee State University (MTSU) and Oak Ridge National Laboratory (ORNL) and I would like to thank everybody who made this possible. First of all, I would like to express my deepest gratitude to Professor Yuri Melnikov and Professor Abdul Khaliq who were my research advisors at MTSU. Without their continuous support and guidance, I would never be able to pass this way and avoid numerous mistakes. I want to express additional thanks to Professor Melnikov for his patience, understanding and involvement. I would also like to thank my dissertation committee and Professor Tibor Koritsanszky for his valuable advices which helped to improve the quality of dissertation.

I would like to extend my sincere appreciation to Dr. John Wallin and Dr. Donald Nelson for creating a very supportive working and research environment in the Computational Science Program and the Department of Mathematical Sciences at MTSU. Finally, I would like to express my special thanks to Dr. Clayton Webster and Dr. Guannan Zhang for giving me an opportunity to work with the leading professionals in the field of applied mathematics. Our joint project during my summer internship in the Computational and Applied Mathematics group at ORNL strongly influenced my future research and inspired me for this dissertation.

ABSTRACT

The multilevel Monte Carlo (MLMC) method has been recently proposed as a variance reduction technique for the efficient estimation of expected values of the quantities of interest associated with solutions of stochastic and random differential equations. By combining the ideas of multigrid discretization and Monte Carlo sampling, it allows to achieve the optimal asymptotical complexity of the estimator for the very large class of problems. The actual cost of the estimator, however, is more problem and solver dependent as the method requires one to solve a large number of decoupled deterministic problems. The efficiency of the estimator is hence strongly influenced by the complexity of the corresponding pathwise integrators. It is the task of this dissertation to study several problems for which the significant reduction in the computational complexity of the MLMC estimator can be achieved by the appropriate problem and level dependent choice of deterministic solvers. Three particular problems are considered: integration of stiff SDEs, estimation of initial guesses for iterative linear solvers and boundary values problems in randomly perturbed domains. The brief description of each problem is given below.

In Chapter II, we consider acceleration of the MLMC method in application to stochastic differential equations (SDEs). SDEs are often used in modeling of time-dependent phenomena at the mesoscopic level. Physical systems at this level are characterized by the presence of the vast range of temporal scales which makes them intrinsically stiff. In stochastic setting, stiffness is a serious issue in numerical treatment of differential systems due to non-trivial interaction between noise and multiscale dynamics. To resolve this issue, we propose the family of split-step implicit integrators which are capable to generate stable solutions without destroying geometry of the true stochastic dynamics. In the context of the MLMC method, the proposed integrators allow to exploit all the levels of the multilevel discretization without the need to explicitly resolve the fastest scale of the dynamics. The efficiency of the proposed technique is illustrated by applying it to stiff stochastic chemical

systems and both qualitative and quantitative results are presented.

Chapter III is devoted to the acceleration of the MLMC method in application to partial differential equations (PDEs) with random input data. As was mentioned above, MLMC requires solving a large number of decoupled deterministic problems corresponding to different realizations of input data. For stationary partial differential equations, these solutions are often constructed by means of iterative process and the choice of initial guess can have a drastic influence on its convergence. It will be shown that the estimation of initial guesses to iterative solvers can be efficiently performed by recycling previously calculated data. For this purpose, we use the kernel based approximation technique and perform the asymptotic cost analysis of the accelerated method to illustrate its superiority.

Finally, partial differential equations in random domains are discussed in Chapter IV. Problems with topological uncertainties appear in many fields ranging from nano-device engineering and analysis of micro electromechanical systems to design of bridges. In many of such problems, only part of the domain is subjected to random perturbations and conventional schemes relying on discretization of the whole domain become inefficient. We study linear PDEs in domains with boundaries comprised of both deterministic and random parts and apply the method of modified potentials with kernels given by the Green's functions defined on the deterministic part of the domain. This approach allows to reduce the size of the original differential problem by reformulating it as a boundary integral equation posed on the random part of the boundary only. The MLMC method is then applied to this modified integral equation leading to significant computational savings. We provide the qualitative analysis of the proposed technique and support it with numerical results.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Understanding of the temporal and spatial response of different systems to the imposed constraints and forcings is essential in many applications of science and engineering. After the invention of calculus in the 17th century mathematical description of such processes is naturally formulated in terms of differential equations. Since then, the field of qualitative and quantitative analysis of ordinary and partial differential equations has reached the high level of maturity and still remains an active research area. There have been discovered, however, numerous applications of practical importance where the classical approach faces serious limitations.

Firstly, most of the real-life processes are characterized by the presence of the vast range of spatial and temporal scales. Standard models are often based on the phenomenological description of the underlying laws and the general principles of energy conservation which makes differential equations a convenient tool for the macroscopic coarse-grained description of the systems of interest. At the same time, macroscopic behavior is the result of numerous microscopic processes and subsystem interactions taking place on much smaller scales. Microscopic fluctuations of these underworld processes are often intrinsically nonlinear and chaotic which directly influences the macroscale evolution of the underlying system. According to [Hak83], description of the liquid at the microscopic level deals with individual atoms and molecules; the mesoscopic level considers clusters of atoms which are large compared to interatomic distances but are smaller than macroscopic patterns; at the macroscpic level, one wish to study the corresponding spatial patterns. Direct microscopic simulation is almost never possible either because of the limited computing resources or due to the poorly understood physics at this level. Mean-field macroscopic dynamics also has limited accuracy. It is the mesoscopic approach which allows to describe the interaction of the small (macro) subsystem with the large (micro) environment. As is

shown in [Kot08], the mesoscopic level is probabilistic in nature and stochastic differential equations (SDEs) provide a faithful model for the description of microscopic fluctuations in a certain mathematical idealization. During the last several decades these ideas became increasingly important and found wide applications in physics [Cho94, Str92], chemistry [Gar09, Gil77, vK92], biology [AK15, Bai75, EK09], neural science [LL10, Tuc89, Wal81] and finance [BS73, HP81, Mer73] to name a few. The abstract mathematical formulation of the problem also attracted a lot of attention since SDEs are forced by irregular processes, have nondifferentiable paths and require a non-standard stochastic calculus. A special interest has been paid to the study of stochastic ordinary differential equations (SODEs) driven by semimartingale processes due to their exceptional practical importance and a number of excellent monographs has been published [App09, Arn74, Fri75, GS72, Kha12, Mao07, Øks03, Pro05]. Stochastic partial differential equations also represent a rapidly growing field of study and an increasing number of monographs and papers appear every year [Cho14, DKM$^+$09, HØUZ10, Kot08, LR15, PZ07, PZ92, PR07, Wal86]. Other related topics include stochastic optimal control [Nis15, ØSB07, Pha09], backward stochastic differential equations [EKPQ97, PP90, PT99], SDEs with Markovian switching [MY06, YZ10] and SDEs driven by non-martingale noise [HLN12, MN68, NR02].

As explained above, stochastic differential equations arise in the modeling of systems which exhibit intrinsically random behavior. Additionally, in engineering practice, uncertainties often stem from the presence of model discrepancies, numerical errors and inexact measurements. For instance, according to deterministic modeling framework, information about internal properties of the system and its external interactions enters the corresponding initial-boundary-value problems in the form of operator coefficients, forcing terms, initial and boundary conditions. Classical models of mathematical physics are constructed under the assumption of the full knowledge and availability of this data. Unfortunately, it is often not precisely known or is very expensive to determine experimentally. The uncer-

tainty associated with such incomplete knowledge can be incorporated into the models by using random fields for the representation of the input data [Chr92, Van88]. Numerous applications of this approach include reliability analysis [GS91b, HM00, KK88, Rac01], sensitivity analysis [CMM09], modeling of heterogeneous media [Kam05, OS98, OS07] and multiphase flows [GD98], problems in structural mechanics [HB98, LC09, Sch07] and fluid mechanics [MKNG01, MRN$^+$02, MK10], safety assessments for nuclear reactors [BHGH92], subsurface stormflow modeling [GTMW96] and others.

The goal of uncertainty quantification is to make predictions about response of the systems in the presence of different forms of uncertainty. It is very uncommon that such behavior can be described analytically and approaches using approximate and computational techniques are required. A review of some of the state-of-the-art techniques can be found, for instance, in [CD15, GWZ14, Ste09, Xiu09] and a basic overview of numerical methods for both stochastic and random differential equations is presented below.

I.1. Overview of numerical methods for stochastic differential equations

This section contains a brief overview of numerical methods for stochastic ordinary differential equations. Development of the similar theory for SPDEs is currently an active research area; some recent results can be found, for instance, in [JK09, DPJR10, GNT05]. While numerical approximations of stochastic partial differential equations are not considered in this work, one may refer to [LPS14] and references therein for the basic introduction to the field.

We start with the formulation and the discussion of the well-posedness of initial-value problems for SODEs driven by semimartingale processes. The elements of the classical Itô calculus for such equations are provided next as a motivation for the construction of numerical approximations. The classical Euler-Maruyama and Milstein methods are also

given to illustrate the convergence properties and restrictions of the numerical schemes derived from the stochastic Taylor series [KP92b, MT04, PBL10]. An overview of more general Itô schemes completes the discussion in this section.

I.1.1. Problem formulation

Consider the following initial-value problem for the system of stochastic ordinary differential equations

$$\begin{cases} dX(t) = f(t,X)dt + \sum_{i=1}^{m} g_i(t,X)dW_i + \sum_{j=1}^{l} \int_{\mathcal{E}} c_j(v,t,X_{t^-}) p_{\varphi_j}(dv,dt), & t \in [t_0,T] \\ X(t_0) = X_0, \end{cases} \tag{I.1}$$

where $X(t) \in \mathbb{R}^d$, $f : [t_0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $g = (g_1, g_2, .., g_m) : [t_0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$, $c = (c_1, c_2, .., c_n) : [t_0,T] \times \mathbb{R}^d \times \mathcal{E} \to \mathbb{R}^{d\times l}$, $W(t)$ is an m-dimensional Wiener process defined on the complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $p_\varphi(dv,dt)$ is a Poisson measure, $\mathcal{E}$ is a finite label set and $\varphi(dv)$ is a measure on $\mathcal{E}$.

The global Lipschitz and the linear growth conditions for the coefficients $f(t,X)$, $g(t,X)$ and $c(v,t,X_{t^-})$ in (I.1) guarantee the well-posedness of the above problem.

Assumption I.1.1 (Global Lipschitz condition)**.**

$$\begin{aligned} \left|f(t,x) - f(t,y)\right|^2 + \sum_{i=1}^{m} \left|g_i(t,x) - g_i(t,y)\right|^2 \\ + \sum_{j=1}^{l} \int_{\mathcal{E}} \left|c(v,t,x) - c(v,t,y)\right|^2 \varphi_j(dv) \le C_1|x-y|^2 \end{aligned}$$

for $t \in [0,T]$ *and all* $x, y \in \mathbb{R}^d$.

Assumption I.1.2 (Linear growth condition)**.**

$$\left|f(t,x)\right|^2 + \sum_{i=1}^{m} \left|g_i(t,x)\right|^2 + \sum_{j=1}^{l} \int_{\mathcal{E}} \left|c(v,t,x)\right|^2 \varphi_j(dv) \le C_2(1+|x|^2)$$

for $t \in [0, T]$ and all $x \in \mathbb{R}^d$.

The following theorem establishes existence and uniqueness of strong solutions of jump-diffusion SDEs under the above assumptions on the coefficients.

Theorem I.1.1 ([PBL10, Theorem 1.9.3])**.** *Suppose that the coefficient functions $f(\cdot)$, $g_i(\cdot)$ and $c_j(\cdot)$ of the SDE* (I.1) *satisfy the Lipschitz condition in Assumption I.1.1 and the linear growth condition in Assumption I.1.2. Then the SDE* (I.1) *admits a unique strong solution. Moreover, the solution $X(t)$ of the SDE* (I.1) *satisfies the estimate*

$$\mathbb{E}\left[\sup_{0\le s\le T}|X(s)|^2\right] \le C\big(1+\mathbb{E}\left[|X(0)|^2\right]\big)$$

with $T < \infty$, where C is a finite positive constant.

I.1.2. Itô calculus and stochastic Taylor series

Let $Z_1(t)$, $Z_2(t)$ be two stochastic processes and consider the corresponding quadratic variation and covariation processes

$$[Z_1]_t \overset{P}{=} \lim_{h\to 0}[Z_1]_{h,t} = \lim_{h\to 0}\sum_k \big(Z_1(t_k) - Z_1(t_{k-1})\big)^2,$$
$$[Z_1, Z_2]_t \overset{P}{=} \lim_{h\to 0}[Z_1, Z_2]_{h,t} = \lim_{h\to 0}\sum_k \big(Z_1(t_k - Z_1(t_{k-1})\big)\big(Z_2(t_k) - Z_2(t_{k-1})\big)$$

for the sequence of time discretizations $t_k = kh : k \in 0, 1, ...$ with a decreasing time step h. The equalities are understood in the sense of convergence in probability.

The semimartingale process $X(t)$ admits the following decomposition

$$X(t) = X(t_0) + A^{i,c}(t) + M^{i,c}(t) + X^{i,d}(t),$$

where $A^{i,c}(t)$ is a continuous process with finite total variation, $M^{i,c}(t)$ is a continuous local martingale and $X^{i,d}(t)$ is the jump part of the process. Then, for the $\mathcal{C}^{(1,2)}$ function

$u : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ and a semimartingale $X(t)$, the Itô formula provides the stochastic variant of the deterministic chain rule [App09, PBL10, Pro05]

$$du(t, X_t) = \frac{\partial u(t, X_t)}{\partial t} dt + \sum_{i=1}^{d} \frac{\partial u(t, X_t)}{\partial x^i} dX_t^{i,c} + \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2 u(t, X_t)}{\partial x^i \partial x^j} d[M^{i,c}, M^{k,c}]_t + \big(u(t, X_t) - u(t-, X_{t-})\big).$$

The above formula agrees with the classical deterministic calculus in that the quadratic variation and covariation equals to zero for deterministic functions with bounded variation. In application to the strong solution of the SDE in (I.1), the Itô formula reads as

$$du(t, X_t) = \left(\frac{\partial u(t, X_t)}{\partial t} + \sum_{i=1}^{d} f_t^i \frac{\partial u(t, X_t)}{\partial x^i} + \frac{1}{2} \sum_{i,j=1}^{d} \sum_{k=1}^{m} g_k^i g_k^j \frac{\partial^2 u(t, X_t)}{\partial x^i \partial x^j} \right) dt + \sum_{i=1}^{d} \sum_{k=1}^{m} g_k^i \frac{\partial u(t, X_t)}{\partial x^i} dW_k + \sum_{j=1}^{l} \int_{\mathcal{E}} \Big(u(t, X_t) - u(t^-, X_{t^-}) \Big) p_{\varphi_j}(dv, dt),$$

where the arguments of the coefficients $f(t, X_t)$ and $g(t, X_t)$ are omitted for the economy of notation. The integral variant of the Itô formula has the form

$$u(t, X_t) = u(0, X_0) + \int_0^t L^0 u(s, X_s) ds + \sum_{k=1}^{m} \int_0^t L^k u(s, X_s) ds + \sum_{j=1}^{l} \int_0^t \int_{\mathcal{E}} L_v^{-j} u(s, X_s) p_{\varphi_j}(dv, ds), \tag{I.2}$$

where the operators L^0, L^k and L^{-j} are defined as

$$L^0 = \frac{\partial}{\partial t} + \sum_{i=1}^{d} f_t^i \frac{\partial}{\partial x^i} + \frac{1}{2} \sum_{i,j=1}^{d} \sum_{k=1}^{m} g_k^i g_k^j \frac{\partial^2}{\partial x^i \partial x^j},$$

$$L^k = \sum_{i=1}^{d} g_k^i \frac{\partial}{\partial x^i} dW_k,$$

$$L_v^{-j} = u(t, X_s + c_j(v, t, X_s)) - u(t, X_s).$$

By repeatedly applying the formula (I.2) to the functions $L^0u(s, X_s)$, $L^ku(s, X_s)$ and $L^{-k}u(s, X_s)$ one gets the generalization of the Itô formula which allows explicit evaluation via the Wagner-Platen expansion.

Theorem I.1.2 ([PBL10, Theorem 4.4.1])**.** *For two given stopping times ρ and τ with $0 \leq \rho \leq \tau \leq T$ a.s., a hierarchical set $\mathcal{A} \in \mathcal{M}$, and a function $u : [0, \infty) \times \mathbb{R} \to \mathbb{R}$, we obtain the corresponding Wagner-Platen expansion*

$$u(t, X_\tau) = \sum_{\alpha \in \mathcal{A}} I_\alpha \big[u_\alpha(\rho, X_\rho)\big]_{\rho,\tau} + \sum_{\alpha \in \mathcal{B}(\mathcal{A})} I_\alpha \big[u_\alpha(\cdot, X_\cdot)\big]_{\rho,\tau} \tag{I.3}$$

assuming that the function u and the coefficients of the SDE (I.1) *are sufficiently smooth and integrable such that the arising coefficient functions u_α and are well defined and the corresponding multiple stochastic integrals exist.*

In the above theorem, a row vector $\alpha = (j_1, j_2, ..., j_n)$ is a multi-index of length $n = n(\alpha)$. The set of all multi-indices α is denoted by

$$\mathcal{M} = \{(j_1, ..., j_n) : j_i \in \{-l, ..., -1, 0, 1, ..., m\}, i \in 1, 2, ..., n\}.$$

The number of negative components of the multi-index is denoted by $s(\alpha)$. Moreover, $\alpha-$ and $-\alpha$ denote the multi-indices obtained by removing the last and the first components from α respectively. Additionally, a subset $\mathcal{A} \in \mathcal{M}$ is hierarchical if $\sup_{\alpha \in \mathcal{A}} n(\alpha) < \infty$ and $-\alpha \in \mathcal{A}$ for each nonempty α. The remainder set is defined as

$$\mathcal{B}(\mathcal{A}) = \{\alpha \in \mathcal{M} \setminus \mathcal{A} : -\alpha \in \mathcal{A}\}.$$

The stochastic integrals are defined recurrently as

$$I_\alpha[u(\cdot)]_{\rho,\tau} = \begin{cases} u(\tau,x) & \text{for } l(\alpha)=0, \\ \int_\rho^\tau I_{\alpha-}[u(\cdot)]_{\rho,z}dz & \text{for } l(\alpha)\geq 1, j_n = 0, \\ \int_\rho^\tau I_{\alpha-}[u(\cdot)]_{\rho,z}dW_{j_n}(z) & \text{for } l(\alpha)\geq 1, j_n \in \{1,...,m\}, \\ \int_\rho^\tau \int_{\mathcal{E}} I_{\alpha-}[u(\cdot)]_{\rho,z^-}p_{\varphi_{j_n}}(dv_{s(\alpha)},dt) & \text{for } l(\alpha)\geq 1, j_n \in \{-1,...,-l\}, \end{cases}$$

where $u(\cdot) = u(t,x,v_1,...,v_{s(\alpha)})$ and the corresponding integrands are given by

$$u_\alpha(t,x,v_1,...,v_{s(\alpha)}) = \begin{cases} u(t,x) & \text{for } l(\alpha)=0, \\ L^{j_1}f_{-\alpha}(t,x,v_1,...,v_{s(-\alpha)}) & \text{for } l(\alpha)\geq 1, j_1 \in \{0,1,...,m\}, \\ L^{-j_1}f_{-\alpha}(t,x,v_1,...,v_{s(-\alpha)}) & \text{for } l(\alpha)\geq 1, j_1 \in \{-1,...,-l\}. \end{cases}$$

The sets of admissible integrands are described in [PBL10].

I.1.3. Euler-Maruyama and Milstein methods

The Wagner-Platen expansion in (I.3) is the stochastic analog of the deterministic Taylor series. When applied to the function $u(t,x) = x$, it provides the basis for the derivation of approximation schemes for the SDE in (I.1). The two simplest schemes are the Euler-Maruyama and the Milstein methods which correspond to the hierarchical sets

$$\mathcal{A}_{EM} = \{\varnothing,-1,0,1\}, \qquad \mathcal{A}_{Mil} = \{\varnothing,-1,0,1,(1,1),(1,-1),(-1,1),(-1,-1)\}$$

in the case of one-dimensional noise terms, i.e., for $m = l = 1$.

Assume that the coefficients $c_j(v,t,X_t) = c_j(t,X_t)$ are mark-independent and consider the following stochastic integrals

$$I^{\rho,\tau}_{(0)} = \int_\rho^\tau dz, \qquad I^{\rho,\tau}_{(i)} = \int_\rho^\tau dW_i(z), \qquad I^{\rho,\tau}_{(-j)} = \int_\rho^\tau \int_{\mathcal{E}} p_{\varphi_j}(dv,dz),$$
$$I^{\rho,\tau}_{(i_1,i_2)} = \int_\rho^\tau \int_\rho^{z_2} dW_{i_1}(z_1)dW_{i_2}(z_2), \qquad I^{\rho,\tau}_{(i_1,-j_2)} = \int_\rho^\tau \int_{\mathcal{E}} \int_\rho^{z_2} dW_{i_1}(z_1)p_{\varphi_{j_2}}(dv,dz_2),$$

$$I^{\rho,\tau}_{(-j_2,i_1)} = \int_\rho^\tau \int_\rho^{z_1} \int_{\mathcal{E}} p_{\varphi_{j_2}}(dv, dz_2) dW_{i_1}(z_1), \tag{I.4}$$
$$I^{\rho,\tau}_{(-j_1,-j_2)} = \int_\rho^\tau \int_{\mathcal{E}} \int_\rho^{z_2} \int_{\mathcal{E}} p_{\varphi_{j_1}}(dv_1, dz_1) p_{\varphi_{j_2}}(dv_2, dz_2).$$

Then the single step of the Euler-Maruyama scheme for the k-th component of the approximate solution Y_t on the interval $t \in [\rho, \tau]$ can be written as

$$Y_\tau = Y_\rho + f(\rho, X_\rho) I^{\rho,\tau}_{(0)} + \sum_{i=1}^m g_i(\rho, X_\rho) I^{\rho,\tau}_{(i)} + \sum_{j=1}^l c_j(\rho, X_\rho) I^{\rho,\tau}_{(-j)}$$

while the Milstein scheme gives

$$\begin{aligned} Y^k_\tau = Y^k_\rho &+ f^k(\rho, X_\rho) I^{\rho,\tau}_{(0)} + \sum_{i=1}^m g^k_i(\rho, X_\rho) I^{\rho,\tau}_{(i)} + \sum_{j=1}^l c^k_j(\rho, X_\rho) I^{\rho,\tau}_{(-j)} \\ &+ \sum_{i_1=1}^m \sum_{i_2=1}^m \sum_{i=1}^d g^i_{i_1} \frac{\partial g^k_{i_2}}{\partial x^i} I^{\rho,\tau}_{(i_1,i_2)} + \sum_{i_1=1}^m \sum_{j_2=1}^l \sum_{i=1}^d g^i_{i_1} \frac{\partial c^k_{j_2}}{\partial x^i} I^{\rho,\tau}_{(i_1,-j_2)} \\ &+ \sum_{i_1=1}^m \sum_{j_2=1}^l \Big(g^k_{i_1}(\rho, Y_\rho + c_{j_2}(\rho, Y_\rho)) - g^k_{i_1}(\rho, Y_\rho) \Big) I^{\rho,\tau}_{(-j_2,i_1)} \\ &+ \sum_{j_1=1}^l \sum_{j_2=1}^l \Big(c^k_{j_1}(\rho, Y_\rho + c_{j_2}(\rho, Y_\rho)) - c^k_{j_1}(\rho, Y_\rho) \Big) I^{\rho,\tau}_{(-j_1,-j_2)}. \end{aligned}$$

Definition I.1.1. *A discrete approximation Y_n converges strongly with order $\gamma \in (0, \infty)$ and weakly with order $\beta \in (0, \infty)$ if there exist finite constants C_1 and C_2 and a positive constant δ_0 (independent of h) such that for each $h \in (0, \delta_0)$*

$$\mathbb{E}\left[|X(t_n) - Y_n|\right] \le C_1 h^\gamma,$$
$$\left|\mathbb{E}\left[b\big(X(t_n)\big) - b\big(Y_n\big)\right]\right| \le C_2 h^\beta,$$

where $b \in \mathcal{C}^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ and all its derivatives of order up to $2(\beta+1)$ have polynomial growth.

In order for the truncated Wagner-Platen expansion (I.3) to converge with a strong order $\gamma \in \{0.5, 1, 1.5, ...\}$ and a weak order $\beta \in \{1, 2, 3, ...\}$, one has to use the hierarchical sets [KP92b, PBL10]

$$\mathcal{A}_\gamma = \left\{ \alpha \in \mathcal{M} : n(\alpha) + z(\alpha) \leq 2\gamma \quad \text{or} \quad n(\alpha) = z(\alpha) = \gamma + \frac{1}{2} \right\},$$
$$\mathcal{A}_\beta = \left\{ \alpha \in \mathcal{M} : n(\alpha) \leq \beta \right\},$$

where $n(\alpha)$ and $z(\alpha)$ denote the length and the number of zero components of the multi-index α. In addition to the above conditions, classical convergence results also require for the coefficients of the Wagner-Platen expansion to be globally Lipschitz and to grow at most linearly, see [PBL10, Theorems 6.4.1, 11.2.1] and [KP92b, Theorems 10.6.3, 14.5.1].

Under the above conditions, the Euler-Maruyama and the Milstein schemes both have the weak order 1 and the strong orders 0.5 and 1 respectively. The main issue in the construction of higher order schemes is the approximation of multiple and mixed stochastic integrals in (I.4) since efficient algorithms for their simulation exist only in the case of one dimensional noise forcing. Multidimensional approximations are usually derived by using Fourier series expansions which converge slowly and often have limited practical value [Wik01]. Approximation of mixed stochastic integrals involving integration over one or several Poisson measures is also problematic since it requires one to explicitly track the jump times and marks of all Poisson processes. Therefore, the first order of strong convergence achieved by the Milstein scheme is usually the best one can get in practical calculations for general multidimensional SDEs.

I.1.4. General Itô schemes

Derivation of the stochastic Taylor series indicates that the higher order of stochastic convergence is achieved by including more terms with multiple stochastic integrals into the scheme. However, the required conditions for the convergence of the truncated expansion

are often too restrictive for many practical problems. For instance, local Lipschitz continuity of the coefficients can be a sufficient condition for the convergence under additional assumptions on the boundedness of moments [HMS02], the linear growth of coefficients [YM08] or the Khasminskii-type condition [Mao16]. At the same time, Hutzenthaler et al. showed in [HJK11] that the classical Euler-Maruyama method diverges in finite time for SDEs with non-globally Lipschitz continuous coefficients with superlinear growth. They proposed a tamed Euler scheme where the drift term is modified so that it is uniformly bounded and proved its convergence under the globally one-sided Lipschitz condition for the drift term and with globally Lipschitz diffusion coefficient [HJK12]. The tamed Euler scheme was also discussed in [DKS14, S$^+$13, ZWH14] and the tamed Milstein method was proposed in [KS14, WG13]. These results motivate the development of numerical methods which do not follow directly from the Wagner-Platen expansion (I.3) but contain terms with multiple stochastic integrals. Such schemes can be considered as a generalization of the classical stochastic expansion.

Implicit modifications of the stochastic Taylor series provide a variant of such generalization [KP92a, TB01, WGW12]. Drift-implicit methods give a similar technique by making implicit only deterministic terms of the expansion. This allows to resolve the issue of fully implicit methods associated with possible singularity of linear systems in nonlinear iterative solvers due to the dependence of their matrices on the stochastic integrals. Convergence and stability properties of the drift-implicit methods were discussed, for example, in [AVZ13, MS13, HMS02, Tal02] for SDEs driven by Brownian motion and in [BLP07, HK05, HK06, sWMX07] for jump-diffusion systems. Additionally, the combination of semi-implicit and implicit steps is used in the family of composite methods to achieve better stability properties [BT01, OAHR11]. The balanced schemes [HG11, KS06, MPS98, WL09] and the split-step methods [HH12, JZYH16, RKVZ15, VK15] form another group of methods which allow to incorporate implicitness into the stochastic part of

the scheme as well.

An alternative to the Wagner-Platen expansion is a large class of derivative-free stochastic Runge-Kutta (SRK) methods introduced in [Rüe82, BB98]. Rößler in [Röß06b, Röß06a] extended these results and applied a colored rooted tree analysis to derive the general order conditions for the coefficients of the scheme. Flexibility of the derived order conditions allows to construct the efficient SRK schemes for which the number of evaluations of the drift and diffusion coefficients is independent of the dimension of the driving stochastic process [Röß10]. On the other side, simulation of multiple stochastic integrals is still required for higher order convergence. Implicit versions of the SRK method also follow naturally from the general derivation of the scheme [BT04, DR09, HHR16]. Other variants of the SRK schemes include methods for jump-diffusion systems [BNR07, BR11], methods for systems with small noise [BRW10, VH12] and explicit S-ROCK stiff integrators [AC08, KB13].

Predictor-corrector schemes [BLP08, HSW07], multistep methods [BW06, BW07], exponential integrators [EL16, KB14], local linearization approximations [CJ08] and schemes with adaptive time-stepping [RW06, SK08] are some of the other methods which can be found in literature. Also, numerical methods for the approximation of continuous time Markov chain models will be discussed later in Chapter II.

I.2. Overview of numerical methods for random differential equations

This section contains an overview of the numerical techniques for random partial differential equations (PDEs). It is assumed that the random fields are time-independent which allows to consider only stationary PDEs without loss of generality. We start with the formulation of the boundary-value problems with random input data and the description of the required function spaces. This is followed by the discussion of techniques for the finite-dimensional representation of random fields. Stochastic spectral methods, which rely

on such representations, are discussed next along with their limitations on the stochastic dimensionality. The section is completed by the brief overview of perturbation techniques for the systems with small uncertainties.

I.2.1. Problem formulation

Let $\mathcal{L}$ be a possibly nonlinear stationary operator defined on a bounded Lipschitz domain $D \subset \mathbb{R}^d$, $d = 1, 2, 3$, with a boundary $\partial D = \partial D_D \cup \partial D_N$. Operator $\mathcal{L}$ has a random coefficient $a(x, \omega)$, $x \in D$, $\omega \in \Omega$, defined on the complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Here Ω denotes the sample space of possible outcomes, $\mathcal{F} \subset 2^{\Omega}$ is the σ-algebra of events, and $\mathbb{P}$ is the complete probability measure on $\mathcal{F}$. Denote by $u(x, \omega)$ the strong solution of the following stochastic boundary-value problem

$$\left\{ \begin{array}{rcll} \mathcal{L}\big(a(x,\omega), u(x,\omega)\big) & = & f(x,\omega) & \text{in } D \times \Omega, \\ \gamma\big(u(x,\omega)\big) & = & g(x,\omega) & \text{on } \partial D \times \Omega, \end{array} \right. \tag{I.5}$$

where γ is a trace operator which defines Dirichlet boundary condition on ∂D_D and Neumann boundary condition on ∂D_N. We require $u(x, \omega)$ to be a Bochner integrable function with values in some Banach space $W(D)$, i.e., $u(x, \omega) \in L^p(\Omega; W(D))$, the function space given by

$$L^p\Big(\Omega; W(D)\Big) := \left\{ u : \Omega \to W(D) \;\middle|\; u \text{ is strongly measurable and } \|u\|_{L^p(\Omega;W(D))} < \infty \right\}$$

with the corresponding norm

$$\|u\|^p_{L^p(\Omega;W(D))} = \left\{ \begin{array}{ll} \displaystyle\int_{\Omega} \|u(\cdot,\omega)\|^p_{W(D)}\, d\mathbb{P}(\omega) & \text{if } 0 < p < \infty, \\ ess\sup_{\omega\in\Omega} \|u(\cdot,\omega)\|_{W(D)} & \text{if } p = \infty. \end{array} \right.$$

For simplicity, we will write $L^p(\Omega)$ instead of $L^p(\Omega; W(D))$ when the particular function space $W(D)$ can be concluded from the context.

It is assumed that all random input fields are defined on the same probability space and are chosen so that the problem in (I.5) is well-posed in the following sense

Assumption I.2.1. *For the fixed values of the random parameter $\omega \in \Omega$, there exist unique realizations of the solution $u(x, \omega) \in W(D)$ such that*

$$\|u(x,\omega)\|_{W(D)} \leq C(\omega) \|f(x,\omega)\|_{W^{-1}(D)},$$

where $W^{-1}(D)$ is the dual space of $W(D)$ and C is a constant which may depend on ω.

I.2.2. Representation of random fields

The solution $u(x, \omega)$ of the boundary value problem in (I.5) is a random function. Dependence on the random variable adds additional dimensions to the problem description which makes the process of numerical discretization a more challenging task in comparison to its deterministic counterpart. In most cases, probabilistic characteristics of the random input fields are also subjected to certain simplifying assumptions due to the lack of real experimental data. It is critically important that the mathematical representation of these random fields facilitates the process of computation while keeping the model realistic.

The Gaussian assumption is one of the most frequently used in practice. This is due to the fact that the Gaussian random quantities are completely characterized by their first two moments which dramatically simplifies analysis of uncertain systems. Furthermore, it arises naturally in applications when only information about the second moments of the random field is available [SZ98]. Simulation of non-Gaussian processes is currently an active research area; some results can be found in [PHQ05, Ste09] and references therein.

The classical approach for the representation of Gaussian random fields is via the Karhunen-Loéve (KL) series [Loe78].

Theorem I.2.1 (Karhunen–Loéve)**.** *Let $a \in L^2(\Omega; L^2(D))$ be the square-integrable stochastic process with continuous covariance function $C(x, y)$. Then*

$$a(x, \omega) = \mathbb{E}\left[a(x, \omega)\right] + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(x)y_i(\omega), \tag{I.6}$$

where

$$y_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_D \Big(u(x, \omega) - \mathbb{E}\left[u(x, \omega)\right]\Big)\phi_i(x)dx$$

and λ_i, $\phi_i(x)$ are the eigenvalues and the eigenfunctions of the integral operator with covariance function as kernel, i.e.,

$$\int_D C(x, y)\phi_i(y)dy = \lambda_i\phi_i(x).$$

Random variables y_i are pairwise uncorrelated, zero-mean and unit-variance. Furthermore, if $a(u, \omega)$ is Gaussian, then y_i are also independent and $y_i \sim N(0, 1)$.

The choice of the covariance eigenfunctions $\phi_i(x)$ for the basis of $L^2(D)$ is optimal in the sense that the error of the truncation of the series in (I.6) is minimized in $L^2(\Omega; L^2(D))$. Comparison of the KL expansion with some of the other methods for the approximation of random fields can be found, for instance, in [Gri06, LK93, PHQ04, Sch97, Ste09].

Representation of random fields by the series in (I.6) requires explicit knowledge of the eigenpairs of the covariance operator and thus cannot be applied for the unknown solution $u(x, \omega)$ of the problem in (I.5). Generalized polynomial chaos (gPC) expansion is another technique which is most commonly employed in practice for this purpose. It allows to represent the arbitrary random function via orthogonal polynomials of independent random variables. The independence of the random variables is an essential requirement which helps to simplify analysis of the problem by seeking its solution in a functional space with the tensor product structure. In this regard, it is hard to overestimate the importance of the

Karhunen–Loéve representation of the input Gaussian fields as a linear combination of i.i.d. variables.

Let us consider the random vector $y = [y^1, ..., y^N]^T : \Omega \to \Gamma$ with a support in the image space $\Gamma(\Omega) = \prod_{k=1}^{N} \Gamma_k \in \mathbb{R}^N$ equipped with a product probability measure $\mu(y)dy = \mu(y_1)dy_1 \times ... \times \mu(y_N)dy_N$, where $\mu_i(y_i)$ is a distribution on Γ_i and $\mathbb{P}(y) = \int_\Gamma \mu(y)dy$ is a probability measure on Γ. Let $\mathcal{P}_i^{d_i}$ denote the space of polynomials of degree at most d_i which are orthonormal with respect to $\mu_i(y_i)dy_i$, i.e.,

$$\mathcal{P}_i^{d_i} := \left\{ v : \Gamma_i \to \mathbb{R} : v \in \text{span}\{\psi_m(y_i)\}_{m=0}^{d_i} \right\} \quad \text{and} \quad \int_{\Gamma_i} \psi_m(y_i)\psi_n(y_i)\mu_i(y_i)dy_i = \delta_{mn},$$

where $\psi_m(y_i)$ is the polynomial of degree m and δ_{mn} is the Kronecker delta function.

Then the space of N-variate orthonormal polynomials in Γ of total degree at most P is defined as

$$\mathcal{P}_N^P = \bigotimes_{|d| \le P} \mathcal{P}_i^{d_i},$$

and the tensor product is over all possible combinations of the multi-index $d = (d_1, ..., d_N)$ satisfying $|d| = \sum_{k=1}^{N} \le P$. The basis functions of $\mathcal{P}_N^P$ are given by

$$\mathcal{P}_N^P = \text{span}\left\{ \Psi_m(y) := \prod_{|m| \le P} \psi_{m_i} \;\middle|\; m = (m_1, ..., m_N),\ m_i \le d_i \right\}$$

and the total number of basis functions is

$$\dim(\mathcal{P}_N^P) = \binom{N+P}{N}.$$

The P-th order gPC approximation of the arbitrary second-order random function of N

independent random variables has the following representation in the above basis [CM47]

$$u(x,\omega) \approx \sum_{m=1}^{M} \hat{u}_m(x)\Psi_m(y), \qquad M = \binom{N+P}{N},$$

where the coefficients $\hat{u}_m(x)$ of the expansion can be determined by projecting $u(x,\omega)$ onto $\mathcal{P}_N^P$ as

$$\hat{u}_m(x) = \mathbb{E}\left[u(x,\omega)\Psi_m(y(\omega))\right] = \int_\Gamma u(x,y)\Psi_m(y)\mu(y)dy. \tag{I.7}$$

I.2.3. Stochastic spectral methods

By analogy with the Karhunen–Loéve series, the formula for the coefficients of the gPC expansion (I.7) is useless for the unknown function $u(x,\omega)$. The goal of the stochastic spectral methods is to determine these coefficients by exploiting orthogonality properties of the underlying spaces.

A typical approach, employed in the family of stochastic Galerkin (SG) methods, is to start with the gPC approximation of the solution

$$u_N^P(x,\omega) = \sum_{m=1}^{M} \hat{u}_m(x)\Psi_m(y)$$

and to satisfy the governing equations (I.5) in the following weak sense

$$\begin{aligned} \int_\Gamma \mathcal{L}(a,x,y)\big(u_N^P(x,y)\big)w(y)\mu(y)dy &= \int_\Gamma f(x,y)w(y)\mu(y)dy \qquad \text{in } D, \\ \int_\Gamma \gamma\big(u_N^P(x,y)\big)w(y)\mu(y)dy &= \int_\Gamma g(x,y)w(y)\mu(y)dy \qquad \text{on } \partial D \end{aligned}$$

for all $w(y) \in \mathcal{P}_N^P$. This gives the system of M coupled deterministic PDEs for the expansion coefficients $\hat{u}_m(x)$. It is worth noting that the choice of approximation in the parameter space Γ is not limited to the generalized Polynomial Chaos. Other approaches using standard

global and piecewise continuous finite element approximations have also been proposed [BTZ04, DBO01].

The stochastic Galerkin procedure is a well-established technique and an extensive literature exists for both theoretical [BTZ04, CD15, DBO01, MK05] and practical [EU10, FST05, GS91a, XLSK01] aspects of its implementation. It is exceptionally efficient for systems with moderate dimensionality of parameter space and shows exponentially fast convergence for equations with coefficients depending analytically on the random inputs. However, its disadvantages are also well known. For complicated operators, the derivation of the deterministic system corresponding to the stochastic Galerkin method can be a non-trivial task. Moreover, the method is intrusive as the conventional codes for deterministic PDEs have to be modified in order to include the projection onto the parameter space.

Instead, in the stochastic collocation (SC) methods, the governing equations in (I.5) are satisfied at a discreet set of points $\{y_m\}_{m=1}^M$ in the parameter space Γ and the solution is then reconstructed in the form of interpolant as

$$\mathcal{I}u(x,y) = \sum_{m=1}^{M} \tilde{u}_m(x)\Phi_m(y),$$

where $\{\Phi_m(y)\}_{m=1}^M$ is the appropriately chosen basis for interpolation in Γ. This approach requires one to solve a number of uncoupled deterministic problems for each collocation point $y_i \in \Gamma$. Since existing software can be used to find solutions of these problems, the SC methods are nonintrusive and can be easily applied to arbitrary linear and non-linear PDEs. While preserving the accuracy of the stochastic Galerkin methods, they also offer some other advantages like the possibility to treat nonindependent and unbounded random variables or nonlinearly parameterized input fields [BNT07, GWZ14, XH05].

Computational efficiency of both stochastic Galerkin and collocation methods rely heavily on the choice of quadrature and collocation points in $\Gamma = \prod_{k=1}^N \Gamma_k$. This choice is often nontrivial due to deteriorating accuracy and exponentially growing complexity of

conventional methods for the numerical integration and interpolation in high-dimensional spaces. Some of the commonly applied techniques include integration/interpolation on tensor product grids with full [BNT07], sparse [NTW08b, Smo63] or adaptive anisotropic structure [MZ09, NTW08a]. The full tensor product grid integration of an $\mathcal{C}^r$ function has the accuracy $O(n^{-r/d})$ and requires n^d function evaluations, where d is the number of dimensions and n is the number of quadrature nodes in each dimension. This is known as the curse of dimensionality since the applicability of the method is limited only to parameter spaces with moderate number of dimensions. Quadrature rules based on sparse and anisotropic grids have better complexity but cannot beat the accuracy of quadratures on full tensor product grids. Thus, they also suffer from the curse of dimensionality.

I.2.4. Perturbation methods

The perturbation techniques have been widely used in engineering practice for the estimation of the response statistics of systems with small uncertainties. The basic idea of the method is to expand all the functions about their mean values and to determine the unknown coefficients by equating the same order terms. For instance, consider the following equation with random operator and right hand side

$$\mathcal{L}\big(x,y\big)u(x,y) = f(x,y), \qquad x \in D,\ y \in \Gamma \subset \mathbb{R}^N. \tag{I.8}$$

By expanding the random input fields and the solution into the Taylor series around the mean value of the random parameter, one gets

$$\left\{\mathcal{L}\big(x,\mu_y\big) + \sum_{i=1}^{N} \frac{\partial \mathcal{L}\big(x,\mu_y\big)}{\partial y_i}\epsilon_i \Delta y_i + \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial^2 \mathcal{L}\big(x,\mu_y\big)}{\partial y_i \partial y_j}\epsilon_i\epsilon_j \Delta y_i \Delta y_j + \ldots\right\}$$
$$\times \left\{u\big(x,\mu_y\big) + \sum_{i=1}^{N} \frac{\partial u\big(x,\mu_y\big)}{\partial y_i}\epsilon_i \Delta y_i + \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial^2 u\big(x,\mu_y\big)}{\partial y_i \partial y_j}\epsilon_i\epsilon_j \Delta y_i \Delta y_j + \ldots\right\}$$

$$= \left\{ f\big(x,\mu_y\big) + \sum_{i=1}^{N} \frac{\partial f\big(x,\mu_y\big)}{\partial y_i}\epsilon_i \Delta y_i + \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial^2 f\big(x,\mu_y\big)}{\partial y_i \partial y_j}\epsilon_i\epsilon_j \Delta y_i \Delta y_j + ... \right\},$$

where ϵ_i are small parameters, $\mu_y = \mathbb{E}\left[y\right]$ and $\Delta y = y - \mu_y$. By multiplying and collecting the terms with the same powers of ϵ_i, we get the following sequence of deterministic equations for the coefficient of $u(x,y)$

$$\begin{aligned}
u &= \mathcal{L}^{-1} f, \\
\frac{\partial u}{\partial y_i} &= \mathcal{L}^{-1}\left(\frac{\partial f}{\partial y_i} - \frac{\partial \mathcal{L}}{\partial y_i} u\right), \\
\frac{\partial^2 u}{\partial y_i \partial y_j} &= \mathcal{L}^{-1}\left(\frac{\partial^2 f}{\partial y_i \partial y_j} - \frac{\partial \mathcal{L}}{\partial y_i}\frac{\partial u}{\partial y_j} - \frac{\partial \mathcal{L}}{\partial y_j}\frac{\partial u}{\partial y_i} - \frac{\partial^2 \mathcal{L}}{\partial y_i \partial y_j} u\right), \qquad i,j = 1,...,N,
\end{aligned}$$

where all operators and functions are evaluated at (x,μ_y).

The first-order approximation of the mean and covariance of $u(x,y)$ can be easily calculated as

$$\begin{aligned}
\mathbb{E}\left[u(x,y)\right] &= u\big(x,\mu_y\big), \\
\mathbb{C}ov\left[u(x,y),u(z,y)\right] &= \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial u\big(x,\mu_y\big)}{\partial y_i}\frac{\partial u\big(z,\mu_y\big)}{\partial y_j}\mathbb{E}\left[\Delta y_i \Delta y_j\right].
\end{aligned}$$

The second-order approximations of the response statistics require the knowledge of the joint probability distribution of the random parameters y_i and thus can be easily computed only for Gaussian fields [KH92, LMB87]. Higher-order approximations have also been discussed, for example, in [Kam13].

Neumann expansion method is another technique which is conceptually similar to the perturbation method. It starts with the first order approximation of the random operator by decomposing it into deterministic and perturbation parts as follows

$$\mathcal{L}(x,y) = \mathcal{L}_0(x) + \Delta\mathcal{L}(x,y)$$

and applies Taylor series expansion to the inverse of this operator, i.e.,

$$\mathcal{L}^{-1}(x,y) = \big(\mathcal{L}_0(x) + \Delta\mathcal{L}(x,y)\big)^{-1} = \Big(\sum_{k=0}^{\infty}(-1)^k\big(\mathcal{L}_0^{-1}(x)\Delta\mathcal{L}(x,y)\big)^k\Big)\mathcal{L}_0^{-1}(x).$$

The convergence of the above series is guaranteed when $\|\mathcal{L}_0(x)\Delta\mathcal{L}(x,y)\| < 1$ in the operator norm. This condition is not very restrictive and can be satisfied for perturbations of arbitrary size by proper rescaling [YSD88].

The Neumann series solution of the problem in (I.8) with deterministic right hand side takes the form

$$\begin{aligned} u(x,y) &= \mathcal{L}^{-1}(x,y)f(x) \\ &= \Big(I - \mathcal{L}_0^{-1}(x)\Delta\mathcal{L}(x,y) + \big(\mathcal{L}_0^{-1}(x)\Delta\mathcal{L}(x,y)\big)^2 - ...\Big)u_0(x), \end{aligned}$$

where $u_0(x)$ solves the problem

$$\mathcal{L}_0(x)u_0(x) = f(x).$$

One of the advantages of the Neumann expansion is that the deterministic operator $\mathcal{L}_0$ has to be inverted only once. However, the method still requires simulation to obtain statistics of the solution in general case [GS91a].

I.3. Monte Carlo methods

This section contains the thorough description of the problem considered in this dissertation. We start with the formulation of the Monte Carlo method and the brief discussion of its convergence rate. The classical techniques for the improvement of the Monte Carlo integration are provided next. The multilevel Monte Carlo method is then formulated as a

variance reduction technique of the control variate type. At the end of the section, we give the contemporary overview of the existing acceleration techniques for the multilevel Monte Carlo method and show how the proposed techniques fit and contribute to this classification.

I.3.1. Formulation and complexity of the Monte Carlo method

Both spectral and perturbation methods can be very efficient for the problems which satisfy certain assumptions regarding the parametric dimensionality and the smoothness or the magnitude of random perturbations. When these conditions are not met, they usually demonstrate extremely poor behaviour. At the same time, Monte Carlo (MC) simulation is a purely statistical technique. Its performance depends solely on the geometry of the underlying random distribution and thus is immune to all of the above mentioned issues.

