Numerical Complexity Analysis of Weak Approximation of Stochastic Differential Equations

Raúl Tempone Olariaga

TRITA-NA-0220

Doctoral Dissertation
Royal Institute of Technology
Department of Numerical Analysis and Computer Science
Numerical Complexity Analysis of Weak Approximation of Stochastic Differential Equations

Raúl Tempone Olariaga

Akademisk avhandling som med tillstånd av Kungl Tekniska Högskolan framlägges till offentlig granskning för avläggande av teknisk doktorsexamen Fredagen den 11 oktober 2002 kl 10.00 i D3, Lindstedtsvägen 5, Entreplan, Tekniska Högskolan, Stockholm.

ISBN 91-7283-350-5, TRITA-NA-0220, ISSN 0348-2952, ISRN KTH/NA/R-02/20,

© Raúl Tempone Olariaga, October 2002
Royal Institute of Technology, Stockholm 2002
Akademisk avhandling som med tillstånd av Kungl Tekniska Högskolan framläggdes till offentlig granskning för avläggande av teknisk doktorsexamen Fredagen den 11 oktober 2002 kl 10.00 i D3, Lindstedtsvägen 5, Entreplan, Tekniska Högskolan, Stockholm.

TRITA-NA-0220
ISSN 0348-2952
ISRN KTH/NA/R-02/20

© Raúl Tempone Olariaga, October 2002

KTH, Stockholm 2002
Abstract

The thesis consists of four papers on numerical complexity analysis of weak approximation of ordinary and partial stochastic differential equations, including illustrative numerical examples. Here by numerical complexity we mean the computational work needed by a numerical method to solve a problem with a given accuracy. This notion offers a way to understand the efficiency of different numerical methods.

The first paper develops new expansions of the weak computational error for Itô stochastic differential equations using Malliavin calculus. These expansions have a computable leading order term in a posteriori form, and are based on stochastic flows and discrete dual backward problems. Beside this, these expansions lead to efficient and accurate computation of error estimates and give the basis for adaptive algorithms with either deterministic or stochastic time steps. The second paper proves convergence rates of adaptive algorithms for Itô stochastic differential equations. Two algorithms based either on stochastic or deterministic time steps are studied. The analysis of their numerical complexity combines the error expansions from the first paper and an extension of the convergence results for adaptive algorithms approximating deterministic ordinary differential equations. Both adaptive algorithms are proven to stop with an optimal number of time steps up to a problem independent factor defined in the algorithm. The third paper extends the techniques to the framework of Itô stochastic differential equations in infinite dimensional spaces, arising in the Heath Jarrow Morton term structure model for financial applications in bond markets. Error expansions are derived to identify different error contributions arising from time and maturity discretization, as well as the classical statistical error due to finite sampling.

The last paper studies the approximation of linear elliptic stochastic partial differential equations, describing and analyzing two numerical methods. The first method generates iid Monte Carlo approximations of the solution by sampling the coefficients of the equation and using a standard Galerkin finite elements variational formulation. The second method is based on a finite dimensional Karhunen-Loève approximation of the stochastic coefficients, turning the original stochastic problem into a high dimensional deterministic parametric elliptic problem. Then, a deterministic Galerkin finite element method, of either h or p version, approximates the stochastic partial differential equation. The paper concludes by comparing the numerical complexity of the Monte Carlo method with the parametric finite element method, suggesting intuitive conditions for an optimal selection of these methods.

2000 Mathematics Subject Classification. Primary 65C05, 60H10, 60H35, 65C30, 65C20; Secondary 91B28, 91B70.

Keywords and phrases. Adaptive methods, a posteriori error estimates, stochastic differential equations, weak approximation, Monte Carlo methods, Malliavin Calculus, HJM model, option price, bond market, stochastic elliptic equation, Karhunen-Loève expansion, perturbation estimates, numerical complexity
Acknowledgments

First I wish to thank my advisor, Anders Szepessy, for all his support, both enthusiastic and continuous, and for his enlightening guidance during my PhD studies. Without his help this dissertation could hardly be written.

I am also grateful to Georgios Zouraris and Kyoung-Sook Moon for many fruitful discussions, scientific and others, and for their nice company at different scientific meetings.

All my good friends at NADA have contributed to make my stay a pleasant one, and I am thankful for that. Among them, I wish to particularly thank Jesper Oppelstrup and Ingrid Melinder. I would like to thank Germund Dalhquist for sharing his exciting views of numerical analysis and finding time for scientific discussions. Special thanks also to my friends Ulf and Christer Andersson, for their infinite patience and invaluable computer consulting.

Ivo Babuška arranged the financial support for my two visits to the Texas Institute for Computational and Applied Mathematics (TICAM) and gave me the opportunity to learn from him there. Many thanks to him and to the people at TICAM for their warm hospitality.

Tomas Björk introduced me to mathematical finance. Here goes my appreciation for sharing his expertise in the area.

Last but not least, I want to specially thank my dear wife Roxana for all her love and patience during these years.

This work was financially supported by the Swedish council for Engineering Science (TFR), grant # 22-148, UdelaR and UdeM in Uruguay and by the National Science Fundation with the grant DMS-9802367. All these grants are gratefully acknowledged.
1 Introduction 3

1.1 Itô Stochastic Differential Equations . . . . . . . . . . . . . . . . . 7
  1.1.1 Weak Approximation of SDEs . . . . . . . . . . . . . . . . . . 7
  1.1.2 Overview of Paper 1 . . . . . . . . . . . . . . . . . . . . . . . 10
  1.1.3 Overview of Paper 2 . . . . . . . . . . . . . . . . . . . . . . . 11
  1.1.4 Weak Approximation of an Infinite Dimensional SDE . . . 12
  1.1.5 Overview of Paper 3 . . . . . . . . . . . . . . . . . . . . . . . 14

1.2 Stochastic Partial Differential Equations . . . . . . . . . . . . . . 15
  1.2.1 Overview of Paper 4 . . . . . . . . . . . . . . . . . . . . . . . 20

List of Papers


A Roxana.
Chapter 1

Introduction

This work studies numerical methods for weak approximation of stochastic differential equations. In particular, it focuses on the numerical complexity of different numerical