Let $u(x,\omega) : D \times \Omega \to W(D)$ be the solution of the stochastic initial-value problem (I.1) or the random boundary-value problem (I.5). With this definition, D can denote both temporal and spatial domain of the corresponding problem and $W(D)$ is a suitable codomain of the random function $u(x,\omega)$. Consider the problem of estimation of the expected value

$$\mathbb{E}\left[f\big(u(x,y(\omega))\big)\right] = \int_{\Omega} f\big(u(x,y(\omega))\big)d\mathbb{P}(\omega) = \int_{\Gamma} f\big(u(x,y)\big)\mu(y)dy \tag{I.9}$$

where $\Gamma(\Omega)$ is the image space of the random vector $y(\omega)$ as discussed in section I.2.2, $\mu(y)$ is a probability density on Γ and $f : W(D) \to \mathbb{R}$ is some continuous functional of a random function $u(x,\omega)$ with a Lipschitz bound

$$\big|f(u) - f(v)\big| \leq c\,\|u - v\|_{W(D)}\,. \tag{I.10}$$

For the sequence $\{y^m\}$ of independent and identically distributed random samples from Γ, consider the following averaged sum

$$\mathbb{E}_{\mathbb{MC}}\left[f\big(u(x,\omega)\big)\right] = \frac{1}{M}\sum_{m=1}^{M} f\big(u(x,\omega^m)\big). \tag{I.11}$$

If the integral in (I.9) exists, then the strong law of large numbers ensures the almost sure convergence of the above sum to the true expectation when $M \to \infty$ [Fis96]. The rate of convergence follows from Chebyshev's inequality which states that for $k > 0$

$$\mathbb{P}\left[\left|\mathbb{E}\left[f(y)\right] - \mathbb{E}_{\mathbb{MC}}\left[f(y)\right]\right| \geq k\sqrt{\mathbb{V}ar\left[\mathbb{E}_{\mathbb{MC}}\left[f(y)\right]\right]}\right] \leq \frac{1}{k^2}$$

or by changing variables and using that $\mathbb{V}ar\left[\mathbb{E}_{\mathbb{MC}}\left[f(y)\right]\right] = \mathbb{V}ar\left[f(y)\right]/M$

$$\mathbb{P}\left[\left|\mathbb{E}\left[f(y)\right] - \mathbb{E}_{\mathbb{MC}}\left[f(y)\right]\right| \geq k\right] \leq \sqrt{\frac{\mathbb{V}ar\left[f(y)\right]}{Mk}}.$$

This shows that the error of the Monte Carlo estimator converges with a fixed polynomial rate $O(M^{-1/2})$ with no regard to the underlying distribution, smoothness and dimensionality of the integrand or correlation between components of the random vector $y \in \Gamma$. The only requirement is the finiteness of the true variance of the estimated functional.

Similar result follows from the Central Limit theorem which states that

$$\left(\mathbb{E}\left[f(y)\right] - \mathbb{E}_{\mathbb{MC}}\left[f(y)\right]\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\mathbb{V}ar\left[f(y)\right]}{M}\right) \quad \text{as } M \to \infty,$$

where the convergence is understood in the sense of distributions. Thus the mean square error of the Monte Carlo estimator reads as

$$\mathbb{E}\left[\left(\mathbb{E}_{\mathbb{MC}}\left[f(y)\right] - \mathbb{E}\left[f(y)\right]\right)^2\right] = \frac{\mathbb{V}ar\left[f(y)\right]}{M} = \epsilon^2. \tag{I.12}$$

Asymptotical complexity of the method

From the discussion in sections I.1 and I.2, it is obvious that solutions of the stochastic and random differential equations are rarely obtainable in a closed analytic form and numerical approximations must be used. When the MC integration is applied to this approximation, we obtain the bias-variance decomposition of the mean square error of the

estimator

$$\mathbb{E}\left[\left(\mathbb{E}_{\mathbb{MC}}\left[f_h\right]-\mathbb{E}\left[f\right]\right)^2\right]=\underbrace{\left(\mathbb{E}\left[f_h-f\right]\right)^2}_{\text{I:=Discretization bias}}+\underbrace{\mathbb{E}\left[\left(\mathbb{E}_{\mathbb{MC}}\left[f_h\right]-\mathbb{E}\left[f_h\right]\right)^2\right]}_{\text{II:=Sampling variance}}$$

$$=\epsilon_I^2+\epsilon_{II}^2=\epsilon^2, \tag{I.13}$$

where $f_h = f(u_h(x,y))$ is the approximate value of the functional $f(u(x,y))$ obtained numerically on the discretization grid with step size h.

In the above formula, the bias error term is induced by the approximation of the solution of the differential equation while the second term is the mean square error of the MC integration as in (I.12). If we suppose that α is the convergence rate of the weak approximation of $u(x,y)$, then the estimate of the bias in (I.13) reads as

$$\left|\mathbb{E}\left[f(u)-f(u_h)\right]\right| = O(h^\alpha)=\epsilon_I.$$

Using this estimate and the bound on the sampling error in (I.12), one can find the size of the discretization grid and the number of Monte Carlo samples

$$h=O(\epsilon_I^{1/\alpha}), \qquad M=\frac{\mathbb{V}ar\left[f_h\right]}{\epsilon_{II}^2}$$

which are sufficient to guarantee the desired level of accuracy.

Assume that the cost of generating a single path of $u(x,y)$ grows polynomially with decreasing h as

$$C_h = O(h^{-\gamma}) = O(\epsilon_I^{-\gamma/\alpha})$$

for some real positive constant γ. Then the ϵ-complexity of the Monte Carlo method has the following asymptotical behavior

$$C_{\mathbb{MC}} = M\cdot C_h = \mathbb{V}ar\left[f_h\right] O\left(\epsilon^{-2-\gamma/\alpha}\right). \tag{I.14}$$

For example, for sufficiently smooth drift and diffusion coefficients, the Euler-Maruyama scheme in section I.1.3 has the parameter values $\gamma = \alpha = 1$ and the ϵ-cost of the MC estimator is $C = O(\epsilon^{-3})$.

I.3.2. Improving the Monte Carlo method

Brute force Monte Carlo method is a simple and powerful but nevertheless very slow simulation technique. Even though it outperforms standard quadrature rules in application to high dimensional and/or non-smooth integrands, its computational complexity can remain too high and impractical. Among the factors which limit the accuracy and efficiency of the MC integration, the major three are the accessibility of the sampling measure, the high discrepancy of the random quadrature nodes and the large variance of the integrand. The overview of some of the available solutions to these problems is given below.

Sampling measure. When sampling from the probability distribution is difficult but the measure itself can be easily evaluated, the *importance sampling* allows to integrate with respect to another, simpler, measure using the identity

$$\mathbb{E}\left[f(y)\right] = \int_{\Gamma} f(y) d\mathbb{P}_{\mu}(y) = \int_{\Gamma} f(y) \frac{d\mathbb{P}_{\mu}(y)}{d\mathbb{P}_{\nu}(y)} d\mathbb{P}_{\nu}(y) = \int_{\Gamma} f(y) \frac{\mu(y)}{\nu(y)} dy,$$

where the probability measure $\mathbb{P}_{\mu}(y) = \int_{\Gamma} \mu(y) dy$ is absolutely continuous with respect to $\mathbb{P}_{\nu}(y) = \int_{\Gamma} \nu(y) dy$ and $d\mathbb{P}_{\mu}(y)/d\mathbb{P}_{\nu}(y)$ is the Radon-Nikodym derivative. The absolute continuity of measures implies that $supp\big(\mu(y)\big) < supp\big(\nu(y)\big)$ and $\nu(y)$ must have heavier tales then $\mu(y)$. Then the modified Monte Carlo estimator takes the form

$$\mathbb{E}_{\mathbb{MC}}\left[f(y)\right] = \frac{1}{M} \sum_{m=1}^{M} \frac{f(y^m)\mu(y^m)}{\nu(y^m)},$$

where the random vectors $\{y^m\}_{m=1}^{M}$ are drawn from the new distribution $\mathbb{P}_{\nu}(y)$. This estimator is unbiased since $\mathbb{E}\left[\mathbb{E}_{\mathbb{MC}}\left[f(y)\right]\right] = \mathbb{E}\left[f(y)\right]$ and have the variance

$$\mathbb{V}ar\left[\mathbb{E}_{\mathbb{MC}}\left[f(y)\right]\right] = \frac{1}{M}\mathbb{V}ar\left[f(y)\frac{\mu(y)}{\nu(y)}\right].$$

The above expression shows that the importance sampling can be also used to reduce the variance of the estimator by the appropriate choice of $\nu(y)$.

Markov Chain Monte Carlo (MCMC) methods represent another group of algorithms which allow to draw samples from the stationary distribution of a suitably constructed Markov chain. The method can be considered then as a tool for approximate sampling from the a-priory unknown distribution and the quality of this approximation improves as the number of samples increases. The construction of an appropriate Markov chain is usually not a problem and a number of algorithms exists. For example, the idea of the classical *Metropolis-Hastings algorithm* is to iteratively explore the sampling space according to some proposal distribution $g\big(y'|y^{m-1}\big)$ which is conditioned on the previous iteration y^{m-1}. The probability of accepting the step is then given by the Metropolis-Hastings ratio

$$\alpha\big(y^m|y'\big) = \min\left\{1, \frac{\mu\big(y'\big)g\big(y^m|y'\big)}{\mu\big(y^m\big)g\big(y'|y^m\big)}\right\},$$

where $\mu(y)$ is the desired sampling distribution. In Bayesian inference, $\mu(y) = \mu\big(y|y_{obs}\big)$ is the posterior distribution conditioned on the available data y_{obs} and, by Bayes' theorem, the Metropolis-Hastings ratio converts to

$$\alpha\big(y^m|y'\big) = \min\left\{1, \frac{\mathcal{L}\big(y_{obs}|y'\big)\mu_0\big(y'\big)g\big(y^m|y'\big)}{\mathcal{L}\big(y_{obs}|y^m\big)\mu_0\big(y^m\big)g\big(y'|y^m\big)}\right\}, \tag{I.15}$$

where $\mathcal{L}\big(y_{obs}|y\big)$ is the likelihood and $\mu_0(y)$ is the given prior distribution. Such construction leads to the Markov chain since every next iteration depends only on the last step. Moreover, as the method proceeds, more and more samples will tend to stay in the high density regions of the target distribution as desired.

Metropolis-Hastings Markov chain obviously converges to its stationary distribution. However, a rapid mixing is also required for the good chain such that the stationary state is

reached quickly and this remains an active research area. In practice, a large initial burn-in period is required before the chain converges to the target distribution and the convergence has to be checked on-the-fly resulting in the extra computational cost and the potentially unbounded running time.

Discrepancy. Immunity to the curse of dimensionality is the consequence of the statistical nature of the Monte Carlo integration which does not utilize the structure of the sampling space and explores it in a totally random manner. As a result, some points appear close to each other while leaving a lot of empty spots with no points at all. A simple argument shows that about $\sqrt{M}$ out of M points lie in clumps [Caf98]. This appears to be a limiting factor in accuracy of the MC integration as is shown by the Koksma–Hlawka inequality.

Theorem I.3.1 (Koksma-Hlawka, [Caf98, Theorem 5.1])**.** *Consider the quadrature rule in* (I.11) *with an arbitrary sequence of nodes* $\{y^m\} \in [0;1]^N$ *applied to any function* $f(y) : [0;1]^N \to \mathbb{R}$ *with bounded Hardy-Krause variation*

$$V[f] = \int_{[0;1]^N} \left| \frac{\partial^N f}{\partial y_1 \dots \partial y_N} \right| dy + \sum_{k=1}^{N} V[f_1^{(i)}]$$

in which $f_1^{(i)}$ *is the restriction of the function* f *to the boundary* $y_i = 1$*. Then the integration error is bounded by*

$$\left| \int_{[0;1]^N} f(y)dy - \frac{1}{M} \sum_{m=1}^{M} f(y^m) \right| \le V[f] D_M^*$$

and D_M^* *is the star discrepancy defined as*

$$D_M^* = \sup_{J \in E^*} \left| \frac{\#\{y^m \in J\}}{M} - Leb(J) \right|,$$

where $Leb(J)$ *is the Lebesgue measure of the set* J *and* E^* *is the collection of all products of the form* $\prod_{k=1}^{N} [0; b_k)$ *with* $0 \le b_k \le 1$.

The Koksma–Hlawka theorem gives the worst-case bound of the error and thus often underestimates the practical accuracy of the numerical integration. However, the discrepancy of the quadrature nodes is indeed indicative of actual performance. It provides the quantitative measure of uniformity of a sequence of points and explains the low performance of the Monte Carlo method with random or pseudo-random sequences.

An alternative is given by the family of *quasi-Monte Carlo* (QMC) methods which use deterministic quasi-random numbers. These low-discrepancy sequences are designed to provide better uniformity than their random counterparts resulting in potentially faster convergence rate which is $O((\log M)^N M^{-1})$. Description of the the most commonly used quasi-random sequences such as Van der Corput, Halton, Sobol and others can be found in [Caf98, DKS13, Fox99] and references therein.

It is worth noting the quasi-Monte Carlo method is a compromise between Monte Carlo and grid-based quadratures. While combining some of their advantages, it also suffers from their limitations. For instance, the convergence rate of the QMC estimator is dimension dependent and the method looses its effectiveness for large values of N, the dimension of the parameter space Γ. It also depends on the smoothness of the integrator which is indicated by the term $V[f]$ in the Koksma-Hlawka inequality. Finally, quasi-Monte Carlo methods are not directly applicable to simulation due to correlations between the points of the quasi-random sequences. This, however, is not a big issue since the method is designed for integration and many of the quantities of interest appear in the form of functionals.

Latin hypercube sampling is another integration technique of Monte Carlo type which is designed to prevent the clustering of points in the sample space. It is conceptually similar to the stratified sampling as it divides the sample space into parts of equal probability to ensure that all portions of the distribution are represented evenly. The method picks the samples from the N-dimensional Latin hypercube with each dimension divided into M regions of equal probability. By its construction, the Latin hypercube has exactly one sample in each

axis–aligned hyperplane and the possible number of ways to place the samples into the cube is equal to $(M!)^{N-1}$, where N denotes the dimension of the sampling space and M is the number of samples. The Latin hypercube estimator is unbiased and has the variance smaller than that of the classical Monte Carlo sampling [Loh96, Ste87]. This is an advantage of the method but its practical application is still limited only to moderate dimensions. Firstly, the method requires the number of samples to be predetermined prior to the actual simulation which complicates the error analysis and the possibly required extension of the number of samples. Secondly, all of the generated samples must be stored in the computer memory which can become a very restrictive requirement for large values of N.

Variance reduction. The estimate in (I.14) indicates that the complexity the MC integration is proportional to the variance of the integrated quantity. It is thus possible to reduce the computational complexity of the method by transforming the integrand in a way which reduces its variance without modifying the final result.

One of such techniques, the importance sampling, has been discussed above. It changes the measure of integration in order to oversample some portions of interest in the sampling space that receive lower probability under the target distribution. A different approach is to introduce artificial correlation between originally independent elements of the sampling sequence such that the variance is reduced and the estimator remains unbiased. Antithetic sampling and the method of control variates fall into this category.

Antithetic sampling approach is based on finding a pair of unbiased and identically distributed estimators, say $\mathbb{E}^1_{\mathbb{MC}}$ and $\mathbb{E}^2_{\mathbb{MC}}$, which are negatively correlated. Consider the following estimators

$$\mathbb{E}^1_{\mathbb{MC}}[f] = \frac{1}{M}\sum_{m=1}^{M} f(y^m), \qquad \mathbb{E}^2_{\mathbb{MC}}[f] = \frac{1}{M}\sum_{m=M+1}^{2M} f(y^m)$$

and

$$\mathbb{E}_{\mathbb{MC}}[f] = \frac{1}{2M}\sum_{m=1}^{M} f(y^m), \qquad \hat{\mathbb{E}}_{\mathbb{MC}}[f] = \frac{\mathbb{E}^1_{\mathbb{MC}}[f] + \mathbb{E}^2_{\mathbb{MC}}[f]}{2}.$$

Clearly, both $\mathbb{E}_{\mathbb{MC}}$ and $\hat{\mathbb{E}}_{\mathbb{MC}}$ are unbiased and have the variances

$$\begin{aligned}\mathbb{V}ar\left[\hat{\mathbb{E}}_{\mathbb{MC}}\right] &= \frac{\mathbb{V}ar\left[\mathbb{E}^1_{\mathbb{MC}}\right] + \mathbb{V}ar\left[\mathbb{E}^2_{\mathbb{MC}}\right]}{4} + \frac{\mathbb{C}ov\left[\mathbb{E}^1_{\mathbb{MC}}, \mathbb{E}^2_{\mathbb{MC}}\right]}{2} \\ &= \frac{(1+\rho)\mathbb{V}ar\left[f\right]}{2M} = (1+\rho)\mathbb{V}ar\left[\mathbb{E}_{\mathbb{MC}}\right],\end{aligned}$$

where ρ is the correlation between $\mathbb{E}^1_{\mathbb{MC}}$ and $\mathbb{E}^2_{\mathbb{MC}}$. Hence the negative correlation $\rho \in [-1, 0)$ ensures that $\mathbb{V}ar\left[\hat{\mathbb{E}}_{\mathbb{MC}}\right] < \mathbb{V}ar\left[\mathbb{E}_{\mathbb{MC}}\right]$ as desired.

The *control variate* technique is used to improve the quality of the estimator by comparing it to another correlated estimator, say $\mathbb{E}^1_{\mathbb{MC}}$, with a known expectation, i.e.,

$$\hat{\mathbb{E}}_{\mathbb{MC}}[f] = \mathbb{E}_{\mathbb{MC}}[f] - \lambda\left(\mathbb{E}^1_{\mathbb{MC}}[g] - \mathbb{E}[g]\right) \tag{I.16}$$

where $f = f(y)$ and $g = g(y)$ are possibly different functions and λ is an arbitrary number. The above estimator is unbiased and has the variance

$$\mathbb{V}ar\left[\hat{\mathbb{E}}_{\mathbb{MC}}\right] = \mathbb{V}ar\left[\mathbb{E}_{\mathbb{MC}}\right] + \lambda^2\mathbb{V}ar\left[\mathbb{E}^1_{\mathbb{MC}}\right] - 2\lambda\mathbb{C}ov\left[\mathbb{E}_{\mathbb{MC}}, \mathbb{E}^1_{\mathbb{MC}}\right],$$

which attains its minimum value

$$\min_{\lambda}\left(\mathbb{V}ar\left[\hat{\mathbb{E}}_{\mathbb{MC}}\right]\right) = \mathbb{V}ar\left[\mathbb{E}_{\mathbb{MC}}\right] - \frac{\left(\mathbb{C}ov\left[\mathbb{E}_{\mathbb{MC}}, \mathbb{E}^1_{\mathbb{MC}}\right]\right)^2}{\mathbb{V}ar\left[\mathbb{E}^1_{\mathbb{MC}}\right]} = (1-\rho^2)\mathbb{V}ar\left[\mathbb{E}_{\mathbb{MC}}\right]$$

at

$$\lambda = \frac{\mathbb{C}ov\left[\mathbb{E}_{\mathbb{MC}}, \mathbb{E}^1_{\mathbb{MC}}\right]}{\mathbb{V}ar\left[\mathbb{E}^1_{\mathbb{MC}}\right]}. \tag{I.17}$$

The coefficient $\rho \in [-1, 1]$ again denotes the correlation between $\mathbb{E}_{\mathbb{MC}}$ and $\mathbb{E}^1_{\mathbb{MC}}$ but, unlike the antithetic sampling, both positive and negative correlations are allowed.

More details about the variance reduction techniques and the methods for the construction of correlated estimators can be found, for instance, in [Ros13].

I.3.3. Formulation and complexity of the multilevel Monte Carlo method

Recall the expression (I.14) for the asymptotical complexity of the Monte Carlo method

$$C_{\mathbb{MC}} = M \cdot C_h = \mathbb{V}ar\left[f_h\right] O\left(\epsilon^{-2-\gamma/\alpha}\right).$$

As is mentioned above, one way to improve this complexity is to apply the control variate technique. For highly correlated estimators, the constant parameter in (I.17) can be set to $\lambda \approx 1$ and the formula in (I.16) converts to

$$\hat{\mathbb{E}}_{\mathbb{MC}}[f(u_{h_1})] = \mathbb{E}\left[g(u)\right] + \mathbb{E}_{\mathbb{MC}}\left[f(u_{h_1}) - g(u)\right],$$

for which we additionally assumed that $\mathbb{E}_{\mathbb{MC}} = \mathbb{E}^1_{\mathbb{MC}}$. The next step in the construction of the control variate is to find the functional $g(u)$ which is correlated to $f(u_{h_1})$ and such that $\mathbb{E}\left[g(u)\right]$ can be easily evaluated. This can be achieved by taking $g(u) = f(u_{h_0})$ for $h_0 > h_1$, i.e., the same functional evaluated at the coarser approximation of the true solution $u(x, y)$. Obviously, it is very unlikely that one can find the exact value of the expectation $\mathbb{E}\left[f(u_{h_0})\right]$. However, it can be estimated with much less effort than that required for the evaluation of $\mathbb{E}_{\mathbb{MC}}\left[f(u_{h_1})\right]$.

Finally, the two-level approximation of $\mathbb{E}\left[f\big(u(x,y)\big)\right]$ can be written in the form

$$\hat{\mathbb{E}}_{\mathbb{MC}}[f(u_{h_1}(x,y))] = \frac{1}{M_0}\sum_{m=1}^{M_0} f\big(u_{h_0}(x,y^m)\big) + \frac{1}{M_1}\sum_{m=1}^{M_1}\Big(f\big(u_{h_1}(x,y^m)\big) - f\big(u_{h_0}(x,y^m)\big)\Big).$$

The difference in the second sum uses the same sequence of random samples for both coarse and fine approximations of $u(x, y)$. Thus, its magnitude is controlled by the strong convergence of the numerical method resulting in the small variance of the Monte Carlo

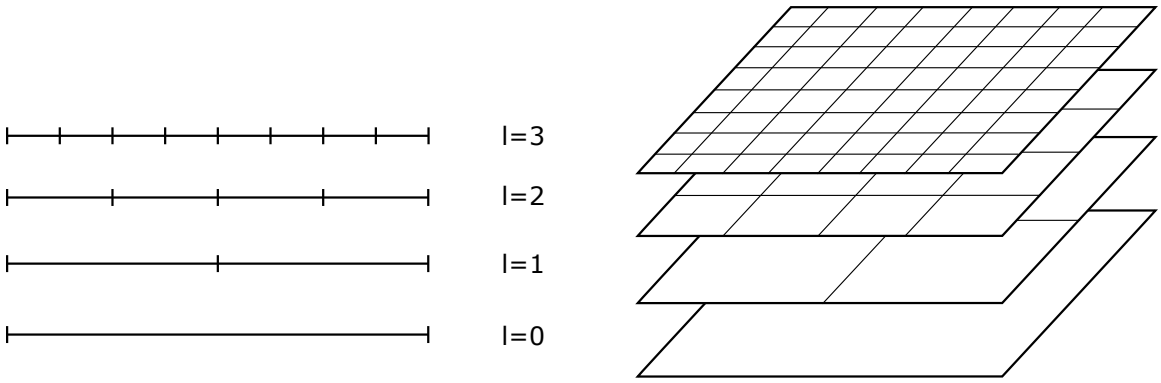


Figure 1: Geometric sequence of nested discretizations.

estimator. In other words, the total desired accuracy of the estimator can be achieved with $M_1 \ll M_0$ and much less effort is required to run the MC simulation on the coarse grid. Moreover, one can find the optimal values of the numbers of samples M_0 and M_1 on coarse and fine levels respectively such that the total cost of the Monte Carlo estimator is minimized.

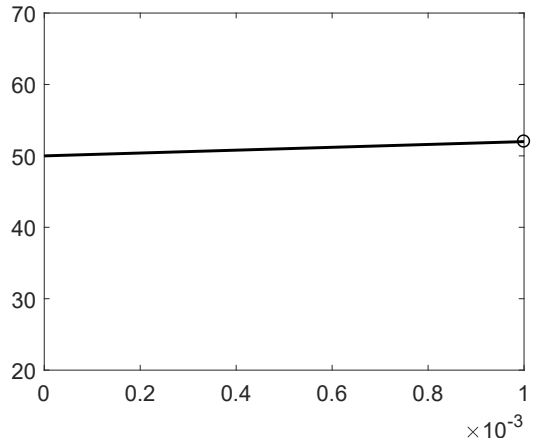
Giles in [Gil08a] extended this idea to the multilevel setting and proved that by the appropriate choice of the numbers of samples at different levels, it is even possible to improve the total order of complexity of the estimator. For instance, consider a geometric sequence of nested discretizations with step sizes

$$h_L < h_{L-1} < ... < h_l < ... < h_0, \qquad h_l = q^{-l} h_0, \tag{I.18}$$

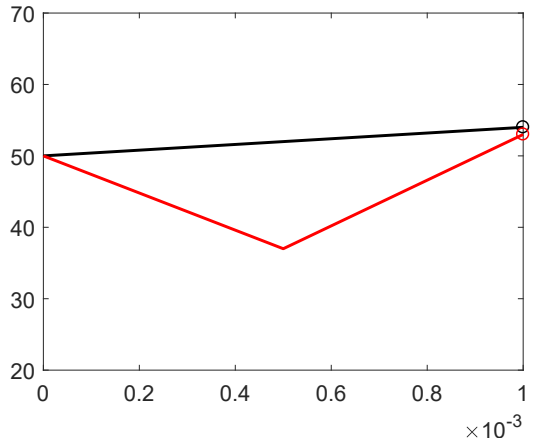
where $q \in \mathbb{N} \setminus 1$ is a refinement parameter. An example of such sequences for one- and two-dimensional domains is depicted in Figure 1.

Denote by $u_l(x, y)$ the approximation of $u(x, y)$ at the level l and let $f_l(y) = f\big(u_l(x, y)\big)$. Then the value of the functional at the finest discretization level L is given as the telescoping series

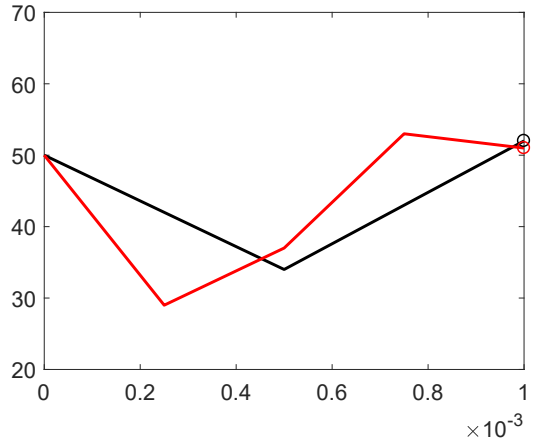
$$f_L(y) = f_0(y) + \sum_{l=1}^{L} \Big(f_l(y) - f_{l-1}(y) \Big), \tag{I.19}$$



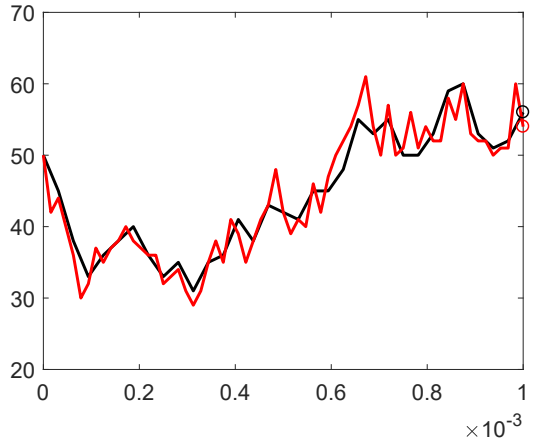
(A) Solution at level 0.



(B) Coupled solutions at level 1.



(C) Coupled solutions at level 2.



(D) Coupled solutions at level 6.

Figure 2: Sample paths of the approximate solution evaluated at different levels in the telescoping sum (I.19).

i.e., it can be represented as a solution at the coarsest mesh plus corrections calculated at finer meshes. An example of such multilevel construction is illustrated in Figure 2. It is worth noting that correction terms in the sum (I.19) must be evaluated at the same realization of the random vector y. This leads to the stronger correlation of the coupled paths at higher levels as is clearly seen in Figures 2B-2D.

Taking advantage of the linearity of expectation, we obtain

$$\mathbb{E}\left[f_L(y)\right] = \mathbb{E}\left[f_0(y)\right] + \sum_{l=1}^{L} \mathbb{E}\left[f_l(y) - f_{l-1}(y)\right].$$

By setting $\Delta_l(y) = f_l(y) - f_{l-1}(y)$, the above expression yields the multilevel Monte-Carlo estimator

$$\mathbb{E}\left[f\big(u(x,y)\big)\right] \approx \mathbb{E}_{\mathbb{ML}}\left[f_L\right] = \sum_{l=0}^{L} \mathbb{E}_{\mathbb{MC}}\left[\Delta_l\right] = \sum_{l=0}^{L} \frac{1}{M_l} \sum_{m_l=1}^{M_l} \Delta_l^{m_l}, \tag{I.20}$$

where $\Delta_0 = u_0(x,y)$ and M_l is the number of random samples generated at each level l.

Asymptotical complexity of the method

Similarly to (I.13), the bias-variance decomposition of the MLMC error reads as

$$\mathbb{E}\left[\Big(\mathbb{E}_{\mathbb{ML}}\left[f_L\right] - \mathbb{E}\left[f\right]\Big)^2\right] = \underbrace{\Big(\mathbb{E}\left[f_L - f\right]\Big)^2}_{\text{I:=Discretization bias}} + \underbrace{\mathbb{E}\left[\Big(\big(\mathbb{E}_{\mathbb{ML}} - \mathbb{E}\big)\left[f_L\right]\Big)^2\right]}_{\text{II:=Sampling variance}}$$
$$= \epsilon_I^2 + \epsilon_{II}^2 = \epsilon^2.$$

Thus, the discretization bias is the same as for the MC estimator. Together with (I.18), this gives the following condition for the finest level in the discretization hierarchy

$$\Big|\mathbb{E}\left[f_L - f\right]\Big| \leq c \cdot h_L^{\alpha} = \epsilon_I \qquad \rightarrow \qquad L \leq c + \log_q\left(h_0 \epsilon_I^{-1/\alpha}\right), \tag{I.21}$$

where c is the generic constant and α is the weak convergence order of the numerical method.

The variance of the MLMC estimator in (I.20) follows from the formula for the variance of the sum of independent Monte Carlo estimators

$$\mathbb{V}ar\left[\mathbb{E}_{\mathbb{ML}}\right] = \sum_{l=0}^{L} \frac{\mathbb{V}ar\left[\Delta_l\right]}{M_l} = \epsilon_{II}^2.$$

Given the desired tolerance ϵ_{II}, the above condition can be satisfied by the taking

$$M_l = \frac{\mathbb{V}ar\left[\Delta_l\right]}{a_l \epsilon_{II}^2} \qquad \text{and} \qquad \sum_{l=0}^{L} a_l = 1,$$

where the coefficients a_l are the weights assigning certain part of the sampling error to each level.

Let $V_l = \mathbb{V}ar\left[\Delta_l\right]$ and denote by C_l the cost of generating the single realization of Δ_l. Consider the cost function of the MLMC method

$$C_{\mathbb{ML}} = \sum_{l=0}^{L} M_l C_l = \sum_{l=0}^{L} \frac{V_l C_l}{a_l \epsilon_{II}^2}.$$

By treating a_l as continuous variables, this cost can be minimized by using the method of Lagrange multipliers with respect to the constraint

$$\sum_{l=0}^{L} a_l = 1.$$

In this case, the Lagrangian function has the form

$$\Lambda(a_0, ..., a_L, \lambda) = \sum_{l=0}^{L} \frac{V_l C_l}{a_l \epsilon_{II}^2} + \lambda \left(\sum_{l=0}^{L} a_l - 1 \right)$$

and the minimization problem has the solution

$$a_l = \frac{(C_l V_l)^{1/2}}{\sum_{k=0}^{L} (C_k V_k)^{1/2}}.$$

Thus, the optimal number of samples at each level is given by

$$M_l = \epsilon_{II}^{-2} \left(\frac{V_l}{C_l} \right)^{1/2} \sum_{k=0}^{L} (C_k V_k)^{1/2} \tag{I.22}$$

and the cost of the MLMC estimator takes the form

$$C_{\mathbb{ML}} = \epsilon_{II}^{-2} \left(\sum_{l=0}^{L} (C_l V_l)^{1/2} \right)^2 . \tag{I.23}$$

To avoid the proliferation of constants, we will write $a \lesssim b$ for two positive values a and b if a/b is uniformly bounded. We will also write $a \simeq b$ if $a \lesssim b$ and $b \lesssim a$.

The following theorem gives the asymptotical complexity of the multilevel Monte Carlo method.

Theorem I.3.2 ([Gil08a])**.** *If there exist independent estimators $\mathbb{E}_{\mathbb{MC}}[\Delta_l]$ based on M_l Monte Carlo samples, each with expected cost C_l and variance V_l, and positive constants α, β, γ such that $\min(\beta, \gamma) \leq 2\alpha$ and*

1. $\left| \mathbb{E}\left[f_l(y) - f\big(u(x,y)\big) \right] \right| \lesssim h_l^{\alpha}$,
2. $V_l \lesssim h_l^{\beta}$,
3. $C_l \lesssim h_l^{-\gamma}$,

then for any $\epsilon < e^{-1}$ there are values L and M_l for which the multilevel estimator (I.20) *has a mean-square-error with bound*

$$\mathbb{E}\left[\Big(\mathbb{E}_{\mathbb{ML}}[f_L(y)] - \mathbb{E}\left[f\big(u(x,y)\big) \right] \Big)^2 \right] < \epsilon^2$$

with a computational complexity $C_{\mathbb{ML}}$ with bound

$$C_{\mathbb{ML}} \lesssim \begin{cases} \epsilon^{-2} & \text{if } \gamma - \beta < 0, \\ \epsilon^{-2} \left| \ln \epsilon \right|^2 & \text{if } \gamma - \beta = 0, \\ \epsilon^{-2 - \frac{\gamma - \beta}{\alpha}} & \text{if } \gamma - \beta > 0. \end{cases}$$

Proof. Let γ, β and q be some positive real constants. Define the function

$$\theta(\gamma, \beta) = \sum_{l=0}^{L} q^{\frac{\gamma - \beta}{2} l}, \tag{I.24}$$

where L is the number of levels as in (I.21). Then

$$\theta(\gamma,\beta) \lesssim \begin{cases} 1 & \text{if } \gamma-\beta<0, \\ \left|\ln\epsilon\right| & \text{if } \gamma-\beta=0, \\ \epsilon^{-\frac{\gamma-\beta}{2\alpha}} & \text{if } \gamma-\beta>0. \end{cases}$$

To show this, consider 3 cases.

Case 1: $\gamma<\beta$.

$$\sum_{l=0}^{L} q^{\frac{\gamma-\beta}{2}l} = \frac{1-q^{\frac{\gamma-\beta}{2}(L+1)}}{1-q^{\frac{\gamma-\beta}{2}}} \le \frac{1}{1-q^{\frac{\gamma-\beta}{2}}} = O(1).$$

Case 2: $\gamma=\beta$.

$$\sum_{l=0}^{L} q^{\frac{\gamma-\beta}{2}l} = L+1 \le c+\log_q\left(h_0\epsilon^{-1/\alpha}\right) = O(\ln\epsilon^{-1}).$$

Case 3: $\gamma>\beta$. Using the inequality in (I.21), we obtain

$$\begin{aligned}\sum_{l=0}^{L} q^{\frac{\gamma-\beta}{2}l} &= \frac{q^{\frac{\gamma-\beta}{2}(L+1)}-1}{q^{\frac{\gamma-\beta}{2}}-1} = \frac{q^{\frac{\gamma-\beta}{2}L}-q^{-\frac{\gamma-\beta}{2}}}{1-q^{-\frac{\gamma-\beta}{2}}} \\ &\le \frac{q^{\frac{\gamma-\beta}{2}L}}{1-q^{-\frac{\gamma-\beta}{2}}} \le \frac{c(qh_0)^{\frac{\gamma-\beta}{2}}}{1-q^{-\frac{\gamma-\beta}{2}}}\epsilon^{-\frac{\gamma-\beta}{2\alpha}} = O(\epsilon^{-\frac{\gamma-\beta}{2\alpha}}).\end{aligned}$$

Now consider the cost of the MLMC method (I.23). Define q as in (I.18), then

$$\begin{aligned}C_{\mathbb{ML}} = \sum_{l=0}^{L}\left\lceil M_l\right\rceil C_l &\le \sum_{l=0}^{L} C_l + \epsilon_{II}^{-2}\left(\sum_{l=0}^{L}(C_l\overline{V}_l)^{1/2}\right)^2 \\ &\lesssim h_0^{-\gamma}\sum_{l=0}^{L} q^{\gamma l} + h_0^{\beta-\gamma}\epsilon_{II}^{-2}\left(\sum_{l=0}^{L} q^{\frac{\gamma-\beta}{2}l}\right)^2.\end{aligned}$$

For some positive constant c, we have from (I.24) that

$$\sum_{l=0}^{L} p^{\gamma l} \leq c\epsilon^{-\frac{\gamma}{\alpha}} \leq c\epsilon^{-2}$$

since we assumed $\alpha \geq \frac{1}{2}\min(\gamma, \beta)$.

If $\epsilon \leq e^{-1}$ then $\ln \epsilon^{-1} \geq 1$ and the result follows from (I.24).

□

I.3.4. Applications of the multilevel Monte Carlo method

It is obvious that the multilevel approach results in the significant reduction of the computational complexity when compared to the cost of the classical Monte Carlo integration in (I.14). For example, in his original work [Gil08a], Giles applied the method to SDEs with Brownian motion forcing and showed that, for Lipschitz continuous functionals, the Euler-Maruyama approximation gives the parameter values $\alpha = \beta = \gamma = 1$ and hence the $O\big(\epsilon^{-2}|\ln \epsilon|^2\big)$ complexity of the estimator which is a large improvement over the $O\big(\epsilon^{-3}\big)$ complexity in (I.14). It was also shown in [Gil08b] that the strong convergence order 1 of the Milstein scheme is the highest needed for the MLMC method to attain the optimal $O(\epsilon^{-2})$ asymptotical complexity when applied to SDEs with scalar and diagonal noise terms. Indeed, by the Lipschitz continuity of $f\big(u(x,y)\big)$ in (I.10), one gets

$$\begin{aligned} V_l &= \mathbb{E}\left[\left|f_l - f_{l-1}\right|^2\right] - \Big(\mathbb{E}\left[f_l - f_{l-1}\right]\Big)^2 \\ &= c^2\mathbb{E}\left[\|u_l - u_{l-1}\|^2_{W(D)}\right] + O\big(h_l^{2\alpha}\big) = O\big(h_l^{\beta}\big) + O\big(h_l^{2\alpha}\big), \end{aligned} \tag{I.25}$$

for which $\beta/2$ is the order of strong convergence of the method. For the Milstein scheme, one still has $\alpha = \gamma = 1$ but now $\beta = 2$ and hence faster decay of the variance V_l.

Application of the multilevel approach to problems with non-globally Lipschitz and discontinuous functionals is more challenging and requires the non-standard strong conver-

gence analysis of the approximation scheme. This issue was addressed in [Avi09, GHM09] where it was shown that the multi-level Monte Carlo method can be rigorously justified for a broad class of non-globally Lipschitz functions of bounded variation at the cost of slight degradation in the order of convergence. Additionally, Altmayer et al. in [AN15] studied quadratures of discontinuous payoffs in a multidimensional Heston model of mathematical finance. They established the specially designed smoothing technique based on Malliavin integration by parts formula and proved that it preserves the computational cost of order $\epsilon^{-2}\left|\log\epsilon\right|$ for their multilevel estimator. The issue of non-smooth functionals can be also treated by applying different approximations on the coarse and fine levels such that

$$\mathbb{E}_{\mathbb{ML}}\left[f_L(y)\right] = \sum_{l=0}^{L} M_l^{-1} \sum_{m_l=1}^{M_l} \left(f_l^f(y^{m_l}) - f_{l-1}^c(y^{m_l})\right) \quad \text{and} \quad \mathbb{E}\left[f_l^f\right] = \mathbb{E}\left[f_l^c\right].$$

Some of the methods which exploit this approach are based on the ideas of conditional expectation, splitting and the change of measure which are discussed in [Gil15] and references therein.

Lévy-driven SDEs can be also treated in the multilevel fashion. Jump-diffusion equations with state-independent finite intensities of jumps represent the simplest variant of Lévy models. For regular Itô schemes, the analysis and implementation of the multilevel discretization of such systems is identical to the pure diffusion case. Alternatively, the jump-adapted versions of the Euler-Maruyama and Milstein schemes were analyzed in [ABB11] and [XG12] respectively. In this case, the slight modification of the classical MLMC method had to be performed in order to account for the nonuniform jump-adapted grids. This results in the extra constant multiple term in the cost of the MLMC estimator in Theorem I.3.2 but does not change its asymptotic complexity.

A very general class of Lévy-driven SDEs was also considered in [GX14, DH11, Der11, Mar10]. Giles et al. in [GX14] adapted MLMC to path-dependent functionals of infinite-activity pure jump exponential Lévy models and showed that the complexity of estimator

is the same as that in Theorem I.3.2. Dereich et al. in [DH11] studied functionals of Lévy processes which are Lipschitz with respect to supremum norm and obtained an $O\left(\epsilon^{(6b)/(4-b)}\right)$ upper bound on the worst case computational complexity. The parameter $b \in [0, 2]$ in their estimate denotes the Blumental–Getoor index which measures the frequency of small jumps, where a large index corresponds to a process which has small jumps at high frequencies. They showed that for small values of b, the complexity of the MLMC estimator is the same as for diffusion SDEs but it decreases dramatically for larger Blumental–Getoot indices. As a remedy for such undesirable behavior, they proposed a model with Gaussian correction of smaller jumps by approximating them by a normal distribution [Der11]. A similar idea was also proposed in [Mar10] where all the jumps with absolute value smaller than the certain threshold where cut away and the variance was recovered by changing the Brownian motion part of the process.

The proper coupling of the coarse and fine paths is one of the key points in the construction of the MLMC estimator as it controls the decay rate β of the level variances V_l. The coupling of paths in the original formulation of the algorithm for Itô SDEs is easily achieved by using the same path of the driving Brownian motion on different levels. The variance decay follows then from the strong convergence of the approximation scheme as is apparent from (I.25). In jump-diffusion models with state-independent finite jump intensities, the fine and coarse paths share the same jump times which also makes their coupling a trivial task. The case of state-dependent intensities requires a special treatment since approximations at different levels may jump at different times which in turn can lead to increase in the variance of the multilevel corrections. The thinning algorithm provides one of the possible solutions to this problem. It is based on the construction of a Poisson process with a constant rate which is an upper bound of the state-dependent rate. This gives a set of candidate jump times, and these are then selected as true jump times with certain probabilities. The appropriate change of measure is additionally used to ensure that acceptance probability is the same for

coarse and fine paths [XG12]. A different approach was proposed in [AH12, AHS14] for the continuous time Markov chain models. The two stochastic processes are constructed in a way which makes them jump together the vast majority of times. The auxiliary process is then introduced to jump the extra times of either of the processes. Additionally, the coupling of stochastically adaptive time-stepping schemes was also discussed in [GLW16, HvSST13] while the coupling of weak schemes was considered in [BN14].

The multilevel Monte Carlo method can be also generalized to stochastic and random partial differential equations. Furthermore, the computational savings of the multilevel discretization will be even greater than for SDEs due to the rapid growth of the cost of computing the level corrections in high spatial dimensions. In most cases, the implementation of the MLMC algorithm for SPDEs remains trivial and the major theoretical efforts are concentrated on the analysis of the variance of the multilevel corrections. The first research in this area was devoted to the study of elliptic PDEs with random coefficeints [BSZ11, CGST11]. Since then there have been published numerous papers for equations of various types including elliptic, parabolic, hyperbolic and mixed systems.

One of the recent extensions of the standard MLMC algorithm is the multi-index Monte Carlo method [HANT16]. It generalizes the idea of nested levels to multidimensional setting by allowing the originally scalar levels to be defined in multiple directions. A motivation for such generalization is given, for instance, by approximation of SPDEs where discretizations in different spatio-temporal dimensions should be indexed independently. The multi-index notation provides an additional flexibility in construction of the telescoping sums of nested approximations and can be viewed as a combination of sparse grids and Monte Carlo sampling. Hence, similarly to sparse-grids, it offers the possibility of dimension-independent complexity for SPDEs and other high-dimensional stochastic applications [HANT16, Gil15]. This can give a substantial improvement over the standard MLMC for which the rate γ of increase in computational cost C_l increases at least linearly as a function of dimension.

I.3.5. Improving the multilevel Monte Carlo method

Theorem I.3.2 gives the asymptotical ϵ-complexity of the estimator under very general assumptions on the growth of costs C_l and the decay of variances V_l in (I.23). The constant of proportionality, however, is more problem and solver dependent. In this regard, it is possible to consider at least three types of the acceleration techniques for the multilevel Monte Carlo method. The brief discussion of each technique is given below.

The first and the most obvious approach is to perform the optimization of the parameters in the discretization hierarchy of the MLMC estimator. Originally, the method was formulated for geometric sequences of uniform grids and the equal splitting between the discretization bias and the variance components of the total mean-square error [Gil08a]. At the same time it was observed numerically that the choice of the mesh refinement ratio can have a serious impact on the performance of the algorithm. The rigorous theoretical study of this problem was accomplished later in [HANvST16] under certain assumptions on the asymptotical models for the weak and strong convergence and the computational cost of pathwise solvers. The optimal sequences of discretization grids and the optimal tolerance splitting between the bias and the statistical error contributions were explicitly determined in terms of the calibrated parameters of the assumed asymptotical models for both geometric and non-geometric hierarchies. However, the authors also mentioned several issues with practical implementation of this approach due to possible constraints on the mesh sizes arising from additional stability requirements or inaccuracy of the asymptotical models at coarse levels. We will propose the remedy for these issues later in Chapter II.

The problem of parameter calibration can be rather challenging by itself due to the well-known issue of large curtosis of standard Monte Carlo estimators at high discretization levels [Gil15]. As a result, the number of samples required for the accurate estimation of parameters can be much larger than their optimal values in (I.22). One of the possible solutions to this problem was studied in [CHAN$^+$15]. The basic idea of the proposed

continuation multilevel Monte Carlo algorithm was to perform the Bayesian estimation of model parameters at deep levels using the available data from coarser levels for the a-priory given discretization hierarchies and the parametric models of the estimated quantities. The provided numerical results indicated significant computational savings for both random and stochastic differential equations. A different approach was considered in [MTV16] in application to continuous time Markov chain models. The authors studied the hybrid Chernoff tau-leap algorithm with automatic switching between approximate and exact path generation. The purpose of the proposed Chernoff-type bound on the time step of the approximate tau-leaping method was to keep the solution in the positive lattice with prescribed exit probability. The hybrid nature of the path simulation preserved positivity of the solution but complicated extrapolation of the estimated quantities from coarse to fine levels. This issue was resolved by using the dual-weighted residual estimation technique which proved to be much more efficient than standard sampling estimators.

A different approach for the acceleration of the MLMC method is based on statistical techniques discussed in section I.3.2 in the context of improving Monte Carlo integration. This includes multilevel variants of the Markov chain Monte Carlo (MCMC) sampling, the quasi-Monte Carlo integration and the methods for additional variance reduction. Such generalizations, however, usually cannot be applied directly and require appropriate modifications of the standard techniques to account for the multilevel structure of the estimator. In practice, this is often done by introducing two discretization hierarchies: one for physical and one for parametric approximations. The variety of the obtained multifidelity models is then combined appropriately to reduce the overall computational burden of the estimator [PWG16, VCNGP15].

This idea is utilized, for instance, in the multilevel variant of the Markov chain Monte Carlo algorithm when evaluation of the likelihood in (I.15) is based upon the solution of the differential equation. In such case, the posterior distribution has to be level dependent

since otherwise the cost of generating the samples at all levels would be dominated by the cost of evaluating the likelihood at the highest level leading to no actual cost gain. Therefore, the construction of the corresponding Markov chains at different levels must be performed judiciously so that sufficient decay of the level corrections is guaranteed without introducing additional bias to the estimator. One of possible solutions to this problem was proposed in [DKST15] for the elliptic differential equation of subsurface flow. In order to construct the unbiased estimator with a pair of correlated Markov chains at each level, the authors considered the sequence of approximations with increasing physical and parametric dimensionality. The coarser of the two chains at each level was constructed using the standard Metropolis-Hastings algorithm and the result was then reused for the candidate step of the finer chain with the extra random modes being generated independently. With appropriately defined two-level acceptance/rejection strategy, such construction guarantees that all samples from the same discretization level across the multilevel hierarchy are generated from the same posterior distribution leading to no additional bias. Several variants of this approach were also studied in [EJMT15] and [HSS13] in the context of multiscale approximation of PDEs and Bayesian inference respectively. It was shown that, for the considered problems in d physical dimensions, the ϵ-complexity of the multilevel MCMC algorithm is $O\left(\epsilon^{-d}\right)$ which gives a substantial improvement over the $O\left(\epsilon^{-d-2}\right)$ complexity of the standard single-level algorithm. An alternative to the Markov chain Monte Carlo sampling, the sequential Monte Carlo method, has been additionaly considered in [BJL$^+$16] and the results also indicated the computational superiority of the multilevel approach.

Similarly to the multilevel MCMC algorithm, the level-dependent dimension truncation strategy was applied in the multilevel quasi-Monte Carlo method for a class of elliptic PDEs with random coefficients [KSS15]. It was illustrated that by the appropriate choice of the functional basis for the representation of the input random field, it is possible to achieve the optimal complexity of the multilevel estimator which is essentially of the order of the single

PDE solve at the finest discretization level. It is worth noting that the standard MLMC method cannot have this property due to the slow convergence of MC integration.

Multifidelity models have been also considered in the context of variance reduction. For example, computationally cheap low-fidelity approximations of the quantities of interest can be used as a control variate for the estimation of expectations of the level corrections. This approach does not improve the order of complexity of the estimator but the results in [FDKI16, GEI15, NT15] indicate that the overall cost of the algorithm is nevertheless superior to both MLMC and MC methods.

A conceptually different antithetic technique was proposed in [GS14, Ric13]. It is based on the idea of using different estimators for the fine and coarse levels as follows

$$\mathbb{E}_{\mathbb{ML}}\left[f_L(y)\right] = \sum_{l=0}^{L} \mathbb{E}_{\mathbb{MC}}\left[f_l^f(y^{m_l}) - f_{l-1}^c(y^{m_l})\right]$$

provided $\mathbb{E}\left[f_l^f\right] = \mathbb{E}\left[f_l^c\right]$. This definition gives the flexibility to construct approximations for which the values of the level corrections $f_l^f - f_{l-1}^c$ are much smaller than the original difference $f_l - f_{l-1}$. Moreover, it is even possible to improve the total order of complexity of the estimator when the choice of approximations leads to the increased rate of decay β of the variances V_l in Theorem I.3.2. In [GS14], this was achieved by setting $f_{l-1}^c = f_{l-1}$ and $f_l^f = (f_l + f_l^a)/2$, where f_l^a is a suitably constructed antithetic twin of the fine approximation such that $\mathbb{E}\left[f_l^a\right] = \mathbb{E}\left[f_l\right]$ and $f_l - f_{l-1}^c = -(f_l^a - f_{l-1}^c)$. Therefore, $f_l^f = (f_l + f_l^a)/2 \approx f_{l-1}^c$ and $f_l^f - f_{l-1}^c$ is a very small number. It was shown that, for a class of smooth functionals, the described construction can result in the $O\left(\epsilon^{-2}\right)$ complexity of the estimator even with low-order pathwise integrators. As a variant of such integrator, the authors proposed to use the truncated Milstein scheme obtained by skipping non-diagonal multiple stochastic integrals. It was illustrated that this choice leads to the significant reduction in both asymptotical and finite-tolerance complexities since, as was mentioned in section I.1.3, the high computational cost of generating stochastic integrals is the main limiting factor for practical application of

higher order schemes.

Finally, as a sampling technique, the MLMC method requires one to solve a large number of deterministic problems. While the asymptotical complexity of the algorithm is controlled solely by the rates of growth and decay of costs C_l and variances V_l respectively (Theorem I.3.2), the actual cost of the estimator is more problem and solver dependent. The efficiency of the method is hence strongly influenced by the complexities of the corresponding pathwise integrators. It is the task of this dissertation to study several problems for which the significant reduction in the computational complexity of the MLMC estimator can be achieved by the appropriate problem and level dependent choice of deterministic solvers. Three particular problems are considered: integration of stiff SDEs, estimation of initial guesses for iterative linear solvers and boundary values problems in randomly perturbed domains. The detailed description of the proposed techniques will be given later in Chapters II-IV while the remainder of this section provides a brief overview of some of the other approaches available in the literature.

One of the difficulties in the practical implementation of the multilevel Monte Carlo method is the selection of the coarsest level due to possible divergence of numerical solutions evaluated at coarse grids. For example, the divergence of the multilevel Monte Carlo Euler method for non-linear stochastic differential equations with superlinear and one-sided Lipschitz continuous drift coefficient was proven in [HJK13]. As a remedy for this problem, the authors proposed to use the tamed Euler scheme which was mentioned earlier in section I.1.4. Alternatively, the truncated Euler-Maruyama method was discussed in [GLMZ16] as another way to resolve this issue. Application of implicit schemes, as discussed in section I.1.4, is also classical in this context.

The issue of stiffness is another serious limiting factor for the integration of differential equations at coarse discretization levels. It often stems from the presence of multiple scales in the systems of interest leading to possible instability and thus divergence of

numerical approximations. It is a well known fact that application of explicit schemes for stable integration of stiff dynamics is possible only on time grids which explicitly resolve the fastest scale of the system. This constraint can significantly decrease the number of admissible discretization levels resulting in substantial reduction of the efficiency of the multilevel estimator. There have been proposed several stabilization techniques which allow to exploit all the levels of the multilevel discretization using both implicit [BHMT16] and explicit integrators [AB13] at coarse levels. In Chapter II, we will elaborate on this idea and propose a similar but different approach as a solution to this problem.

The choice of suitable patwhise integrators has been also discussed previously in the literature. It was shown in [DR15] that it is possible to reduce the computational cost of the multilevel estimator by a constant factor $(p/p')^2$ when the cheap approximation with low weak order p' on levels $l = 0, 1, ..., L-1$ is combined with a more expensive approximation of higher order p on the finest level L. The provided analysis required that the growth of costs is the same as the decay of variances ($\beta = \gamma$ in Theorem I.3.2) for both approximations. This, however, is not a severe restriction since simulation of multiple stochastic integrals is not necessary for the construction of higher weak order schemes. Possibility of using weak convergence properties of the numerical schemes for the construction of efficient MLMC estimators was also studied in [MSS15]. The authors introduced the novel modified equation analysis as an alternative to strong-approximation theory which is classically used to derive the condition (iii) in Theorem I.3.2. Their result is completely based on the weak approximation theory and the strong approximation is only needed to relate the original and modified equations and not the numerical methods. The impact of choosing the particular time-stepping method on the efficiency of the estimator was also discussed in the paper. It was mentioned that integrators based on splitting methods can reduce the total cost of the MLMC method by an order of magnitude when compared to the standard Euler-Maruyama scheme as patwise integrator for Langevin equation. Similar result was obtained later in

[KMS$^+$16] in application to atmospheric dispersion modelling. In Chapters III-IV, we will follow this path and consider acceleration of the multilevel Monte Carlo method by selection of suitable integrators for stationary PDEs subjected to random input data and topological uncertainties.

CHAPTER II

STIFF STOCHASTIC CHEMICAL KINETICS

Stochastic differential equations (SDEs) provide a convenient mathematical model for the description of time-dependent physical phenomena at the mesoscopic level [Kot08, Gil07]. Physical systems at this level are characterized by the presence of a vast range of spatial and temporal scales. Additionally, it is often the case that dynamical behavior of the most practical interest happens on the slowest of scales. Extraction of the effective, lower dimensional, dynamics can be achieved, for instance, by applying the model reduction techniques to eliminate the fast variables through the process of averaging [PS08, GKS04, WLVE05]. Within this framework, the reduced system is obtained by averaging the slow dynamics over the invariant ergodic measure generated by the fast variables. In practice, such ergodic measure is rarely obtainable in a closed form but can be estimated empirically as an ensemble average of solutions from an auxiliary micro solver [VE03, GK08, GKK06, SV05]. A similar approach is given by the quasi-steady state or partial equilibrium approximations which are based on the assumption that the fast reactions rapidly reach equilibrium state with a relaxation time which is much smaller than the time scale of slow reactions [Gou05, CGP05a, RA03, MHR07, HN14]. These ideas were implemented in the so-called slow-scale and nested algorithms for the simulation of stochastic chemical kinetics [CGP05c, CGP05b, CP08, ELVE07, ELVE05].

Model reduction techniques have proven their efficiency for a large class of problems. However, it is not clear if they can be extended to the multilevel Monte Carlo setting. In this case, one still has to rely on conventional solvers. As was mentioned previously in section I.3.5, the choice of these solvers for stiff systems has to be performed carefully due to their possible instability at low levels of the multilevel hierarchy. For stiff ODEs, it is a well known result that implicit integrators can generate stable solutions with time steps of the order of the slow scale of the system while most of explicit methods must

resolve the fastest scale to be stable [HW96]. Efficiency of implicit solvers stems from their damping effect which prevents the rapid transients to escape from the stable slow solution manifold. Motivated by these results, several variants of implicit and semi-implicit stochastic integrators with enhanced stability properties have been proposed in the literature [APR09, Sch12, WGW12, TB01, HHR16, RPCG03, HK06, RKVZ15]. However, contrary to deterministic case, random perturbations represent an important geometrical feature of stochastic dynamics and overdamping of these fluctuations is not desirable. This problem was raised in [RPCG03, CPRG04] in application to chemical kinetics driven by a discrete stochastic process. For a carefully chosen test equation, it was shown that the variance of the numerical solution calculated with the drift implicit Euler method is underestimated. As was later explained in [LAE08], this happens due to the inability of implicit methods to resolve the invariant quasi-equilibrium distribution of fast variables which plays the role of a slow manifold for SDEs. Moreover, it will be shown in section II.2 that the damping property of implicit integrators has a negative impact on the MLMC method as well by reducing the number of allowed coarse levels even when the solver is stable.

It is clear from the above discussion that the requirements for the construction of efficient stiff stochastic solvers are twofold since one has to treat destabilizing factors without destroying geometry of the true dynamics. As an alternative to the model reduction techniques, several approaches which do not require explicit separation of scales have been proposed in the literature [CPRG04, RPCG03, CR11, LL12]. It was shown in [CPRG04] that the trapezoidal tau-leaping method is able to recover the stationary variance of the true solution for the carefully chosen test system. However, as was pointed out in [LAE08], this feature is the consequence of the linearity of the test problem. Moreover, the method is only A-stable and has impractically large relaxation time to the stationary state. An interlacing strategy proposed in [RPCG03, CR11] does not have this issue. It allows to recover the variance of stochastic systems by interlacing implicit integrators with short bursts

of explicit methods which restore the overly damped stochastic fluctuations. In section II.2, we will adopt this idea to the multilevel Monte Carlo setting by considering a composite implicit-explicit integrator which is cheap, stable and is able to resolve the variance of the true solution at low levels with a reasonable accuracy.

As an implicit part of our composite integrator, we consider the family of split-step methods. The splitting of the original differential system into several parts provides the flexibility to choose a combination of subsystem solvers with the most desirable set of properties. The classical application of this approach is in geometric integration of dynamical systems in both deterministic [MQ02] and stochastic [AVZ15] settings. Another application is the integration of stiff systems where the split-step methods appeared as an attempt to construct stochastically implicit integrators. The first method of this type was introduced by Higham in [HMS02] as a modification of the classical Euler-Maruyama method. Further developments include results on the split-step Milstein method [WL09, WL10, HH12, VK15] for SDEs driven by Wiener noise and results for jump-diffusion systems [HK05, HK06, AHL10]. In this chapter, we propose the novel two-stage splitting technique which inherits all advantages of implicit integrators without destroying geometry of the true stochastic dynamics. It will be shown that, as a part of the proposed composite integrator, it allows to utilize more levels of the multilevel hierarchy leading to potential computational savings of the MLMC estimator.

II.1. Problem setting

Consider a thermally equilibrated chemical system which consists of N well-stirred molecular species $\{S_i, i = 1, ..., N\}$ interacting through R reactions channels $\{\mathcal{R}_r, r = 1, ..., R\}$

$$\mathcal{R}_1 : \quad \nu_{11}^- S_1 + \nu_{21}^- S_2 + ... + \nu_{N1}^- S_N \xrightarrow{c_1} \nu_{11}^+ S_1 + \nu_{21}^+ S_2 + ... + \nu_{N1}^+ S_N,$$

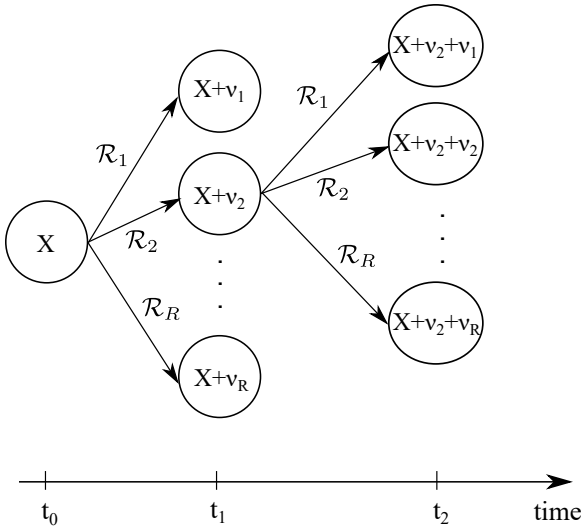


Figure 3: Realization of the Markov chain (II.1) describing evolution of the chemical system.

$$\begin{aligned}
\mathcal{R}_2: &\quad \nu_{12}^- S_1 + \nu_{22}^- S_2 + ... + \nu_{N2}^- S_N \xrightarrow{c_2} \nu_{12}^+ S_1 + \nu_{22}^+ S_2 + ... + \nu_{N2}^+ S_N, \\
&\quad \vdots \\
\mathcal{R}_R: &\quad \nu_{1R}^- S_1 + \nu_{2R}^- S_2 + ... + \nu_{NR}^- S_N \xrightarrow{c_R} \nu_{1R}^+ S_1 + \nu_{2R}^+ S_2 + ... + \nu_{NR}^+ S_N.
\end{aligned}$$

At any time instance this system can be in exactly one of the states $X(t) = (X_1, X_2, ..., X_N)^T$, with X_i denoting the number of molecules of type S_i. Under a reasonable assumptions that collisions between molecules occur in a random manner, evolution of the state of the system can be modeled as a stochastic jump process [Gil77, AK15]

$$\begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_N(t) \end{pmatrix} = \begin{pmatrix} X_1(0) \\ X_2(0) \\ \vdots \\ X_N(0) \end{pmatrix} + \begin{pmatrix} \nu_{11} & \nu_{12} & \dots & \nu_{1R} \\ \nu_{21} & \nu_{22} & \dots & \nu_{2R} \\ \vdots & & \ddots & \vdots \\ \nu_{N1} & \nu_{N2} & \dots & \nu_{NR} \end{pmatrix} \begin{pmatrix} N_1(t) \\ N_2(t) \\ \vdots \\ N_R(t) \end{pmatrix}, \tag{II.1}$$

where each stoichiometric column vector $\nu_r = [\nu_{1r}^+ - \nu_{1r}^-, ..., \nu_{Nr}^+ - \nu_{Nr}^-]^T$ denotes the change in molecular populations from the reaction channel $\mathcal{R}_r$, and markovian processes $N_r(t)$

count the number of corresponding reactions in the time interval $[0, t]$.

Since the counting processes $N_r(t)$ are markovian, evolution of the state of the system also represents a continuous time Markov chain with transition probabilities

$$\mathbb{P}\left[X(t+dt) - X(t) = \nu_r | \mathcal{F}_t\right] = a_r(X(t))dt + o(dt), \qquad r = 1, .., R$$

and with associated (chemical) master equation (CME)

$$\frac{\partial \mathbb{P}\left[X(t) = x\right]}{\partial t} = -\sum_{r=1}^{R} a_r(x)\mathbb{P}\left[X(t) = x\right] + \sum_{r=1}^{R} a_r(x - \nu_r)\mathbb{P}\left[X(t) = x - \nu_r\right], \quad \text{(II.2)}$$

$$\mathbb{P}\left[X(0) = x\right] = p_x. \quad \text{(II.3)}$$

It describes the change in time of the probability mass function of every element x from the state space of the system. According to the stochastic law of mass action, propensities $a_r(x)$, which define transition intensities of the Markov chain, are proportional to reaction-rate constants c_r and the number of distinct combinations of molecules of source species in corresponding reactions

$$a_r(x) = c_r \prod_{i=1}^{N} \binom{x_i}{\nu_{ri}^-} = c_r \prod_{i=1}^{N} \frac{x_i!}{(x_i - \nu_{ri}^-)!}, \qquad r = 1, .., R.$$

Although the master equation provides complete statistical description of the system at any time instance, it admits analytical solutions only in very special simple cases and rarely can be solved numerically. On the other hand, approximate solutions of the CME can be extracted directly from (II.1) by considering ensembles of individual state paths. This is a relatively simple task after the counting processes $N_r(t)$ are reformulated in terms of the unit rate Poisson processes $\mathcal{P}(t)$ by means of the random time change formula [AK15, Theorem 1.10]. As a result, equation (II.1) converts to the following stochastic integral

equation

$$X(t) = X(0) + \sum_{r=1}^{R} \nu_r \mathcal{P}\left(\int_0^t a_r(X(s))ds\right). \tag{II.4}$$

The above equation has equivalent differential form via Poisson random measure $\mu(dt \times da)$ generated by compound Poisson process $\{t_i, a_i\}$ with jump times t_i and labels a_i

$$dX(t) = \sum_{r=1}^{R} \nu_r \int_0^{A(R,t)} \mathbf{1}_{\{a\in(A(r-1,t);A(r,t)]\}} \mu(dt \times da). \tag{II.5}$$

The state dependent functions $A(r,t) = \sum_{i=1}^{r} a_i(X(t))$, partitioning the label set, are responsible for the choice of a particular reaction similarly to the acceptance-rejection sampling strategy as is shown in Figure 4A. Taking into account that holding times of the Poisson jump process are exponentially distributed random variables, the state paths of the system can be generated with the stochastic simulation algorithm (SSA) [Gil77] as follows

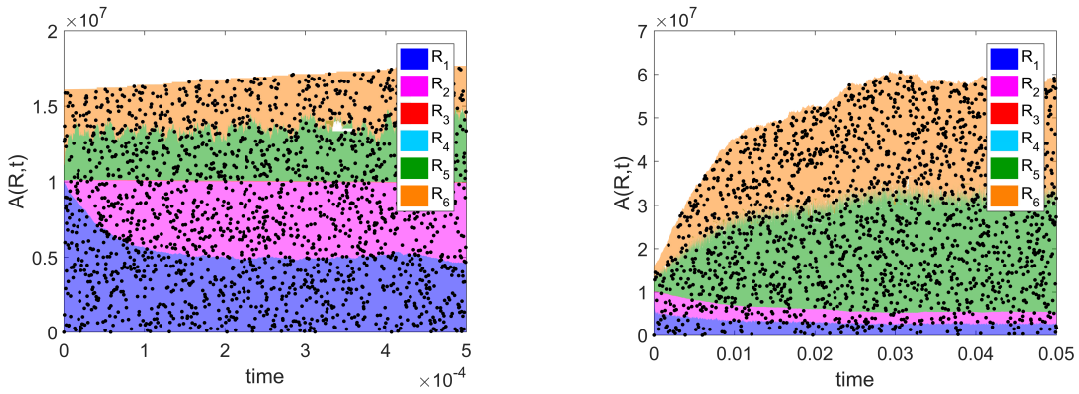
$$X(t_n + \tau) = X(t_n) + \nu_r,$$

where the holding time and the index of the next reaction are calculated as

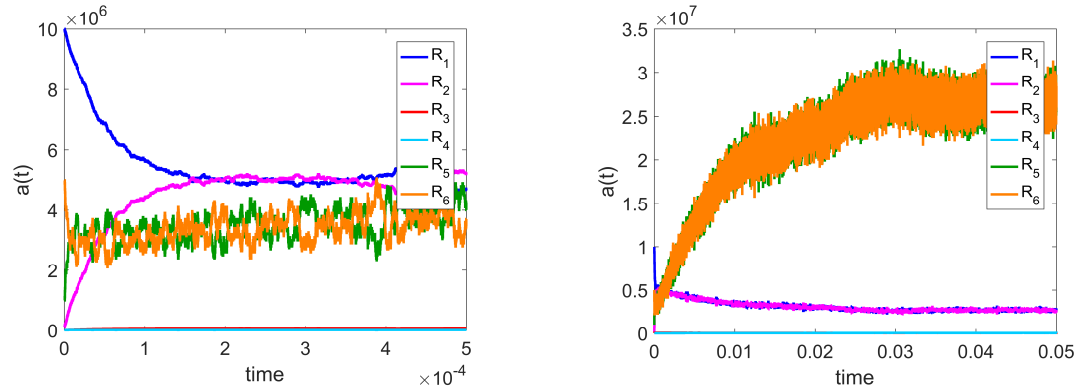
$$\tau = -\frac{1}{A(R,t_n)} \ln r_1, \tag{II.6}$$

$$r = \min\{r : A(r,t_n) > r_2 A(R,t_n)\}. \tag{II.7}$$

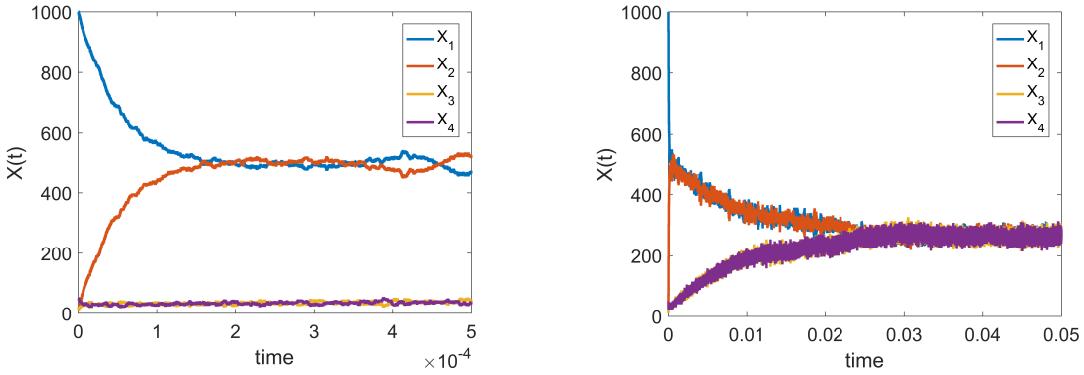
In the above formulas, r_1 and r_2 are two random numbers drawn from the standard uniform distribution. There also exist several modifications of this algorithm such as the first reaction method and the next reaction method [Gil77, GB00].



(A) Cumulative propensities (shaded regions) and realization of compound Poisson process (points).



(B) Sample propensities of the compound Poisson process in Fig. 4A.



(C) Sample paths of the compound Poisson process in Fig. 4A.

Figure 4: Realization of the compound Poisson process for the set of reactions in (II.8).

II.1.1. Stochastic stiffness

The time scales of a chemical system are defined by the relative magnitudes of its propensity functions. When probabilities, or equivalently propensities, of some reactions are relatively large, the SSA inevitably spends most of the time tracking the impact of such reactions.

It is known that the long time behavior of a chemical network which admits a complex-balanced equilibrium is uniquely determined by its stationary distribution [AK15, Theorem 3.7]. In systems with severe separation of time scales, existence of the stationary distribution can be assumed when the relaxation time of species or their combinations involved in fast reactions is much smaller than the time scale of slower reactions. It is obvious that sampling of intermediate fast species from their quasi stationary distributions becomes redundant on time intervals which are larger than appropriate relaxation times. This explains inefficiency of SSA in application to such systems.

The situation described above is clearly seen in Figure 4 which illustrates dynamics of the following reaction network [ELVE05]

$$S_1 \underset{10^4}{\overset{10^4}{\rightleftharpoons}} S_2 \underset{10^2}{\overset{10^2}{\rightleftharpoons}} S_3 \underset{10^5}{\overset{10^5}{\rightleftharpoons}} S_4. \tag{II.8}$$

This system has three pairs of reversible reactions evolving on three different time scales. Indeed, one can observe from Figure 4 that each pair of reactions relaxes to a certain stable state in time τ_{relax} which has the order of the corresponding time scale: $\tau_{relax} \simeq 10^{-5}$ for the fastest third pair, $\tau_{relax} \simeq 10^{-4}$ for the first pair and $\tau_{relax} \simeq 10^{-2}$ for the slowest second pair of reactions. It is worth noting that while all the species are involved in fast reactions and are in fact fast variables, the molecular complexes $S_1 + S_2$ and $S_3 + S_4$ evolve on the slow time scale of the second reaction. It is clear that the SSA is not able to recognize and take advantage of this feature.

II.1.2. Approximate path simulation

Stochastic simulation algorithm produces exact solutions to the equations (II.4)-(II.5) because it updates the state of the system every time when any reaction fires. This explains its inefficiency in application to stiff systems but it can be rather slow even for reduced order models in situations when extremely high level of accuracy is not required. Approximate tau-leaping methods are more suitable in this case since they allow to use time steps which are larger then the time of a single reaction.

Consider the variant of the integral equation (II.4) over the time interval $t \in [t_n; t_{n+1}]$

$$\begin{aligned} X(t_{n+1}) &= X(t_n) + \sum_{r=1}^{R} \nu_r \mathcal{P}\left(\int_{t_n}^{t_{n+1}} a_r(X(s))ds\right) \\ &= X(t_n) + \sum_{r=1}^{R} \nu_r \int_{t_n}^{t_{n+1}} a_r(X(s))ds + \sum_{r=1}^{R} \nu_r \overline{\mathcal{P}}\left(\int_{t_n}^{t_{n+1}} a_r(X(s))ds\right) \end{aligned} \tag{II.9}$$

with the driving stochastic processes which are martingales

$$\overline{\mathcal{P}}\left(\int_{t_n}^{t_{n+1}} a_r(X(s))ds\right) \;=\; \mathcal{P}\left(\int_{t_n}^{t_{n+1}} a_r(X(s))ds\right) \;-\; \int_{t_n}^{t_{n+1}} a_r(X(s))ds.$$

The major difficulty in construction of accurate numerical solutions to the above equation is associated with approximation of the Poisson processes. First of all, calculation of the corresponding intensities requires evaluation of the state dependent integrals which severely limits the range of practically implementable schemes. It utilizes the straightforward explicit approximation of the integrals

$$\overline{\mathcal{P}}\left(\int_{t_n}^{t_{n+1}} a_r(X(s))ds\right) \approx \overline{\mathcal{P}}\Big(a_r(X(t_n))\tau\Big), \qquad \tau = t_{n+1} - t_n.$$

This approach produces reasonably accurate results for slow reactions with propensities which are not expected to change values dramatically inside the interval of integration. The

error analysis of tau-leaping schemes can be found, for instance, in [Li07, AGK11, RPCG05]. On the other side, by leaping over the large number of fast reactions one can destroy stability of the numerical solution even when exact values of propensities are given. The reason is that large intensities yield potentially large values of the generated Poisson increments. Statistical independence of these increments means that different reaction channels separately update populations of the common fast species leading to unstable solutions.

The issue of numerical stability is induced by the stiffness of reaction networks with largely separated time scales. In deterministic setting, it is efficiently resolved by incorporating implicitness into numerical schemes. Construction of implicit integrators for stochastic systems is more intricate. However, taking advantage of the flexibility in approximation of deterministic integrals, we can rewrite the equation (II.9) in the form

$$\begin{aligned} X(t_{n+1}) = X(t_n) + \sum_{r=1}^{R} \nu_r \left(\int_{t_n}^{t_{n+1}-\theta\tau} a_r(X(s))ds + \int_{t_{n+1}-\theta\tau}^{t_{n+1}} a_r(X(s))ds \right) \\ + \sum_{r=1}^{R} \nu_r \overline{\mathcal{P}} \left(\int_{t_n}^{t_{n+1}} a_r(X(s))ds \right), \qquad \theta \in [0;1] \end{aligned} \tag{II.10}$$

which gives the well-known family of drift implicit theta methods

$$\begin{aligned} Y_{n+1} &= Y_n + \sum_{r=1}^{R} \nu_r \Big((1-\theta)a_r(Y_n) + \theta a_r(Y_{n+1}) \Big)\tau + \sum_{r=1}^{R} \nu_r \overline{\mathcal{P}}\Big(a_r(Y_n)\tau \Big) \\ &= Y_n + \theta \sum_{r=1}^{R} \nu_r a_r(Y_{n+1})\tau + \sum_{r=1}^{R} \nu_r \Big[\mathcal{P}\Big(a_r(Y_n)\tau \Big) - \theta a_r(Y_n)\tau \Big]. \end{aligned} \tag{II.11}$$

The classical explicit [Gil01], implicit [RPCG03] and trapezoidal [CPRG04] tau-leaping methods correspond to the parameter values $\theta = 0$, $\theta = 1$ and $\theta = \frac{1}{2}$ respectively

$$\text{Explicit:} \qquad Y_{n+1} = Y_n + \sum_{r=1}^{R} \nu_r \mathcal{P}(a_r(Y_n)\tau),$$

$$\text{Implicit:} \qquad Y_{n+1} = Y_n + \sum_{r=1}^{R} \nu_r a_r(Y_{n+1})\tau + \sum_{r=1}^{R} \nu_r \Big[\mathcal{P}\Big(a_r(Y_n)\tau \Big) - a_r(Y_n)\tau \Big],$$

Trapezoidal: $$Y_{n+1} = Y_n + \frac{1}{2}\sum_{r=1}^{R}\nu_r a_r(Y_{n+1})\tau + \sum_{r=1}^{R}\nu_r\Big[\mathcal{P}\Big(a_r(Y_n)\tau\Big) - \frac{1}{2}a_r(Y_n)\tau\Big].$$

Theta methods are implicit in their deterministic part but use explicit approximation of driving stochastic processes. Split-step methods form another class of integrators which allow to incorporate implicitness into stochastic part of the system as well. This is achieved by treating deterministic and stochastic parts separately. The classical split-step method starts with computing the deterministic predictor of the solution which is then used in approximation of intensities of the driving Poisson processes. The numerical scheme reads as

$$\begin{aligned}\hat{Y}_n &= Y_n + \sum_{r=1}^{R}\nu_r a_r(\hat{Y}_n)\tau, \\ \tilde{Y}_n &= \hat{Y}_n + \sum_{r=1}^{R}\nu_r\overline{\mathcal{P}}\left(a_r(\hat{Y}_n)\tau\right) = Y_n + \sum_{r=1}^{R}\nu_r\mathcal{P}\left(a_r(\hat{Y}_n)\tau\right).\end{aligned} \tag{II.12}$$

It is seen that split-step scheme is very similar to the explicit tau-leaping approximation in a way it treats stochastic integrals. Combined with enhanced stability properties, it makes this scheme very useful and easily implementable. But this similarity with explicit methods also has undesirable implications which will be explained in the next section.

In this chapter, we propose the modified two-stage split-step method which combines ideas of the theta method (II.11) and the split-step method (II.12). The proposed scheme has the form

$$\begin{aligned}\hat{Y}_n &= Y_n + \left((1-\eta_1)\sum_{r=1}^{R}\nu_r a_r(Y_n) + \eta_1\sum_{r=1}^{R}\nu_r a_r(\hat{Y}_n)\right)(1-\theta)\tau, \\ \tilde{Y}_n &= \hat{Y}_n + \sum_{r=1}^{R}\nu_r\overline{\mathcal{P}}\left(a_r(\hat{Y}_n)\tau\right), \\ Y_{n+1} &= \tilde{Y}_n + \left((1-\eta_2)\sum_{r=1}^{R}\nu_r a_r(\tilde{Y}_n) + \eta_2\sum_{r=1}^{R}\nu_r a_r(Y_{n+1})\right)\theta\tau,\end{aligned} \tag{II.13}$$

with parameters $\nu_1, \nu_2, \theta \in [0; 1]$. When $\theta = 0$, $\nu_1 = 1$ this scheme converts to the classical split-step method while for the parameter values $\theta = 1$, $\nu_2 = 1$ it gives the drift implicit tau-leaping scheme.

We also consider the modified variant of the above scheme where the parameters $\theta_r \in [0; 1]$ are chosen independently for every reaction channel

$$\begin{aligned}
\hat{Y}_n &= Y_n + \left((1-\eta_1) \sum_{r=1}^{R} (1-\theta_r)\nu_r a_r(Y_n) + \eta_1 \sum_{r=1}^{R} (1-\theta_r)\nu_r a_r(\hat{Y}_n) \right) \tau, \\
\tilde{Y}_n &= \hat{Y}_n + \sum_{r=1}^{R} \nu_r \overline{\mathcal{P}} \left(a_r(\hat{Y}_n)\tau \right), \\
Y_{n+1} &= \tilde{Y}_n + \left((1-\eta_2) \sum_{r=1}^{R} \theta_r \nu_r a_r(\tilde{Y}_n) + \eta_2 \sum_{r=1}^{R} \theta_r \nu_r a_r(Y_{n+1}) \right) \tau.
\end{aligned} \tag{II.14}$$

It will be shown in the next section that, with the proper choice of the parameters θ_r, this modified scheme shows very good results in application to chemical system with several well-separated time scales.

II.1.3. Numerical stability

In this section, we perform comparative linear stability analysis of the theta tau-leaping method (II.11) and the split-step scheme (II.13) and show superiority of the latter in application to stiff chemical systems.

Stiff systems are characterized by the presence of multiple time scales with the fastest scale being stable. In deterministic setting, this means existence of the attractive invariant manifold which slaves the fast variables of the system. Efficient stiff integrators must ensure that generated numerical solutions stay close to this invariant manifold. For instance, implicit methods achieve this by damping fast components of the solution which try to push the system out from its stable state.

In the context of stiff chemical systems, stability of the fastest scale means existence of

the stationary distribution of the corresponding fast variables. By analogy with deterministic systems, it would be very beneficial if one can skip over the stable fast states while capturing their stochastic influence on slow species. Therefore, the ability to correctly resolve the stationary distribution of fast species is an essential requirement in construction of stiff stochastic integrators. The classical theta tau-leaping methods are not well suited for this purpose.

In order to see this, consider the reversible isomerization reaction which serves as analog of the deterministic Dahlquist's test equation [CPRG04]

$$S_1 \underset{c_2}{\overset{c_1}{\rightleftharpoons}} S_2. \tag{II.15}$$

This reaction has linear propensities $a_1 = c_1 X_1$, $a_2 = c_2 X_2$ with stoichiometric vectors $\nu_1 = (-1, 1)^T$ and $\nu_2 = (1, -1)^T$. The stationary distribution of (II.15) is the binomial distribution

$$\mathbb{P}\left[X^* = x\right] = \frac{x_T!}{x_T!(x_T - x)!} q^x (1-q)^{x_T - x}$$

with mean and variance

$$\begin{aligned} \mathbb{E}\left[X^*\right] &= q x_T = \frac{c_2}{c_1 + c_2} x_T, \\ \mathbb{V}ar\left[X^*\right] &= q(1-q) x_T = \frac{c_1 c_2}{(c_1 + c_2)^2} x_T, \end{aligned}$$

where $X^* = X_1(\infty)$, $q = \dfrac{c_2}{c_1 + c_2}$ and x_T denotes the fixed total number of molecules in the system, i.e., $x_T = X_1 + X_2$.

Let Y_n denote the numerical approximation of $X_1(t_n)$ and consider the theta method (II.11) applied to the test system (II.15)

$$Y_{n+1} = Y_n + \Big(c_2(x_T - Y_{n+1}) - c_1 Y_{n+1} - c_2(x_T - Y_n) + c_1 Y_n \Big) \theta\tau$$

$$+\mathcal{P}\Big(c_2(x_T-Y_n)\tau\Big)-\mathcal{P}\Big(c_1Y_n\tau\Big)$$
$$=Y_n+\frac{1}{1+(c_1+c_2)\theta\tau}\left[\mathcal{P}\Big(c_2(x_T-Y_n)\tau\Big)-\mathcal{P}\Big(c_1Y_n\tau\Big)\right].$$

The mean and the variance of the solution obtained with the theta method can be easily calculated as

$$\mathbb{E}\left[Y_{n+1}\right]=\mathbb{E}\left[\mathbb{E}\left[Y_{n+1}|Y_n\right]\right]=\mathbb{E}\left[Y_n+\frac{c_2(x_T-Y_n)\tau-c_1Y_n\tau}{1+(c_1+c_2)\theta\tau}\right]$$
$$=\frac{1-(c_1+c_2)(1-\theta)\tau}{1+(c_1+c_2)\theta\tau}\mathbb{E}\left[Y_n\right]+\frac{c_2x_T\tau}{1+(c_1+c_2)\theta\tau}$$

and

$$\mathbb{V}ar\left[Y_{n+1}\right]=\mathbb{V}ar\left[\mathbb{E}\left[Y_{n+1}|Y_n\right]\right]+\mathbb{E}\left[\mathbb{V}ar\left[Y_{n+1}|Y_n\right]\right]$$
$$=\mathbb{V}ar\left[Y_n+\frac{c_2(x_T-Y_n)\tau-c_1Y_n\tau}{1+(c_1+c_2)\theta\tau}\right]+\mathbb{E}\left[\frac{c_1Y_n\tau+c_2(x_T-Y_n)\tau}{(1+(c_1+c_2)\theta\tau)^2}\right]$$
$$=\left(\frac{1-(c_1+c_2)(1-\theta)\tau}{1+(c_1+c_2)\theta\tau}\right)^2\mathbb{V}ar\left[Y_n\right]+\frac{(c_1-c_2)\tau}{(1+(c_1+c_2)\theta\tau)^2}\mathbb{E}\left[Y_n\right]$$
$$+\frac{c_2x_T\tau}{(1+(c_1+c_2)\theta\tau)^2}.$$

These recurrences have the solutions

$$\mathbb{E}\left[Y_n\right]=A^n\mathbb{E}\left[Y_0\right]+\frac{1-A^n}{1-A}B,$$
$$\mathbb{V}ar\left[Y_n\right]=A^{2n}\mathbb{V}ar\left[Y_0\right]+\frac{C}{A}\frac{1-A^n}{1-A}\left(\mathbb{E}\left[Y_n\right]-\frac{B}{1-A}\right)+\frac{1-A^{2n}}{1-A^2}\left(\frac{BC}{1-A}+D\right),$$

where the coefficients A, B, C and D are given by

$$A=\frac{1-(c_1+c_2)(1-\theta)\tau}{1+(c_1+c_2)\theta\tau},\qquad B=\frac{c_2x_T\tau}{1+(c_1+c_2)\theta\tau},$$

$$C = \frac{(c_1 - c_2)\tau}{(1 + (c_1 + c_2)\theta\tau)^2}, \qquad D = \frac{c_2 x_T \tau}{(1 + (c_1 + c_2)\theta\tau)^2}.$$

The following condition ensures the global stability of the mean and the variance of the numerical solution obtained with the theta method (II.11)

$$|A| = \left| \frac{1 - (c_1 + c_2)(1 - \theta)\tau}{1 + (c_1 + c_2)\theta\tau} \right| < 1. \tag{II.16}$$

Assuming that $|A| < 1$ and letting $n \to \infty$, we get the mean and the variance of the stationary distribution generated by the theta method

$$\begin{aligned} \mathbb{E}\left[Y_\infty\right] &= \frac{B}{1 - A} = \mathbb{E}\left[X^*\right], \\ \mathbb{V}ar\left[Y_\infty\right] &= \frac{1}{1 - A^2}\left(\frac{BC}{1 - A} + D\right) = \frac{2}{2 + (c_1 + c_2)(2\theta - 1)\tau}\mathbb{V}ar\left[X^*\right]. \end{aligned} \tag{II.17}$$

It is seen that, when the stability condition (II.16) is satisfied, theta method recovers the stationary mean of the true solution for all values of the parameter θ. Moreover, the trapezoidal method with $\theta = \frac{1}{2}$ is able to recover the stationary variance as well. However, if $\theta < \frac{1}{2}$, the theta method underdamps the variance while for $\theta > \frac{1}{2}$, the stationary variance is overdamped.

Table 1 shows propagation coefficients of the error and amplifiers of the stationary variance of the scheme (II.11) for several values of parameter θ. Due to the linearity of the test system (II.15), propagation of the error in the mean of the solution has the same behavior as if the theta scheme was applied to the corresponding deterministic reaction rate equation. For the same reason, explicit tau-leaping method with $\theta = 0$ is only conditionally stable for $\tau < \dfrac{2}{c_1 + c_2}$. Trapezoidal and implicit tau-leaping schemes are unconditionally stable for all values of the time step. But, by analogy with deterministic systems, trapezoidal scheme is only A-stable since its propagation coefficient approaches -1 for increasing time steps. This severely limits its application to stiff systems even though trapezoidal scheme

Table 1: Propagation coefficients of the error and amplifiers of the stationary variance of the theta method (II.11) applied to the test system (II.15). The parameter λ is set to $\lambda = c_1 + c_2$.

θ	Propagation coefficients	Variance amplifiers
θ	$\dfrac{1-\lambda(1-\theta)\tau}{1+\lambda\theta\tau}$	$\dfrac{2}{2+\lambda(2\theta-1)\tau}$
0	$1-\lambda\tau$	$\dfrac{2}{2-\lambda\tau}$
$\dfrac{1}{2}$	$\dfrac{2-\lambda\tau}{2+\lambda\tau}$	1
1	$\dfrac{1}{1+\lambda\tau}$	$\dfrac{2}{2+\lambda\tau}$

correctly resolves the stationary variance. The reason is that the A-stability implies that the relaxation time of the mean value can be much larger then the time scale separation of the system. On the other side, the implicit scheme is L-stable but it is also not suitable for stiff systems since the stationary distribution becomes more atomic for larger time steps.

Now consider the split-step method (II.13) applied to the system (II.15)

$$\begin{aligned}
\hat{Y}_n &= \frac{1-(1-\eta_1)(c_1+c_2)(1-\theta)\tau}{1+\eta_1(c_1+c_2)(1-\theta)\tau}Y_n + \frac{c_2 x_T(1-\theta)\tau}{1+\eta_1(c_1+c_2)(1-\theta)\tau},\\
\tilde{Y}_n &= (1+\tau(c_1+c_2))\hat{Y}_n + \mathcal{P}\left(c_2(x_T-\hat{Y}_n)\tau\right) - \mathcal{P}\left(c_1\hat{Y}_n\tau\right) - c_2\tau x_T,\\
Y_{n+1} &= \frac{1-(1-\eta_2)(c_1+c_2)\theta\tau}{1+\eta_2(c_1+c_2)\theta\tau}\tilde{Y}_n + \frac{c_2 x_T\theta\tau}{1+\eta_2(c_1+c_2)\theta\tau}.
\end{aligned}$$

Equations for the mean and the variance take the form

$$\mathbb{E}\left[Y_{n+1}\right] = \left(\frac{1-(1-\eta_1)(c_1+c_2)(1-\theta)\tau}{1+\eta_1(c_1+c_2)(1-\theta)\tau}\right)\left(\frac{1-(1-\eta_2)(c_1+c_2)\theta\tau}{1+\eta_2(c_1+c_2)\theta\tau}\right)\mathbb{E}\left[Y_n\right]$$

$$+\frac{1-(1-\eta_1-\eta_2)(c_1+c_2)\theta(1-\theta)\tau}{\Big(1+\eta_1(c_1+c_2)(1-\theta)\tau\Big)\Big(1+\eta_2(c_1+c_2)\theta\tau\Big)}c_2x_T\tau,$$

$$\mathbb{V}ar\left[Y_{n+1}\right]=\left(\frac{1-(1-\eta_1)(c_1+c_2)(1-\theta)\tau}{1+\eta_1(c_1+c_2)(1-\theta)\tau}\right)^2\left(\frac{1-(1-\eta_2)(c_1+c_2)\theta\tau}{1+\eta_2(c_1+c_2)\theta\tau}\right)^2\mathbb{V}ar\left[Y_n\right]$$

$$+\left(\frac{1-(1-\eta_1)(c_1+c_2)(1-\theta)\tau}{1+\eta_1(c_1+c_2)(1-\theta)\tau}\right)\left(\frac{1-(1-\eta_2)(c_1+c_2)\theta\tau}{1+\eta_2(c_1+c_2)\theta\tau}\right)^2(c_1-c_2)\tau\mathbb{E}\left[Y_n\right]$$

$$+\left(\frac{1-(1-\eta_2)(c_1+c_2)\theta\tau}{1+\eta_2(c_1+c_2)\theta\tau}\right)^2\left(1+\frac{(c_1-c_2)(1-\theta)\tau}{1+\eta_1(c_1+c_2)(1-\theta)\tau}\right)c_2x_T\tau$$

and the global stability condition reads as

$$\left|\left(\frac{1-(1-\eta_1)(c_1+c_2)(1-\theta)\tau}{1+\eta_1(c_1+c_2)(1-\theta)\tau}\right)\left(\frac{1-(1-\eta_2)(c_1+c_2)\theta\tau}{1+\eta_2(c_1+c_2)\theta\tau}\right)\right|<1.$$

If the above stability condition is satisfied then, similarly to (II.17), the stationary moments of the solution generated by the split-step method (II.13) have the following values

$$\mathbb{E}\left[Y(\infty)\right]=\mathbb{E}\left[X^*\right],$$

$$\mathbb{V}ar\left[Y(\infty)\right]$$

$$=\frac{2\tau(c_1+c_2)}{\left(\frac{1+\eta_2(c_1+c_2)\theta\tau}{1-(1-\eta_2)(c_1+c_2)\theta\tau}\right)^2-\left(\frac{1-(1-\eta_1)(c_1+c_2)(1-\theta)\tau}{1+\eta_1(c_1+c_2)(1-\theta)\tau}\right)^2}\mathbb{V}ar\left[X^*\right].$$

One can see that the split-step method also preserves the stationary mean of the true solution. The stationary variance depends on the choice of implicitness parameters θ, η_1, η_2 and has more difficult behavior.

Table 2 provides propagation coefficients and variance amplifiers of the scheme (II.13) for different values of the parameters. The first interesting observation from this table is that the two-stage split-step scheme with $\eta_1=0$, $\eta_2=1$ reproduces stability properties of the theta tau-leaping method (II.11) for the corresponding values of the parameter θ. The second

Table 2: Propagation coefficients of the error and amplifiers of the stationary variance of the split-step method (II.12) applied to the test system (II.15). The parameter λ is set to $\lambda = c_1 + c_2$.

θ	η_1	η_2	Propagation coefficients	Variance amplifiers
θ	0	1	$\dfrac{1-\lambda(1-\theta)\tau}{1+\lambda\theta\tau}$	$\dfrac{2}{2+\lambda(2\theta-1)\tau}$
θ	1	1	$\dfrac{1}{1+\lambda\tau+\theta(1-\theta)(\lambda\tau)^2}$	$\dfrac{2\lambda\tau}{(1+\theta\lambda\tau)^2-\dfrac{1}{(1+(1-\theta)\lambda\tau)^2}}$
0	0	1	$1-\lambda\tau$	$\dfrac{2}{2-\lambda\tau}$
1	0	1	$\dfrac{1}{1+\lambda\tau}$	$\dfrac{2}{2+\lambda\tau}$
$\frac{1}{2}$	0	1	$\dfrac{2-\lambda\tau}{2+\lambda\tau}$	1
0	1	-	$\dfrac{1}{1+\lambda\tau}$	$\dfrac{2(1+\lambda\tau)^2}{2+\lambda\tau}$

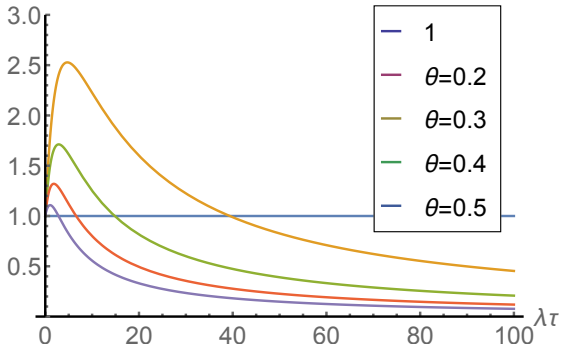


Figure 5: Variance amplifiers of the split-step scheme (II.13) with $\eta_1 = \eta_2 = 1$ for different values of the parameter θ. The parameter λ is set to $\lambda = c_1 + c_2$

observation is that the classical split-step scheme ($\theta = 0$, $\eta_1 = 1$) is not able to preserve stability of the stationary variance leading to possibly unbounded pathwise solutions.

Figure 5 illustrates the variance amplifiers of the split-step scheme eq. (II.13) with $\eta_1 = \eta_2 = 1$ for different values of the parameter $\theta \in (0; 1)$. It is worth noting that, for this choice of the parameters η_1 and η_2, the method is L-stable in the sense that propagation coefficient tends to 0 for large values of time step. It is also clear that by appropriate choice of the parameter θ it is possible to get the scheme which is able to reproduce the exact stationary variance of the solution, i.e., we want to have

$$\frac{2\lambda\tau}{(1+\theta\lambda\tau)^2 - \dfrac{1}{(1+(1-\theta)\lambda\tau)^2}} = 1.$$

The above equation does not have a closed form solution in terms of elementary functions. However, for large values of $\lambda\tau$, the fraction in the denominator can be neglected which gives the following approximate solution

$$\theta = \sqrt{\frac{2}{\lambda\tau}} - \frac{1}{\lambda\tau}, \qquad \lambda = c_1 + c_2.$$

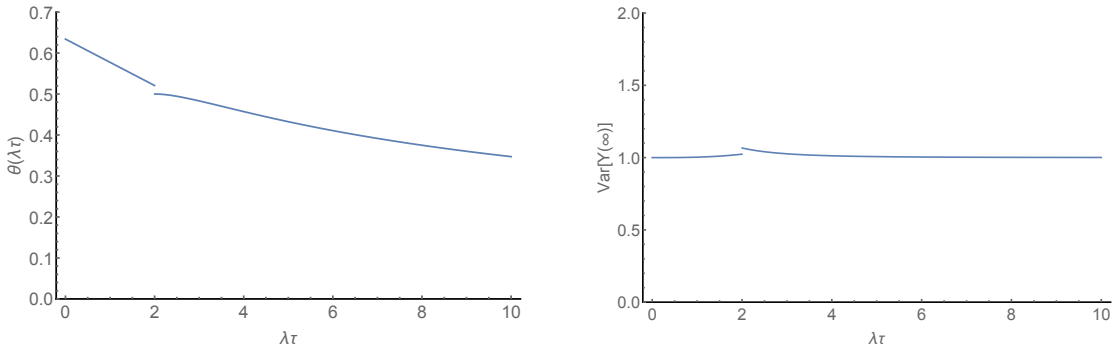


Figure 6: Optimal values of the parameter θ according to (II.18) and corresponding values of the variance amplifier for the split-step scheme.

For small values of $\lambda\tau$, one can search for the solution in the form

$$\theta = \theta_0 + \theta_1\lambda\tau + \theta_2(\lambda\tau)^2 + \theta_3(\lambda\tau)^3 + ...$$

Substituting this asymptotic expansion into the original equation and solving for the coefficients, we get

$$\theta_0 = \frac{3-\sqrt{3}}{2}, \quad \theta_1 = \frac{-9+5\sqrt{3}}{6}, \quad \theta_2 = \frac{108-187\sqrt{3}}{24}, \quad \theta_3 = \frac{-531+307\sqrt{3}}{36}, \quad ...$$

As can be seen from Figure 6, it is appropriate to choose the following representation for the parameter θ

$$\theta(\lambda\tau) = \begin{cases} \theta_0 + \theta_1\lambda\tau, & \text{if } \lambda\tau \leq 2, \\ \sqrt{\dfrac{2}{\lambda\tau}} - \dfrac{1}{\lambda\tau}, & \text{if } \lambda\tau > 2 \end{cases} \tag{II.18}$$

with $\lambda = c_1 + c_2$.

II.2. Description of the acceleration technique. Selection of the coarsest level

Let $X(t)$ be the solution of the stochastic integral equation (II.4) and $f(X(t))$ be the corresponding functional of interest. Given the sequence of nested time discretizations with step sizes

$$\tau_L < \tau_{L-1} < ... < \tau_l < ... < \tau_0 = T,$$

consider the problem of evaluation of the expectation of $f(X(t))$ with the multilevel Monte Carlo estimator of the form

$$\mathbb{E}\left[f\big(X(t)\big)\right] \approx \mathbb{E}_{\mathbb{ML}}\left[f_L\right] = \frac{1}{M_0}\sum_{m_0=1}^{M_0} f_0^{m_0} + \sum_{l=1}^{L}\frac{1}{M_l}\sum_{m_l=1}^{M_l}\left(f_l^{m_l} - f_{l-1}^{m_l}\right), \tag{II.19}$$

where $f_l = f(Y_l)$ and Y_l is the numerical approximation of $X(t)$ at level l obtained with any of the methods described in section II.1.2 .

The overall cost of the above estimator is given by (I.23) as

$$C_{\mathbb{ML}} = \epsilon_{II}^{-2}\left(\left(C_0' V_0'\right)^{1/2} + \sum_{l=1}^{L}(C_l V_l)^{1/2}\right)^2, \tag{II.20}$$

where C_0', V_0' denote the pathwise cost and the variance of the functional f_0 at the coarsest level and C_l, V_l denote the costs and the variances of the level corrections $(f_l - f_{l-1})$.

In the formulas (II.19) and (II.20), we have implicitly assumed that the coarsest level $l = 0$ corresponds to the largest possible time step $\tau_0 = T$. However, it has been mentioned previously that this choice can be far from optimal. To show this, let $C_{\mathbb{ML}}(L_0)$ denote the computational cost of the MLMC method as a function of the coarsest level $L_0 \geq 0$ and consider the difference

$$\Delta C_{\mathbb{ML}} = C_{\mathbb{ML}}(L_0+1) - C_{\mathbb{ML}}(L_0)$$
$$= \epsilon_{II}^{-2}\left[\left(\sqrt{C'_{L_0+1}V'_{L_0+1}} + \sum_{l=L_0+2}^{L}\sqrt{C_l V_l}\right)^2 - \left(\sqrt{C'_{L_0}V'_{L_0}} + \sum_{l=L_0+1}^{L}\sqrt{C_l V_l}\right)^2\right]$$
$$= \epsilon_{II}^{-1}\left(\sqrt{C'_{L_0+1}V'_{L_0+1}} - \sqrt{C'_{L_0}V'_{L_0}} - \sqrt{C_{L_0+1}V_{L_0+1}}\right)\left(\sqrt{C_{\mathbb{ML}}(L_0)} + \sqrt{C_{\mathbb{ML}}(L_0+1)}\right).$$

The construction of the multilevel estimator suggests that the cost $C_{\mathbb{ML}}(L_0)$ must be monotonically increasing function of L_0. This means that the difference $\Delta C_{\mathbb{ML}}$ must be strictly positive and the following condition must be satisfied for all $L_0 \geq 0$

$$\sqrt{C'_{L_0}V'_{L_0}} + \sqrt{C_{L_0+1}V_{L_0+1}} < \sqrt{C'_{L_0+1}V'_{L_0+1}}. \tag{II.21}$$

In practice, this assertion is often violated at coarse levels and the above condition converts to the stopping criterion for the greedy estimation of the coarsest level. To apply it, one can start with some sufficiently high level and move down the discretization hierarchy by iteratively checking (II.21). This procedure does not require the knowledge of parametric models for the costs and the variances and relies solely on the empirically observed values of C_l and V_l. The stability constraint on the largest allowed time step of the numerical scheme is also automatically satisfied by using (II.21) and no a-priory analysis is required.

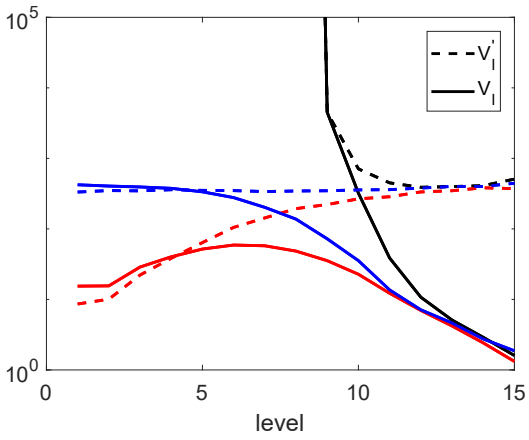
II.2.1. Comparative study of numerical schemes

As a benchmark problem to study the impact of the choice of pathwise integrators on the overall complexity of the MLMC algorithm, we will use the reaction network in (II.8)

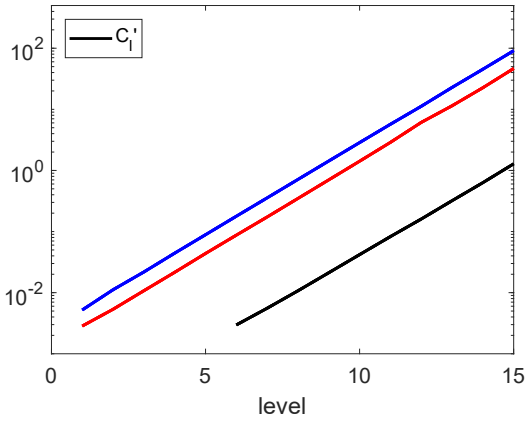
$$S_1 \underset{10^4}{\overset{10^4}{\rightleftharpoons}} S_2 \underset{10^2}{\overset{10^2}{\rightleftharpoons}} S_3 \underset{10^5}{\overset{10^5}{\rightleftharpoons}} S_4$$

with the terminal time $T = 10^{-3}$ and the functional $f(X(t)) = \big|X_3(T) - X_2(T)\big|$.

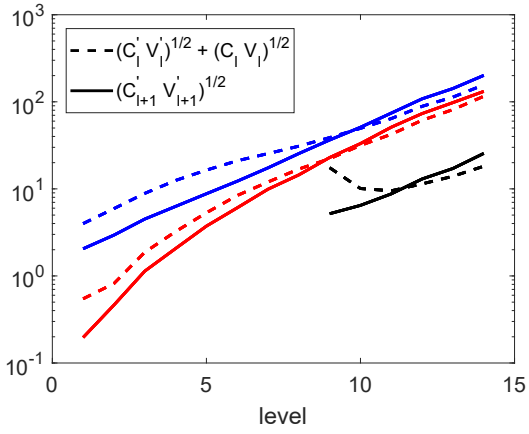
For example, Figure 7 illustrates comparative performance of three particular numerical



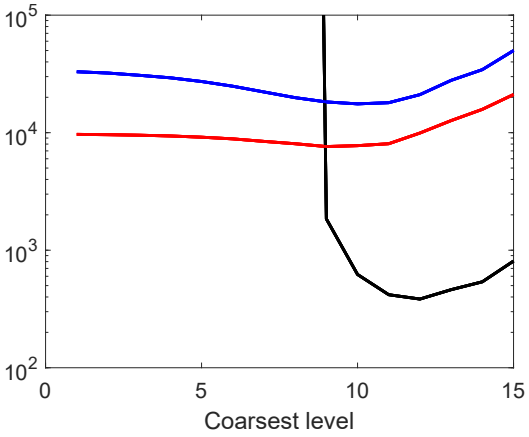
(A) Variances.



(B) Costs.



(C) Stopping criterion.



(D) Total cost of the estimator.

Figure 7: Comparison of pathwise integrators.

Black color is used for explicit scheme, red color is used for implicit scheme and blue color is used for the split-step scheme. The highest level is fixed to $L = 15$.

schemes: the classical explicit and implicit tau-leap methods and the proposed split-step scheme with adaptive choice of the parameter θ in (II.18). Figure 7D clearly shows that none of these methods allow to utilize all levels of the hierarchy since, as we move from the level 15 to the level 0, the cost of the estimator eventually starts to grow. Explanation for this undesirable behavior can be found in Figures 7A-7C. Figure 7A indicates that the variances V_l' and V_l of the functionals and the level corrections have the same order of magnitude at low discretization levels. For explicit method (black color) this is the result of instability of numerical approximation which leads to unbounded values of V_l' and V_l. Implicit scheme (red color) is stable but, as the analysis in the previous section shows, it severely overdamps the variance of the numerical solution for large time steps which again gives that $V_l' \approx V_l$ at low levels. The variance V_l' of the split-step method (blue color) remains nearly constant as predicted but V_l still growths due to the decreasing correlation between coupled paths for increasing step size.

The condition (II.21) can remain true only if $V_{L_0+1} \ll V_{L_0}'$ in order to compensate the large cost of generating the coupled paths which is defined as $C_{L_0+1} = C_{L_0}' + C_{L_0+1}'$. The failure of this requirement can be seen in Figure 7C. The solid lines correspond to the contribution of the coarsest level to the total cost of the estimator while the dashed lines show the extra cost associated with moving down the discretization hierarchy. Intersection of these lines indicate the failure of (II.21) and it is not difficult to see that this occurs when $V_{L_0+1} \approx V_{L_0}'$. As was mentioned above, for explicit integrator, this is the consequence of its instability at low levels. An obvious remedy for this issue would be to substitute it with implicit scheme at these levels in the following way

$$\mathbb{E}_{\mathbb{ML}}\left[f_L\right] = \underbrace{\sum_{l=L_0}^{L_{int}-1} \mathbb{E}_{\mathbb{MC}}\left[\Delta_l^{imp}\right]}_{\text{implicit levels}} + \underbrace{\mathbb{E}_{\mathbb{MC}}\left[f_{L_{int}+1}^{exp} - f_{L_{int}}^{imp}\right]}_{\text{interface level}} + \underbrace{\sum_{l=L_{int}+1}^{L} \mathbb{E}_{\mathbb{MC}}\left[\Delta_l^{exp}\right]}_{\text{explicit levels}}, \quad \text{(II.22)}$$

where $\Delta_l = f_l - f_{l-1}$, L_{int} is the interface level determined from (II.21) as the last stable

explicit level and f_L^{imp}, f_L^{exp} correspond to the solutions obtained with implicit and explicit methods respectively. However, Figure 8 shows that such naive combination of integrators does not give any improvement over the purely explicit approach. This is due to the large difference in costs of explicit and implicit methods which gives the large jump in the overall complexity of the estimator when passing the interface level $l = L_{int}$ (green line in Figure 8D). A possible solution to this problem is proposed below.

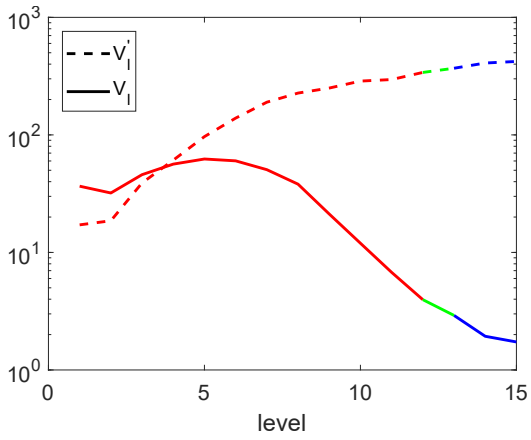
II.2.2. Criteria for the selection of pathwise integrators. Composite scheme

Based on the above mentioned, one can distinguish three desired properties of pathwise integrators which lead to the reduction of the complexity of the MLMC estimator:

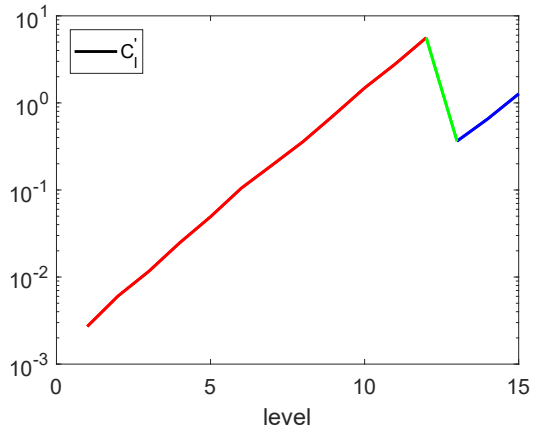
1. $V_l' \approx const$, i.e., the variance of the solution must be stable,
2. $V_l \ll V_l'$, i.e., the variance of the corrections must be smaller than that of the solution,
3. $C_l' < C_{l+1}'$, i.e., the cost at low levels must be smaller than the cost at high levels.

As is seen in Figures 7 and 8, the first and second properties are violated by all considered schemes while the last property is violated by the naive combination of explicit and implicit integrators. As a remedy for such undesirable behavior, we propose a different combination of implicit-explicit solvers which possesses all of the required features.

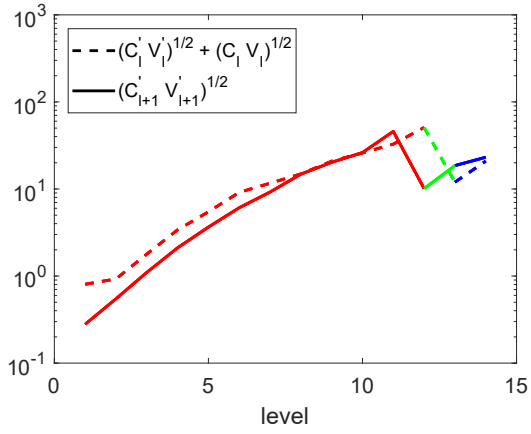
The proposed technique can be briefly describe as follows. By analogy with (II.22), we divide all levels into explicit, implicit and interface parts. The coarsest stable level $l = L_{int} + 1$ in the explicit part is again determined by repeatedly applying the stopping criterion in (II.21). However, the implicit integrator is constructed more wisely to ensure both decay of the cost and stability of the variance. Starting from the interface level $l = L_{int}$, we divide the original time interval $[0, T]$ into two subintervals and apply implicit and explicit schemes for $t \in [0, T_1]$ and $t \in [T_1, T]$ respectively. The discretization of the first subinterval is coarsened at lower levels $l < L_{int}$ in a regular manner in order to guarantee



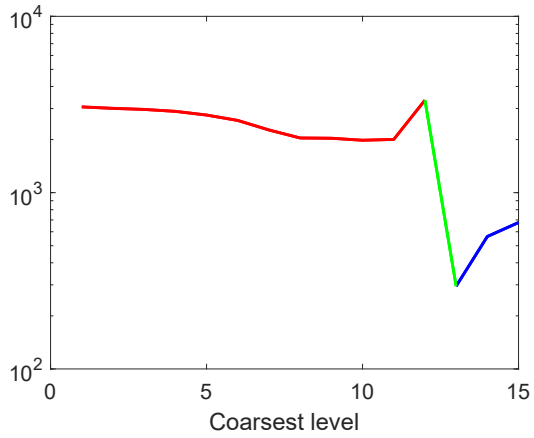
(A) Variances.



(B) Costs.



(C) Stopping criterion.



(D) Total cost of the estimator.

Figure 8: Performance of MLMC with stabilized integrator.

Blue and red colors correspond to the levels which use explicit and implicit solvers respectively, the interface level is highlighted with green color. The highest level is fixed to $L = 15$.

the decay of the cost. The second subinterval keeps the same discretization for all $l \leq L_{int}$ and is responsible for the correct behavior of the variances V_l' and V_l.

The sample paths of this composite integrator for explicit, implicit and interface levels are depicted in Figure 9. Additionally, Figure 10 demonstrates its performance in the multilevel Monte Carlo setting. One can see that the variances V_l' are stable, $V_l \ll V_l'$ as desired and the costs C_l' are strictly increasing functions. The proposed approach allows to use five additional levels when compared to the pure explicit integrator (blue line in Figure 10D) which is a large improvement over the naive stabilized algorithm, see Figure 8.

II.3. Numerical results

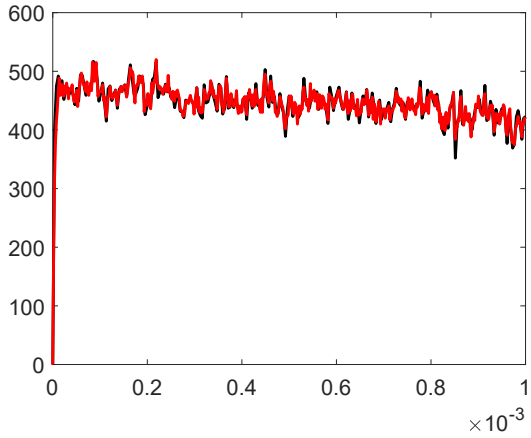
The proposed acceleration technique is not based on any a-priory given information about the system of interest. Therefore, the complexity analysis of the accelerated estimator cannot give a-priory estimates as well. The goal of this section is to provide the empirical justification of the proposed method and the quantitative study of its performance.

Example 1. For the first example, consider the linear chemical network (II.8) which was used as the benchmark problem in previous sections

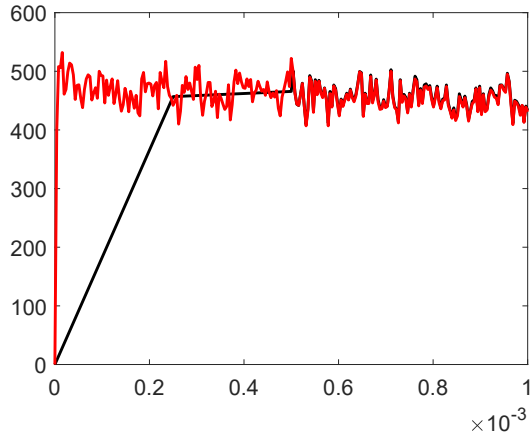
$$S_1 \underset{10^4}{\overset{10^4}{\rightleftharpoons}} S_2 \underset{10^2}{\overset{10^2}{\rightleftharpoons}} S_3 \underset{10^5}{\overset{10^5}{\rightleftharpoons}} S_4.$$

The initial condition is $X(0) = [1000, 10, 10, 50]^T$, the terminal time is fixed to $T = 10^{-3}$ and the functional is given by $f(X(t)) = \big|X_3(T) - X_2(T)\big|$.

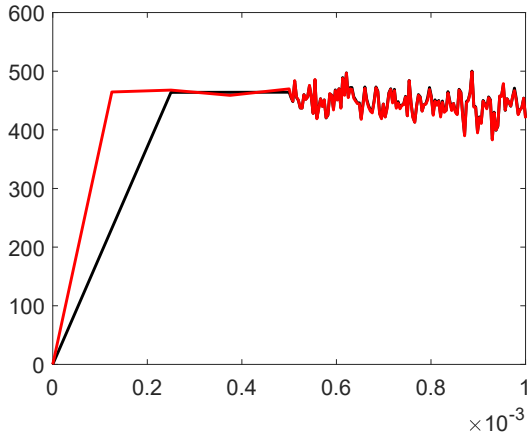
Figure 11 demonstrates performance of the multilevel Monte Carlo method in application to this system for two variants of the composite integrator with implicit part given by the classical implicit tau-leap scheme and the proposed split-step method respectively. One can see from Figure 11A that explicit tau-leap scheme cannot be applied for levels $l < 11$



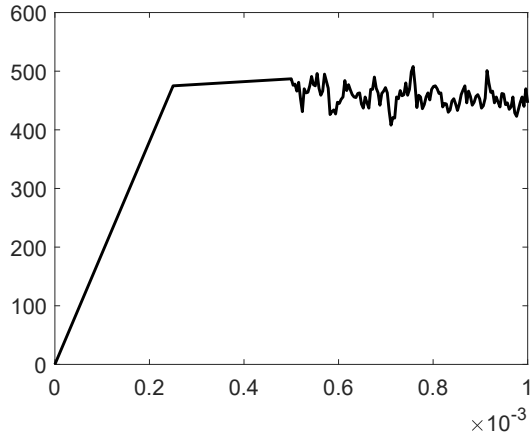
(A) Explicit level $l > L_{int}$.



(B) Interface level $l = L_{int}$.

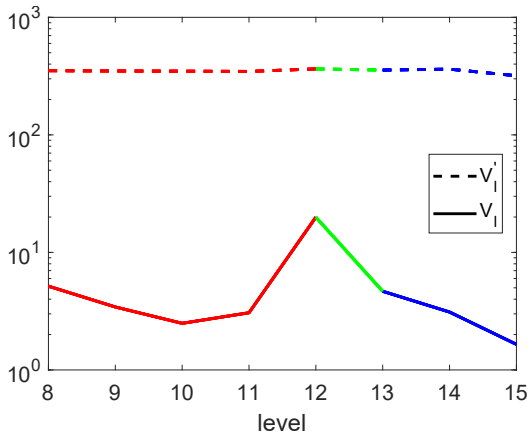


(C) Implicit level $l < L_{int}$.

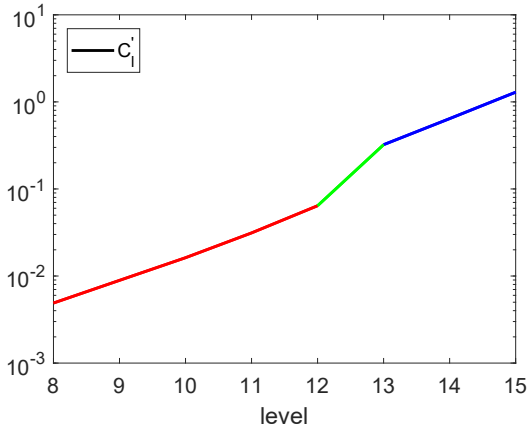


(D) Base level $l = L_0$.

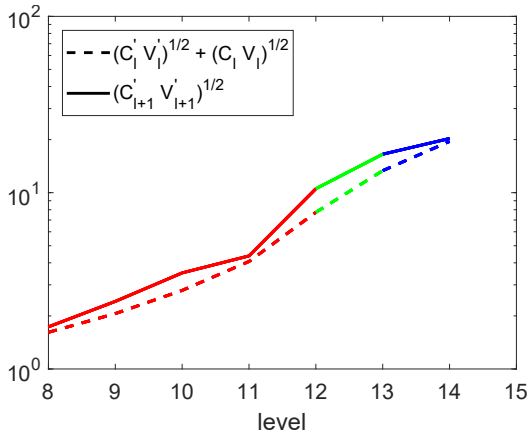
Figure 9: Sample paths of the composite integrator.



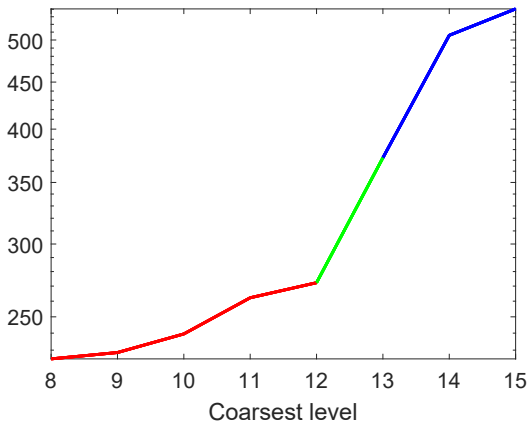
(A) Variances.



(B) Costs.



(C) Stopping criterion.



(D) Total cost of the estimator.

Figure 10: Performance of MLMC with composite integrator.

Blue and red colors correspond to the levels which use explicit and implicit solvers respectively, the interface level is highlighted with green color. The highest level is fixed to $L = 15$.

while the composite integrators can explore four additional levels. Figure 11B shows that this results in the significant speedup of the accelerated algorithm over the standard MLMC method. It is obvious that the relative speedup will decay for increasing tolerance ϵ of the estimator as the computational savings of the improved algorithm remain finite and the ϵ-complexity growths with no bound. However, the cost gain always remains a positive quantity as desired.

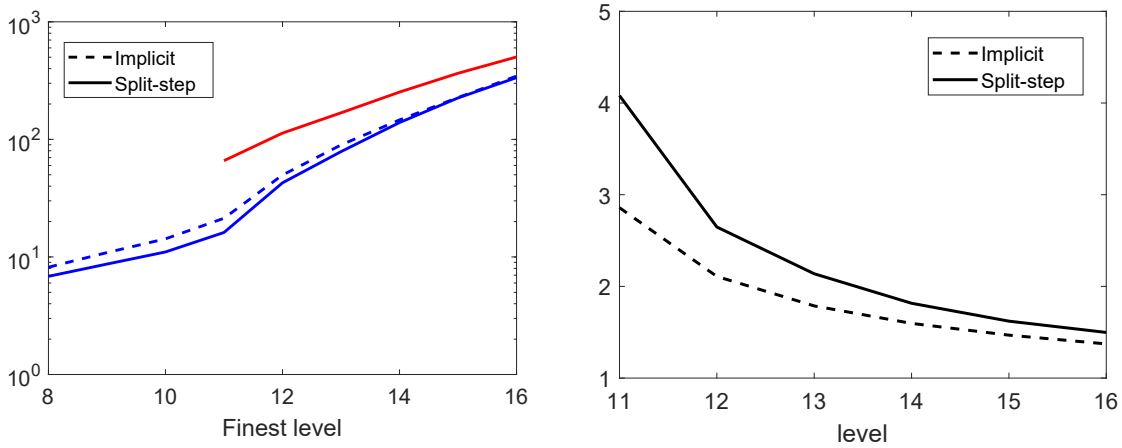
Example 2. In the second example, we consider the following stiff 3-species 6-reaction system with nonlinear reaction intensities

$$S_1 + S_2 \underset{10^3}{\overset{10^3}{\rightleftharpoons}} S_3,$$
$$S_1 + S_3 \underset{10}{\overset{10^{-5}}{\rightleftharpoons}} S_2,$$
$$S_2 + S_3 \underset{10^6}{\overset{1}{\rightleftharpoons}} S_1.$$

The initial condition is $X(0) = [900, 1000, 950000]^T$, the terminal time is fixed to $T = 10^{-4}$ and the functional is given by $f(X(t)) = \big|X_2(T) - X_1(T)\big|$.

Figure 12 illustrates performance of the MLMC method for this system. One can see that the accelerated MLMC algorithm results in even bigger computational savings than in Example 1. This happens due to the higher stiffness of this chemical network which prevents application of the explicit scheme at levels $l < 12$.

It is worth noting that, in this case, the split-step method is less efficient than the classical implicit scheme. Nevertheless, both variants of the composite integrator perform exceptionally well which fully justifies application of the proposed acceleration technique.

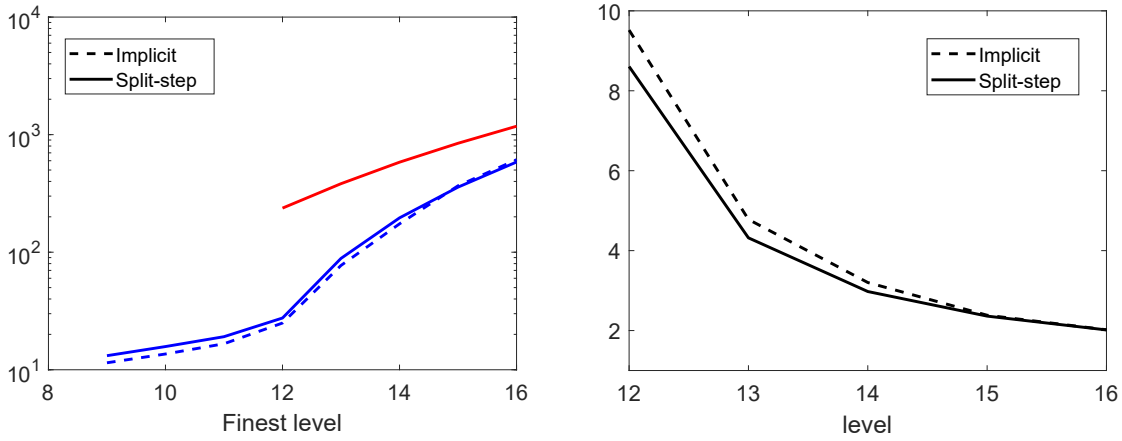


(A) Total cost of the estimator.

(B) Speedup.

Figure 11: Performance of the MLMC algorithm for the problem in Example 1.

Red color is used for the standard MLMC with explicit integrator. Dashed and solid lines correspond to the implicit and the split-step tau-leap methods as the implicit part of the composite integrator respectively. The lowest level of the explicit integrator is fixed to $L_{int} = 11$. The lowest level of the composite integrator is fixed to $L_0 = 8$.



(A) Total cost of the estimator.

(B) Speedup.

Figure 12: Performance of the MLMC algorithm for the problem in Example 2.

Red color is used for the standard MLMC with explicit integrator. Dashed and solid lines correspond to the implicit and the split-step tau-leap methods as the implicit part of the composite integrator respectively. The lowest level of the explicit integrator is fixed to $L_{int} = 12$. The lowest level of the composite integrator is fixed to $L_0 = 9$.

CHAPTER III

PARTIAL DIFFERENTIAL EQUATIONS WITH RANDOM INPUT DATA

In this chapter, we consider partial differential equations which are stochastically parameterized by the random input data. As was mentioned in previous chapters, the multilevel Monte Carlo method requires one to solve a large number of decoupled deterministic problems corresponding to different realizations of this data. For stationary PDEs, these solutions are often constructed by means of iterative process and the choice of initial guess can have a drastic influence on its convergence. The goal of this chapter is to study the acceleration of the multilevel Monte-Carlo method by supplying initial guesses to iterative solvers via recycling of the previously calculated data. Similar ideas have been paid attention in the literature before. For instance, recycling of Krylov subspace vectors in solving sequences of linear systems was discussed in [PdSM$^+$06, EES00, EEU07]. Later this approach was adopted for the construction of multilevel and stochastic preconditioners to the linear systems that arise from stochastic Galerkin approximations [GP11, JCL07]. Reusing solutions to previous systems as initial guesses to linear solvers is another obvious strategy which has been applied to the stochastic Galerkin (SG) and stochastic collocation (SC) methods in [GP11, GK96, PG00]. Recently an improvement to this approach was proposed in [GJWZ16] where the authors took advantage of the hierarchical construction of sparse grids and associated polynomial spaces of increasing fidelity to efficiently interpolate initial guesses.

Efficiency of acceleration techniques which are posed in the context of SG and SC methods is the consequence of the special intelligent choice of quadrature or interpolation nodes. Unfortunately, this very feature makes them impractical in the context of Monte Carlo methods where the locations of collocation points are generated randomly. In this chapter, we propose the acceleration technique that does not require any information about the structure of collocation points and does not depend on any particular form of the input parameter or

the range of the parameter space. The only required information is the appropriate metric of similarity between different instances of the parameter which may be finite or even infinite dimensional. This technique may be successfully applied to any problem involving large number of repeated solvings of iterative linear systems. Possible applications include but are not limited to classical and multilevel Monte Carlo methods, Markov Chain Monte Carlo method, stochastic optimization algorithms and other methods involving random iterates.

III.1. Problem setting

Recall the problem setting in section I.2. Let $\mathcal{L}$ be a possibly nonlinear stationary operator defined on a bounded Lipschitz domain $D \subset \mathbb{R}^d$, $d = 1, 2, 3$, with a boundary $\partial D = \partial D_D \cup \partial D_N$. Operator $\mathcal{L}$ has a random coefficient $a(x, \omega)$, $x \in D$, $\omega \in \Omega$, defined on the complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Here Ω denotes the sample space of possible outcomes, $\mathcal{F} \subset 2^{\Omega}$ is the σ-algebra of events, and $\mathbb{P}$ is the complete probability measure on $\mathcal{F}$. Denote by $u(x, \omega)$ the strong solution of the following stochastic boundary-value problem

$$\begin{cases} \mathcal{L}\big(a(x,\omega), u(x,\omega)\big) &= f(x,\omega) \quad \text{in } D \times \Omega, \\ \gamma\big(u(x,\omega)\big) &= g(x,\omega) \quad \text{on } \partial D \times \Omega, \end{cases} \tag{III.1}$$

where γ is a trace operator which defines Dirichlet boundary condition on ∂D_D and Neumann boundary condition on ∂D_N. We require $u(x, \omega)$ to be a Bochner integrable function with values in some Banach space $W(D)$, i.e., $u(x, \omega) \in L^p(\Omega; W(D))$, the function space given by

$$L^p\Big(\Omega; W(D)\Big) := \left\{ u : \Omega \to W(D) \;\middle|\; u \text{ is strongly measurable and } \|u\|_{L^p(\Omega;W(D))} < \infty \right\}$$

with the corresponding norm

$$\|u\|^p_{L^p(\Omega;W(D))} = \begin{cases} \displaystyle\int_\Omega \|u(\cdot,\omega)\|^p_{W(D)}\, d\mathbb{P}(\omega) & \text{if } 0 < p < \infty, \\ ess\sup_{\omega\in\Omega} \|u(\cdot,\omega)\|_{W(D)} & \text{if } p = \infty. \end{cases}$$

For simplicity, we will write $L^p(\Omega)$ instead of $L^p(\Omega; W(D))$ when the particular function space $W(D)$ can be concluded from the context.

It is assumed that all random input fields are defined on the same probability space and are chosen so that the problem in (I.5) is well-posed in the following sense

Assumption III.1.1. *For the fixed values of the random parameter $\omega \in \Omega$, there exist unique realizations of the solution $u(x,\omega) \in W(D)$ such that*

$$\|u(x,\omega)\|_{W(D)} \le C(\omega)\, \|f(x,\omega)\|_{W^{-1}(D)}\,,$$

where $W^{-1}(D)$ is the dual space of $W(D)$ and C is a constant which may depend on ω.

Example III.1.1 (Linear elliptic system)**.** *Given $f \in L^2\big(\Omega; H^{-1}(D)\big)$, find $u(x,\omega) \in L^2\big(\Omega; H^1_0(D)\big)$ such that*

$$\begin{cases} -\nabla\Big(a(x,\omega)\nabla u(x,\omega)\Big) &= f(x,\omega) \qquad \textit{in } D\times\Omega, \\ u(x,\omega) &= 0 \qquad \textit{on } \partial D\times\Omega \end{cases} \tag{III.2}$$

with $a(\cdot,\omega)$ bounded and coercive almost surely, i.e.,

$$\mathbb{P}\left[\omega\in\Omega : 0 < a_{min}(\omega) \le a(x,\omega) \le a_{max}(\omega) < \infty \quad, \forall x \in \overline{D}\right] = 1.$$

Example III.1.2 (Steady Navier-Stokes system)**.** *Let $\boldsymbol{H}^1(D) := \Big(H^1(D)\Big)^d$. Given $\boldsymbol{f} \in L^2\big(\Omega; \boldsymbol{H}^{-1}(D)\big)$ and $\boldsymbol{g} \in L^2\big(\Omega; \boldsymbol{H}^{1/2}(D)\big)$, find the pair $\Big(\boldsymbol{u}(x,\omega), p(x,\omega)\Big) \in L^2\big(\Omega; \boldsymbol{H}^1_g(D)\times L^2_0(D)\big)$ such that*

$$\begin{cases} -\nu(\omega)\nabla^2 \boldsymbol{u} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} + \nabla p &= \boldsymbol{f}(x,\omega) \qquad \textit{in } D\times\Omega, \\ \nabla\cdot\boldsymbol{u} &= 0 \\ \boldsymbol{u} &= \boldsymbol{g}(x,\omega) \qquad \textit{on } \partial D\times\Omega, \end{cases} \tag{III.3}$$

where $\boldsymbol{H}^1_g := \left\{\boldsymbol{u}\in\boldsymbol{H}^1(D) : \boldsymbol{u} = \boldsymbol{g} \textit{ on } \partial D\right\}$, $L^2_0(D) := \left\{p\in L^2(D) : \int_D p = 0\right\}$ *and* $\nu(\omega) > 0$ *is the spatially uniform uncertain viscosity.*

Note that the elliptic regularity and the Babuška-Brezzi inf-sup condition imply the existence of solutions for the problems in (III.2) and (III.3) respectively. However, the uniqueness result for the problem in (III.3) is guaranteed only if $\kappa_c \|\mathbf{f}(\cdot,\omega_f)\| < \nu^2$, i.e., for sufficiently small data **f** or sufficiently large viscosity ν [GR86, Gun89, Lio96]. Assuming that these conditions hold, Assumption (III.1.1) is satisfied and the considered problems are well-posed.

As was mentioned previously in section I.2.2, stochastic fields are often approximated with a finite number of uncorrelated, or even independent, random variables. For instance, piecewise constant approximation and Karhunen-Loeve (KL) expansion represent examples of truncated random fields which are most frequently employed in practice. It is worth noting that, unlike stochastic collocation and stochastic Galerkin methods, the finite-dimensional noise representation is not required in the Monte Carlo sampling as well as in the proposed acceleration technique. As a matter of fact, we will only need some similarity measure between realizations of random fields in a functional space which may be finite or infinite dimensional. However, finite stochastic dimensionality is assumed in order to simplify the analysis.

III.2. Acceleration of iterative solvers

Consider a hierarchical family of nested finite element discretizations

$$W_0 \subset W_1 \subset ... \subset W_k \subset ... \subset W(D),$$

where each W_k corresponds to the finite dimensional space of continuous piecewise polynomial functions defined on the triangulation τ_k of the domain D and $h_k := \max_{\tau \in \tau_k} diam(\tau)$ is the maximum mesh spacing parameter.

Let $u_k(\omega)$ denote the finite element projection of the solution $u(x, \omega)$ onto W_k. Then solution at the finest discretization level L is given as the telescoping series

$$u_L(\omega) = u_0(\omega) + \sum_{l=1}^{L} \Big(u_l(\omega) - u_{l-1}(\omega) \Big).$$

By setting $\Delta_l(y) = u_l(y) - u_{l-1}(y)$, the above expression yields the multilevel Monte-Carlo estimator

$$\mathbb{E}\left[u(x,\omega)\right] \approx \mathbb{E}_{\mathbb{ML}}\left[u_L\right] = \sum_{l=0}^{L} \frac{1}{M_l} \sum_{m_l=1}^{M_l} \Delta_l^{m_l}, \tag{III.4}$$

where $\Delta_0 = u_0(x, \omega)$ and M_l is the number of random samples generated at each level l.

Multilevel approximation allows to balance the cost and the sampling error of Monte Carlo estimators between different levels in a way which minimizes the computational cost of the method. The bulk of this cost goes into solving a large number of decoupled deterministic problems corresponding to different samples from the space of parameters. When solutions of deterministic problems are obtained with iterative solvers, one can improve the total performance of the algorithm by providing a fast and simple way for the estimation of good initial guesses. In this chapter, we will utilize a repeated sampling framework of the Monte Carlo method to devise a learning algorithm for prediction of initial guesses using the dynamically changing training dataset of the computed data.

The proposed method can be briefly described as follows. Consider the problem in (III.1) with a random coefficient $a(x, \omega)$ ranging over the compact set $\mathcal{A}$ of a suitable normed metric space X. A solution of this problem can be viewed as a function

$$u : a \to u(a)$$

which maps X to $W(D)$. We can supplement this solution map with an initial guess interpolated from the previously computed data. Due to the random design of the Monte Carlo method, realizations of the coefficient $a(\omega)$ have no structure or order between their relative locations in X. This motivates the use of scattered data interpolation techniques to get the desired prediction. Typically, such interpolants are constructed in reproducing kernel Hilbert spaces which suggest the model function of the form

$$\overset{a}{u}(a(\omega)) = \sum_i K\Big(\|a(\omega) - a(\omega_i)\|_X \Big)\alpha_i, \tag{III.5}$$

where $K\Big(\|a(\omega) - a(\omega_i)\|_X \Big) : X \times X \to \mathbb{R}$ is a positive-definite reproducing kernel acting as a measure of similarity in X [BKO$^+$96, Wen05, Fas07].

It is worth noting that we do not require very high accuracy of the interpolant in (III.5) since the goal is to estimate the initial guess and not the solution itself. In our case, the crucial issue is the cost/quality relationship of the interpolation procedure. Therefore, we will limit our attention to methods which allow explicit formulation, i.e., do not require a solution of linear systems.

The Shepard approximation is the classical example of such method. It has the form

$$\overset{a}{u}(a(\omega)) = \sum_i \frac{\Phi_\delta\left(\|a(\omega) - a(\omega_i)\|_X\right)}{\sum_j \Phi_\delta\left(\|a(\omega) - a(\omega_j)\|_X\right)} u(a(\omega_i)), \tag{III.6}$$

where $\Phi_\delta(\cdot) = \Phi(\cdot/\delta) : X \times X \to \mathbb{R}$ is a symmetric positive-definite kernel function with a support of radius δ and $K(\cdot) = \Phi_\delta(\cdot) / \sum_j \Phi_\delta(\cdot)$ is a moving least squares reproducing kernel (MLSRK) which provides a local polynomial reproduction of degree 0, i.e., it can be considered as locally fitting constant to the data [Wen05, LL04].

The Shepard approximation formula can be interpreted as a weighted average of the function values at a set of given nodes. In particular, for the uniform kernel $\Phi_\delta(a) = I(\|a\|_X/\delta \leq 1)$ with a variable support radius $\delta = \left\|a - a_{[k_{NN}]}\right\|_X$, it converts to the k-nearest neighbor (k-NN) estimator

$$\overset{a}{u}(a(\omega)) = \sum_i \frac{I\left(\dfrac{\|a(\omega)-a(\omega_i)\|_X}{\left\|a(\omega)-a_{[k_{NN}]}\right\|_X} \leq 1\right)}{\sum_j I\left(\dfrac{\|a(\omega)-a(\omega_j)\|_X}{\left\|a(\omega)-a_{[k_{NN}]}\right\|_X} \leq 1\right)} u(a(\omega_i)) = \frac{1}{k_{NN}} \sum_{i=1}^{k_{NN}} u(a(\omega_i)), \quad \text{(III.7)}$$

where $a_{[k_{NN}]}$ is the k_{NN}-th closest to $a(\omega)$ point and realizations of the coefficient are reordered according to increasing values of $\|a(\omega) - a(\omega_i)\|_X$.

Application of the formulas (III.5)-(III.7) requires calculation of distances in a functional space X. In a general case, this can be a serious limitation of the algorithm because such calculations are usually computationally expensive. Fortunately, finite dimensional representations of random fields often provide an isometric map to the discrete space of coefficients of such representations. This makes an evaluation of distances a trivial task. Consider the following example.

Example III.2.1 (Karhunen-Loeve expansion)**.** *Any second order random field with a continuous covariance function can be represented as a Karhunen-Loeve (KL) series of the form*

$$a(x,\omega) = \mathbb{E}\left[a(x,\omega)\right] + \sum_{k=1}^{\infty} \phi_k(x) y^k(\omega), \quad \text{(III.8)}$$

where $y^k(\omega)$ are uncorrelated random variables with variance λ_k and λ_k, ϕ_k are the dominant eigenvalues and eigenfunctions of the covariance operator. By virtue of the Parseval's identity, the space $L^2(D)$ of functions represented by the above series is isometrically isomorphic to the space l^2 of coefficients $\{y^k\}_{k=1}^{\infty}$.

Finite-dimensional representations of random fields are obtained by truncation of the series (III.8). *This means that the classical Euclidean distance in the image space* $\Gamma(\Omega) = \prod_{k=1}^{s} \Gamma^k \in \mathbb{R}^s$ *of the random vector* $[y^1, ..., y^s] : \Omega \to \Gamma$ *can be used as a measure of similarity between different realizations of the random field* $a(x, \omega) \in L^2\Big(\Omega, L^2(D)\Big)$.

In addition to the evaluation of distances, the search for nearest neighbors (NN) in space X is also required to apply formulas (III.6)-(III.7) because of the compact support of kernel functions. There are different techniques for the exact NN search based on partitioning trees, locality sensitive hashing, graph methods or approximate nearest neighbor matching [ML14]. Existing codes are highly scalable and can perform NN search to the very large data sets in parallel. We found the FLANN (Fast Library for Approximate Nearest Neighbor Search) library [ML] very useful in the framework of our acceleration method.

Our suggested implementation of the accelerated MLMC algorithm is given in Algorithm 1. Note that efficient NN search algorithms perform preprocessing of the data to build the appropriate data structures. Insertion of the new point is costly and may deteriorate the performance of the method. Therefore, we suggest that data structures must be rebuilt only after generation of a sufficient number of points.

The proposed acceleration technique can be successfully applied to both linear and nonlinear problems. To show this, we provide two examples of the application of Algorithm 1.

Example III.2.2. *Consider the linear elliptic problem* (III.2) *with a random coefficient* $a(x, \omega)$. *For each sampled* $\omega \in \Omega$, *finite element solutions* $u_l(x, \omega) = \sum_{i=1}^{N_l} c_{l,i}\phi_{l,i}$ *of corresponding deterministic problems are obtained by solving a linear system*

$$A_l c_l = f_l,$$

where N_l *is the number of nodes in discretization at level* l, $\phi_{l,i}\big|_{i=1}^{N_l}$ *is the basis of* W_l *and*

$$f_{l,i} = \int_D f \cdot (\phi_l)_i dx \qquad \text{for } i = 1, .., N_l,$$

Algorithm 1: The accelerated MLMC algorithm

Input: Desired tolerance ϵ of the estimator

Result: $\mathbb{E}\left[u(x,\omega)\right] \approx \sum_{l=0}^{L} M_l^{-1} \sum_{m_l=1}^{M_l} \left(u_l^{m_l} - u_{l-1}^{m_l}\right)$

Define discretization at level $l = 0$
Initialize the number of levels L
Initialize the number of samples per level M_l
Set $M_l^{add} = M_l$
while $\sum_{l=0}^{L} M_l^{add} \neq 0$ **do**
  for $l = 0$ **to** L **do**
    for $m_l = 1$ **to** M_l^{add} **do**
      Sample $\omega \in \Omega$
      Compute the initial guess according to (III.6):

$$\overset{a}{u}{}_{l-1}^{m_l} = \sum_i \frac{\Phi_\delta\left(\|a(\omega) - a(\omega_i)\|_X\right)}{\sum_j \Phi_\delta\left(\|a(\omega) - a(\omega_j)\|_X\right)} u^i$$

      Compute the solution $u_{l-1}^{m_l}$ with $\overset{a}{u}{}_{l-1}^{m_l}$ as initial guess
      Compute the solution $u_l^{m_l}$ with $u_{l-1}^{m_l}$ as initial guess
    end
    Update the MC estimator at level l
    Estimate the sample variance $\overline{\sigma}_l$ at level l
    Estimate the computational cost $\mathbf{C}_l$ at level l
  end
  Update the MLMC estimator
  Update the number of levels L
  Update the number of samples per level M_l
  Find the number of extra samples per level M_l^{add}
end

$$A_{l,ii'} = \int_D a(\omega) \cdot \nabla \phi_{l,i} \cdot \nabla \phi_{l,i'} dx \qquad \textit{for } i, i' = 1, .., N_l.$$

This means that resulting stiffness matrices A_l are sparse, symmetric and positive definite. In this case, the canonical choice of the linear solver is the conjugate gradient (CG) method [Saa03] which defines iterative process for $\tilde{c}_l = c_l - c_l^{(0)}$ as

$$\begin{aligned} \tilde{c}_l^{(0)} &= 0, \\ \tilde{c}_l^{(k+1)} &= \tilde{c}_l^{(k)} + \frac{p_k^T r_{k-1} p_k}{p_k^T A_l p_k}, \end{aligned}$$

where $r_k = f_l - A_l \left(\tilde{c}_l^{(k)} + c_l^{(0)} \right)$ is the residual at k-th iteration, vectors p_k are mutually conjugate with respect to A_l and $p_0 = r_0$. The convergence of this recursion is determined by the choice of initial guess $c_l^{(0)}$ and conditioning of the matrix A_l. Therefore, for the given matrix A_l, we can accelerate the CG method by proposing initial guess as in Algorithm 1.

We will use the above linear elliptic system as the benchmark problem for the theoretical analysis of our acceleration technique. Additionally, we will consider the following example to show numerically that the method can be successfully applied to nonlinear problems as well.

Example III.2.3. *Consider the Navier-Stokes system* (III.3) *with a random forcing term $\boldsymbol{f}(x, \omega)$. Let $(\boldsymbol{V}_l, S_l)$ denote the pair of div-stable velocity and pressure finite element spaces with basis functions $\boldsymbol{v}_i(x)$, $i = 1, .., N_l$, and $q_j(x)$, $j = 1, .., P_l$, respectively. According to the composite method [Gun89], coefficients $\alpha_i(x)$, $i = 1, .., N_l$, and $\beta_j(x)$, $j = 1, .., P_l$, of the finite element representations for velocity and pressure fields can be found by solving the sequence of linear systems*

$$\begin{bmatrix} A_l^{(k+1)} & B_l^{(k+1)} \\ \left(B_l^{(k+1)} \right)^T & 0 \end{bmatrix} \begin{bmatrix} \alpha_l^{(k+1)} \\ \beta_l^{(k+1)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}_l^{(k+1)} \\ 0 \end{bmatrix},$$

where components of the matrices and the right-hand side are defined as

$$A_{l,ii'}^{(k+1)} = \int_D \Big(\nu \nabla \boldsymbol{v}_{l,i'} : \nabla \boldsymbol{v}_{l,i} + \sigma \boldsymbol{v}_{l,i'} \cdot \nabla \boldsymbol{u}_l^{(k)}(\omega) \cdot \boldsymbol{v}_{l,i} + \boldsymbol{u}_l^{(k)}(\omega) \cdot \nabla \boldsymbol{v}_{l,i'} \cdot \boldsymbol{v}_{l,i} \Big) dx$$

$$\textit{for } i, i' = 1, .., N_l,$$

$$B_{l,ji}^{(k+1)} = -\int_D q_{l,j} (\nabla \cdot \boldsymbol{v}_{l,i}) d\boldsymbol{x} \qquad \textit{for } j = 1, .., P_l \textit{ and } i = 1, .., N_l.$$

$$\boldsymbol{f}_{l,i}^{(k+1)} = \int_D \left(\boldsymbol{f}(\omega) \cdot \boldsymbol{v}_{l,i} + \sigma \boldsymbol{u}_l^{(k)}(\omega) \cdot \nabla \boldsymbol{u}_l^{(k)}(\omega) \cdot \boldsymbol{v}_{l,i} \right) d\boldsymbol{x} \qquad \textit{for } i = 1, .., N_l,$$

Note that the coefficient matrix and the right-hand side depend on $\boldsymbol{u}_l^{(k)}$. Thus, a different linear system has to be solved at each iteration with any available solver. These linear solvers are not accelerated by our technique; instead, a good initial guess can reduce the number of nonlinear iterations.

The composite method converts to the Newton method when $\sigma = 1$ and to the simple iteration method when $\sigma = 0$. It is well known that the Newton method converges quadratically to the unique solution; however, contrary to linear systems, convergence is not guaranteed for arbitrary initial guesses. Moreover, it is found in practice that the radius of convergence is decreasing for increasing Reynolds number. This issue can be resolved by applying the simple iteration method first and then switching to the Newton method when residual of the solution becomes small enough. This approach improves the stability of the method, but it comes at a price of slower convergence. By choosing an appropriate initial guess, it is not only possible to reduce the number of iterations but also to improve the effective convergence rate of the composite technique if the initial guess is close to the basin of attraction of the Newton method.

To quantify the performance of the accelerated method, we have to choose an appropriate efficiency metric. For the MLMC method, it can be trivially defined as

$$C_{\mathbb{ML}} = \sum_{l=0}^{L} \sum_{m_l=1}^{M_l} C_l^{m_l},$$

where $C_l^{m_l}$ is the time cost of the corresponding deterministic problems. According to the usual implementation of the finite element method, this cost can be decomposed into two components:

- the cost C_l^{sys} of assembling the system of equations,
- the cost $C_l^S \sum_{m_l=1}^{M_l} J_l^{m_l}$ of an iterative solver, where C_l^S is the cost of one iteration and $J_l^{m_l}$ is the number of iterations at the level l corresponding to $\omega_{m_l} \in \Omega$.

Therefore, the cost function of the MLMC method can be written in the form

$$C_{\mathbb{ML}} = \sum_{l=0}^{L} \left(M_l C_l^{sys} + C_l^S \sum_{m_l=1}^{M_l} J_l^{m_l} \right) = \sum_{l=0}^{L} M_l \Big(C_l^{sys} + C_l^S J_l \Big) = \sum_{l=0}^{L} M_l C_l, \qquad \text{(III.9)}$$

where J_l and C_l are the average number of iterations and the average computational cost of the finite element method at the given level l.

The number of iterations required to achieve the desired accuracy of an iterative solver can be reduced by providing a good initial guess. This means that the cost gain from the proposed acceleration technique can be quantified as

$$\Delta C_{\mathbb{ML}} = \sum_{l=0}^{L} C_l^S \sum_{m_l=1}^{M_l} \left(J_l^{m_l} - \overset{a}{J}_l^{m_l} \right) - C^{predict}, \qquad \text{(III.10)}$$

where $\overset{a}{J}_l^{m_l}$ is the number of iterations of the accelerated method and $C^{predict}$ is the additional cost of applying Algorithm 1.

Obviously, the cost $C^{predict}$ must be small enough to make the computational savings possible, i.e., the cost gain (III.10) must be positive. The proposed acceleration technique satisfies this condition. The next section provides theoretical justification of this assertion.

III.3. Complexity analysis

III.3.1. Error component analysis

The accuracy of the MLMC estimator is controlled by two components: spatial discretization error and sampling error. However, there are a lot of other additional sources of errors due to the inexact computation of integrals, approximation of the boundary of the domain, approximation of the input data, etc. The finite tolerance of numerical linear solvers is one of such sources of error. To understand its implication for the approximation of $\mathbb{E}\left[u(x,\omega)\right]$, consider the following error estimate

$$\left\|\mathbb{E}_{\mathbb{ML}}\left[\tilde{u}_L\right]-\mathbb{E}\left[u\right]\right\|_{L^2(\Omega;\tilde{W}(D))} \tag{III.11}$$

$$\leq \underbrace{\left\|\mathbb{E}\left[u_L-u\right]\right\|_{\tilde{W}(D)}}_{\text{I := Discretization error}}+\underbrace{\left\|\mathbb{E}_{\mathbb{ML}}\left[u_L\right]-\mathbb{E}\left[u_L\right]\right\|_{L^2(\Omega)}}_{\text{II := Sampling error}}+\underbrace{\left\|\mathbb{E}_{\mathbb{ML}}\left[\tilde{u}_L-u_L\right]\right\|_{L^2(\Omega)}}_{\text{III := Solver error}},$$

where $\tilde{u}_L \in \tilde{W}(D)$ is the actually computed approximation of the true solution solution $u(x,\omega)$ and $\tilde{W}(D)$ is a suitable spatial function space. In the above expression, the first component is the spatial error, the second component is the sampling error and the third error component represents additional bias introduced by inexact computation of MC trajectories. To achieve the desired accuracy ϵ, it is sufficient to balance the total error between all three components in the following way

$$\left\|\mathbb{E}_{\mathbb{ML}}\left[\tilde{u}_L\right]-\mathbb{E}\left[u\right]\right\|_{L^2(\Omega;\tilde{W}(D))} \leq \epsilon_I+\epsilon_{II}+\epsilon_{III}=\epsilon. \tag{III.12}$$

If we apply an iterative solver to the linear systems arising after discretization, then the errors I, II and III can be equilibrated by the proper choice of the spatial discretization, the number of MC samples and the number of solver iterations respectively.

As it has been mentioned, in the following all the analysis will be conducted for the finite element discretization of the linear elliptic system from Examples III.1.1 and III.2.2 with the conjugate gradient (CG) method used as a linear solver. For this case, one can easily estimate the spatial, sampling and CG solver errors.

Spatial discretization error. Let the finite element spaces be defined as the function spaces of piecewise polynomials of degree at most p based on a triangulation τ_k of the spatial domain D with mesh spacing parameter $h_k := \max_{\tau\in\tau_k} diam(\tau)$. For elliptic equations with homogeneous Dirichlet boundary conditions, the classical result from the finite element approximation theory gives the estimate [Ste08, BR08]

$$\|u_L(x,\omega) - u(x,\omega)\|_{H^s(D)} \leq c_f h_L^{p+1-s} \|u(x,\omega)\|_{H^{p+1}(D)}, \tag{III.13}$$

which holds a.e. in Ω. Here $H^s(D)$ denotes the Sobolev space of functions with square-integrable derivatives up to the order s and $c_f > 0$ is a constant independent of ω and h_k. We will investigate L^2 norm of the error, i.e. $s = 0$ and

$$\tilde{W}(D) = H^0(D) \subset W(D).$$

It follows that

$$\begin{aligned}\|\mathbb{E}\left[u_L - u\right]\|^2_{L^2(D)} &= \int_D \left| \int_\Omega \Big(u_L - u\Big) d\mathbb{P}(\omega) \right|^2 dx \leq \int_D \int_\Omega \Big|u_L - u\Big|^2 d\mathbb{P}(\omega)\, dx \\ &= \int_\Omega \left(\int_D \Big|u_L - u\Big|^2 dx \right) d\mathbb{P}(\omega) = \mathbb{E}\left[\|u_L - u\|^2_{L^2(D)} \right].\end{aligned}$$

Combining with (III.12), this yields the estimate

$$I := \|\mathbb{E}\left[u_L - u\right]\|_{H^0(D)} \leq c_f h_L^{p+1} \|u\|_{L^2(\Omega;H^{p+1}(D))} = c_{fe} h_L^\alpha = \epsilon_I, \tag{III.14}$$

where $\alpha = p + 1$.

According to (I.21), the condition (III.14) will be satisfied if the number of levels is fixed to be

$$L = \left\lceil \log_q \left(h_0 (c_{fe} \epsilon_I^{-1})^{1/\alpha} \right) \right\rceil \leq 1 + \log_q \left(h_0 (c_{fe} \epsilon_I^{-1})^{1/\alpha} \right).$$

Sampling error. Let $V[u] = \mathbb{V}ar\left[u(x,\omega)\right]$ and $\overline{V}[u] = \int_D V[u] dx$. Then for any M and for any $u(x,\omega) \in L^2(\Omega; L^2(D))$, it holds that

$$\begin{aligned}
\left\| \mathbb{E}_{\mathbb{MC}}\left[u_L\right] - \mathbb{E}\left[u_L\right] \right\|^2_{L^2(\Omega)} &= \mathbb{E}\left[\left\| \mathbb{E}_{\mathbb{MC}}\left[u_L\right] - \mathbb{E}\left[u_L\right] \right\|^2_{L^2(D)} \right] \\
&= \frac{1}{M^2} \mathbb{E}\left[\int_D \left| \sum_{m=1}^{M} \left(u_L^m - \mathbb{E}\left[u_L\right] \right) \right|^2 dx \right] \\
&= \frac{1}{M^2} \int_D \sum_{m=1}^{M} \mathbb{V}ar\left[u_L^m\right] dx + \frac{1}{M^2} \int_D \sum_{m=1}^{M} \sum_{\substack{m'=1 \\ m \neq m'}}^{M} \mathbb{C}ov\Big(u_L^m, u_L^{m'}\Big) dx.
\end{aligned}$$

By virtue of the independence of i.i.d. samples u_L^m, we have that $\mathbb{C}ov\Big(u_L^m, u_L^{m'}\Big) = 0$ and

$$\left\| \mathbb{E}_{\mathbb{MC}}\left[u_L\right] - \mathbb{E}\left[u_L\right] \right\|_{L^2(\Omega; L^2(D))} = \sqrt{\frac{\overline{V}[u_L]}{M}}. \tag{III.15}$$

Based on this result, one gets the error of the MLMC estimator as follows

$$\begin{aligned}
\left\| \mathbb{E}_{\mathbb{ML}}\left[u_L\right] - \mathbb{E}\left[u_L\right] \right\|^2_{L^2(\Omega)} &= \left\| \sum_{l=0}^{L} \frac{1}{M_l} \sum_{m_l=1}^{M_l} \left[u_l^{m_l} - u_{l-1}^{m_l} \right] - \mathbb{E}\left[u_L\right] \right\|^2_{L^2(\Omega)} \\
&= \sum_{l=0}^{L} \left\| \frac{1}{M_l} \sum_{m_l=1}^{M_l} \left(\Delta_l^{m_l} - \mathbb{E}\left[\Delta_l\right] \right) \right\|^2_{L^2(\Omega)} \\
&= \sum_{l=0}^{L} \frac{\overline{V}_l}{M_l}.
\end{aligned}$$

Thus, the sampling error of the MLMC method can be estimated as

$$II^2 := \left\| \mathbb{E}_{\mathbb{ML}} \left[u_L \right] - \mathbb{E} \left[u_L \right] \right\|^2_{L^2(\Omega)} = \sum_{l=0}^{L} \frac{\overline{V}_l}{M_l} = \epsilon^2_{II}, \tag{III.16}$$

where $\overline{V}_l = \overline{V}[\Delta_l[$ and Δ_l is defined in (III.4).

According to (I.22), the condition (III.16) will be satisfied if the number of MC samples at each discretization level is set to

$$M_l = \epsilon_{II}^{-2} \left(\frac{\overline{V}_l}{C_l} \right)^{1/2} \sum_{k=0}^{L} (C_k \overline{V}_k)^{1/2}.$$

Linear solver error. Finally, consider the error of the CG iterative linear solver. To work with different spatial discretizations, we will require the following definition.

Definition III.3.1 ([BR08])**.** *Let $u(x) \in L^2(D)$. Consider the finite element space W_l equipped with the mesh-dependent inner product*

$$(v, w)_l = h_l^2 \sum_{i=1}^{n_l} v(p_i) w(p_i),$$

where $\{p_i\}_{i=1}^{n_l}$ is the set of internal vertices. The mesh-dependent norm induced by this inner-product is given by

$$\|v\|_l = h_l \|v\|_2 = h_l \left(\sum_{i=1}^{n_l} v^2(p_i) \right)^{1/2},$$

where $\|\cdot\|_2$ is the usual Euclidean norm.

Moreover, there exist positive constants c_1 and c such that

$$\frac{1}{c_1} \|v\|_{L^2(D)} \leq \|v\|_l \leq c \|v\|_{L^2(D)}, \qquad \forall v \in W_l.$$

Denote the error of the linear solver at the discretization level l as $e_l = \tilde{u}_l - u_l$, where

u_l is the true solution and $\tilde{u}_l$ is the actually computed solution. With the tolerance of the solver fixed, we may assume that $\|e_l\|_2 = \|e_{l-1}\|_2 = \|e_0\|_2$.

Using the definition (III.3.1), we arrive at

$$\begin{aligned}
&\|\mathbb{E}_{\mathbb{ML}}\left[\tilde{u}_L - u_L\right]\|^2_{L^2(\Omega;L^2(D))} = \sum_{l=0}^{L} \mathbb{E}\left[\|(\tilde{u}_l - \tilde{u}_{l-1}) - (u_l - u_{l-1})\|^2_{L^2(D)}\right] \\
&= \sum_{l=0}^{L} \mathbb{E}\left[\|e_l - e_{l-1}\|^2_{L^2(D)}\right] \leq \sum_{l=0}^{L} \left(\mathbb{E}\left[\|e_l\|^2_{L^2(D)} + \|e_{l-1}\|^2_{L^2(D)}\right]\right) \\
&\leq c_1 \sum_{l=0}^{L} \left(\mathbb{E}\left[\|e_l\|^2_l + \|e_{l-1}\|^2_{l-1}\right]\right) = c_1 \sum_{l=0}^{L} (h_l + h_{l-1})\mathbb{E}\left[\|e_l\|^2_2\right] \\
&= c_1(q+1)\sum_{l=0}^{L} h_l \mathbb{E}\left[\|e_l\|^2_2\right],
\end{aligned}$$

where q is the refinement parameter from (I.18).

The error of an iterative linear solver may be considered as a function of three parameters: the initial guess vector, the conditioning of the stiffness matrix and the number of iterations. For the CG solver, these parameters are related by the well-known estimate [Ste08, Saa03]

$$\|e\|_E \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^J \|\overset{a}{e}\|_E$$

which is computed in the energy norm $\|e\|_E = \sqrt{(Ae, e)}$, where A is the stiffness matrix with the condition number κ, $\overset{a}{e}_l = \overset{a}{u}_l - u_l$ is the initial error and J is the number of iterations.

Let λ_{min} and λ_{max} denote the minimum and maximum eigenvalues of the matrix A respectively. From the properties of Rayleigh quotients, we have

$$\sqrt{\lambda_{min}}\,\|e\|_2 \leq \|e\|_E \leq \sqrt{\lambda_{max}}\,\|e\|_2,$$

yielding in the following error estimate

$$\|e\|_2 \leq \frac{1}{\sqrt{\lambda_{min}}}\|e\|_E \leq \frac{2}{\sqrt{\lambda_{min}}}\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^J \|\overset{a}{e}\|_E \leq 2\sqrt{\kappa}\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^J \|\overset{a}{e}\|_2. \quad \text{(III.17)}$$

Now recall the (1,1)-Pade approximation of the exponential function

$$R_{1,1}(x) = \frac{1 + \frac{1}{2}x}{1 - \frac{1}{2}x}.$$

It can be shown that $R_{1,1}(-x) \leq e^{-x}$ which gives the useful bound

$$\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \leq \exp\left(-\frac{2}{\sqrt{\kappa}}\right).$$

Together with (III.17), this gives the estimate

$$\|e\|_2 \leq 2\sqrt{\kappa} \exp\left(-2J\left(\sqrt{\kappa}\right)^{-1}\right) \|\overset{a}{e}\|_2 . \tag{III.18}$$

It is well known that the spectral condition number of a stiffness matrix arising from the finite element discretization of elliptic problems has the following asymptotical behavior [Ste08]

$$1 \leq \kappa_l \leq c_2^2 \, h_l^{-2}. \tag{III.19}$$

Taking into account this result, the bound in (III.18) converts to

$$\|e_l\|_2 \leq 2c_2 h_l^{-1} \exp\left(-2c_2^{-1} h_l J_l\right) \|\overset{a}{e}_l\|_2 .$$

Finally, we obtain the following bound for the solver error

$$\begin{aligned} III^2 &:= \|\mathbb{E}_{\mathbb{ML}}\left[\tilde{u}_L - u_L\right]\|^2_{L^2(\Omega)} \\ &\leq 4(q+1)^2 c_1^2 \sum_{l=0}^{L} \mathbb{E}\left[c_2^2 \exp\left(-4c_2^{-1} h_l J_l\right) \|\overset{a}{e}_l\|_2^2\right] = \epsilon_{III}^2. \end{aligned} \tag{III.20}$$

The condition in (III.20) can be satisfied pathwise if the number of CG iterations at each

discretization level is given by

$$J_l = \max\left(\frac{c_2}{2h_l}\ln\frac{2c_1c_2(q+1)\left\|\overset{a}{e}_l\right\|_2}{b_l\epsilon_{III}}, 0\right), \quad \text{for } l = 0, .., L \qquad \text{and} \qquad \sum_{l=0}^{L} b_l = 1,$$

where coefficients b_l are the weights assigning certain part of the solver error to each level. In the best case, the choice of these coefficients should satisfy some optimality condition. However, we have already assumed that the tolerance of the solver is fixed and is the same at each level. Therefore, it is natural to set $b_l = 1/L$ which simplifies the above expression to

$$J_l = \max\left(\frac{c_2}{2h_l}\ln\frac{2c_1c_2(q+1)L\left\|\overset{a}{e}_l\right\|_2}{\epsilon_{III}}, 0\right), \qquad l = 0, .., L. \tag{III.21}$$

III.3.2. Complexity analysis of the standard algorithm

From the above error analysis, it follows that accuracy of the MLMC method is controlled by three parameters:

1. the number of levels $L = \left\lceil \log_q\left(h_0(c_{fe}\epsilon_I^{-1})^{1/\alpha}\right)\right\rceil$,
2. the number of samples per level $M_l = \left\lceil \epsilon_{II}^{-2}\left(\frac{\overline{V}_l}{C_l}\right)^{1/2}\sum_{k=0}^{L}(C_k\overline{V}_k)^{1/2}\right\rceil$ and
3. the number of CG iterations per sample $J_l = \left\lceil \frac{c_2}{2h_l}\ln\frac{2c_1c_2(q+1)L\left\|\overset{a}{e}_l\right\|_2}{\epsilon_{III}}\right\rceil$.

By analogy with (I.23), the optimal cost of the MLMC estimator is defined as

$$C_{\mathbb{ML}} = \sum_{l=0}^{L}\left\lceil M_l\right\rceil C_l \le \sum_{l=0}^{L} C_l + \epsilon_{II}^{-2}\left(\sum_{l=0}^{L}(C_l\overline{V}_l)^{1/2}\right)^2.$$

Efficient finite element codes assemble the stiffness matrix and the right-hand side of the system with a time complexity which is proportional to the number of degrees of freedom, i.e., $C_l^{sys} = O\left(h_l^{-d}\right)$, where d is the spatial dimension of the problem. Also,

the dominating operations in the CG algorithm are matrix-vector products which require $C_l^S = O\left(h_l^{-d}\right)$ operations per iteration since the stiffness matrices are sparse. Thus, the cost of the CG algorithm behaves asymptotically as $O\left(h_l^{-(d+1)}\left|\ln \epsilon_{III}\right|\right)$ because $J_l = O\left(h_l^{-1}\left(\left|\ln \epsilon_{III}\right| + \ln L\right)\right)$ and $L = O\left(\left|\ln \epsilon_I\right|\right)$. One may conclude that the cost of the CG linear solver is asymptotically dominant and the ϵ-cost of the finite element method at the discretization level l behaves as

$$C_l = C_l^{sys} + C_l^S J_l = O\left(h_l^{-\gamma}\left|\ln \epsilon_{III}\right|\right), \tag{III.22}$$

where $\gamma = d + 1$ and d is the spatial dimension.

According to Theorem I.3.2 and taking into account (III.22), the ϵ-cost of the MLMC method applied to the problem in (III.2) is given by

$$C_{\mathbb{ML}} \simeq \left|\ln \epsilon\right| \begin{cases} \epsilon^{-2} & \text{if } \gamma - \beta < 0, \\ \epsilon^{-2}\left|\ln \epsilon\right|^2 & \text{if } \gamma - \beta = 0, \\ \epsilon^{-2-\frac{\gamma-\beta}{\alpha}} & \text{if } \gamma - \beta > 0 \end{cases} \tag{III.23}$$

with $\gamma = d + 1$, where d is the spatial dimension of the problem.

Note that the common logarithm term in the above estimate appears due to the finite tolerance ϵ_{III} of the conjugate gradient method as can be seen from the cost estimate in (III.22).

III.3.3. Complexity analysis of the accelerated algorithm

Consider the cost of the MLMC estimator (III.9) using the zero initial guess for the CG solver

$$C_{\mathbb{ML}} = \sum_{l=0}^{L}\left(M_l C_l^{sys} + C_l^S \sum_{m_l=1}^{M_l} J_l^{m_l}\right)$$

$$= \sum_{l=0}^{L} \left(M_l C_l^{sys} + \frac{c_2 C_l^S}{2h_l} \sum_{m_l=1}^{M_l} \ln \frac{2c_1 c_2 (q+1) L \left\| u_l^{m_l} \right\|_2}{\epsilon_{III}} \right),$$

where $J_l^{m_l}$ was defined in (III.21) and $\left\| e_l^{m_l} \right\|_2 = \left\| u_l^{m_l} - 0 \right\|_2 = \left\| u_l^{m_l} \right\|_2$. Then the cost gain in (III.10) can be redefined as

$$\begin{aligned} \Delta C_{\mathbb{ML}} &= \sum_{l=0}^{L} C_l^S \sum_{m_l=1}^{M_l} \left(J_l^{m_l} - \overset{a}{J}_l^{m_l} \right) - C^{predict} \\ &= \frac{c_2}{2} \sum_{l=0}^{L} C_l^S h_l^{-1} \sum_{m_l=1}^{M_l} \ln \frac{\left\| u_l^{m_l} \right\|_2}{\left\| \overset{a}{e}_l^{m_l} \right\|_2} - C^{predict}, \end{aligned} \tag{III.24}$$

where the initial residual of the linear system $\left\| \overset{a}{e}_l^{m_l} \right\|_2$ is given by the error of the interpolant.

The magnitude of the error $\left\| \overset{a}{e}_l^{m_l} \right\|_2$ is determined by approximation properties of the corresponding interpolant. To simplify the analysis, we will assume that the input random fields allow finite dimensional representation as in Example III.2.1. In this case, the convergence analysis can be performed in the image space $\Gamma(\Omega) = \prod_{k=1}^{s} \Gamma^k \in \mathbb{R}^s$ of the random input vectors equipped with the usual Euclidian distance metric.

The approximation power of the Shepard's formula in (III.6) is not uniform and depends on the local distribution of interpolation nodes. To study the global approximation properties of the method, one needs the uniform distance measure in Γ. The classical choice of such measure for the data in $\Gamma \in \mathbb{R}^s$ is the fill distance

$$\delta_\Gamma = \sup_{y \in \Gamma} \min_{1 \le j \le n} \left\| y - y_j \right\|_2, \qquad y, y_j \in \Gamma,$$

which defines the radius of the largest data-free ball in Γ.

The convergence analysis of the Shepard approximation is complicated by the obvious fact that the fill distance δ_Γ is a function of random variables $y_i(\omega) \in \Gamma$. This issue can be easily resolved if one notes that the problem of function interpolation for a random data design can be considered in the framework of regression analysis. From this point of view,

the formulas (III.6) and (III.7) represent Nadaraya-Watson and k-NN smoothers respectively [Nad64, Gyo02].

It is well known that the theoretical time complexity of the Shepard's formula is $O(n\log(n))$ which is very inefficient compared to the $O(\log(n))$ complexity of the nearest neighbour estimate. Therefore, we will consider the nearest neighbour approximation only.

In order to guarantee the convergence, the regression function estimate in (III.7) must be consistent for the considered distribution of the training data $y_i\big|_{i=1}^{n} \in \Gamma$. Stone [Sto77] gave the conditions on the weights which provide the universal consistency of the local averaging estimates and showed that the k-NN estimate is universally consistent in $\mathbb{R}^s$ provided $k \to \infty$ and $k/n \to 0$.

The following theorem establishes the convergence rate of the $k-NN$ estimate.

Theorem III.3.1 ([Gyo02])**.** *Assume that Γ is bounded and function $u_l(y)$ is Lipschitz continuous in Γ, i.e.,*

$$\|u_l(y) - u_l(z)\|_2 \leq C\,\|y - z\|_2\,, \qquad y, z \in \Gamma.$$

Let $\overset{a}{u}$ denote the k-NN estimate and assume that available data is noiseless. Then

$$\mathbb{E}\left[\|\overset{a}{e}_l\|_2^2\right] = \mathbb{E}\left[\|\overset{a}{u}_l - u_l\|_2^2\right] \leq c_3^2\left(\frac{k_{NN}}{n}\right)^{2/s},$$

where n is the number of training points and c_3 is some positive constant.

It is interesting to note that the convergence rate of the Shepard approximation formula cannot exceed $n^{1/s}$ [Far86]. Thus, the nearest neighbor estimator represents an optimal stable interpolation operator, insofar as the approximation power of the estimator is concerned.

Theorem III.3.2. *Consider the finite element approximation of the linear elliptic system in* (III.2) *as in Example III.2.2. The expected ϵ-cost of the accelerated MLMC method in*

Algorithm 1 is given by

$$\mathbb{E}\left[C_{\mathbb{ML}}^{NN}\right] \simeq \left(1 - \frac{1}{s}\right) C_{\mathbb{ML}},$$

where s is the dimension of the stochastic parameter space and $C_{\mathbb{ML}}$ is the computational cost of the standard MLMC method in (III.23)*.*

Proof. Assume that each iteration of the CG solver has $C_l^S \simeq h_l^{-d}$ complexity and consider expectation of the cost gain (III.10)

$$\begin{aligned}\mathbb{E}\left[\Delta C_{\mathbb{ML}}\right] &= \mathbb{E}\left[\frac{c_2}{2}\sum_{l=0}^{L} C_l^S h_l^{-1} \sum_{m_l=1}^{M_l} \ln \frac{\|u_l^{m_l}\|_2}{\|\overset{a}{e}_l^{m_l}\|_2} - C^{predict}\right] \\ &\simeq \sum_{l=0}^{L} h_l^{-(d+1)} \sum_{m_l=1}^{M_l} \ln \mathbb{E}\left[\|\overset{a}{e}_l^{m_l}\|_2^2\right]^{-1} - \mathbb{E}\left[C^{predict}\right].\end{aligned}$$

Using the result of Theorem III.3.1 and the average complexity of the nearest neighbor search $\mathbb{E}\left[C^{predict}\right] \simeq \ln n$, we obtain

$$\mathbb{E}\left[\Delta C_{\mathbb{ML}}\right] \simeq \sum_{l=0}^{L} h_l^{-(d+1)} \sum_{m_l=1}^{M_l} \ln n_{l,m_l}^{1/s} - \sum_{l=0}^{L}\sum_{m_l=1}^{M_l} \ln n_{l,m_l},$$

where $n_{l,m_l} = \sum_{p=0}^{l-1} M_p + m_l - 1$ denotes the number of points in the dataset for m_l-th sample at level l.

The number of samples at each level have the following asymptotic behavior

$$M_l = \frac{\overline{V}_l}{a_l \epsilon_{II}^2} = \epsilon_{II}^{-2}\left(\sum_{k=0}^{L}(C_k \overline{V}_k)^{1/2}\right)\left({C_l}^{-1}\overline{V}_l\right)^{1/2} \simeq \epsilon_{II}^{-2}\theta(\gamma,\beta) h_l^{\frac{\gamma+\beta}{2}}, \tag{III.25}$$

where $\theta(\gamma,\beta)$ is defined in (I.24) and the cost C_l of the finite element method is given in (III.22).

From the Stirling's formula, we have for $l = 0$

$$\sum_{m_0=1}^{M_0} \ln(m_0) = \ln(M_0!) = M_0\Big(\ln M_0 - 1\Big) + O\Big(\ln M_0\Big) \simeq M_0 \ln\left(\epsilon_{II}^{-2}\theta(\gamma,\beta)\right).$$

Similarly, we obtain for $l \geq 1$

$$\begin{aligned}\ln n_{l,m_l} &= \ln\left(\sum_{k=0}^{l-1} M_k + m_l - 1\right) = \ln\left(\sum_{k=0}^{l-1} M_k\right) + \ln\left(1 + \frac{m_l - 1}{\sum_{k=0}^{l-1} M_k}\right) \\ &\simeq \ln\left(\sum_{k=0}^{l-1} M_k\right) \simeq \ln\left(\epsilon_{II}^{-2}\theta(\gamma,\beta)\right) + \underbrace{\ln \sum_{k=0}^{l-1} h_k^{-\frac{\gamma+\beta}{2}}}_{O(1)} \simeq \ln\left(\epsilon_{II}^{-2}\theta(\gamma,\beta)\right).\end{aligned}$$

Therefore, we get

$$\sum_{m_l=1}^{M_l} \ln n_{l,m_l} \simeq M_l \ln\left(\epsilon_{II}^{-2}\theta(\gamma,\beta)\right) \qquad \text{for } l = 0,..,L.$$

Together with (III.25), this provides the following estimate

$$\begin{aligned}\mathbb{E}\left[\Delta C_{\mathbb{ML}}\right] &\simeq \ln\left(\epsilon_{II}^{-2}\theta(\gamma,\beta)\right) \sum_{l=0}^{L} \left[M_l\left(\frac{1}{s}h_l^{-(d+1)} - 1\right)\right] \\ &\simeq \epsilon_{II}^{-2}\theta(\gamma,\beta) \ln\left(\epsilon_{II}^{-2}\theta(\gamma,\beta)\right) \sum_{l=0}^{L} \left[h_l^{\frac{\gamma+\beta}{2}}\left(\frac{1}{s}h_l^{-\gamma} - 1\right)\right] \\ &\simeq \epsilon_{II}^{-2}\theta(\gamma,\beta) \ln\left(\epsilon_{II}^{-2}\theta(\gamma,\beta)\right)\left(\frac{\theta(\gamma,\beta)}{s} - O(1)\right) \\ &\simeq \frac{1}{s}\epsilon_{II}^{-2}\theta^2(\gamma,\beta) \ln\left(\epsilon_{II}^{-2}\theta(\gamma,\beta)\right) \simeq \frac{1}{s}\epsilon_{II}^{-2}\theta^2(\gamma,\beta)\left|\ln\epsilon\right|,\end{aligned}$$

where the last relation holds because $\ln\left(\epsilon^{-2}\left|\ln\epsilon\right|\right) = O(\left|\ln\epsilon\right|)$

Finally, we get from (I.24) that

$$\mathbb{E}\left[\Delta C_{\mathbb{ML}}\right] \simeq \frac{1}{s}\left|\ln \epsilon\right| \begin{cases} \epsilon^{-2} & \text{if } \gamma - \beta < 0, \\ \epsilon^{-2}\left|\ln \epsilon\right|^2 & \text{if } \gamma - \beta = 0, \\ \epsilon^{-2-\frac{\gamma-\beta}{\alpha}} & \text{if } \gamma - \beta > 0 \end{cases}$$

and the result follows from (III.23). □

III.4. Numerical results

This section provides a numerical illustration of the proposed acceleration technique. We consider three test problems with different sources of uncertainty. However, the geometry of the domain and the properties of the input random fields are shared by all test cases in order to simplify the comparison and analysis of results. In particular, the problems are set in a square domain $D = [-1, 1]^2$ and we use homogeneous Dirichlet boundary conditions for all components of the solution.

Uncertainty in the input data is modeled using the random field $a(X, \omega)$ with the separable exponential covariance

$$\mathbb{C}ov\big(a\big)(x, z) = \sigma^2 \exp\left(-\frac{|x_1 - z_1| + |x_2 - z_2|}{L_c}\right),$$

where L_c is the correlation length and σ is the variance. We set $\sigma = 0.2$ in all examples below. Eigenfunctions of the above covariance kernel are given by $\phi_n(x) = \phi_i^1(x_1)\phi_k^2(x_2)$ where we have for $m = \{1, 2\}$

$$\phi_i^m(x_m) = \begin{cases} \dfrac{\sin\left(\omega_i x\right)}{\sqrt{1 - \sin(2\omega_i)/(2\omega_i)}} & \text{if } i \text{ even}, \\ \dfrac{\cos\left(\omega_i x\right)}{\sqrt{1 + \sin(2\omega_i)/(2\omega_i)}} & \text{if } i \text{ odd}. \end{cases}$$

The corresponding eigenvalues are given by $\lambda_n = \lambda_i^1 \lambda_k^2$ with $\lambda_i^m = \dfrac{2L_c}{1 + L_c^2 \psi_i}$ and values ψ_i are determined from the following transcendental equations

$$\begin{cases} \psi_i + L_c^{-1} \tan(\psi_i) = 0 & \text{if } i \text{ even}, \\ L_c^{-1} - \psi_i \tan(\psi_i) = 0 & \text{if } i \text{ odd}. \end{cases}$$

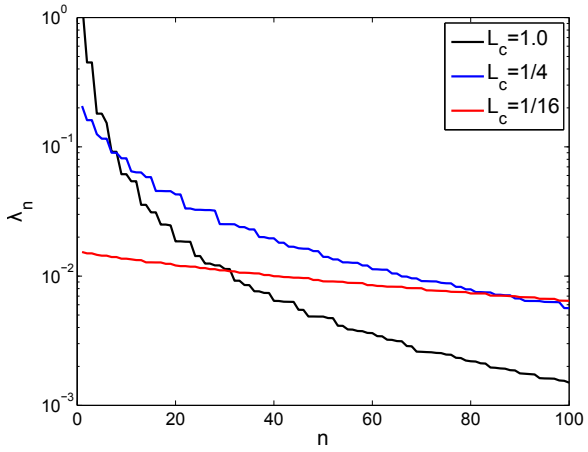


Figure 13: Eigenvalues of the two-dimensional exponential covariance.

Given the set of eigenfunctions of the covariance kernel, we use the Karhunen-Loeve approximation of the random field

$$a(x, \omega) \approx 0.5 + \sum_{n=1}^{s} \sqrt{\lambda_n} y_n(\omega) \phi_n(x), \tag{III.26}$$

where s represents the stochastic dimension of the problem, eigenvalues λ_n are ordered in the descending order and random variables $y_n(\omega)$ are independent, have zero mean, unit variance and are uniformly distributed in the interval $[-\sqrt{3}, \sqrt{3}]$.

The regularity of the input data and the required tolerance of the solution are the main parameters which control the computational cost of the multilevel Monte Carlo method. Regularity determines the stochastic dimensionality of the problem which affects

the effectiveness of the proposed acceleration technique. The impact of the tolerance on the computational cost is obvious. Smaller tolerance results in higher cost and smaller relative savings. To understand the actual influence of these parameters on the performance of the algorithm, we have to carry out a series of numerical tests with different correlation lengths, stochastic dimensions, spatial discretizations and tolerances.

Example 1. Linear elliptic equation. In the above, we used the linear elliptic equation as the benchmark problem for our analysis. It is natural to consider the same problem for the numerical support of the obtained theoretical results. For this purpose, we will study two different examples.

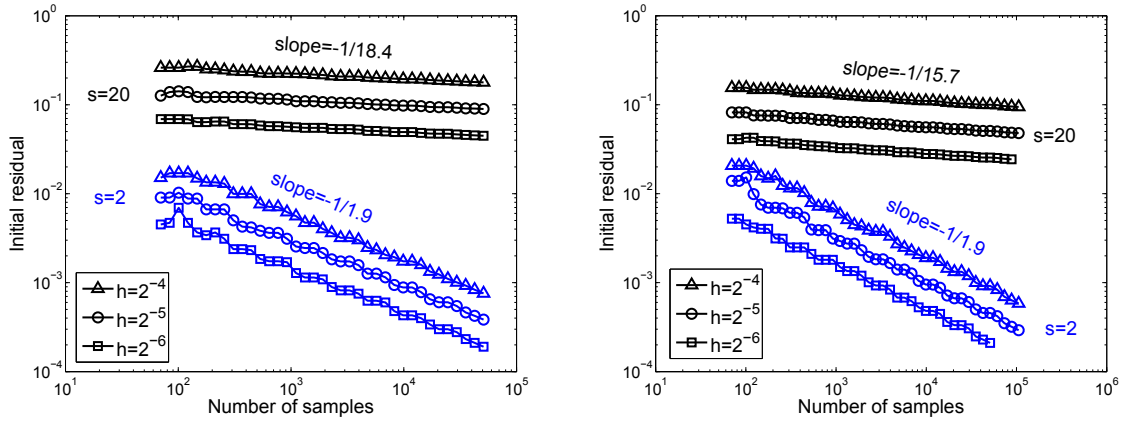
The first example represents the elliptic system with the lognormal random coefficient and the deterministic forcing term

$$\begin{cases} -\nabla\Big(\exp\Big(a(x,\omega)\Big)\nabla u(x,\omega)\Big) &= 1 \qquad \text{in } D\times\Omega, \\ u(x,\omega) &= 0 \qquad \text{on } \partial D\times\Omega \end{cases} \tag{III.27}$$

while for the second example, the coefficient is deterministic and the right-hand side is random

$$\begin{cases} -\nabla^2 u(x,\omega) &= a(x,\omega) \qquad \text{in } D\times\Omega, \\ u(x,\omega) &= 0 \qquad \text{on } \partial D\times\Omega. \end{cases} \tag{III.28}$$

First of all, we have to determine the accuracy of the nearest neighbor estimator and test out its convergence rate. Figure 14 illustrates the convergence of the Euclidean norm of the residual for the NN approximation of initial guesses for selected spatial discretizations and stochastic dimensions. The correlation length of the random field and the tolerance of the solution are set to $L_c = 1/4$ and $\epsilon = 10^{-5}$ respectively. One may observe that the convergence is linear in the log-log scale and the rate of convergence is independent of



(A) Example 1. Random diffusion coefficient. (B) Example 2. Random forcing term.

Figure 14: Average initial residual of the CG linear solver for the problems (III.27) (left) and (III.28) (right) with the nearest neighbor prediction of true solutions.

Table 3: Computed convergence rates of initial residuals for $h = 0.2L_c$.

L_c	$1/s$					
	0.33	0.20	0.14	0.10	0.07	0.05
1	0.34	0.21	0.16	0.12	0.09	0.07
$1/4$	0.34	0.21	0.16	0.11	0.08	0.06
$1/16$	0.43	0.27	0.21	0.14	0.10	0.08

the spatial discretization. Moreover, the slopes of the lines are proportional to $1/s$ which is the theoretical rate from Theorem III.3.1. The actual convergence rates computed for different correlation lengths and stochastic dimensions are given in Table 3. It is clear that the obtained results confirm the theory.

Exponential decay of initial residual suggests that the number of CG iterations is a logarithmically decreasing function of the number of Monte Carlo samples. Indeed, Figure 15 illustrates that decay of iterations is linear in logarithmic scale with the rate which is inversely proportional to the dimension of parameter space. Evidently, the observed number of iterations and their decay rates increase for decreasing mesh size which empirically shows

Table 4: Total relative savings of iterations, %

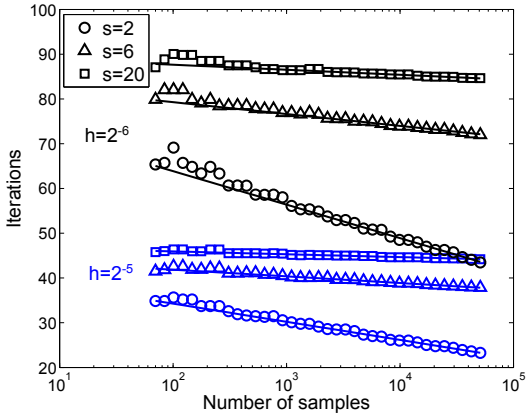
s		3	5	7	10	15	20
$100/s$		33.3	20.0	14.3	10.0	6.7	5.0
$L_c = 1$	$\epsilon = 10^{-4}$	32.4	20.5	15.7	11.5	9.9	9.4
	$\epsilon = 10^{-5}$	24.9	15.3	11.8	8.6	6.7	6.4
	$\epsilon = 10^{-6}$	19.6	12.2	8.7	7.3	5.5	4.3
$L_c = 1/4$	$\epsilon = 10^{-4}$	36.9	20.4	14.4	10.6	7.1	5.9
	$\epsilon = 10^{-5}$	30.6	17.2	12.9	8.3	5.8	4.9
	$\epsilon = 10^{-6}$	18.7	11.6	8.8	6.6	4.5	3.8
$L_c = 1/16$	$\epsilon = 10^{-4}$	51.0	36.8	26.1	19.2	13.9	10.4
	$\epsilon = 10^{-5}$	48.4	26.7	18.7	13.4	9.8	6.8
	$\epsilon = 10^{-6}$	37.4	22.0	15.1	10.6	6.8	4.6

the correctness of the formula (III.21). At the same time, Figure 16 shows that relative savings of iterations are insensitive to the level of discretization which leads to the result of the theorem III.3.2 since we have

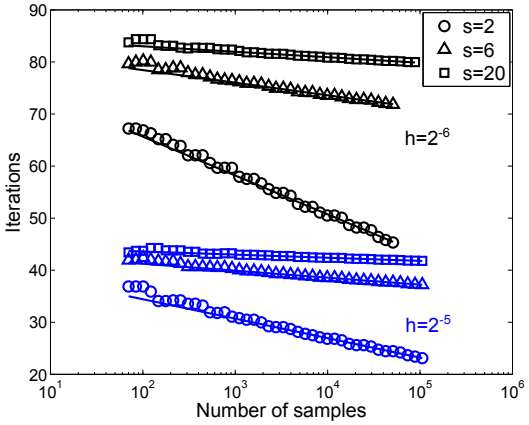
$$\frac{\Delta C_{\mathbb{ML}}}{C_{\mathbb{ML}}} = \frac{\Delta C_0 + \Delta C_1 + ... + \Delta C_L}{C_0 + C_1 + ... + C_L} \lesssim \frac{s^{-1}C_0 + s^{-1}C_1 + ... + s^{-1}C_L}{C_0 + C_1 + ... + C_L}$$

and $\Delta C_{\mathbb{ML}} \simeq \frac{1}{s} C_{\mathbb{ML}}$.

One may predict that relative computational savings are better for higher tolerances of the CG solver since $\Delta C_{\mathbb{ML}}$ does not depend on the tolerance and $C_{\mathbb{ML}}$ is bigger for smaller values of ϵ_{III} as follows from the formula (III.24). Table 4 provides empirical verification of this prediction; it contains total relative savings of iterations estimated for the problem with the random coefficient for different tolerances and correlation lengths. Additionally, the dependence of the savings on the stochastic dimension of the problem is depicted graphically in Figure 17. These empirical results show that, for the given input data and the fixed tolerance of the solution, savings of the proposed acceleration technique are

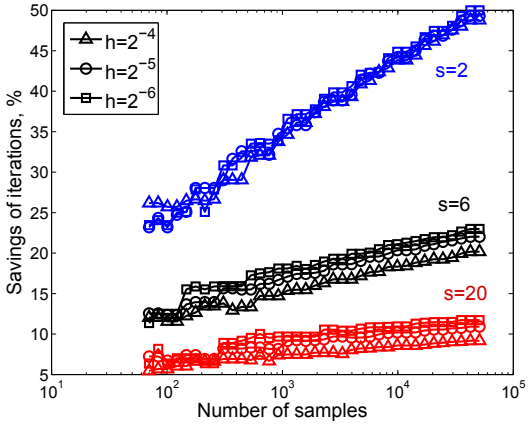


(A) Example 1. Random diffusion coefficient.

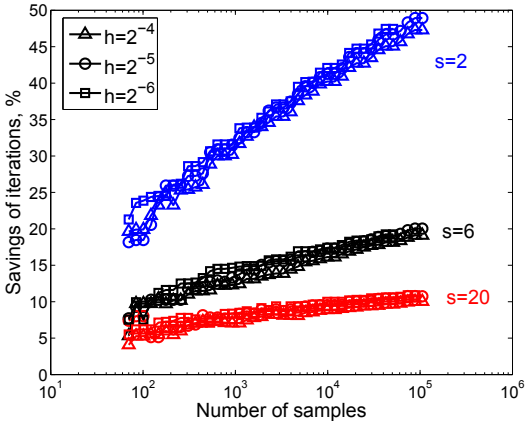


(B) Example 2. Random forcing term.

Figure 15: Average number of iterations of the CG solver for different spatial discretizations and stochastic dimensions and fixed tolerance $\epsilon = 10^{-5}$.



(A) Example 1. Random diffusion coefficient.



(B) Example 2. Random forcing term.

Figure 16: Average relative savings of iterations of the CG solver for different stochastic dimensions and spatial discretizations and fixed tolerance $\epsilon = 10^{-5}$.

controlled by the approximation power of the nearest neighbor estimator which is inversely proportional to the dimension of the parameter space. This, in fact, shows the correctness of Theorem III.3.2.

Finally, Figure 18 compares the gain in the CPU cost from the nearest neighbor approximation of initial guesses to the actual cost of making these predictions. Clearly, the cost of the NN estimator is smaller than the obtained computational savings by several orders of magnitude which fully justifies the proposed acceleration technique.

Example 2. Stationary Navier-Stokes system. The goal of this numerical example is to demonstrate that the proposed acceleration method can be successfully applied to problems other than linear elliptic equations. In particular, here we consider the stationary Navier-Stokes system

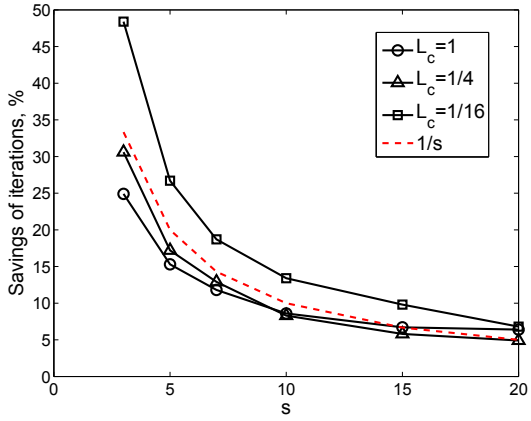
$$\begin{cases} -\nu\nabla^2\boldsymbol{u} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} + \nabla p &= \mathbf{f}(x,\omega_f) \qquad \text{in } D\times\Omega, \\ \nabla\cdot\boldsymbol{u} &= 0 \\ \boldsymbol{u} &= 0 \qquad\qquad \text{on } \partial D\times\Omega, \end{cases} \tag{III.29}$$

which is an example of the nonlinear saddle point problem.

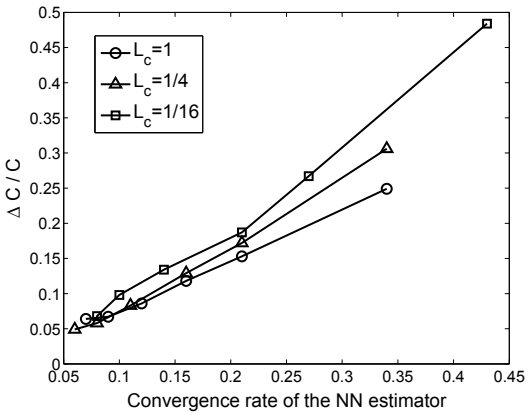
For the homogeneous boundary conditions, the flow is generated only by the random forcing term $\mathbf{f}(x,\omega_f)$ which must induce no flow through the boundary in order to satisfy conservation of mass. This can be achieved if the force function is given by a divergence free vector field. By Helmholtz decomposition of vector fields, this suggests that $\mathbf{f}$ is represented as a curl of a vector potential which gives the following expression in two dimensions

$$\mathbf{f} = \nabla\times\Big(0,0,\psi\Big)^T = \Big(0,0,\frac{\partial\psi}{\partial y} - \frac{\partial\psi}{\partial x}\Big)^T$$

for some scalar function $\psi(x)$ so that $\nabla\times\mathbf{f} = \Big(0,0,-\nabla^2\psi\Big)^T$. This leads to the definition

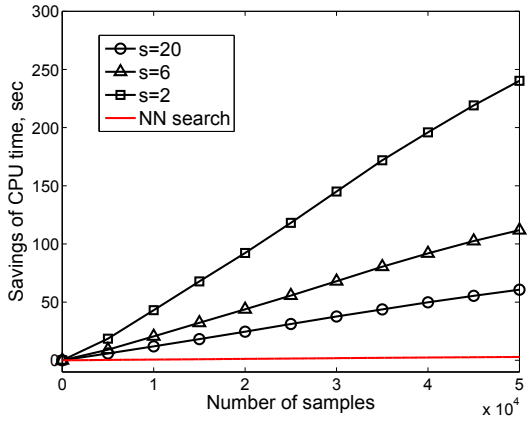


(A) Savings vs stochastic dimension

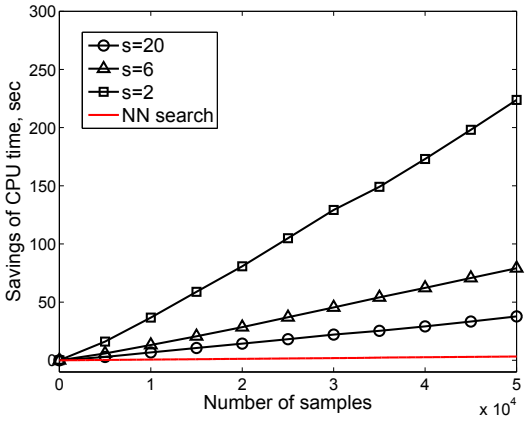


(B) Savings vs convergence rates from Table 3

Figure 17: Total savings of iterations for different correlation lengths and tolerance $\epsilon = 10^{-5}$.



(A) Example 1. Random diffusion coefficient.



(B) Example 2. Random forcing term.

Figure 18: Average savings of CPU time for different stochastic dimensions and fixed spatial discretization $h = 2^{-5}$ and tolerance $\epsilon = 10^{-5}$.

of $\psi(x)$ by

$$\begin{cases} \nabla\psi(x,\omega) &= -a(x,\omega) \qquad \text{in } D\times\Omega, \\ \psi(x,\omega) &= 0 \qquad\qquad\quad \text{on } \partial D\times\Omega. \end{cases}$$

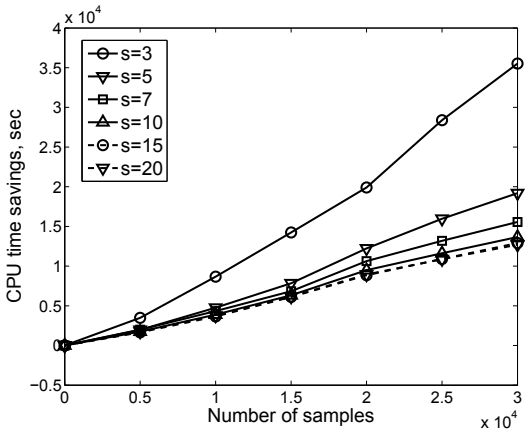
It is convenient to define the inverse operator $\mathcal{L} : \mathbf{H}^{-1}(\Omega) \rightarrow \mathbf{H}^1_0(\Omega)$ which maps the rotational $a(x,\omega)$ of the force term $\mathbf{f}(x,\omega)$ to the "stream function" $\psi(x,\omega)$. Then the forcing function takes the form

$$\mathbf{f}(x,\omega) \approx \mathbf{f}_0 + \sum_{n=1}^{s} \sqrt{\lambda_n} y_n(\omega)\mathbf{f}_n, \qquad \mathbf{f}_i(x) = \nabla \times \Big(0, 0, \mathcal{L}[\phi_i(x)]\Big)^T,$$

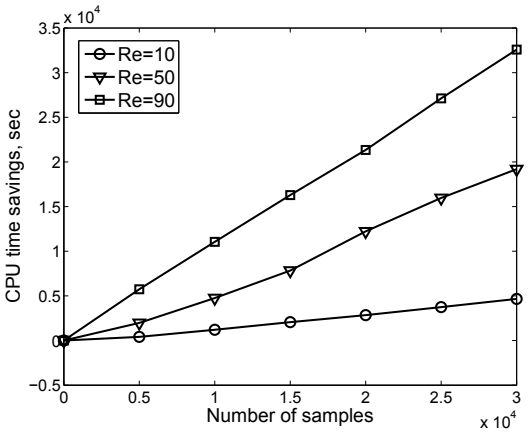
where $\phi_i(x)$ are the eigenfunctions of the exponential covariance operator which form the basis of the KL expansion in (III.26).

It is important to note that correlation length of the input random field cannot be arbitrary because it is related to the Reynolds number. It is physically unlikely that highly irregular and noisy data will produce the flow with low Reynolds number which is required to guarantee the existence of the stationary solution. Therefore, we choose $L_c = 1$.

Nonlinearity of the problem is resolved by the composite technique which combines simple iteration and Newton methods [Gun89]. To illustrate the effectiveness of the acceleration technique, we have accomplished a series of numerical simulations for different dimensions of the input random field and selected Reynolds numbers. The obtained results are presented graphically in Figure 19. It is seen that the nearest neighbor acceleration technique results in high computational savings especially for convective dominated flows where nonlinearity plays crucial role. Additionally, Figure 20 illustrates computational savings as a function of the stochastic dimension. One can observe that relative savings have similar behavior to that of the linear elliptic system.



(A) Different stochastic dimensions and fixed Reynolds number $Re = 50$.



(B) Different Reynolds numbers and fixed stochastic dimension $s = 7$.

Figure 19: Cumulative CPU time savings of the accelerated MLMC method for the problem (III.29). Tolerance of the solution is set to $\epsilon = 10^{-5}$.

s	$Re = 10$	$Re = 50$	$Re = 90$
3	43.4	66.9	69.2
5	22.6	37.6	40.4
7	14.1	29.9	29.9
10	10.5	25.2	28.1
15	8.4	23.3	27.1
20	7.8	23.0	25.5

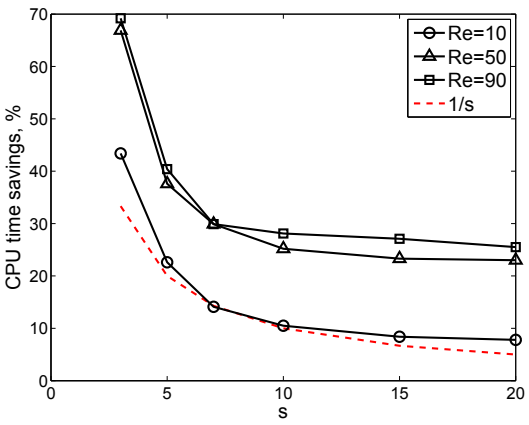


Figure 20: Total relative CPU time savings of the accelerated MLMC method for the problem (III.29). Tolerance of the solution is set to $\epsilon = 10^{-5}$.

CHAPTER IV

PARTIAL DIFFERENTIAL EQUATIONS IN RANDOM DOMAINS

Problems with topological uncertainties appear in many applied fields ranging from nano-device engineering and analysis of micro electromechanical systems (MEMS) [AA07, ZWD04] to design of bridges [CF09]. Other applications include flow over rough surfaces [TX06, ZPTK13], surface imaging [Tso05], corrosion or wear of surfaces, homogenization of random heterogeneous media [SSPD14] and even modelling of blood flow [PIT12].

Despite its practical importance, relatively few results for partial differential equations in random domains are available in the literature. The existing numerical techniques can be fitted into several categories based on the method used for modeling and quantification of uncertainty of random domain representations. Perturbation techniques form the first category of methods. They describe random geometry using Eulerian coordinates in the sense that points in the interior of the domain are held fixed and geometric uncertainty is modeled by random perturbation of the boundary. In this approach, one starts with specification of the deterministic reference boundary ∂D_{ref} and the random perturbation field $\chi(\omega)$ acting on it

$$\chi(x,\omega) : \partial D_{ref} \to \partial D(\omega), \qquad x \in \partial D_{ref}.$$

Since uncertainty in the problem propagates only through stochastic perturbation of the reference boundary, this approach naturally leads to boundary integral representation of solutions to boundary value problems (BVP) under consideration. The key point in this approach is that boundary integral equations (BIE) are formulated on the unperturbed nominal boundary and variation of random geometry is quantified using methods of the "shape calculus". For instance in [Hon05], the "shape Taylor expansion" was applied to unknown densities of integral equations of the classical boundary element method (BEM) and the solution of the BVP was constructed in the form of a polynomial chaos expansion.

The resulting system was then projected on a homogeneous chaos space giving a single matrix equation for coefficients of this expansion. Alternatively, in [HSS08], the solution of the problem itself was represented in the form of the asymptotic shape Taylor series. Using second moment analysis, the mean and the variance of this representation were then determined as solutions of auxiliary deterministic boundary values problems.

The second group of methods utilizes Lagrangian description of random geometries, i.e., stochastic perturbation is applied to every point in the physical domain. This requires explicit knowledge of the random mapping function

$$\chi(x,\omega) : D_{ref} \to D(\omega), \qquad x \in D_{ref}$$

which is usually defined by specifying a random displacement field $g(x,\omega)$ relating the points in the reference domain to the points in each realization of the random domain according to the rule

$$\xi(\omega) = x + g(x,\omega), \qquad x \in D_{ref}, \quad \xi \in D(\omega).$$

When such transformation is established, there are two ways to put it into computational framework. Firstly, via the chain rule, it is possible to transform the original problem in a random domain to a problem with random coefficients posed on a deterministic reference domain. Since uncertainty usually enters the problem through the random perturbation of the boundary, this approach requires solution of a series of auxiliary partial differential equations to extend the random displacement mapping to the whole domain. The resulting problem with random coefficients can then be attacked numerically with available conventional solvers. This approach was developed in [XT06] in conjunction with stochastic Galerkin method and then applied in [TX06, KAF14, HPS14] both to stationary and to time dependent PDEs. Additionally, the domain mapping method followed by the stochastic collocation approximation in random space was considered in [CC14, CCNT15]. An equiva-

lent Lagrangian approach was also proposed in [AA07], where the mapping to the reference domain was applied together with stochastic spectral boundary element approximation of the underlying problem. In this case, however, extension of the random perturbation field to the whole domain was not required due to boundary integral formulation of the problem.

Contrary to the domain mapping method which generates a new problem with a different governing stochastic operator, one can apply the random displacement field directly to the mesh-based representation of random geometries. In this case, deformations are employed on the reference mesh producing the new mesh with random coordinates of nodes but the same fixed connectivity. This procedure does not change the underlying equations enabling reusability of existing deterministic solvers. This idea was proposed in [MNK11] where it was combined with a polynomial chaos approximation in random space. Later it was also studied in [HPS14] in the context of the quasi-Monte Carlo method applied to problems posed on domains with random interfaces.

Another group of numerical techniques which can be found in the literature is presented by fictitious domain methods. Within this approach, the problem is formulated on a larger deterministic domain containing all realizations of a random boundary. The attractive feature of these methods is that the enclosing domain can be chosen arbitrarily allowing for simple discretizations such as, for example, tensor product grids which do not have to conform with random boundaries or interfaces inside the reference domain. However, in order to enforce the boundary conditions of the original problem, additional variables must be added to the new problem formulation on a fictitious domain. A possible way to achieve this is by introducing a Lagrange multiplier acting on a random boundary and transforming the original problem into a new saddle-point problem. For instance, this technique was used in [CK07] together with spectral stochastic Galerkin method. It was also mentioned in [NSM07, NCSM08] in the context of the extended stochastic finite element method (X-SFEM). The X-SFEM was also applied in [SSPD14, LDM13, NC10] to problems with

random interfaces by enriching the traditional finite element basis with a suitably constructed enrichment functions to capture the solution discontinuity at the material interface. In X-SFEM, the a-priory information about location of random boundaries is incorporated into the problem implicitly via the level-set representation of geometries. This technique was also applied in [NCS11] where the solution was constructed with the Proper Generalized Decomposition method using the separated representation of the indicator function of the uncertain domain defined in terms of the level-set function. This approach, however, was limited only to Neumann boundary conditions imposed on the random boundary.

The methods described above have proven their efficiency for solving problems on random domains. Nevertheless, they have certain limitations. For example, the main drawback of the perturbation techniques, shared by all methods using Eulerian description, is that they are applicable only to small random perturbations. The methods which use Lagrangian description do not have this issue but they propagate uncertainty from the boundary to the whole domain. This feature is not desirable when random displacement field is applied only to small part of the boundary. In fictitious domain methods, this issue is resolved via boundary supported Lagrange multiplier. In the resulting discrete saddle-point problem, the information on the geometry of the random domain is encoded only in the matrix coupling the primal variable and the Lagrange multiplier which gives significant computational savings.

In this chapter, we propose the novel computational framework based on the Green's function formulation of the boundary integral equation method. It is conceptually similar to the family of fictitious domain methods, however, no fictitious boundary is required. Our motivation for this approach is clear. Firstly, as it was mentioned above, topological uncertainties usually enter the boundary value problems through random perturbations of the boundary making the BIE method the natural choice for building approximate solutions. Additionally, in many problems, only certain (often relatively small) part of the boundary is

subjected to random perturbations. This significantly reduces efficiency of fully discrete formulations of conventional solvers due to the necessity in discretization of the whole physical domain. In this regard, semi-analitical approximations hold a vast potential. Namely, we propose to construct solutions in the form of potentials with suitable Green's functions as kernels. As will be shown later, this allows to reformulate the problem as an integral equation on the random part of the domain only. The first step towards the practical application of this approach was done in [Mel77] where the so-called method of "modified potentials" was introduced. Later it was successfully applied to various stationary and time-dependent problems [MR14, MHM96]. It is also worth noting that importance of Green's functions has been recognized in various uncertainty quantification and uncertainty reduction techniques [CET15, BST13, NO93, NTWW96, MS96]. Here we apply the multilevel Monte Carlo (MLMC) method for statistical discretization. However, any method of collocation type can be trivially adopted to the proposed numerical technique.

IV.1. Problem setting

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space with a set of outcomes Ω, a sigma algebra of events $\mathcal{F}$ and a probability measure $\mathbb{P}$ defined on it. For each outcome $\omega \in \Omega$, define $D(\omega)$ to be a realization of a random domain with a boundary comprised of deterministic and random parts $\partial D(\omega) := \partial D_1 \cup \partial D_2(\omega)$. We are concerned with solutions of the following boundary value problem

$$\begin{aligned} Lu(x,\omega) &= f(x) && \text{for } x \in D(\omega), \\ B_1 u(x,\omega) &= b_1(x) && \text{for } x \in \partial D_1, \\ B_2 u(x,\omega) &= b_2(x) && \text{for } x \in \partial D_2(\omega), \end{aligned}$$

where L is a linear partial differential operator subjected to boundary conditions determined by linear operators B_1 and B_2. We require $u(x,\omega)$ to be a Bochner integrable function with values in some Banach space $W(D)$, i.e., $u(x,\omega) \in L^p(\Omega; W(D))$, the function space given by

$$L^p\Big(\Omega; W(D)\Big) := \left\{ u : \Omega \to W(D) \;\middle|\; u \text{ is strongly measurable and } \|u\|_{L^p(\Omega;W(D))} < \infty \right\}$$

with the corresponding norm

$$\|u\|^p_{L^p(\Omega;W(D))} = \begin{cases} \displaystyle\int_\Omega \|u(\cdot,\omega)\|^p_{W(D)}\, d\mathsf{P}(\omega) & \text{if } 0 < p < \infty, \\ ess\sup_{\omega\in\Omega} \|u(\cdot,\omega)\|_{W(D)} & \text{if } p = \infty. \end{cases} \tag{IV.1}$$

For simplicity, we will write $L^p(\Omega)$ instead of $L^p(\Omega; W(D))$ when the particular function space $W(D)$ can be concluded from the context.

For the sake of simplicity we will limit our attention to the simple case of the linear elliptic equation of the form

$$\begin{aligned} -\nabla^2 u(x,\omega) &= f(x) && \text{for } x \in D(\omega), \\ \alpha_1 u(x,\omega) + \beta_1 \frac{\partial u(x,\omega)}{\partial n} &= b_1(x) && \text{for } x \in \partial D_1, \\ \alpha_2 u(x,\omega) + \beta_2 \frac{\partial u(x,\omega)}{\partial n} &= b_2(x) && \text{for } x \in \partial D_2(\omega). \end{aligned} \tag{IV.2}$$

With this choice, we do not lose in generality because the proposed solution strategy can be easily extended to more general linear operators.

Denote by D_1 the reference deterministic domain containing all realizations of the random boundary $\partial D_2(\omega)$. This definition is similar to that used in the family of fictitious domain methods. However, we explicitly require that the deterministic part of the boundary ∂D_1 is also the boundary of D_1. This definition is depicted graphically in Figure 21.

Now, due to linearity of the operators in (IV.2), one can represent the solution $u(x,\omega)$ as a superposition of two functions

$$u(x,\omega) = u_1(x) + u_2(x,\omega). \tag{IV.3}$$

The first function $u_1(x)$ represents the deterministic component of the solution $u(x,\omega)$ and satisfies the following boundary value problem defined on the reference domain D_1

$$\begin{aligned} -\nabla^2 u_1(x) &= f(x) && \text{for } x \in D_1, \\ \alpha_1 u_1(x) + \beta_1 \frac{\partial u_1(x)}{\partial n} &= b_1(x) && \text{for } x \in \partial D_1. \end{aligned} \tag{IV.4}$$

The above problem can be solved with any conventional analytic or numerical technique. The conditions on the required regularity of the input data $f(x)$ and $b_1(x)$ are thus determined by the choice of the solver and are out of the scope of this chapter. We only require that the problem (IV.4) is well defined such that the solution $u_1(x)$ exists and is unique.

The random component $u_2(x,\omega)$ of the solution $u(x,\omega)$ can be determined from the following homogeneous boundary value problem

$$\begin{aligned} -\nabla^2 u_2(x,\omega) &= 0 && \text{for } x \in D(\omega), \\ \alpha_1 u_2(x,\omega) + \beta_1 \frac{\partial u_2(x,\omega)}{\partial n} &= 0 && \text{for } x \in \partial D_1, \\ \alpha_2 u_2(x,\omega) + \beta_2 \frac{\partial u_2(x,\omega)}{\partial n} &= \phi(x) && \text{for } x \in \partial D_2(\omega). \end{aligned} \tag{IV.5}$$

with the boundary condition on $\partial D_2(\omega)$ defined by the trace of the deterministic component $u_1(x)$

$$\phi(x) = b_2(x) - \alpha_2 u_1(x) - \beta_2 \frac{\partial u_1(x)}{\partial n}. \tag{IV.6}$$

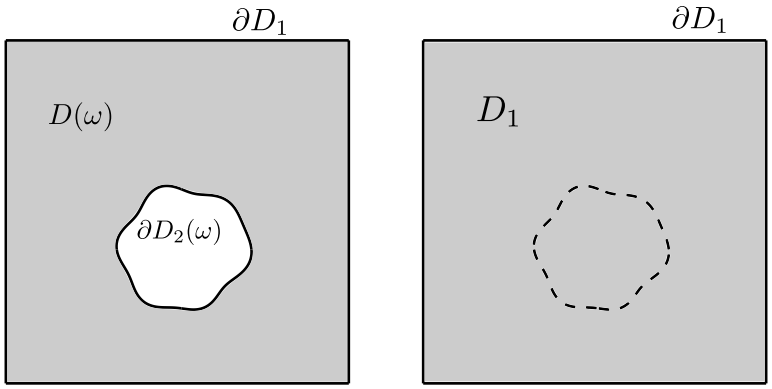


Figure 21: Realization of the random domain $D(\omega)$ (left) and the corresponding deterministic domain D_1 (right).

IV.2. Description of the acceleration technique. Dimension reduction

Since $u_1(x)$ is deterministic, the only source of uncertainty in $u(x,\omega)$ is given by the solution of the problem in (IV.5). It is simpler than the original problem in (IV.2) allowing to apply solution methods which are very well suited for this particular type of problems. One of such methods, namely the method of modified potentials, is proposed in this chapter. Firstly, define the Green's function corresponding to the boundary value problem (IV.4) via the following differential system

$$\begin{aligned} -\nabla^2 G_1(x,\xi) &= \delta(\xi) \qquad \text{for } x,\xi \in D_1, \\ \alpha_1 G_1(x,\xi) + \beta_1 \frac{\partial G_1(x,\xi)}{\partial n_x} &= 0 \qquad \text{for } x \in \partial D_1, \end{aligned} \tag{IV.7}$$

where $\delta(\xi)$ is the Dirac measure of unit mass at point ξ. With this function at hand and taking advantage of the special structure of the problem in (IV.5), we can construct its solution $u_2(x,\omega)$ in the form of a single-layer Green's potential [Jas63, Sym63]

$$u_2(x) = \int_{\partial D_1} G_1(x,y)\nu_1(y)dl(y) + \int_{\partial D_2(\omega)} G_1(x,y(\omega))\nu_2(y(\omega))dl(y(\omega))$$

$$= \int_{\partial D_2(\omega)} G_1(x, y(\omega))\nu_2(y(\omega))dl(y(\omega)) \tag{IV.8}$$
$$= \int_0^1 G_1(x, \xi_\omega(t))\mu(\xi_\omega(t))dt,$$

where $\mu(\xi_\omega(t)) = \nu_2(\xi_\omega(t))\,|\xi'_\omega(t)|$ and $\xi_\omega(t)$ defines a parameterization of the random boundary curve for fixed $\omega \in \Omega$. In the above expression, the first integral over the deterministic part of the boundary vanishes because the Green's function $G_1(x, \xi)$ satisfies the homogeneous boundary conditions on ∂D_1 by definition.

Using (IV.6) and the jump conditions of the derivative of the single-layer potential on the boundary, the unknown density $\mu(\xi_\omega(t))$ of the potential in (IV.8) can be obtained from the following boundary integral equation

$$-\frac{1}{2}\frac{\mu(\xi_\omega(s))}{|\xi'_\omega(s)|} + \int_0^1 \left(\alpha_2 + \beta_2 \frac{\partial}{\partial n_{\xi_\omega(s)}}\right) G_1(\xi_\omega(s), \xi_\omega(t))\mu(\xi_\omega(t))dt = \phi(\xi_\omega(s)), \quad s \in [0; 1], \tag{IV.9}$$

which is a Fredholm equation of the second kind. The pure Neumann problem ($\alpha_2 = 0$, $\beta_2 = 1$) yields the similar equation

$$-\frac{1}{2}\frac{\mu(\xi_\omega(s))}{|\xi'_\omega(s)|} + \int_0^1 \frac{\partial G_1(\xi_\omega(s), \xi_\omega(t))}{\partial n_{\xi_\omega(s)}}\mu(\xi_\omega(t))dt = \phi(\xi_\omega(s)), \quad s \in [0; 1]. \tag{IV.10}$$

It is well known that equations of the second kind are well-posed and numerous numerical techniques have proposed for their efficient solution [Atk97]. However, in the case of Dirichlet boundary conditions ($\alpha_2 = 1$, $\beta_2 = 0$), equation (IV.9) converts to the Fredholm equation of the first kind

$$\int_0^1 G_1(\xi_\omega(s), \xi_\omega(t))\mu(\xi_\omega(t))dt = \phi(\xi_\omega(s)), \quad s \in [0; 1], \tag{IV.11}$$

which is intrinsically ill-posed and must be treated with special care. For this reason, the

traditional approach in solving the Dirichlet problem with methods of potential theory is based on representing the solution in the form of the potential of the double-layer which gives equations of the second kind. The use of the single-layer potentials, however, has advantage of automatically satisfying equation on the deterministic boundary. As a result, when the length of the boundary ∂D_1 is large, the method of modified potentials can lead to significant computational savings compared to traditional approaches relying on discretization of the whole boundary and/or domain. It is well known that equations of the first kind with logarithmically singular kernels admit unique solutions when conformal radius of the boundary is not equal to one[YS88, Che93]. Therefore, we assume that all boundaries satisfy this condition.

Of course, efficiency of the proposed method relies on the availability of the Green's functions satisfying the system (IV.7) for the specific geometry of the domain. This becomes a serious limiting factor when the method is applied to deterministic problems because analytic expressions of the Green's functions can be obtained only for very simple domains. Numerical estimation of Green's functions can be impractical since this requires one to solve the complementary boundary value problem which represents a challenging task by itself. However, in the case of uncertain domains, the complementary problem has to be solved only once and the value of the Green's function at any field point is then readily available through a simple matrix vector product which can be done very efficiently.

IV.2.1. Discretization scheme

Spatial discretization The boundary integral equation (IV.9) is the classical equation of potential theory. It has been extensively studied in the literature and the large database of methods has been already collected [AS91, Atk97, GJS85, JSSE97, Slo00, Che93, Che94]. We do not pretend to give an exhaustive literature review nor we propose the new method. In fact, BIE (IV.9) can be solved with any available technique without modifying the proposed

solution strategy.

For the sake of completeness, we present here the quadrature technique proposed in [SB92] for the first kind Fredholm equations with logarithmic kernels on closed curves. It is a fully discrete method of qualocation type based on the composite quadrature rule, i.e., both the integral operator and the Galerkin projection are approximated with suitable quadratures.

We start with the boundary integral operator

$$(A\mu)(s) = \int_0^1 G_1(\xi_\omega(s), \xi_\omega(t))\mu(\xi_\omega(t))dt$$

and approximate it with the trapezoidal rule on the uniform grid with step $h = 1/N$ for some integer N

$$(A\mu)(s) \approx (A_h\mu)(s) = h\sum_{k=0}^{N-1} G_1(\xi_\omega(s), \xi_\omega(kh))\mu(\xi_\omega(kh)), \qquad s \in [0;1]. \qquad \text{(IV.12)}$$

The case of nonuniform grid is obtained trivially.

We then project this approximation on the test space S_h of 1-periodic smoothest splines of order r with the discrete inner product

$$(v, w)_h = Q_h(v\overline{w}),$$

where

$$Q_h g = h\sum_{k=0}^{N-1}\sum_{j=1}^{J} w_j g((k+\zeta_j)h), \qquad 0 < \zeta_1 < \zeta_2 < ... < \zeta_J < 1$$

and

$$\sum_{j=1}^{J} w_j = 1, \quad w_j > 0, \quad \text{for } 1 \le j \le J.$$

The problem now can be formulated as follows: find μ_h such that

$$(A_h \mu_h, \chi)_h = (\phi, \chi)_h, \qquad \forall \chi \in S_h.$$

For $r = 2$, the basis $(v_0, ..., v_{N-1})$ of S_h is represented by the classical hat functions

$$v_k(s) = \begin{cases} 1 - |s - kh|/h, & \text{if } |x - kh| \le h, \\ 0, & \text{otherwise.} \end{cases} \tag{IV.13}$$

Given the basis, one can write the discrete formulation of the problem: find μ_h such that

$$\sum_{k=0}^{N-1} a_{l,k} \mu_h(\xi_\omega(kh)) = (f, v_l)_h, \qquad l = 0, ..., N-1, \tag{IV.14}$$

where

$$a_{l,k} = h^2 \sum_{k'=0}^{N-1} \sum_{j=1}^{J} w_j G_1\Big(\xi_\omega\big((k' + \zeta_j)h\big), \xi_\omega(kh)\Big) \overline{v_l}\big((k' + \zeta_j)h\big).$$

It was shown in [SB92] that the following choices of the quadrature points and the weights are optimal in terms of stability of the approximation

$$\begin{aligned} &J = 2, \\ &\zeta_1 = \frac{1}{6}, \qquad \zeta_2 = \frac{5}{6}, \\ &w_1 = \frac{1}{2}, \qquad w_2 = \frac{1}{2}, \end{aligned} \tag{IV.15}$$

$$\begin{aligned} &J = 1, \\ &\zeta_1 = \frac{1}{6}, \\ &w_1 = 1. \end{aligned} \tag{IV.15'}$$

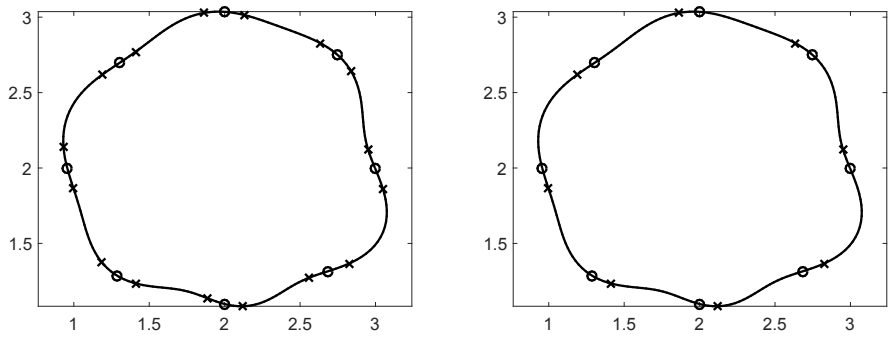


Figure 22: Collocation points (circles) and quadrature points (crosses) according to (IV.15) (left) and (IV.15') (right).

The collocation points $\xi_\omega(kh)$ and the quadrature points $\xi_\omega\big((k'+\zeta_j)h\big)$ for the case of hat basis functions $v_k(s)$ in (IV.13) are depicted in Figure 22.

The choice of nodes in (IV.15) gives the $O(h^3)$ order of uniform convergence [SB92]. The scheme with nodes in (IV.15') has only $O(h^2)$ accuracy but the linear system in (IV.14) converts to much simpler form

$$h\sum_{k=0}^{N-1} G_1\Big(\xi_\omega\big((k'+\zeta)h\big),\xi_\omega(kh)\Big)\mu_h(\xi_\omega(kh)) = f\big((k'+\zeta)h\big), \qquad k'=0,...,N-1. \tag{IV.16}$$

After the density of the potential is determined from the linear system in (IV.14) or (IV.16), one can calculate the solution $u_2(x,\omega)$ at any field point by evaluating the integral with the quadrature rule (IV.12).

IV.2.2. Evaluation of Green's functions for arbitrary domains

As it was mentioned previously, the proposed numerical technique relies heavily on the ability to evaluate Green's functions for domains of arbitrary shapes. We outline here several methods which allow to do this in a computationally attractive way. Following the logic of this chapter, we will consider Green's functions for the classical two-dimensional

Laplace operator only, i.e., we will limit our attention to the problem in (IV.7)

$$-\nabla^2 G_1(x,\xi) = \delta(\xi) \qquad \text{for } x,\xi \in D_1, \tag{IV.17}$$

$$\alpha_1 G_1(x,\xi) + \beta_1 \frac{\partial G_1(x,\xi)}{\partial n_x} = 0 \qquad \text{for } x \in \partial D_1. \tag{IV.18}$$

The case of more general operators can be treated similarly.

Analytical Green's functions.

Our discussion will be incomplete if we do not mention special cases when the Green's function can be obtained in a closed analytic form. ***Fundamental solution*** is, probably, the most important example of such solution since it can be used as a *particular solution* for the system in (IV.17)-(IV.18). It satisfies the equation (IV.17) in the domain with no boundaries and has the form

$$G^*(x,\xi) = -\frac{1}{2\pi}\ln r, \qquad r = \sqrt{(x_1-\xi_1)^2 + (x_2-\xi_2)^2}. \tag{IV.19}$$

The general form of the Green's function can be written as a sum of the particular solution (IV.19) and a "corrector" function aiming to satisfy the boundary condition in (IV.18)

$$G_1(x,\xi) = G^*(x,\xi) + \psi(x,\xi). \tag{IV.20}$$

Defined in this way, the corrector function is called the regular component of the Green's function. It solves the following complementary system

$$-\nabla^2 \psi(x,\xi) = 0 \qquad \text{for } x,\xi \in D_1, \tag{IV.21}$$

$$\alpha_1 \psi(x,\xi) + \beta_1 \frac{\partial \psi(x,\xi)}{\partial n_x} = -\alpha_1 G^*(x,\xi) - \beta_1 \frac{\partial G^*(x,\xi)}{\partial n_x} \qquad \text{for } x \in \partial D_1. \tag{IV.22}$$

In the case of simple geometries possessing certain symmetry properties, the above system admits the closed form analytic solution which can be constructed via the method of images. We list several examples of Green's function obtained in this way below; more examples can be found, for instance, in [Duf01, Mel11]. Additionally, the infinite product representation of Green's functions arising from applying the method of images is discussed in [MM12, Res13].

Dirichlet problem in the upper half-plane $D_1(x,y) = \{x, y \geq 0\}$

$$G_1(x,\xi) = -\frac{1}{2\pi} \ln \sqrt{\frac{(x_1-\xi_1)^2 + (x_2-\xi_2)^2}{(x_1-\xi_1)^2 + (x_2+\xi_2)^2}}.$$

Neumann problem in the upper half-plane $D_1(x,y) = \{x, y \geq 0\}$

$$G_1(x,\xi) = -\frac{1}{2\pi} \ln \sqrt{\big((x_1-\xi_1)^2 + (x_2-\xi_2)^2\big)\big((x_1-\xi_1)^2 + (x_2+\xi_2)^2\big)}.$$

Dirichlet problem in the quarter plane $D_1(r,\varphi) = \{0 \leq r < \infty, 0 \leq \varphi \leq \pi/2\}$

$$G_1\big(x,\xi\big) = \frac{1}{4\pi} \ln \prod_{n=1}^{2} \frac{r^2 - 2r\rho\cos(\varphi - (n\pi - \varsigma)) + \rho^2}{r^2 - 2r\rho\cos(\varphi - ((n-1)\pi + \varsigma)) + \rho^2},$$

$$(x_1, x_2) = r(\cos(\varphi), \sin(\varphi)), \quad (\xi_1, \xi_2) = \rho(\cos(\varsigma), \sin(\varsigma)).$$

Dirichlet-Neumann problem in the quarter plane $D_1(r,\varphi) = \{0 \leq r < \infty, 0 \leq \varphi \leq \pi/2\}$ with Dirichlet condition along $y = 0$ and Neumann condition along $x = 0$

$$G_1\big(x,\xi\big) =$$

$$\frac{1}{4\pi} \ln \left(\frac{r^2 - 2r\rho\cos(\varphi - (2\pi - \varsigma)) + \rho^2}{r^2 - 2r\rho\cos(\varphi - \varsigma) + \rho^2} \cdot \frac{r^2 - 2r\rho\cos(\varphi - (n\pi - \varsigma)) + \rho^2}{r^2 - 2r\rho\cos(\varphi - ((n-1)\pi - \varsigma)) + \rho^2} \right),$$

$$(x_1, x_2) = r(\cos(\varphi), \sin(\varphi)), \quad (\xi_1, \xi_2) = \rho(\cos(\varsigma), \sin(\varsigma)).$$

Dirichlet problem in the infinite wedge $D_1(r,\varphi) = \{0 \le r < \infty, 0 \le \varphi \le \pi/4\}$

$$G_1\big(x,\xi\big) = \frac{1}{4\pi}\ln\prod_{n=1}^{4}\frac{r^2 - 2r\rho\cos(\varphi - (\frac{n\pi}{2} - \varsigma)) + \rho^2}{r^2 - 2r\rho\cos(\varphi - (\frac{(n-1)\pi}{2} + \varsigma)) + \rho^2},$$

$$(x_1, x_2) = r(\cos(\varphi), \sin(\varphi)), \quad (\xi_1, \xi_2) = \rho(\cos(\varsigma), \sin(\varsigma)).$$

Dirichlet-Neumann problem in the infinite wedge $D_1(r,\varphi) = \{0 \le r < \infty, 0 \le \varphi \le \pi/4\}$ with Dirichlet condition along the wedge width $\varphi = 0$ and Neumann condition along the wedge with $\varphi = \pi/4$

$G_1\big(x,\xi\big) =$

$$\frac{1}{4\pi}\ln\prod_{n=1}^{2}\left(\frac{r^2 - 2r\rho\cos(\varphi - (\frac{(2n-1)\pi}{2} + \varsigma)) + \rho^2}{r^2 - 2r\rho\cos(\varphi - (\frac{(2n-1)\pi}{2} - \varsigma)) + \rho^2}\cdot\frac{r^2 - 2r\rho\cos(\varphi - (n\pi - \varsigma)) + \rho^2}{r^2 - 2r\rho\cos(\varphi - ((n-1)\pi + \varsigma)) + \rho^2}\right),$$

$(x_1, x_2) = r(\cos(\varphi), \sin(\varphi)), \quad (\xi_1, \xi_2) = \rho(\cos(\varsigma), \sin(\varsigma)).$

Dirichlet problem in the disk $D_1(r,\varphi) = \{0 \le r < a, 0 \le \varphi \le 2\pi\}$

$$G_1\big(x,\xi\big) = \frac{1}{4\pi}\ln\frac{a^4 - 2r\rho a^2\cos(\varphi - \varsigma) + r^2\rho^2}{a^2(r^2 - 2r\rho\cos(\varphi - \varsigma) + \rho^2)},$$

$$(x_1, x_2) = r(\cos(\varphi), \sin(\varphi)), \quad (\xi_1, \xi_2) = \rho(\cos(\varsigma), \sin(\varsigma)).$$

It is worth noting that presented analytic expressions can be used to build asymptotic approximations of the true Green's function when both the source point ξ and the observation point x are located near the boundary. One can also use them as building blocks for composite kernels.

Direct numerical approximation of Green's functions.

The advantage of having the explicit expression for the inverse operator is most prominent in problems with fixed geometry and multiple boundary conditions or right hand sides. However, having the analytic expression for the Green's function is an exception rather than a rule. Therefore, it is not surprising that the benefits of numerical utilization of Green's functions have been already recognized in the early era of digital computing [BL58, GK69].

By its definition, the Green's function of the boundary value problem is the inverse of the corresponding differential operator. In fact, any numerical method implicitly constructs the approximate inverse operator of the original problem and thus every numerical solution can be formulated in terms of the corresponding approximate Green's functions. For instance, it was shown in [Tot70] that the finite element solution $u_h(x)$ of the boundary value problem with homogeneous boundary conditions has the form

$$u_h(x) = \int_D G_h(x,\xi) f(\xi) d\xi,$$

where $G_h(x,\xi)$ is the FE-Green's function, i.e., the projection of the exact Green's function on the finite element space V_h. One can write the FE-Green's function in terms of the basis functions $(v_1, ..., v_M) \in V_h$ as

$$G_h(x,\xi) = \sum_{i=1}^{M} g_i(\xi) v_i(x).$$

Solving for the coefficients $g_i(\xi)$ we get the following

Theorem IV.2.1. *([Har13]) Let K be the stiffness matrix of the linear system arising after the finite element disretization. The FE-Green's function has the form*

$$G_h(x,\xi) = \mathbf{v}(x)^T K^{-1} \mathbf{v}(\xi), \tag{IV.23}$$

where $\mathbf{v}(x) = (v_1(x), ..., v_M(x))$ *are the basis functions of the FE-space* V_h*.*

It is interesting to note that when the basis $(v_1, ..., v_M) \in V_h$ is given by the eigenfunctions of the differential Sturm-Liouville problem, i.e., in the case of the spectral finite element method, the formula (IV.23) converts to

$$G_h(x, \xi) = \mathbf{v}(x)^T \Lambda^{-1} \mathbf{v}(\xi) = \sum_{i=1}^{M} \frac{v_i(x) v_i(\xi)}{\lambda_i}, \tag{IV.24}$$

where Λ is the diagonal matrix with corresponding eigenvalues. One can immediately recognize in (IV.24) the M-term truncation of the classical eigenfunction representation of the Green's function which is guaranteed to exist by the Mercer's theorem [FM16].

The spectral representation in (IV.24) might be useful when the corresponding eigenvalue problem admits the analytical solution. When this is not the case, one gets the tradeoff between inversion of the stiffness matrix in (IV.23) and numerical approximation of the eigenfunctions and eigenvalues in (IV.24). Green's functions which have analytical spectral representation can be found, for instance, in [Duf01]. As an example, we provide the Green's function for the

Dirichlet problem in the rectangle $D_1(x, y) = \{0 \leq x \leq a, 0 \leq y \leq b\}$

$$G_1\big(x, \xi\big) = 4ab \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{\sin\left(\frac{m\pi\xi_1}{a}\right) \sin\left(\frac{n\pi\xi_2}{b}\right) \sin\left(\frac{m\pi x_1}{a}\right) \sin\left(\frac{n\pi x_2}{b}\right)}{n^2\pi^2a^2 + m^2\pi^2b^2}. \tag{IV.25}$$

Numerical approximation of the regular part of Green's functions.

Representations in (IV.23)-(IV.24) converge very slowly since the Green's functions generally do not belong to the function spaces approximated by the $span(v_1(x), ..., v_M(x))$. For instance, solutions to the Poisson equation are usually constructed in $H^1(D_1)$, the space of square integrable functions with square integrable first derivatives, but the solution to the

problem in (IV.17)-(IV.18) is not in $H^1(D_1)$ since the delta function $\delta(\xi) \notin H^{-1}$ for $d \geq 2$, where d is the physical dimension of the problem. Analogously, analytical expansions like (IV.25) do not have uniform error estimates which seriously limits their practical utilization. In certain cases, one can obtain uniformly convergent spectral representations by partial summation of the series leading to explicit extraction of the singularity. Justification of this approach with practical examples can be found in [Mel98]. For instance the series in (IV.25) is transformed to the following form

$$G_1\big(x,\xi\big) = \frac{1}{2\pi}\ln\left[\frac{E(z-\zeta^*)E(z+\zeta^*)E(z_1+\zeta_1^*)E(z_2+\zeta_2^*)}{E(z-\zeta)E(z+\zeta)E(z_1+\zeta_1)E(z_2+\zeta_2)}\right] - \frac{2}{\pi}\sum_{n=1}^{\infty} S_n(x_1,\xi_1)\sin\left(\frac{n\pi\xi_2}{b}\right)\sin\left(\frac{n\pi x_2}{b}\right), \tag{IV.26}$$

where $z = x_1+ix_2$, $z_1 = (x_1+a)+ix_2$, $z_2 = (x_1-a)+ix_2$, $\zeta = \xi_1+i\xi_2$, $\zeta_1 = (\xi_1+a)+i\xi_2$, $\zeta_2 = (\xi_1-a)+i\xi_2$, $\zeta_1^* = (\xi_1+a)-i\xi_2$, $\zeta_2^* = (\xi_1-a)-i\xi_2$, $E(z) = \left|e^{\pi z/b}-1\right|$ and

$$S_n(x_1,\xi_1) = \frac{e^{n\pi x_1/b}\sinh(n\pi(\xi_1-a)/b) - e^{-n\pi x_1/b}\sinh(n\pi(\xi_1+a)/b)}{2ne^{2n\pi a/b}\sinh(n\pi a/b)}.$$

The remainder term $R_M(x,\xi)$ of the M-term truncation of the expansion in (IV.26) has the estimate

$$|R_M(x,\xi)| \leq \frac{b}{2\pi}\left(\ln\left(1-e^{-\pi a/b}\right) - \sum_{n=1}^{N}\frac{e^{-n\pi a/b}}{n}\right)$$

which reveals the extremely high rate of convergence.

Similarly, the corrector function $\psi(x,\xi)$ obtained in (IV.20) after extracting singularity from the Green's function is harmonic everywhere in the domain D_1 and thus can be efficiently approximated with any conventional numerical method. For instance, the finite element approximation has the form

$$\psi_h(x,\xi) = \mathbf{v}(x)^T K^{-1}\mathbf{g}^*(\xi),$$

where K is the same stiffness matrix as in (IV.23) and $\mathbf{g}^*(\xi)$ encodes the trace of the fundamental solution on the boundary.

One can also construct the regular part of the Green's function in the form of the single layer potential

$$\psi(x,\cdot) = \int_{\partial D_1} G^*(x,y)\nu_1(y)dl(y) = \int_0^1 G^*(x,\xi_1(t))\mu^{\psi}(t)dt,$$

where $G^*(x,y)$ is the fundamental solution of the differential operator, $\mu^{\psi}(t) = \nu_1(\xi_1(t))\,|\xi_1'(t)|$ and $\xi_1(t)$ defines a parameterization of the boundary ∂D_1 of the Green's function domain. It is natural to build the approximate solution of the above equation with the same method used for the approximation of the original integral equation, e.g., with the scheme given in section IV.2.1. We will use this approach in the subsequent sections. An example of the approximate Green's function is given in Figure 23.

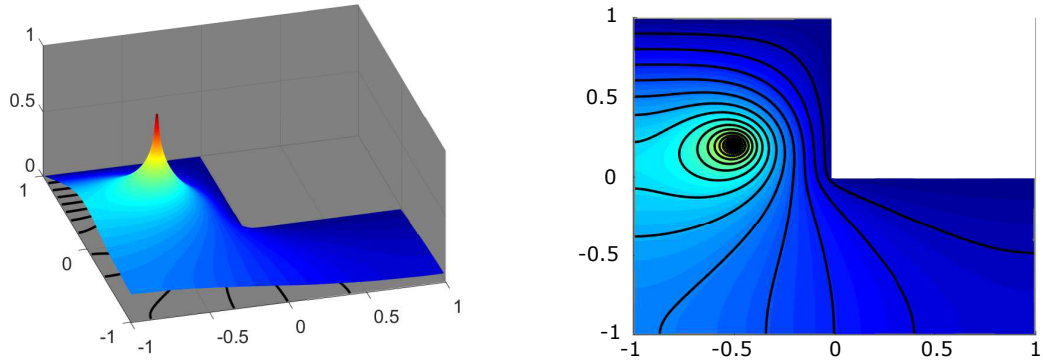
IV.3. Complexity analysis

IV.3.1. Error component analysis

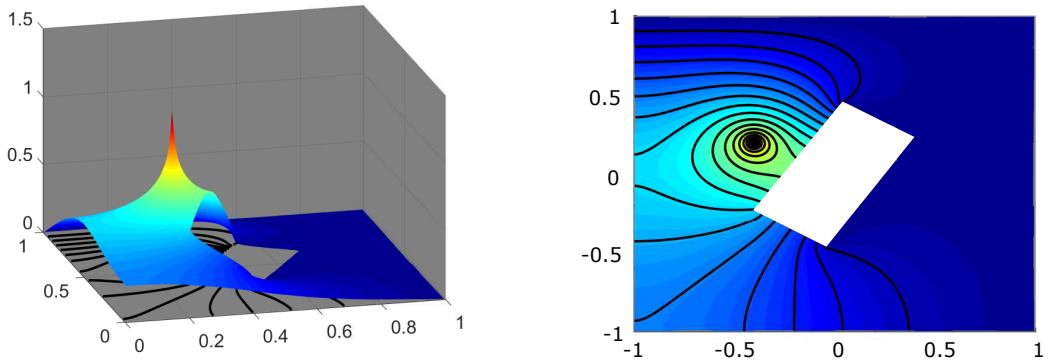
Consider the following decomposition of the total error of the MLMC estimator

$$\left\|\mathbb{E}_{\mathbb{ML}}\left[\tilde{u}_L\right] - \mathbb{E}\left[u\right]\right\|_{L^2(\Omega;\tilde{W}(D))} \leq \underbrace{\left\|\mathbb{E}\left[\tilde{u}_L - u\right]\right\|_{\tilde{W}(D)}}_{\text{I := Discretization error}} + \underbrace{\left\|\left(\mathbb{E}_{\mathbb{ML}} - \mathbb{E}\right)\left[\tilde{u}_L\right]\right\|_{L^2(\Omega)}}_{\text{II := Sampling error}}, \quad \text{(IV.27)}$$

where u is the exact value of the potential (IV.8) and $\tilde{u}_L$ is its approximations at the discretization level L obtained by solving the boundary integral equation with exact or approximate Green's kernel. In above, the error components I and II correspond to the spatial approximation error and the sampling error. To achieve the desired accuracy ϵ, it is



(A) L-shaped domain with a combination of Dirichlet and Neumann boundary conditions.



(B) Multi-connected domain with a combination of Dirichlet and Neumann boundary conditions.

Figure 23: Numerical Green's functions.

sufficient to balance the total error between all components in the following way

$$\|\mathbb{E}_{\mathbb{ML}}\left[\tilde{u}_L\right] - \mathbb{E}\left[u\right]\|_{L^2(\Omega;\tilde{W}(D))} \leq \epsilon_I + \epsilon_{II} = \epsilon. \tag{IV.28}$$

The error analysis for each of the components is provided below.

Spatial discretization error. The Jensen's inequality gives

$$\|\mathbb{E}\left[u_L(x,\omega) - u(x,\omega)\right]\|_{\tilde{W}(D)} \leq \mathbb{E}\left[\|u_L(x,\omega) - u(x,\omega)\|_{\tilde{W}(D)}\right] = \epsilon_I \tag{IV.29}$$

and the estimate of the error component I can be derived from the convergence properties of the spatial discretization scheme for fixed $\omega \in \Omega$. For instance, it was shown in [SB92] that the scheme in section IV.2.1 admits the following estimate for the error in approximation of $\mu \in H^t([0,1])$

$$\|\mu_l - \mu\|_{H^s([0,1])} \leq c h_l^{t-s} \|\mu\|_{H^t([0,1])}$$

provided that $s > \frac{1}{2}$, $s + \frac{1}{2} < t \leq s + \alpha$ and the right hand side of integral equation is continuous and 1-periodic. In the case of optimal regularity, i.e., for $\mu \in H^t([0,1])$ with $t > \alpha + 1/2$, the following error bound is valid

$$\sup_{t\in[0,1]} |\mu_l(t) - \mu(t)| = \|\mu_l - \mu\|_{L^\infty([0,1])} \leq c h_l^{\alpha} \|\mu\|_{H^t([0,1])} \tag{IV.30}$$

due to embedding of H^s ($s > 1/2$) in C_p, the space of 1-periodic continuous functions. The order of convergence is $\alpha = 3$ and $\alpha = 2$ for the schemes with the quadrature nodes and weights as in (IV.15) and (IV.15') respectively.

Consider the error in the approximation of the single-layer potential (IV.8) at the level l

$$|\tilde{u}_l(x) - u(x)| \leq |u_l(x) - u(x)| + |\tilde{u}_l(x) - u_l(x)|,$$

where the first term on the right side gives the error of the numerical scheme with exact Green's function and the second term is the error due to approximation of the Green's kernel itself. We get for the first component that

$$\begin{aligned} u_l(x) - u(x) &= h_l \sum_{k=0}^{N_l-1} G_1(x, \xi_\omega(kh))\mu(kh_l) - \int_0^1 G_1(x, \xi_\omega(t))\mu(t)dt \\ &= \int_0^1 G_1(x, \xi_\omega(t))(\mu_l(t) - \mu(t))dt + R(x), \end{aligned}$$

where $R(x)$ is the error of the trapezoidal rule in (IV.12). Since $R(x) = O(h_l^3)$ for periodic

functions and using (IV.30), one gets

$$|u_l(x) - u(x)| \leq \left| \int_0^1 G_1(x, \xi_\omega(t))(\mu_l(t) - \mu(t))dt \right| + O(h_l^3) \tag{IV.31}$$
$$\leq \|\mu_l - \mu\|_{L^\infty([0,1])} \int_0^1 |G_1(x, \xi_\omega(t))|dt + O(h_l^3) = c \|\mu_l - \mu\|_{L^\infty([0,1])} .$$

Similarly, the Green's kernel is a harmonic and thus analytic function at any internal point of the domain which gives the estimate for the error in the approximation of the derivatives of the potential

$$|u_l^{(i)}(x) - u^{(i)}(x)| \leq \left| \int_0^1 G_1^{(i)}(x, \xi_\omega(t))\big(\mu_l(t) - \mu(t)\big)dt \right| + O(h_l^3)$$
$$\leq \|\mu_l - \mu\|_{L^\infty([0,1])} \int_0^1 |G_1^{(i)}(x, \xi_\omega(t))|dt + O(h_l^3) = c \|\mu_l - \mu\|_{L^\infty([0,1])} ,$$

where $i = (i_1, i_2)$ is a multi-index and $f^{(i)} = \dfrac{\partial^{|i|} f}{\partial x_1^{i_1} \partial x_2^{i_2}}$.

Approximation of the Green's kernel can be performed with any of the methods presented in section IV.2.2. We consider the Green's function as in (IV.20) and construct the regular part in the form of the single layer potential

$$\psi(x, \cdot) = \int_{\partial D_1} G^*(x, y)\nu_1(y)dl(y) = \int_0^1 G^*(x, \xi_1(t))\mu^\psi(t)dt,$$

where $G^*(x, y)$ is the fundamental solution of the differential operator, $\mu^g(t) = \nu_1(\xi(t)) \, |\xi'(t)|$ and $\xi(t)$ defines a parameterization of the boundary of the Green's function domain. It is natural to build the approximate solution of the above equation with the same method used for the approximation of the original integral equation which, by analogy with (IV.31), gives the error estimate

$$\left| \tilde{G}_1(x, \cdot) - G_1(x, \cdot) \right| = |\psi_l(x, \cdot) - \psi(x, \cdot)| = c_1 \left\| \mu_l^\psi - \mu^\psi \right\|_{L^\infty([0,1])} = ch_l^\alpha,$$

where h_l is the step size of the grid approximating the boundary ∂D_1 of the Green's function domain. This gives the following estimate

$$|\tilde{u}_l(x) - u_l(x)| \leq h_l \cdot \sup_k |\mu(kh_l)| \cdot \sum_{k=0}^{N_l-1} \left| \tilde{G}_1(x, \xi(kh_l)) - G_1(x, \xi(kh_l)) \right|$$
$$= c_1 \cdot \sup_k |\mu(kh_l)| \cdot h_l^\alpha \cdot (N_l \cdot h_l) = c h_l^\alpha$$

since $N_l \cdot h_l = O(1)$.

It is trivial to obtain the estimates for the errors in Sobolev norm

$$I := \mathbb{E}\left[\|u_l(x,\omega) - u(x,\omega)\|_{H^s(D)} \right]$$
$$= \mathbb{E}\left[\left(\sum_{|i|\leq s} \int_D |u_l^{(i)}(x,\omega) - u^{(i)}(x,\omega)|^2 dx \right)^{1/2} \right] \leq c h_l^\alpha = \epsilon_I. \tag{IV.32}$$

According to (I.21), the condition (IV.29) will be satisfied if the number of levels is fixed to be

$$L = \left\lceil \log_q \left(h_0 (c_1 \epsilon_I^{-1})^{1/\alpha} \right) \right\rceil \leq c + \log_q \left(h_0 \epsilon_I^{-1/\alpha} \right).$$

Sampling error. Let $\tilde{W}(D) = H^s(D)$ be a Sobolev space. Then by using definition in (IV.1), one obtains

$$\left\| \left(\mathbb{E}_{\mathbb{MC}} - \mathbb{E} \right)[u] \right\|^2_{L^2(\Omega;\tilde{W}(D))} = \mathbb{E}\left[\left\| \left(\mathbb{E}_{\mathbb{MC}} - \mathbb{E} \right)[u] \right\|^2_{H^s(D)} \right]$$
$$= \frac{1}{M^2} \mathbb{E}\left[\left\| \sum_{m=1}^{M} \left(u^m - \mathbb{E}[u] \right) \right\|^2_{H^s(D)} \right]$$
$$= \frac{1}{M^2} \mathbb{E}\left[\sum_{i=0}^{s} \int_D \left| \sum_{m=1}^{M} \left(u^{(i),m} - \mathbb{E}\left[u^{(i)} \right] \right) \right|^2 dx \right]$$

$$= \frac{1}{M^2}\sum_{i=0}^{s}\int_D \sum_{m=1}^{M} \mathbb{V}ar\Big(u^{(i),m}\Big)dx + \frac{1}{M^2}\sum_{i=0}^{s}\int_D \sum_{m=1}^{M}\sum_{\substack{m'=1\\ m\neq m'}}^{M} \mathbb{C}ov\Big(u^{(i),m}, u^{(i),m'}\Big)dx.$$

By virtue of the independence of i.i.d. samples u^m, we have that $\mathbb{C}ov\Big(u^{(i),m}, u^{(i),m'}\Big) = 0$ and

$$\left\|\left(\mathbb{E}_{\mathbb{MC}} - \mathbb{E}\right)[u]\right\|_{L^2(\Omega;H^s(D))} = \sqrt{\frac{\overline{V}(u)}{M}}, \tag{IV.33}$$

where $\overline{V}(u) = \mathbb{E}\left[\|u - \mathbb{E}[u]\|^2_{H^s(D)}\right]$.

Based on this result, and due to independence of MC estimators at each level, one gets the error of the MLMC estimator as follows

$$\begin{aligned}
\left\|\left(\mathbb{E}_{\mathbb{ML}} - \mathbb{E}\right)[u_L]\right\|^2_{L^2(\Omega;H^s(D))} &= \left\|\mathbb{E}[u_L] - \sum_{l=0}^{L}\mathbb{E}_{\mathbb{MC}}[u_l - u_{l-1}]\right\|^2_{L^2(\Omega;H^s(D))} \\
&= \left\|\sum_{l=0}^{L}\Big(\mathbb{E}[\Delta_l] - \mathbb{E}_{\mathbb{MC}}[\Delta_l]\Big)\right\|^2_{L^2(\Omega;H^s(D))} \\
&= \mathbb{E}\left[\left\|\sum_{l=0}^{L}\Big(\mathbb{E}_{\mathbb{MC}}[\Delta_l] - \mathbb{E}[\Delta_l]\Big)\right\|^2_{H^s(D)}\right] \\
&= \sum_{l=0}^{L}\mathbb{E}\left[\|\mathbb{E}_{\mathbb{MC}}[\Delta_l] - \mathbb{E}[\Delta_l]\|^2_{H^s(D)}\right] = \sum_{l=0}^{L}\frac{\overline{V}_l}{M_l},
\end{aligned}$$

where $\overline{V}_l = \overline{V}(\Delta_l)$ and $\Delta_l = u_l - u_{l-1}$.

Thus, the sampling error of the MLMC method can be estimated as

$$II^2 := \left\|\left(\mathbb{E}_{\mathbb{ML}} - \mathbb{E}\right)[u_L]\right\|^2_{L^2(\Omega;H^s(D))} = \sum_{l=0}^{L}\frac{\overline{V}_l}{M_l} = \epsilon^2_{II}. \tag{IV.34}$$

According to (I.22), the above condition will be satisfied if the number of MC samples

at each discretization level is set to

$$M_l = \epsilon_{II}^{-2} \left(\frac{\overline{V}_l}{C_l} \right)^{1/2} \sum_{k=0}^{L} (C_k \overline{V}_k)^{1/2}.$$

IV.3.2. Complexity analysis of the standard algorithm

From the above error analysis, it follows that accuracy of the MLMC method is controlled by two parameters:

1. the number of levels $L = \left\lceil \log_q \left(h_0 (c_1 \epsilon_I^{-1})^{1/\alpha} \right) \right\rceil$,
2. the number of samples per level $M_l = \left\lceil \epsilon_{II}^{-2} \left(\frac{\overline{V}_l}{C_l} \right)^{1/2} \sum_{k=0}^{L} (C_k \overline{V}_k)^{1/2} \right\rceil$.

By analogy with (I.23), the optimal cost of the MLMC estimator is defined as

$$C_{\mathbb{ML}} = \sum_{l=0}^{L} \left\lceil M_l \right\rceil C_l \leq \sum_{l=0}^{L} C_l + \epsilon_{II}^{-2} \left(\sum_{l=0}^{L} (C_l \overline{V}_l)^{1/2} \right)^2.$$

Consider the cost C_l of the solver at each level $l \in [0, L]$. It consists of three components:

1. the cost C_l^a of assembling the matrix of the linear system,
2. the cost C_l^s of solving the linear system

   and
3. the cost C_l^e of evaluating the potential.

Let N_l denote the number of degrees of freedom at the level $l \in [0, L]$ and C_{fs} be the cost of evaluating the fundamental solution at a single point. Then the cost of assembling the matrix is given by $C_l^a = O(N_l^2 C_{fs})$. For direct linear solvers, the cost of solving the resulting system of equations is equal to $C_l^s = O(N_l^\gamma)$ for some $\gamma \in (2, 3]$. Finally, evaluation of the potential at a single point requires evaluation of N_l fundamental solutions and the dot

product of two vectors of size N_l which gives $C_l^e = O(N_l P C_{fs})$, where P is the number of evaluation points.

One may conclude that the cost of the linear solver is asymptotically dominant and the total cost at each level $l \in [0, L]$ behaves as

$$C_l = C_l^a + C_l^s + C_l^e = O\left(N_l^\gamma\right) \tag{IV.35}$$

for some $\gamma \in (2, 3]$.

According to Theorem I.3.2 and taking into account (IV.35) and that $N_l = O(h_l^{-1})$, the ϵ-cost of the MLMC method applied to the linear elliptic equation in (IV.5)-(IV.6) is given by

$$C_{\mathbb{ML}} \simeq \begin{cases} \epsilon^{-2} & \text{if } \gamma - \beta < 0, \\ \epsilon^{-2} \left| \ln \epsilon \right|^2 & \text{if } \gamma - \beta = 0, \\ \epsilon^{-2 - \frac{\gamma - \beta}{\alpha}} & \text{if } \gamma - \beta > 0. \end{cases} \tag{IV.36}$$

with $\gamma \in (2, 3]$.

Note that, in this case, the equation is formulated on the whole boundary $\partial D = \partial D_1 \cup \partial D_2$.

IV.3.3. Complexity analysis of the accelerated algorithm

For the level $l \in [0, L]$, let $N_{1,l}$, $N_{2,l}$ be the number of degrees of freedom corresponding to the fixed and random parts of the boundary ∂D_1 and ∂D_2 respectively such that $N_l = N_{1,l} + N_{2,l}$.

Analytical Green's kernel. Denote by C_{gf} the cost of evaluating the analytical Green's function at a single point. Then the cost of assembling the matrix is given by $\tilde{C}_l^a =$

$O(N_{2,l}^2 C_{gf})$. The cost of solving the system of linear equations is $\tilde{C}_l^s = O(N_{2,l}^\gamma)$ for some $\gamma \in (2, 3]$ and the cost of evaluation of the potential at P points is $\tilde{C}_l^e = O(PN_{2,l}C_{gf})$. In this case, the cost of the linear solver is also asymptotically dominant and the total cost at each level $l \in [0, L]$ behaves as

$$\tilde{C}_l = \tilde{C}_l^a + \tilde{C}_l^s + \tilde{C}_l^e = O\left(N_{2,l}^\gamma\right). \tag{IV.37}$$

By comparing (IV.37) to (IV.35) it follows that

$$\tilde{C}_l \simeq \left(\frac{N_{2,l}}{N_l}\right)^\gamma C_l \simeq \left(\frac{|\partial D_2|}{|\partial D|}\right)^\gamma C_l,$$

where $|\partial D|$ denotes the length of the boundary ∂D.

Theorem IV.3.1. *Consider the boundary integral formulation of the linear elliptic system in* (IV.5)-(IV.6)*. Suppose that the kernels of integral operators are given by the analytical Green's function defined on the deterministic part of the domain. The ϵ-cost of the accelerated MLMC method is given by*

$$\tilde{C}_{\mathbb{ML}} \simeq \left(\frac{|\partial D_2|}{|\partial D|}\right)^\gamma C_{\mathbb{ML}},$$

where $C_{\mathbb{ML}}$ is defined in eq. (IV.36) *as the cost of the standard formulation of the boundary integral equation method.*

Proof. By its definition, the computational cost of the accelerated MLMC method is

$$\tilde{C}_{\mathbb{ML}} = \sum_{l=0}^{L} \left\lceil M_l \right\rceil \tilde{C}_l \simeq \left(\frac{|\partial D_2|}{|\partial D|}\right)^\gamma \sum_{l=0}^{L} \left\lceil M_l \right\rceil C_l \simeq \left(\frac{|\partial D_2|}{|\partial D|}\right)^\gamma C_{\mathbb{ML}}.$$

□

Approximate Green's kernel. By our assumption, the same method is used to represent the solution of the original differential system and the regular part of Green's kernel. Therefore, the cost of evaluating the numerical Green's function at a single point is the same as the cost of evaluating the potential over the boundary ∂D_1, i.e., $C_{gf} = O(N_{1,l}C_{fs})$. Then the cost of assembling the matrix is given by $\tilde{C}_l^a = O(N_{2,l}^2 C_{gf}) = O(N_{2,l}^2 N_{1,l} C_{fs})$. The cost of solving the system of linear equations is $\tilde{C}_l^s = O(N_{2,l}^\gamma)$ for some $\gamma \in (2, 3]$ and the cost of evaluation of the potential at P points is $\tilde{C}_l^e = O(PN_{2,l}C_{gf}) = O(PN_{2,l}N_{1,l}C_{fs})$.

Lemma IV.3.1. *Let the costs $\tilde{C}_l^a$, $\tilde{C}_l^s$ and $\tilde{C}_l^e$ be defined as above. Then*

$$\tilde{C}_l \simeq \varkappa(r)\, h_l^{-3},$$

where $r = \dfrac{N_{2,l}}{N_{1,l}}$ *and*

$$\varkappa(r) \lesssim \begin{cases} r^{\gamma-1} & \text{if } r \ll 1, \\ \dfrac{r^{\gamma-1}}{(r+1)^\gamma} & \text{if } r = O(1), \\ r^{-1} & \text{if } r \gg 1. \end{cases}$$

Proof. It happens that the cost of assembling the system is asymptotically dominant, i.e.,

$$\tilde{C}_l = \tilde{C}_l^a + \tilde{C}_l^s + \tilde{C}_l^e = O\left(N_{2,l}^2 N_{1,l} C_{fs}\right). \tag{IV.38}$$

By comparing (IV.38) to (IV.35) it follows that

$$\tilde{C}_l \simeq \frac{N_{2,l}^2 N_{1,l}}{N_l^\gamma} C_l = \frac{N_{2,l}^2 N_{1,l}}{(N_{1,l} + N_{2,l})^\gamma} C_l.$$

Consider 3 cases.

Case 1: $\dfrac{N_{2,l}}{N_{1,l}} = r \ll 1.$

$$\frac{N_{2,l}^2 N_{1,l}}{(N_{1,l}+N_{2,l})^\gamma} \simeq \frac{N_{2,l}^2 N_{1,l}}{N_{1,l}^\gamma} = \left(\frac{N_{2,l}}{N_{1,l}}\right)^{\gamma-1} N_{2,l}^{3-\gamma} \simeq r^{\gamma-1} N_{2,l}^{3-\gamma}$$

Case 2: $\dfrac{N_{2,l}}{N_{1,l}} = r = O(1).$

$$\frac{N_{2,l}^2 N_{1,l}}{(N_{1,l}+N_{2,l})^\gamma} \simeq \frac{r^{\gamma-1}}{(r+1)^\gamma} N_{2,l}^{3-\gamma}$$

Case 3: $\dfrac{N_{2,l}}{N_{1,l}} = r \gg 1.$

$$\frac{N_{2,l}^2 N_{1,l}}{(N_{1,l}+N_{2,l})^\gamma} \simeq \frac{N_{2,l}^2 N_{1,l}}{N_{2,l}^\gamma} = r^{-1} N_{2,l}^{3-\gamma}$$

The result follows from $N_{2,l} = O(h_l^{-1})$ and $C_l = O(h_l^{-\gamma})$. □

The following theorem establishes the computational cost of the MLMC method for the method of boundary integrals with numerically approximated Green's kernels.

Theorem IV.3.2. *Consider the boundary integral formulation of the linear elliptic system in* (IV.5)-(IV.6)*. Suppose that the kernels of integral operators are given by the approximate Green's function defined on the deterministic part of the domain. The ϵ-cost of the accelerated MLMC method is given by*

$$\tilde{C}_{\mathbb{ML}} \simeq \varkappa(r) \begin{cases} \epsilon^{-2} & \text{if } \beta > 3, \\ \epsilon^{-2} \left| \ln \epsilon \right|^2 & \text{if } \beta = 3, \\ \epsilon^{-2-\frac{\gamma-\beta}{\alpha}} & \text{if } \beta < 3, \end{cases}$$

where $\varkappa(r)$ is defined in Lemma IV.3.1.

Proof. Proof is the same as for (IV.36) and Theorem IV.3.1. □

If the Gaussian elimination is used for the linear solver, then $\gamma = 3$ in (IV.36) and the result of the above theorem is equivalent to $\tilde{C}_{\mathbb{ML}} \simeq \varkappa(r)\, C_{\mathbb{ML}}$. It is clear that the coefficient $\varkappa(r)$ is always less than one which indicates computational savings for all values of $r = N_{2,l}/N_{1,l}$. If $\gamma < 3$, then asymptotical complexity of the proposed technique may not be optimal. However, for the finite value of the tolerance ϵ, the value of $\varkappa(r)$ is usually small enough to guarantee $\tilde{C}_{\mathbb{ML}} < C_{\mathbb{ML}}$.

IV.4. Numerical results

In this section, we test the convergence and the cost properties of the proposed discretization scheme for the fixed deterministic boundary. Consider the problem

$$\begin{aligned} -\nabla^2 u(x) &= 0 && \text{for } x \in D, \\ u(x) &= 0 && \text{for } x \in \partial D_1, \\ u(x) &= G_1(x,\xi) && \text{for } x \in \partial D_2, \end{aligned}$$

where D is the square domain with a single aperture depicted in Figure 24. ∂D_1 is the boundary of the rectangle while ∂D_2 is the boundary of the aperture which has the following parametrization

$$\big(x(t,\omega), y(t,\omega)\big) = \big(x_c(\omega), y_c(\omega)\big) + R(t,\omega)\big(\cos(t), \sin(t)\big), \quad t \in [0, 2\pi).$$

The radius of the aperture is defined as

$$R(t,\omega) = \overline{R}(t) + \sigma_r \sum_{n=1}^{s} \Big(a_n(\omega)\cos(2\pi n t) + b_n(\omega)\sin(2\pi n t)\Big)$$

with the mean radius $\overline{R}(t)$ and the random coefficients $a_n(\omega) = \mathcal{U}(-\sqrt{3}, \sqrt{3})$, $b_n(\omega) = \mathcal{U}(-\sqrt{3}, \sqrt{3})$. Coefficient σ_r controls intensity of the random perturbation. Coordinates of

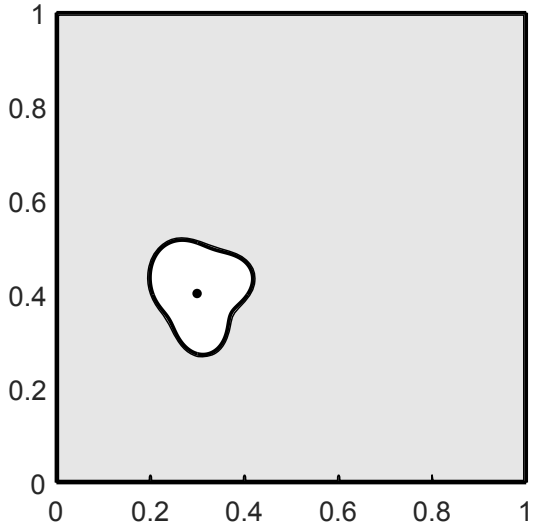


Figure 24: Geometry of the domain in Example 1.

the center are also random variables

$$x_c(\omega) = \overline{x}_c + \sigma_x \mathcal{U}(-1, 1),$$
$$y_c(\omega) = \overline{y}_c + \sigma_y \mathcal{U}(-1, 1),$$

where $(\overline{x}_c, \overline{y}_c)$ is the mean location of the center and coefficients σ_x, σ_y control deviation from the mean.

For this particular example, we set $(\overline{x}_c, \overline{y}_c) = (0.3, 0.4)$, $\sigma_x = \sigma_y = 0$, $\overline{R} = 0.1$, $\sigma_r = 0.01$ and $s = 8$. The function $G_1(x, \xi)$ is the Green's function (IV.26) for the square bounded by ∂D_1 and the location of the source is $\xi = (\overline{x}_c, \overline{y}_c)$.

The choice of the boundary condition on ∂D_2 suggests $G_1(x, \xi)$ as the analytical solution of the above problem. We can use it to test the convergence properties of the scheme in section IV.2.1 applied to the boundary integral equation with exact and approximate kernels. Figure 25 shows discretization of the boundary for both cases. For BIE with exact kernel, only the boundary of the aperture has to be discretized while for BIE with approximate kernel, it is necessary to discretize both boundaries.

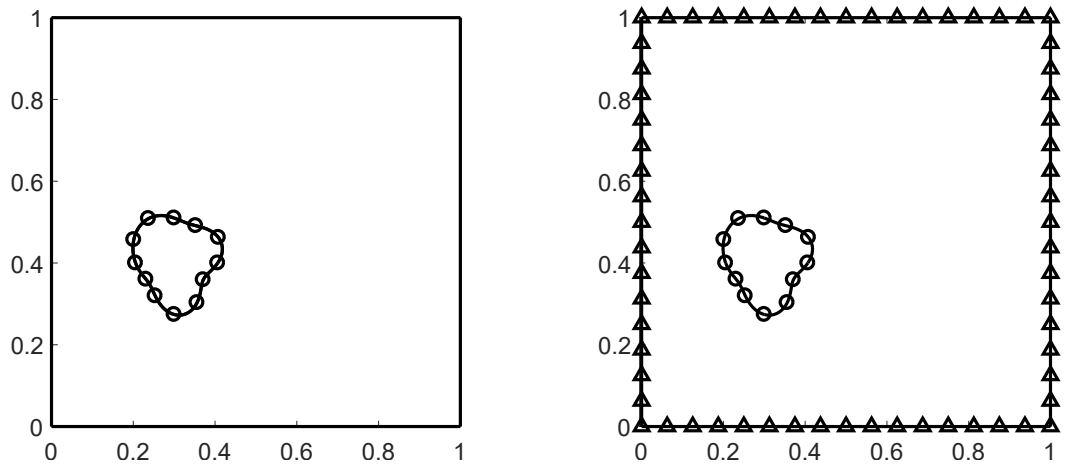


Figure 25: Discretization of the boundary for BIE with analytical (left) and numerical (right) Green's kernels.

Convergence properties of the scheme (IV.14) with the quadrature nodes in (IV.15') are illustrated in Tables 5 and 6. Table 5 provides the errors of the numerical solution of the boundary integral equation with analytical Green's kernel as well as the errors of the solution of the corresponding boundary value problem. The reference density of the potential μ is evaluated with the higher order scheme which uses the quadrature nodes in (IV.15). The apparent rates of convergence are defined as $\alpha_h = \dfrac{\log(e_l/e_{l-1})}{\log(h_l/h_{l-1})}$. It is seen that the errors have the order of convergence $\alpha = 2$ in all norms as predicted by the analysis. Table 6 provides the same information in application to BIE with approximate Green's kernel. Obviously, the errors have greater values but the difference is not large. The computed rates of convergence also satisfy the predicted values.

Results in Tables 5 and 6 are also presented graphically in Figures 27 and 28. Figure 27 illustrates the supremum norm of the error along the isocontours depicted in Figure 26. Superiority of the analytical Green's kernel is obvious near the deterministic boundary ∂D_1 but both approaches show good results far from the boundaries. Figure 28 confirms that approximation of the Green's kernel does not change the order of the numerical scheme and does not have serious impact on the accuracy of the numerical solution.

Table 5: Convergence with analytical Green's kernel.

N_2	$\|\mu - \mu_h\|_{L^\infty}$	rate	$\|u - u_h\|_{L^\infty}$	rate	$\|u - u_h\|_{H^1}$	rate
8	5.074e-1	-	3.323e-2	-	9.462e-2	-
20	7.008e-2	1.87	4.506e-3	2.36	2.992e-2	1.56
68	5.215e-3	2.06	1.820e-4	3.33	2.058e-3	2.64
260	3.533e-4	2.01	4.898e-6	2.34	1.312e-5	3.87
1028	2.258e-5	2.00	3.133e-7	2.00	8.176e-7	2.00
4100	1.428e-6	1.99	1.970e-8	2.00	5.140e-8	2.00

N_2 denotes the number of degrees of freedom on the random part of the boundary ∂D_2 and u is the solution of the boundary value problem in the form of the single layer potential (IV.8) with density μ.

Table 6: Convergence with numerical Green's kernel.

N_1	N_2	$\|\mu - \mu_h\|_{L^\infty}$	rate	$\|u - u_h\|_{L^\infty}$	rate	$\|u - u_h\|_{H^1}$	rate
48	8	5.071e-1	-	3.323e-2	-	1.048e-1	-
104	20	7.004e-2	1.87	4.507e-3	2.36	3.460e-2	1.53
328	68	5.212e-3	2.07	1.820e-4	3.33	2.462e-3	2.64
1232	260	3.530e-4	2.01	4.901e-6	2.34	1.335e-5	3.98
4848	1028	2.257e-5	2.00	3.135e-7	2.00	8.319e-7	1.99
19336	4100	1.422e-6	1.99	1.971e-8	2.00	5.230e-8	2.00

N_1 and N_2 denote the number of degrees of freedom on deterministic and random parts of the boundary ∂D_1 and ∂D_2 respectively. u is the solution of the boundary value problem in the form of the single layer potential (IV.8) with density μ.

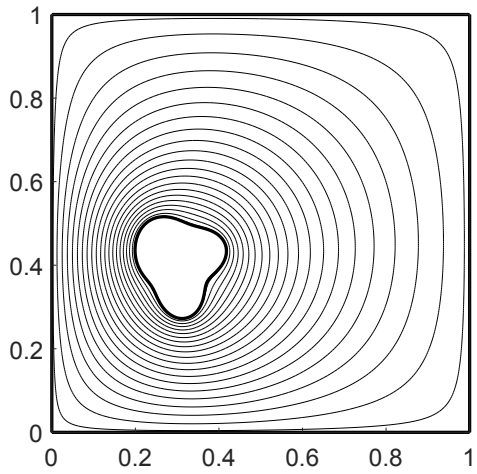


Figure 26: Isocontours of the aperture.

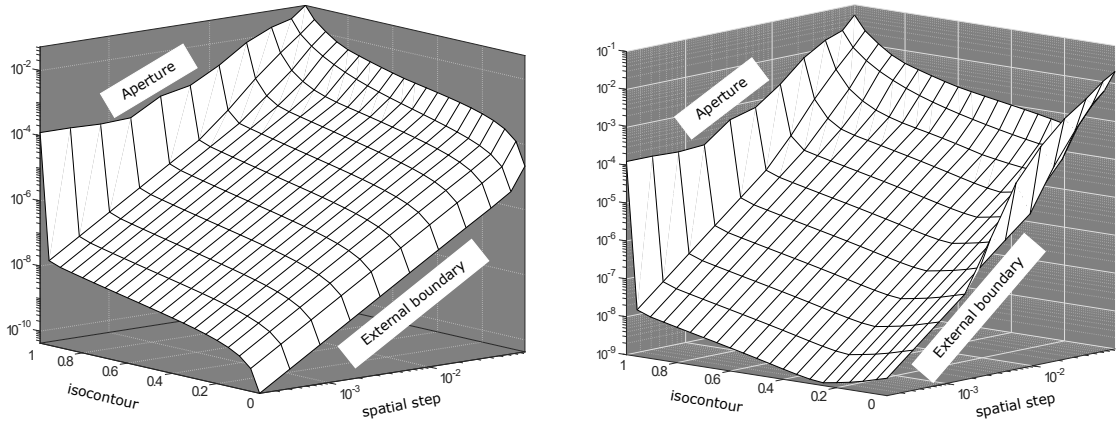


Figure 27: Maximum errors of the spatial approximation along the isocontours of the boundary for BIE with analytical (left) and numerical (right) Green's kernels.

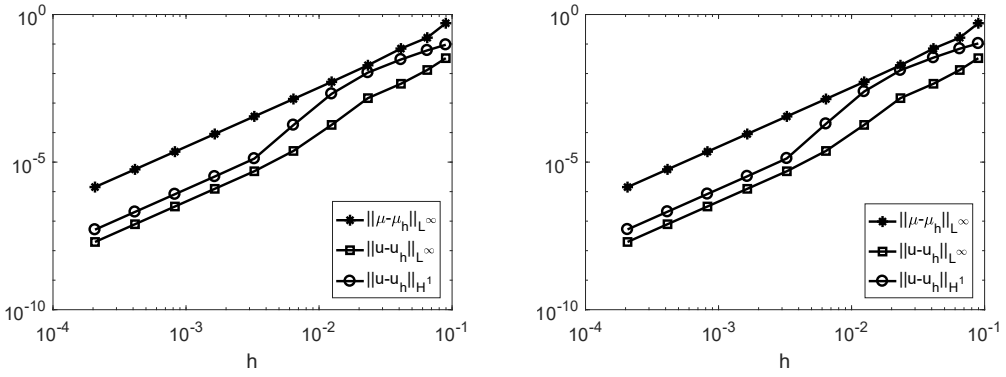


Figure 28: Convergence of the spatial approximation with analytical (left) and numerical (right) Green's kernels.

The costs C^a of assembling the matrix and C^s of solving the resulting system are illustrated in Figure 30. As predicted, the cost C^s is asymptotically dominant for BIE with analytical Green's kernel while the cost C^a dominates in the case of numerical Green's function. To verify the cost analysis of the proposed acceleration technique, we fix the discretization on the boundary ∂D_1 and sequentially refine the grid on the boundary ∂D_2. Tables 7 and 8 provide the computational savings for various values of the ratios $\frac{N}{N_2} \simeq \frac{|\partial D|}{|\partial D_2|}$ and $\frac{N_1}{N_2} \simeq \frac{|\partial D_1|}{|\partial D_2|}$. It is seen that the proposed technique is computationally efficient for all tested values of this ratio. Computational savings in solving the linear system are exceptionally good for both analytical and numerical Green's kernels. Also, as expected, the cost gain of assembling the system is much better in the first case. In any case, the total cost of the proposed scheme is smaller than the cost of the standard approach for both variants of the Green's kernel. This is clearly seen in Figure 29 which shows the computational speedup of the accelerated algorithm for boundary integral equations with both analytical and approximate kernels.

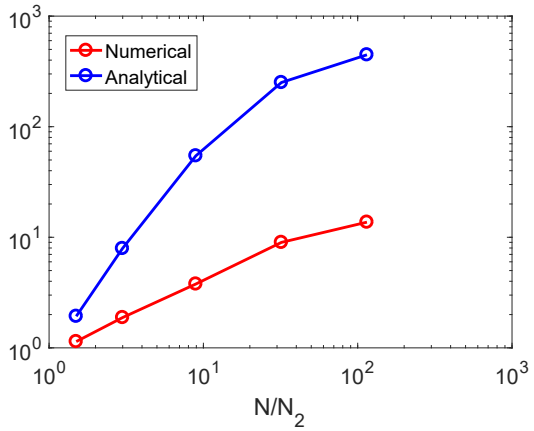


Figure 29: Computational speedup of the accelerated algorithm for the problem in Example 1. Blue and red colors correspond to the boundary integral equations with analytical and numerical Green's kernels.

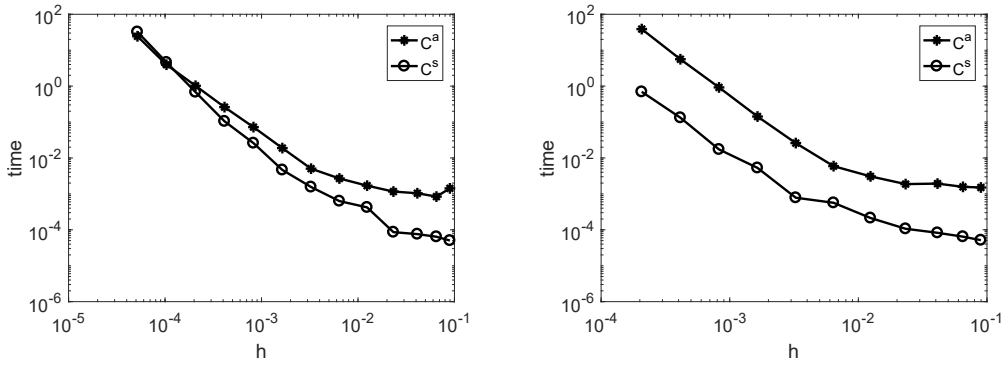


Figure 30: The costs C^a and C^s of assembling and solving the linear system for the scheme with analytical (left) and numerical (right) Green's kernels.

Table 7: Computational gain for BIE with analytical Green's kernel.

N_1	N_2	N/N_2	$C^a/\tilde{C}^a$	$C^s/\tilde{C}^s$	$C_{\mathbb{ML}}/\tilde{C}_{\mathbb{ML}}$
4096	36	114.78	205.74	4044.82	446.94
4096	132	32.03	116.43	1367.12	250.97
4096	516	8.94	27.37	189.69	54.56
4096	2052	2.99	3.79	17.87	7.92
4096	8196	1.50	0.85	2.92	1.93

Table 8: Computational gain for BIE with numerical Green's kernel.

N_1	N_2	N_1/N_2	$C^a/\tilde{C}^a$	$C^s/\tilde{C}^s$	$C_{\mathbb{ML}}/\tilde{C}_{\mathbb{ML}}$
4096	36	113.78	5.92	6379.26	13.70
4096	132	31.03	3.73	2105.12	8.99
4096	516	7.94	1.60	228.94	3.78
4096	2052	1.99	0.68	17.93	1.88
4096	8196	0.50	0.35	2.93	1.14

N_1 and N_2 denote the number of degrees of freedom on deterministic and random parts of the boundary ∂D_1 and ∂D_2 respectively. $N = N_1 + N_2$ is the total number of degrees of freedom. C^a, C^s and $C_{\mathbb{ML}}$ denote the cost of assembling the system, the cost of solving the system and the total cost of the estimator for the standard MLMC algorithm. $\tilde{C}^a$, $\tilde{C}^s$ and $\tilde{C}_{\mathbb{ML}}$ denote the same quantities for the accelerated algorithm.

CHAPTER V

CONCLUSIONS AND FUTURE WORK

Multilevel Monte Carlo method is a recently proposed technique which has already proven its efficiency for a broad class of problems with a large number of stochastic dimensions. Many of these problems have internal structure which can be efficiently combined with that of the multilevel Monte Carlo method to significantly improve the overall performance of the algorithm. Three of such problems were studied in this dissertation.

In the first problem, we considered the issue of stochastic stiffness which can drastically reduce the number of admissible discretizations in the multilevel hierarchy. We established the necessary criteria for the pathwise integrators for such systems and proposed the novel implicit-explicit composite integrator which satisfies these requirements. The provided numerical experiments indicated the high empirical efficiency of the proposed technique. Nevertheless, the optimal selection of the parameters of the composite integration remains an open question. The rigorous theoretical study of this problem is still required and can be considered as a possible future research in this direction.

The repetitive sampling framework of the Monte Carlo integration was used in the second problem for the acceleration of iterative solvers arising after discretization of the stationary PDEs with random input data. The proposed technique for the estimation of initial guesses for such solvers proved to be efficient for linear problems with low to moderate stochastic dimensionality and for nonlinear problems with solutions near instability. As a possible extension of this approach, it would be interesting to perform theoretical analysis for more general nonlinear problems and to study the optimal distribution of the tolerance between iterative solvers at different levels.

The last project was devoted to the acceleration of the MLMC method in application to PDEs in randomly perturbed domains. In this case, the reduction of the cost of the MLMC estimator was achieved by reformulating the original boundary value problem as a boundary

integral equation on the random part of the domain only. The proposed technique proved to be extremely efficient for domains with known analytical Green's functions but also showed very good results for general domains. An obvious extension of this approach would be to consider problems with more general stationary and time-dependent operators.

BIBLIOGRAPHY

[AA07] Nitin Agarwal and N.R. Aluru. A stochastic Lagrangian approach for geometrical uncertainties in electrostatics. *Journal of Computational Physics*, 226(1):156 – 179, 2007.

[AB13] Assyr Abdulle and Adrian Blumenthal. Stabilized multilevel Monte Carlo method for stiff stochastic differential equations. *Journal of Computational Physics*, 251:445 – 460, 2013.

[ABB11] Assyr Abdulle, Adrian Blumenthal, and Evelyn Buckwar. The multilevel Monte Carlo method for stochastic differential equations driven by jump–diffusion processes. Technical report, ANMC-EPFL, 2011.

[AC08] Assyr Abdulle and Stephane Cirilli. S-ROCK: Chebyshev methods for stiff stochastic differential equations. *SIAM Journal on Scientific Computing*, 30(2):997–1014, 2008.

[AGK11] David F. Anderson, Arnab Ganguly, and Thomas G. Kurtz. Error analysis of tau-leap simulation methods. *The Annals of Applied Probability*, 21(6):2226–2262, 2011.

[AH12] David F. Anderson and Desmond J. Higham. Multilevel Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics. *Multiscale Modeling & Simulation*, 10(1):146–179, 2012.

[AHL10] Assyr Abdulle, Yucheng Hu, and Tiejun Li. Chebyshev methods with discrete noise: the tau-ROCK methods. *Journal of Computational Mathematics*, 28(2):195–217, 2010.

[AHS14] David F. Anderson, Desmond J. Higham, and Yu Sun. Complexity of multilevel Monte Carlo tau-leaping. *SIAM Journal on Numerical Analysis*, 52(6):3106–3127, 2014.

[AK15] David F. Anderson and Thomas Kurtz. *Stochastic Analysis of Biochemical Systems*, volume 1.2 of *Stochastics in Biological Systems*. Springer International Publishing, 2015.

[AN15] Martin Altmayer and Andreas Neuenkirch. Multilevel Monte Carlo Quadrature of Discontinuous Payoffs in the Generalized Heston Model Using Malliavin Integration by Parts. *SIAM Journal on Financial Mathematics*, 6(1):22–52, 2015.

[App09] David Applebaum. *Lévy Processes and Stochastic Calculus*, volume 166 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2009.

[APR09] Sk. Safique Ahmad, Nigam Chandra Parida, and Soumyendu Raha. The fully implicit stochastic α-method for stiff stochastic differential equations. *Journal of Computational Physics*, 228(22):8263–8282, 2009.

[Arn74] Ludwig Arnold. *Stochastic differential equations*. John Wiley & Sons, Inc., 1974.

[AS91] Kendall E Atkinson and Ian H Sloan. The numerical solution of first-kind logarithmic-kernel integral equations on smooth open arcs. *Mathematics of computation*, 56(193):119–139, 1991.

[Atk97] Kendall E Atkinson. The numerical solution of boundary integral equations. In *Institute of mathematics and its applications conference series*, volume 63, pages 223–260. Oxford University Press, 1997.

[Avi09] Rainer Avikainen. On irregular functionals of SDEs and the Euler scheme. *Finance and Stochastics*, 13(3):381–401, 2009.

[AVZ13] A. Abdulle, G. Vilmart, and K. C. Zygalakis. Mean-square A–stable diagonally drift-implicit integrators of weak second order for stiff Itô stochastic differential equations. *BIT Numerical Mathematics*, 53(4):827–840, 2013.

[AVZ15] Assyr Abdulle, Gilles Vilmart, and Konstantinos C. Zygalakis. Long time accuracy of Lie–Trotter splitting methods for Langevin dynamics. *SIAM Journal on Numerical Analysis*, 53(1):1–16, 2015.

[Bai75] Norman TJ Bailey. *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 1975.

[BB98] K. Burrage and P.M. Burrage. General order conditions for stochastic Runge–Kutta methods for both commuting and non-commuting stochastic ordinary differential equation systems. *Applied Numerical Mathematics*, 28(2):161 – 177, 1998.

[BHGH92] R.J Breeding, J.C Helton, E.D Gorham, and F.T Harper. Summary description of the methods used in the probabilistic risk assessments for NUREG-1150. *Nuclear Engineering and Design*, 135(1):1 – 27, 1992.

[BHMT16] Chiheb Ben Hammouda, Alvaro Moraes, and Raúl Tempone. Multilevel hybrid split-step implicit tau-leap. *Numerical Algorithms*, pages 1–34, 2016.

[BJL+16] Alexandros Beskos, Ajay Jasra, Kody Law, Raul Tempone, and Yan Zhou. Multilevel sequential Monte Carlo samplers. *Stochastic Processes and their Applications*, 2016. In press.

[BKO+96] T. Belytschko, Y. Krongauz, D. Organ, M. Fleming, and P. Krysl. Meshless methods: An overview and recent developments. *Computer Methods in Applied Mechanics and Engineering*, 139(14):3 – 47, 1996.

[BL58] J. M. Berger and G. J. Lasher. The use of discrete Green's functions in the numerical solution of Poisson's equation. *Illinois J. Math.*, 2(4A):593–607, 11 1958.

[BLP07] Nicola Bruti-Liberati and Eckhard Platen. Approximation of jump diffusions in finance and economics. *Computational Economics*, 29(3):283–312, 2007.

[BLP08] Nicola Bruti-Liberati and Eckhard Platen. Strong predictor-corrector Euler methods for stochastic differential equations. *Stochastics and Dynamics*, 08(03):561–581, 2008.

[BN14] Denis Belomestny and Tigran Nagapetyan. Multilevel path simulation for weak approximation schemes. *arXiv preprint arXiv:1406.2581*, 2014.

[BNR07] Maya Briani, Roberto Natalini, and Giovanni Russo. Implicit–explicit numerical schemes for jump–diffusion processes. *Calcolo*, 44(1):33–57, 2007.

[BNT07] Ivo Babuška, Fabio Nobile, and Raúl Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 45(3):1005–1034, 2007.

[BR08] Susanne C. Brenner and Scott L. Ridgway. *The Mathematical Theory of Finite Element Methods*, volume 15 of *Texts in Applied Mathematics*. New York: Springer, 2008.

[BR11] Evelyn Buckwar and Martin G. Riedler. Runge-Kutta methods for jump-diffusion differential equations. *Journal of Computational and Applied Mathematics*, 236(6):1155 – 1182, 2011.

[BRW10] Evelyn Buckwar, Andreas Rößler, and Renate Winkler. Stochastic Runge-Kutta methods for Itô SODEs with small noise. *SIAM Journal on Scientific Computing*, 32(4):1789–1808, 2010.

[BS73] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[BST13] David A. Barajas-Solano and Daniel M. Tartakovsky. Computing Green's functions for flow in heterogeneous composite media. *International Journal for Uncertainty Quantification*, 3(1):39–46, 2013.

[BSZ11] Andrea Barth, Christoph Schwab, and Nathaniel Zollinger. Multi-level Monte Carlo Finite Element method for elliptic PDEs with stochastic coefficients. *Numerische Mathematik*, 119(1):123–161, 2011.

[BT01] Kevin Burrage and Tianhai Tian. The composite Euler method for stiff stochastic differential equations. *Journal of Computational and Applied Mathematics*, 131(12):407 – 426, 2001.

[BT04] Kevin Burrage and Tianhai Tian. Implicit stochastic Runge–Kutta methods for stochastic differential equations. *BIT Numerical Mathematics*, 44(1):21–39, 2004.

[BTZ04] Ivo Babuška, Raúl Tempone, and Georgios E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM Journal on Numerical Analysis*, 42(2):800–825, 2004.

[BW06] Evelyn Buckwar and Renate Winkler. Multistep methods for SDEs and their application to problems with small noise. *SIAM Journal on Numerical Analysis*, 44(2):779–803, 2006.

[BW07] Evelyn Buckwar and Renate Winkler. Improved linear multi-step methods for stochastic ordinary differential equations. *Journal of Computational and Applied Mathematics*, 205(2):912 – 922, 2007.

[Caf98] Russel E. Caflisch. Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*, 7:1–49, 1998.

[CC14] Julio E Castrillon-Candas. A sparse grid collocation method for parabolic pdes with random domain deformations. *arXiv preprint arXiv:1408.6818*, 2014.

[CCNT15] Julio E Castrillon-Candas, Fabio Nobile, and Raul F Tempone. Analytic regularity and collocation approximation for elliptic PDEs with random domain deformations. *arXiv preprint arXiv:1312.7845*, 2015.

[CD15] Albert Cohen and Ronald DeVore. Approximation of high-dimensional parametric PDEs. *Acta Numerica*, 24:1–159, 2015.

[CET15] J. B. Collins, Donald Estep, and Simon Tavener. A posteriori error estimation for a cut cell finite volume method with uncertain interface location. *International Journal for Uncertainty Quantification*, 5(5):415–432, 2015.

[CF09] Claudio Canuto and Davide Fransos. Numerical solution of partialdifferential equations in random domains: An application to wind engineering. *Communications in Computational Physics*, 5(2-4):515–531, 2009.

[CGP05a] Yang Cao, Dan Gillespie, and Linda Petzold. Multiscale stochastic simulation algorithm with stochastic partial equilibrium assumption for chemically reacting systems. *Journal of Computational Physics*, 206(2):395 – 411, 2005.

[CGP05b] Yang Cao, Daniel T. Gillespie, and Linda R. Petzold. Accelerated stochastic simulation of the stiff enzyme-substrate reaction. *The Journal of Chemical Physics*, 123(14), 2005.

[CGP05c] Yang Cao, Daniel T. Gillespie, and Linda R. Petzold. The slow-scale stochastic simulation algorithm. *The Journal of Chemical Physics*, 122(1), 2005.

[CGST11] K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science*, 14(1):3, 2011.

[CHAN+15] Nathan Collier, Abdul-Lateef Haji-Ali, Fabio Nobile, Erik von Schwerin, and Raúl Tempone. A continuation multilevel Monte Carlo algorithm. *BIT Numerical Mathematics*, 55(2):399–432, 2015.

[Che93] R.S.-C. Cheng. On using a modified Nyström method to solve the 2-D potential problem. *J. Integral Equations Applications*, 5(2):167–194, 06 1993.

[Che94] R.S.-C. Cheng. Some numerical results using the modified Nyström method to solve the 2-D potential problem. *Engineering Analysis with Boundary Elements*, 14(4):335 – 342, 1994.

[Cho94] Alexandre J. Chorin. *Vorticity and Turbulence*, volume 103 of *Applied Mathematical Sciences*. Springer-Verlag New York, 1994.

[Cho14] Pao-Liu Chow. *Stochastic Partial Differential Equations*. Advances in Apllied Mathematics. Chapman and Hall/CRC, 2014.

[Chr92] G. Christakos. *Random Field Models in Earth Sciences*. Academic Press, San Diego, CA, 1992.

[CJ08] F. Carbonell and J. C. Jimenez. Weak local linear discretizations for stochastic differential equations with jumps. *J. Appl. Probab.*, 45(1):201–210, 2008.

[CK07] Claudio Canuto and Tomas Kozubek. A fictitious domain approach to the numerical solution of PDEs in stochastic domains. *Numerische Mathematik*, 107(2):257–293, 2007.

[CM47] Robert H Cameron and William T Martin. The orthogonal development of non-linear functionals in series of Fourier–Hermite functionals. *Annals of Mathematics*, 48(2):385–392, 1947.

[CMM09] Thierry Crestaux, Olivier Le Matre, and Jean-Marc Martinez. Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering & System Safety*, 94(7):1161 – 1172, 2009.

[CP08] Yang Cao and Linda Petzold. Slow-scale tau-leaping method. *Computer Methods in Applied Mechanics and Engineering*, 197(4344):3472 – 3479, 2008.

[CPRG04] Yang Cao, Linda R. Petzold, Muruhan Rathinam, and Daniel T. Gillespie. The numerical stability of leaping methods for stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 121(24):12169–12178, 2004.

[CR11] Ioana Cipcigan and Muruhan Rathinam. Uniform convergence of interlaced Euler method for stiff stochastic differential equations. *Multiscale Modeling & Simulation*, 9(3):1217–1252, 2011.

[DBO01] Manas K. Deb, Ivo M. Babuka, and J.Tinsley Oden. Solution of stochastic partial differential equations using Galerkin finite element techniques. *Com-*

puter Methods in Applied Mechanics and Engineering, 190(48):6359 – 6372, 2001.

[Der11] Steffen Dereich. Multilevel Monte Carlo algorithms for Lévy–driven SDEs with Gaussian correction. *Ann. Appl. Probab.*, 21(1):283–311, 02 2011.

[DH11] Steffen Dereich and Felix Heidenreich. A multilevel Monte Carlo algorithm for Lévy-driven stochastic differential equations. *Stochastic Processes and their Applications*, 121(7):1565 – 1587, 2011.

[DKM$^+$09] Robert C. Dalang, Davar Khoshnevisan, Carl Mueller, David Nualart, and Yimin Xiao. *A Minicourse on Stochastic Partial Differential Equations*, volume 1962 of *Lecture Notes in Mathematics*. Springer-Verlag Berlin Heidelberg, 2009.

[DKS13] Josef Dick, Frances Y. Kuo, and Ian H. Sloan. High-dimensional integration: The quasi-Monte Carlo way. *Acta Numerica*, 22:133–288, 005 2013.

[DKS14] Konstantinos Dareiotis, Chaman Kumar, and Sotirios Sabanis. On tamed Euler approximations of SDEs driven by Lévy noise with applications to delay equations. *arXiv preprint arXiv:1403.0498*, 2014.

[DKST15] T. J. Dodwell, C. Ketelsen, R. Scheichl, and A. L. Teckentrup. A Hierarchical Multilevel Markov Chain Monte Carlo Algorithm with Applications to Uncertainty Quantification in Subsurface Flow. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):1075–1108, 2015.

[DPJR10] Giuseppe Da Prato, Arnulf Jentzen, and Michael Röckner. A mild Itô formula for SPDEs. *arXiv preprint arXiv:1009.3526*, 2010.

[DR09] Kristian Debrabant and Andreas Rößler. Diagonally drift-implicit Runge-Kutta methods of weak order one and two for it sdes and stability analysis. *Applied Numerical Mathematics*, 59(3):595 – 607, 2009.

[DR15] Kristian Debrabant and Andreas Rößler. On the acceleration of the multi-level Monte Carlo method. *J. Appl. Probab.*, 52(2):307–322, 06 2015.

[Duf01] Dean G Duffy. *Greens functions with applications*. CRC Press, 2001.

[EES00] Michael Eiermann, Oliver G. Ernst, and Olaf Schneider. Analysis of acceleration strategies for restarted minimal residual methods. *Journal of Computational and Applied Mathematics*, 123(12):261 – 292, 2000.

[EEU07] Michael Eiermann, OliverG. Ernst, and Elisabeth Ullmann. Computational aspects of the stochastic finite element method. *Computing and Visualization in Science*, 10(1):3–15, 2007.

[EJMT15] Yalchin Efendiev, Bangti Jin, Presho Michael, and Xiaosi Tan. Multilevel Markov chain Monte Carlo method for high-contrast single-phase flow problems. *Communications in Computational Physics*, 17(1):259–286, 001 2015.

[EK09] Stewart N Ethier and Thomas G Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 2009.

[EKPQ97] N. El Karoui, S. Peng, and M. C. Quenez. Backward stochastic differential equations in finance. *Mathematical Finance*, 7(1):1–71, 1997.

[EL16] Utku Erdoğan and Gabriel J Lord. A new class of exponential integrators for stochastic differential equations with multiplicative noise. *arXiv preprint arXiv:1608.07096*, 2016.

[ELVE05] Weinan E, Di Liu, and Eric Vanden-Eijnden. Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates. *The Journal of Chemical Physics*, 123(19), 2005.

[ELVE07] Weinan E, Di Liu, and Eric Vanden-Eijnden. Nested stochastic simulation algorithms for chemical kinetic systems with multiple time scales. *Journal of Computational Physics*, 221(1):158 – 180, 2007.

[EU10] Oliver G. Ernst and Elisabeth Ullmann. Stochastic Galerkin Matrices. *SIAM Journal on Matrix Analysis and Applications*, 31(4):1848–1872, 2010.

[Far86] Reinhard Farwig. Rate of convergence of Shepard's global interpolation formula. *Mathematics of Computation*, 46(174):577–590, 1986.

[Fas07] Gregory F Fasshauer. *Meshfree approximation methods with MATLAB*. World Scientific Publishing Co., Inc., 2007.

[FDKI16] Hillary Fairbanks, Alireza Doostan, Christian Ketelsen, and Gianluca Iaccarino. A low-rank control variate for Multilevel Monte Carlo simulation of high-dimensional uncertain systems. *arXiv preprint arXiv:1611.02213*, 2016.

[Fis96] George Fishman. *Monte Carlo: Concepts, Algorithms, and Applications*. Springer Series in Operations Research and Financial Engineering. Springer-Verlag New York, 1996.

[FM16] Gregory E Fasshauer and Michael McCourt. *Kernel-based approximation methods using Matlab*. World Scientific, 2016.

[Fox99] Bennett L. Fox. *Strategies for Quasi-Monte Carlo*, volume 22 of *International Series in Operations Research & Management Science*. Springer US, 1999.

[Fri75] Avner Friedman. *Stochastic differential equations and applications*. Academic Press, 1975.

[FST05] Philipp Frauenfelder, Christoph Schwab, and Radu Alexandru Todor. Finite elements for elliptic problems with stochastic coefficients. *Computer Methods in Applied Mechanics and Engineering*, 194(25):205 – 228, 2005.

[Gar09] Crispin Gardiner. *Stochastic methods: a handbook for the natural and social sciences*, volume 13 of *Springer Series in Synergetics*. Springer-Verlag Berlin Heidelberg, 2009.

[GB00] Michael A. Gibson and Jehoshua Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry A*, 104(9):1876–1889, 2000.

[GD98] R. Ghanem and S. Dham. Stochastic finite element analysis for multiphase flow in heterogeneous porous media. *Transport in Porous Media*, 32(3):239–262, 1998.

[GEI15] G Geraci, M Eldred, and G Iaccarino. A multifidelity control variate approach for the multilevel Monte Carlo technique. Technical report, Stanford University, 2015.

[GHM09] Michael B. Giles, Desmond J. Higham, and Xuerong Mao. Analysing multi–level Monte Carlo for options with non-globally Lipschitz payoff. *Finance and Stochastics*, 13(3):403–413, 2009.

[Gil77] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[Gil01] Daniel T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001.

[Gil07] Daniel T. Gillespie. Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry*, 58(1):35–55, 2007.

[Gil08a] Michael B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[Gil08b] Mike Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 343–358. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[Gil15] Michael B Giles. Multilevel Monte Carlo methods. *Acta Numerica*, 24:259–328, 2015.

[GJS85] Ivan G. Graham, Stephen Joe, and Ian H. Sloan. Iterated Galerkin versus iterated collocation for integral equations of the second kind. *IMA Journal of Numerical Analysis*, 5(3):355–369, 1985.

[GJWZ16] D. Galindo, P. Jantsch, C. G. Webster, and G. Zhang. Accelerating stochastic collocation methods for partial differential equations with random input data. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):1111–1137, 2016.

[GK69] S. Gallof and B. Kaplan. Numerical solution and utilization of Green's functions. *International Journal for Numerical Methods in Engineering*, 1(2):169–175, 1969.

[GK96] Roger G. Ghanem and Robert M. Kruger. Numerical solution of spectral stochastic finite element systems. *Computer Methods in Applied Mechanics and Engineering*, 129(3):289 – 303, 1996.

[GK08] Dror Givon and Ioannis G. Kevrekidis. Multiscale integration schemes for jump-diffusion systems. *Multiscale Modeling & Simulation*, 7(2):495–516, 2008.

[GKK06] Dror Givon, Ioannis G. Kevrekidis, and Raz Kupferman. Strong convergence of projective integration schemes for singularly perturbed stochastic differential systems. *Commun. Math. Sci.*, 4(4):707–729, 12 2006.

[GKS04] Dror Givon, Raz Kupferman, and Andrew Stuart. Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity*, 17(6):R55–R127, 2004.

[GLMZ16] Qian Guo, Wei Liu, Xuerong Mao, and Weijun Zhan. Multi-level Monte Carlo methods with the Truncated Euler-Maruyama Scheme for Stochastic Differential Equations. *arXiv preprint arXiv:1611.07833*, 2016.

[GLW16] Michael B. Giles, Christopher Lester, and James Whittle. Non-nested adaptive timesteps in Multilevel Monte Carlo computations. In *Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Leuven, Belgium, April 2014*, pages 303–314. Springer International Publishing, 2016.

[GNT05] Mihai Gradinaru, Ivan Nourdin, and Samy Tindel. Ito's- and Tanaka's-type formulae for the stochastic heat equation: The linear case. *Journal of Functional Analysis*, 228(1):114 – 143, 2005.

[Gou05] John Goutsias. Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems. *The Journal of Chemical Physics*, 122(18), 2005.

[GP11] Andrew D. Gordon and Catherine E. Powell. On solving stochastic collocation systems with algebraic multigrid. *IMA Journal of Numerical Analysis*, 2011.

[GR86] V Girault and PA Raviart. *Finite element methods for Navier-Stokes equations: Theory and algorithms*, volume 5. Springer-Verlag, 1986.

[Gri06] Mircea Grigoriu. Evaluation of Karhunen–Loéve, spectral, and sampling representations for stochastic processes. *Journal of Engineering Mechanics*, 132(2):179–189, 2006.

[GS72] Iosif I. Gihman and Anatolij V. Skorohod. *Stochastic Differential Equations*, volume 72 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 2. Folge*. Springer-Verlag Berlin Heidelberg, 1972.

[GS91a] Roger G. Ghanem and Pol Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer New York, 1991.

[GS91b] Roger G. Ghanem and Pol D. Spanos. Spectral stochastic finite element formulation for reliability analysis. *Journal of Engineering Mechanics*, 117(10):2351–2372, 1991.

[GS14] Michael B. Giles and Lukasz Szpruch. Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *Ann. Appl. Probab.*, 24(4):1585–1620, 08 2014.

[GTMW96] J. P. Gwo, L. E. Toran, M. D. Morris, and G. V. Wilson. Subsurface stormflow modeling with sensitivity analysis using a Latin-hypercube sampling technique. *Ground Water*, 34(5):811–818, 1996.

[Gun89] M. D. Gunzburger. *Finite element methods for viscous incompressible flows: A guide to theory, practice, and algorithms*, volume 91. Boston: Academic Press, 1989.

[GWZ14] Max D. Gunzburger, Clayton G. Webster, and Guannan Zhang. Stochastic finite element methods for partial differential equations with random input data. *Acta Numerica*, 23:521–650, 2014.

[GX14] Mike Giles and Yuan Xia. Multilevel Monte Carlo for exponential Lévy models. *arXiv preprint arXiv:1403.5309*, 2014.

[Gyo02] L Gyorfi. *A Distribution-Free Theory of Nonparametric Regression*. Springer Series in Statistics. Springer New York, 2002.

[Hak83] Hermann Haken. *Advanced Synergetics: Instability Hierarchies of Self-Organizing Systems and Devices*, volume 20 of *Springer Series in Synergetics*. Springer-Verlag Berlin Heidelberg, 1983.

[HANT16] Abdul-Lateef Haji-Ali, Fabio Nobile, and Raúl Tempone. Multi–index Monte Carlo: when sparsity meets sampling. *Numerische Mathematik*, 132(4):767–806, 2016.

[HANvST16] Abdul-Lateef Haji-Ali, Fabio Nobile, Erik von Schwerin, and Raúl Tempone. Optimization of mesh hierarchies in multilevel Monte Carlo samplers. *Stochastics and Partial Differential Equations Analysis and Computations*, 4(1):76–112, 2016.

[Har13] Friedel Hartmann. *Green's Functions and Finite Elements*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[HB98] J. E. Hurtado and A. H. Barbat. Monte Carlo techniques in computational stochastic mechanics. *Archives of Computational Methods in Engineering*, 5(1):3, 1998.

[HG11] Lin Hu and Siqing Gan. Convergence and stability of the balanced methods for stochastic differential equations with jumps. *International Journal of Computer Mathematics*, 88(10):2089–2108, 2011.

[HH12] Amir Haghighi and S. Mohammad Hosseini. A class of split-step balanced methods for stiff stochastic differential equations. *Numerical Algorithms*, 61(1):141–162, 2012.

[HHR16] Amir Haghighi, Seyed Mohammad Hosseini, and Andreas Rößler. Diagonally drift-implicit Runge-Kutta methods of strong order one for stiff stochastic differential systems. *Journal of Computational and Applied Mathematics*, 293:82 – 93, 2016.

[HJK11] Martin Hutzenthaler, Arnulf Jentzen, and Peter E. Kloeden. Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 467(2130):1563–1576, 2011.

[HJK12] Martin Hutzenthaler, Arnulf Jentzen, and Peter E. Kloeden. Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. *Ann. Appl. Probab.*, 22(4):1611–1641, 08 2012.

[HJK13] Martin Hutzenthaler, Arnulf Jentzen, and Peter E. Kloeden. Divergence of the multilevel Monte Carlo Euler method for nonlinear stochastic differential equations. *Ann. Appl. Probab.*, 23(5):1913–1966, 2013.

[HK05] Desmond Higham and Peter. Kloeden. Numerical methods for nonlinear stochastic differential equations with jumps. *Numerische Mathematik*, 101(1):101–119, 2005.

[HK06] Desmond Higham and Peter Kloeden. Convergence and stability of implicit methods for jump-diffusion systems. *Int. J. Numer. Anal. Model*, 3(2):125–140, 2006.

[HLN12] Yaozhong Hu, Fei Lu, and David Nualart. Feynman-Kac formula for the heat equation driven by fractional noise with hurst parameter H $<$ 1/2. *Ann. Probab.*, 40(3):1041–1068, 2012.

[HM00] Achintya Haldar and Sankaran Mahadevan. *Reliability assessment using stochastic finite element analysis*. John Wiley & Sons, 2000.

[HMS02] Desmond J. Higham, Xuerong Mao, and Andrew M. Stuart. Strong Convergence of Euler-Type Methods for Nonlinear Stochastic Differential Equations. *SIAM Journal on Numerical Analysis*, 40(3):1041–1063, 2002.

[HN14] Xiaoying Han and Habib N. Najm. Dynamical structures in stochastic chemical reaction systems. *SIAM Journal on Applied Dynamical Systems*, 13(3):1328–1351, 2014.

[Hon05] Riki Honda. Stochastic BEM with spectral approach in elastostatic and elastodynamic problems with geometrical uncertainty. *Engineering Analysis with Boundary Elements*, 29(5):415 – 427, 2005.

[HØUZ10] H. Holden, Bernt Øksendal, Jan Ubøe, and Tusheng Zhang. *Stochastic Partial Differential Equations: A Modeling, White Noise Functional Approach*. Universitext. Springer-Verlag New York, 2010.

[HP81] J.Michael Harrison and Stanley R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3):215 – 260, 1981.

[HPS14] Hemut Harbrecht, Michael Peters, and Markus Siebenmorgen. Numerical solution of elliptic diffusion problems on random domains. *Preprint, Mathematisches Institut, Universitt Basel, Switzerland*, 2014.

[HSS08] Helmut Harbrecht, Reinhold Schneider, and Christoph Schwab. Sparse second moment analysis for elliptic problems in stochastic domains. *Numerische Mathematik*, 109(3):385–414, 2008.

[HSS13] Viet Ha Hoang, Christoph Schwab, and Andrew M Stuart. Complexity analysis of accelerated MCMC methods for Bayesian inversion. *Inverse Problems*, 29(8):085010, 2013.

[HSW07] Jialin Hong, Rudolf Scherer, and Lijin Wang. Predictor-corrector methods for a linear stochastic oscillator with additive noise. *Mathematical and Computer Modelling*, 46(56):738 – 764, 2007.

[HvSST13] Hakon Hoel, Erik von Schwerin, Anders Szepessy, and Raul Tempone. Implementation and analysis of an adaptive multilevel Monte Carlo algorithm. *Monte Carlo Methods and Applications*, 20(1):1–41, 2013.

[HW96] Ernst Hairer and Gerhard Wanner. *Solving Ordinary Differential Equations II, Stiff and Differential-Algebraic Systems*. Springer-Verlag, Berlin, 1996.

[Jas63] M. A. Jaswon. Integral equation methods in potential theory. I. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 275(1360):23–32, 1963.

[JCL07] Chao Jin, Xiao-Chuan Cai, and Congming Li. Parallel domain decomposition methods for stochastic elliptic equations. *SIAM Journal on Scientific Computing*, 29(5):2096–2114, 2007.

[JK09] A. Jentzen and P. E. Kloeden. The numerical approximation of stochastic partial differential equations. *Milan Journal of Mathematics*, 77(1):205–244, 2009.

[JSSE97] Y. Jeon, I.H. Sloan, E.P. Stephan, and J. Elschner. Discrete qualocation methods for logarithmic-kernel integral equations on a piecewise smooth boundary. *Advances in Computational Mathematics*, 7(4):547–571, 1997.

[JZYH16] Fengze Jiang, Xiaofeng Zong, Chao Yue, and Chengming Huang. Double–implicit and split two–step Milstein schemes for stochastic differential equations. *International Journal of Computer Mathematics*, 93(12):1987–2011, 2016.

[KAF14] A. Kundu, S. Adhikari, and M. I. Friswell. Stochastic finite elements of discretely parameterized random systems on domains with boundary uncertainty. *International Journal for Numerical Methods in Engineering*, 100(3):183–221, 2014.

[Kam05] Marcin Marek Kaminski. *Computational Mechanics of Composite Materials: Sensitivity, Randomness and Multiscale Behaviour*. Engineering Materials and Processes. Springer-Verlag London, 2005.

[Kam13] Marcin Marek Kaminski. *The stochastic perturbation method for computational mechanics*. John Wiley & Sons, 2013.

[KB13] Yoshio Komori and Kevin Burrage. Strong first order S–ROCK methods for stochastic differential equations. *Journal of Computational and Applied Mathematics*, 242:261 – 274, 2013.

[KB14] Yoshio Komori and Kevin Burrage. A stochastic exponential Euler scheme for simulation of stiff biochemical reaction systems. *BIT Numerical Mathematics*, 54(4):1067–1085, 2014.

[KH92] M. Kleiber and T. Hien. *The Stochastic Finite Element Method: Basic Perturbation Technique and Computer Implementation*. John Wiley, New York, 1992.

[Kha12] Rafail Khasminskii. *Stochastic Stability of Differential Equations*, volume 66 of *Stochastic Modelling and Applied Probability*. Springer-Verlag Berlin Heidelberg, 2012.

[KK88] Armen Der Kiureghian and Jyh-Bin Ke. The stochastic finite element method in structural reliability. *Probabilistic Engineering Mechanics*, 3(2):83 – 91, 1988.

[KMS+16] Grigoris Katsiolides, Eike H Müller, Robert Scheichl, Tony Shardlow, Michael B Giles, and David J Thomson. Multilevel Monte Carlo and Improved Timestepping Methods in Atmospheric Dispersion Modelling. *arXiv preprint arXiv:1612.07717*, 2016.

[Kot08] Peter Kotelenez. *Stochastic Ordinary and Stochastic Partial Differential Equations*, volume 58 of *Stochastic Modelling and Applied Probability*. Springer-Verlag New York, 2008.

[KP92a] P. E. Kloeden and E. Platen. Higher-order implicit strong numerical schemes for stochastic differential equations. *Journal of Statistical Physics*, 66(1):283–314, 1992.

[KP92b] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23 of *Stochastic Modelling and Applied Probability*. Springer-Verlag Berlin Heidelberg, 1992.

[KS06] Christian Kahl and Henri Schurz. Balanced Milstein methods for ordinary SDEs. *Monte Carlo Methods and Applications*, 12(2):143–170, 2006.

[KS14] Chaman Kumar and Sotirios Sabanis. On tamed Milstein schemes of SDEs driven by Lévy noise. *arXiv preprint arXiv:1407.5347*, 2014.

[KSS15] Frances Y. Kuo, Christoph Schwab, and Ian H. Sloan. Multi-level Quasi-Monte Carlo finite element methods for a class of elliptic PDEs with random coefficients. *Foundations of Computational Mathematics*, 15(2):411–449, 2015.

[LAE08] Tiejun Li, Assyr Abdulle, and Weinan E. Effectiveness of implicit methods for stiff stochastic differential equations. *Communications in Computational Physics*, 3(2):295–307, 2008.

[LC09] Jie Li and Jianbing Chen. *Stochastic dynamics of structures*. John Wiley & Sons, 2009.

[LDM13] Christapher Lang, Alireza Doostan, and Kurt Maute. Extended stochastic FEM for diffusion problems with uncertain material interfaces. *Computational Mechanics*, 51(6):1031–1049, 2013.

[Li07] Tiejun Li. Analysis of explicit tau-leaping schemes for simulating chemically reacting systems. *Multiscale Modeling & Simulation*, 6(2):417–436, 2007.

[Lio96] Pierre-Louis Lions. *Mathematical Topics in Fluid Mechanics. Volume 1: Incompressible Models*. New York: Oxford University Press, 1996.

[LK93] Chun-Ching Li and A. Der Kiureghian. Optimal discretization of random fields. *Journal of Engineering Mechanics*, 119(6):1136–1154, 1993.

[LL04] S Li and WK Liu. *Meshfree particle methods*. Springer Berlin Heidelberg, 2004.

[LL10] Carlo Laing and Gabriel J Lord. *Stochastic Methods in Neuroscience*. Oxford University Press, 2010.

[LL12] Haokai Lu and Peng Li. Stochastic projective methods for simulating stiff chemical reacting systems. *Computer Physics Communications*, 183(7):1427 – 1442, 2012.

[LMB87] Wing Kam Liu, A. Mani, and Ted Belytschko. Finite element methods in probabilistic mechanics. *Probabilistic Engineering Mechanics*, 2(4):201 – 213, 1987.

[Loe78] M. Loeve. *Probability Theory II*, volume 46 of *Graduate Texts in Mathematics*. Springer-Verlag New York, 1978.

[Loh96] Wei-Liem Loh. On Latin hypercube sampling. *Ann. Statist.*, 24(5):2058–2080, 1996.

[LPS14] Gabriel James Lord, Catherine Powell, and Tony Shardlow. *An Introduction to Computational Stochastic PDEs*. Cambridge Texts in Applied Mathematics. Cambridge University Press, United Kingdom, 2014.

[LR15] Wei Liu and Michael Röckner. *Stochastic Partial Differential Equations: An Introduction*. Universitext. Springer International Publishing, 2015.

[Mao07] Xuerong Mao. *Stochastic differential equations and applications*. Woodhead Publishing, 2007.

[Mao16] Xuerong Mao. Convergence rates of the truncated Euler-Maruyama method for stochastic differential equations. *Journal of Computational and Applied Mathematics*, 296:362 – 375, 2016.

[Mar10] Henning Marxen. The multilevel Monte Carlo method used on a Lévy driven SDE. *Monte Carlo Methods and Applications*, 16(2):167–190, 2010.

[Mel77] Yu.A. Melnikov. Some applications of the Greens' function method in mechanics. *International Journal of Solids and Structures*, 13(11):1045 – 1058, 1977.

[Mel98] Yuri Melnikov. *Influence functions and matrices*. CRC Press, 1998.

[Mel11] Yu. A. Melnikov. *Green's Functions and Infinite Products: Bridging the Divide*. Birkhäuser Boston, 2011.

[Mer73] Robert C. Merton. Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 4(1):141–183, 1973.

[MHM96] YA Melnikov, S Hughes, and S McDaniel. Boundary element approach based on Green's functions. *WIT Transactions on Modelling and Simulation*, 15, 1996.

[MHR07] Ethan A. Mastny, Eric L. Haseltine, and James B. Rawlings. Two classes of quasi-steady-state model reductions for stochastic kinetics. *The Journal of Chemical Physics*, 127(9), 2007.

[MK05] Hermann G. Matthies and Andreas Keese. Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Computer*

Methods in Applied Mechanics and Engineering, 194(1216):1295 – 1331, 2005. Special Issue on Computational Methods in Stochastic Mechanics and Reliability Analysis.

[MK10] Olivier Le Maitre and Omar M Knio. *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*. Scientific Computation. Springer Netherlands, 2010.

[MKNG01] Olivier Le Maitre, Omar M Knio, Habib N Najm, and Roger G Ghanem. A stochastic projection method for fluid flow: I. Basic formulation. *Journal of Computational Physics*, 173(2):481 – 511, 2001.

[ML] M. Muja and D.G. Lowe. Flann: Fast library for approximate nearest neighbors. *[Online]. Available: http://www.cs.ubc.ca/research/flann*.

[ML14] M. Muja and D.G. Lowe. Scalable nearest neighbor algorithms for high dimensional data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(11):2227–2240, Nov 2014.

[MM12] Yu. A. Melnikov and Max Y. Melnikov. *Green's Functions. Construction and Applications*. Berlin, Boston: De Gruyter, 2012.

[MN68] Benoit B. Mandelbrot and John W. Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM Review*, 10(4):422–437, 1968.

[MNK11] P. Surya Mohan, Prasanth B. Nair, and Andy J. Keane. Stochastic projection schemes for deterministic linear elliptic partial differential equations on random domains. *International Journal for Numerical Methods in Engineering*, 85(7):874–895, 2011.

[MPS98] G. N. Milstein, E. Platen, and H. Schurz. Balanced implicit methods for stiff stochastic systems. *SIAM Journal on Numerical Analysis*, 35(3):1010–1019, 1998.

[MQ02] Robert I. McLachlan and G. Reinout W. Quispel. Splitting methods. *Acta Numerica*, 11:341–434, 1 2002.

[MR14] Yu. A. Melnikov and V. Reshniak. A semi-analytical approach to greens functions for heat equation in regions of irregular shape. *Engineering Analysis with Boundary Elements*, 46:108 – 115, 2014.

[MRN+02] Olivier Le Maitre, Matthew T. Reagan, Habib N. Najm, Roger G. Ghanem, and Omar M. Knio. A stochastic projection method for fluid flow: II. Random process. *Journal of Computational Physics*, 181(1):9 – 44, 2002.

[MS96] G. D. Manolis and R. P. Shaw. Boundary integral formulation for 2d and 3d thermal problems exibiting a linearly varying stochastic conductivity. *Computational Mechanics*, 17(6):406–417, 1996.

[MS13] Xuerong Mao and Lukasz Szpruch. Strong convergence and stability of implicit numerical methods for stochastic differential equations with non–globally Lipschitz continuous coefficients. *Journal of Computational and Applied Mathematics*, 238:14 – 28, 2013.

[MSS15] Eike H. Müller, Rob Scheichl, and Tony Shardlow. Improving multilevel Monte Carlo for stochastic differential equations with application to the Langevin equation. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 471(2176), 2015.

[MT04] Grigori Noah Milstein and Michael V. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Springer-Verlag Berlin Heidelberg, 2004.

[MTV16] Alvaro Moraes, Raúl Tempone, and Pedro Vilanova. Multilevel hybrid Chernoff tau-leap. *BIT Numerical Mathematics*, 56(1):189–239, 2016.

[MY06] Xuerong Mao and Chenggui Yuan. *Stochastic differential equations with Markovian switching*. World Scientific, 2006.

[MZ09] Xiang Ma and Nicholas Zabaras. An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *Journal of Computational Physics*, 228(8):3084 – 3113, 2009.

[Nad64] E. Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.

[NC10] A. Nouy and A. Clment. eXtended Stochastic Finite Element Method for the numerical simulation of heterogeneous materials with random material interfaces. *International Journal for Numerical Methods in Engineering*, 83(10):1312–1344, 2010.

[NCS11] A. Nouy, M. Chevreuil, and E. Safatly. Fictitious domain method and separated representations for the solution of boundary value problems on uncertain parameterized domains. *Computer Methods in Applied Mechanics and Engineering*, 200(4546):3066 – 3082, 2011.

[NCSM08] A. Nouy, A. Clment, F. Schoefs, and N. Mos. An extended stochastic finite element method for solving stochastic partial differential equations on random domains. *Computer Methods in Applied Mechanics and Engineering*, 197(5152):4663 – 4682, 2008.

[Nis15] Makiko Nisio. *Stochastic Control Theory*, volume 72 of *Probability Theory and Stochastic Modelling*. Springer Japan, 2015.

[NO93] Shlomo P. Neuman and Shlomo Orr. Prediction of steady state flow in nonuniform geologic media by conditional moments: Exact nonlocal formalism, effective conductivities, and weak approximation. *Water Resources Research*, 29(2):341–364, 1993.

[NR02] David Nualart and Aurel Rascanu. Differential equations driven by fractional brownian motion. *Collectanea Mathematica*, 53(1):55–81, 2002.

[NSM07] Anthony Nouy, Franck Schoefs, and Nicolas Mos. X-SFEM, a computational technique based on X-FEM to deal with random shapes. *European Journal of Computational Mechanics*, 16(2):277–293, 2007.

[NT15] Fabio Nobile and Francesco Tesei. A Multi Level Monte Carlo method with control variate for elliptic PDEs with log-normal coefficients. *Stochastic Partial Differential Equations: Analysis and Computations*, 3(3):398–444, 2015.

[NTW08a] F. Nobile, R. Tempone, and C. G. Webster. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2411–2442, 2008.

[NTW08b] F. Nobile, R. Tempone, and C. G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345, 2008.

[NTWW96] S. P. Neuman, D. Tartakovsky, T. C. Wallstrom, and C. L. Winter. Correction to prediction of steady state flow in nonuniform geologic media by conditional moments: Exact nonlocal formalism, effective conductivities, and weak approximation by shlomo p. neuman and shlomo orr. *Water Resources Research*, 32(5):1479–1480, 1996.

[OAHR11] M.A. Omar, A. Aboul-Hassan, and S.I. Rabia. The composite Milstein methods for the numerical solution of Itô stochastic differential equations. *Journal of Computational and Applied Mathematics*, 235(8):2277 – 2299, 2011.

[Øks03] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Universitext. Springer-Verlag Berlin Heidelberg, 2003.

[OS98] Martin Ostoja-Starzewski. Random field models of heterogeneous materials. *International Journal of Solids and Structures*, 35(19):2429 – 2455, 1998.

[OS07] Martin Ostoja-Starzewski. *Microstructural randomness and scaling in mechanics of materials*. CRC Press, 2007.

[ØSB07] Bernt Øksendal and Agnés Sulem-Bialobroda. *Applied Stochastic Control of Jump Diffusions*. Universitext. Springer-Verlag Berlin Heidelberg, 2007.

[PBL10] Eckhard Platen and Nicola Bruti-Liberati. *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*, volume 64 of *Stochastic Modelling and Applied Probability*. Springer-Verlag Berlin Heidelberg, 2010.

[PdSM+06] Michael L. Parks, Eric de Sturler, Greg Mackey, Duane D. Johnson, and Spandan Maiti. Recycling Krylov subspaces for sequences of linear systems. *SIAM Journal on Scientific Computing*, 28(5):1651–1674, 2006.

[PG00] M.F Pellissetti and R.G Ghanem. Iterative solution of systems of linear equations arising in the context of stochastic finite elements. *Advances in Engineering Software*, 31(89):607 – 616, 2000.

[Pha09] Huyên Pham. *Continuous-time Stochastic Control and Optimization with Financial Applications*, volume 61 of *Stochastic Modelling and Applied Probability*. Springer-Verlag Berlin Heidelberg, 2009.

[PHQ04] K.K. Phoon, H.W. Huang, and S.T. Quek. Comparison between Karhunen-Loéve and wavelet expansions for simulation of Gaussian processes. *Computers & Structures*, 82(1314):985–991, 2004.

[PHQ05] K.K. Phoon, H.W. Huang, and S.T. Quek. Simulation of strongly non-Gaussian processes using Karhunen-Loéve expansion. *Probabilistic Engineering Mechanics*, 20(2):188 – 198, 2005.

[PIT12] Sang Woo Park, Marcos Intaglietta, and Daniel M. Tartakovsky. Impact of endothelium roughness on blood flow. *Journal of Theoretical Biology*, 300:152 – 160, 2012.

[PP90] E. Pardoux and S. Peng. Adapted solution of a backward stochastic differential equation. *Systems & Control Letters*, 14(1):55 – 61, 1990.

[PR07] Claudia Prévôt and Michael Röckner. *A Concise Course on Stochastic Partial Differential Equations*, volume 1905 of *Lecture Notes in Mathematics*. Springer-Verlag Berlin Heidelberg, 2007.

[Pro05] Philip E. Protter. *Stochastic Integration and Differential Equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag Berlin Heidelberg, 2005.

[PS08] Grigorios A Pavliotis and Andrew Stuart. *Multiscale methods: averaging and homogenization*. Springer New York, 2008.

[PT99] Etienne Pardoux and Shanjian Tang. Forward-backward stochastic differential equations and quasilinear parabolic pdes. *Probability Theory and Related Fields*, 114(2):123–150, 1999.

[PWG16] B. Peherstorfer, K. Willcox, and M. Gunzburger. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. Technical report, Aerospace Computational Design Laboratory, Massachusetts Institute of Technology, 2016.

[PZ92] Guiseppe Da Prato and Jerzy Zabczyk. *Stochastic Equations in Infinite Dimensions*, volume 45 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 1992.

[PZ07] S. Peszat and J. Zabczyk. *Stochastic Partial Differential Equations with Lévy Noise*, volume 113 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 2007.

[RA03] Christopher V. Rao and Adam P. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *The Journal of Chemical Physics*, 118(11):4999–5010, 2003.

[Rac01] Rüdiger Rackwitz. Reliability analysis - a review and some perspectives. *Structural Safety*, 23(4):365 – 395, 2001.

[Res13] Viktor Reshniak. Some further developments in the infinite product representation of elementary functions. *Global Journal of Science Frontier Research*, 13(4), 2013.

[Ric13] LF Ricketson. Three improvements to multi-level Monte Carlo simulation of SDE systems. *arXiv preprint arXiv:1309.1922*, 2013.

[RKVZ15] V. Reshniak, A.Q.M. Khaliq, D.A. Voss, and G. Zhang. Split-step Milstein methods for multi-channel stiff stochastic differential systems. *Applied Numerical Mathematics*, 89:1 – 23, 2015.

[Ros13] S. M. Ross. *Simulation*. Academic press, San Diego, CA, 5th edition, 2013.

[Röß06a] Andreas Rößler. Rooted tree analysis for order conditions of stochastic Runge-Kutta methods for the weak approximation of stochastic differential equations. *Stochastic Analysis and Applications*, 24(1):97–134, 2006.

[Röß06b] Andreas Rößler. Runge–Kutta methods for Itô stochastic differential equations with scalar noise. *BIT Numerical Mathematics*, 46(1):97–110, 2006.

[Röß10] Andreas Rößler. Strong and weak approximation methods for stochastic differential equations–some recent developments. In *Recent Developments in Applied Probability and Statistics: Dedicated to the Memory of Jürgen Lehn*, pages 127–153. Physica-Verlag HD, Heidelberg, 2010.

[RPCG03] Muruhan Rathinam, Linda R. Petzold, Yang Cao, and Daniel T. Gillespie. Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. *The Journal of Chemical Physics*, 119(24):12784–12794, 2003.

[RPCG05] Muruhan Rathinam, Linda R. Petzold, Yang Cao, and Daniel T. Gillespie. Consistency and stability of tau-leaping schemes for chemical reaction systems. *Multiscale Modeling & Simulation*, 4(3):867–895, 2005.

[Rüe82] W. Rüemelin. Numerical treatment of stochastic differential equations. *SIAM Journal on Numerical Analysis*, 19(3):604–613, 1982.

[RW06] Werner Römisch and Renate Winkler. Stepsize control for mean-square numerical methods for stochastic differential equations with small noise. *SIAM Journal on Scientific Computing*, 28(2):604–625, 2006.

[S+13] Sotirios Sabanis et al. A note on tamed Euler approximations. *Electron. Commun. Probab*, 18(47):1–10, 2013.

[Saa03] Yousef Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, 2003.

[SB92] Ian H. Sloan and B.J. Burn. An unconventional quadrature method for logarithmic-kernel integral equations equations on closed curves. *J. Integral Equations Applications*, 4(1):117–151, 03 1992.

[Sch97] G.I. Schuëller. A state-of-the-art report on computational stochastic mechanics. *Probabilistic Engineering Mechanics*, 12(4):197 – 321, 1997.

[Sch07] G.I. Schuëller. On the treatment of uncertainties in structural mechanics and analysis. *Computers & Structures*, 85(56):235 – 243, 2007.

[Sch12] Henri Schurz. *Stochastic Differential Equations and Processes: SAAP, Tunisia, October 7-9, 2010*, chapter Basic Concepts of Numerical Analysis of Stochastic Differential Equations Explained by Balanced Implicit Theta Methods, pages 1–139. Springer Berlin Heidelberg, 2012.

[SK08] Vassilios Sotiropoulos and Yiannis N. Kaznessis. An adaptive time step scheme for a system of stochastic differential equations with multiple multiplicative noise: Chemical langevin equation, a proof of concept. *The Journal of Chemical Physics*, 128(1):014103, 2008.

[Slo00] Ian H Sloan. Qualocation. *Journal of Computational and Applied Mathematics*, 125(12):461 – 478, 2000. Numerical Analysis 2000. Vol. VI: Ordinary Differential Equations and Integral Equations.

[Smo63] Sergey A. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. In *Dokl. Akad. Nauk SSSR*, volume 4, pages 240–243, 1963.

[SSPD14] Dimitris Savvas, George Stefanou, Manolis Papadrakakis, and George Deodatis. Homogenization of random heterogeneous media with inclusions of arbitrary shape modeled by XFEM. *Computational Mechanics*, 54, 11 2014.

[Ste87] Michael Stein. Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, 29(2):143–151, 1987.

[Ste08] Olaf Steinbach. *Numerical Approximation Methods for Elliptic Boundary Value Problems*. New York: Springer, 2008.

[Ste09] George Stefanou. The stochastic finite element method: Past, present and future. *Computer Methods in Applied Mechanics and Engineering*, 198(912):1031 – 1051, 2009.

[Sto77] CJ Stone. Consistent nonparametric regression. *The annals of statistics*, 5(4):595–620, 1977.

[Str92] Rouslan L. Stratonovich. *Nonlinear Nonequilibrium Thermodynamics I: Linear and Nonlinear Fluctuation-Dissipation Theorems*, volume 57 of *Springer Series in Synergetics*. Springer-Verlag Berlin Heidelberg, 1992.

[SV05] A. Samant and D. G. Vlachos. Overcoming stiffness in stochastic simulation stemming from partial equilibrium: A multiscale Monte Carlo algorithm. *The Journal of Chemical Physics*, 123(14), 2005.

[sWMX07] La sheng Wang, Changlin Mei, and Hong Xue. The semi–implicit Euler method for stochastic differential delay equation with jumps. *Applied Mathematics and Computation*, 192(2):567 – 578, 2007.

[Sym63] G. T. Symm. Integral equation methods in potential theory. II. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 275(1360):33–46, 1963.

[SZ98] P.D. Spanos and B.A. Zeldin. Monte Carlo treatment of random fields: a broad perspective. *Applied Mechanics Reviews*, 51(3):219–237, 1998.

[Tal02] D Talay. Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Related Fields*, 8(2):163–198, 2002.

[TB01] Tianhai Tian and Kevin Burrage. Implicit Taylor methods for stiff stochastic differential equations. *Applied Numerical Mathematics*, 38(1):167 – 185, 2001.

[Tot70] H Tottenham. Basic principles. In *Finite Element Techniques in Structural Mechanics*. Southampton Univ. Press, Southampton, 1970.

[Tso05] Tien T Tsong. *Atom-Probe Field Ion Microscopy: Field Ion Emission, and Surfaces and Interfaces at Atomic Resolution*. Cambridge University Press, 2005.

[Tuc89] H. Tuckwell. *Stochastic Processes in the Neurosciences*. Society for Industrial and Applied Mathematics, 1989.

[TX06] Daniel M. Tartakovsky and Dongbin Xiu. Stochastic analysis of transport in tubes with rough walls. *Journal of Computational Physics*, 217(1):248 – 259, 2006.

[Van88] E. Vanmarcke. *Random Fields: Analysis and Synthesis*. The MIT Press, Cambridge, MA, 1988.

[VCNGP15] F. Vidal-Codina, N.C. Nguyen, M.B. Giles, and J. Peraire. A model and variance reduction method for computing statistical outputs of stochastic elliptic partial differential equations. *Journal of Computational Physics*, 297:700 – 720, 2015.

[VE03] Eric Vanden-Eijnden. Numerical techniques for multi-scale dynamical systems with stochastic effects. *Communications in Mathematical Sciences*, 1(2):385–391, 06 2003.

[VH12] A. Valinejad and S. Mohammad Hosseini. A stepsize control algorithm for SDEs with small noise based on stochastic Runge–Kutta Maruyama methods. *Numerical Algorithms*, 61(3):479–498, 2012.

[vK92] Nicolaas G. van Kampen. *Stochastic processes in physics and chemistry*. Elsevier, 1992.

[VK15] David A. Voss and Abdul Q. M. Khaliq. Split–step Adams-Moulton Milstein methods for systems of stiff stochastic differential equations. *International Journal of Computer Mathematics*, 92(5):995–1011, 2015.

[Wal81] John B. Walsh. A stochastic model of neural response. *Advances in Applied Probability*, 13(2):231–281, 1981.

[Wal86] John B. Walsh. An introduction to stochastic partial differential equations. In *École d'Été de Probabilités de Saint Flour XIV-1984*, pages 265–439. Springer, 1986.

[Wen05] Holger Wendland. *Scattered data approximation*. Cambridge University Press, 2005.

[WG13] Xiaojie Wang and Siqing Gan. The tamed Milstein method for commutative stochastic differential equations with non-globally Lipschitz continuous coefficients. *Journal of Difference Equations and Applications*, 19(3):466–490, 2013.

[WGW12] Xiaojie Wang, Siqing Gan, and Desheng Wang. A family of fully implicit Milstein methods for stiff stochastic differential equations with multiplicative noise. *BIT Numerical Mathematics*, 52(3):741–772, 2012.

[Wik01] Magnus Wiktorsson. Joint Characteristic Function and Simultaneous Simulation of Iterated Itô Integrals for Multiple Independent Brownian Motions. *The Annals of Applied Probability*, 11(2):470–487, 2001.

[WL09] Peng Wang and Zhenxin Liu. Split-step backward balanced Milstein methods for stiff stochastic systems. *Applied Numerical Mathematics*, 59(6):1198 – 1213, 2009.

[WL10] P. Wang and Y. Li. Split-step forward methods for stochastic differential equations. *Journal of Computational and Applied Mathematics*, 233(10):2641 – 2651, 2010.

[WLVE05] E Weinan, Di Liu, and Eric Vanden-Eijnden. Analysis of multiscale methods for stochastic differential equations. *Communications on Pure and Applied Mathematics*, 58(11):1544–1585, 2005.

[XG12] Yuan Xia and Michael B. Giles. Multilevel path simulation for jump–diffusion SDEs. In *Monte Carlo and Quasi-Monte Carlo Methods 2010*, pages 695–708. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[XH05] Dongbin Xiu and Jan S. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM Journal on Scientific Computing*, 27(3):1118–1139, 2005.

[Xiu09] Dongbin Xiu. Fast numerical methods for stochastic computations: a review. *Communications in computational physics*, 5(2-4):242–272, 2009.

[XLSK01] D Xiu, D Lucor, CH Su, and G Karniadakis. Stochastic modeling of flow-structure interactions using generalized polynomial chaos. *ASME. J. Fluids Eng.*, 124(1):51–59, 2001.

[XT06] Dongbin Xiu and Daniel M. Tartakovsky. Numerical methods for differential equations in random domains. *SIAM Journal on Scientific Computing*, 28(3):1167–1185, 2006.

[YM08] Chenggui Yuan and Xuerong Mao. A Note on the Rate of Convergence of the Euler-Maruyama Method for Stochastic Differential Equations. *Stochastic Analysis and Applications*, 26(2):325–333, 2008.

[YS88] Y. Yan and I.H. Sloan. On integral equations of the first kind with logarithmic kernels. *J. Integral Equations Applications*, 1(4):549–580, 12 1988.

[YSD88] Fumio Yamazaki, Masanobu Shinozuka, and Gautam Dasgupta. Neumann expansion for stochastic finite element analysis. *Journal of Engineering Mechanics*, 114(8):1335–1354, 1988.

[YZ10] George Yin and Chao Zhu. *Hybrid Switching Diffusions*, volume 63 of *Stochastic Modelling and Applied Probability*. Springer-Verlag New York, 2010.

[ZPTK13] Mohsen Zayernouri, Sang-Woo Park, Daniel M. Tartakovsky, and George Em Karniadakis. Stochastic smoothed profile method for modeling random roughness in flow problems. *Computer Methods in Applied Mechanics and Engineering*, 263:99 – 112, 2013.

[ZWD04] Zhenhai Zhu, Jacob White, and Alper Demir. A stochastic integral equation method for modeling the rough surface effect on interconnect capacitance. In

Proceedings of the 2004 IEEE/ACM International conference on Computer-aided design, pages 887–891. IEEE Computer Society, 2004.

[ZWH14] Xiaofeng Zong, Fuke Wu, and Chengming Huang. Convergence and stability of the semi-tamed Euler scheme for stochastic differential equations with non-Lipschitz continuous coefficients. *Applied Mathematics and Computation*, 228:240 – 250, 2014.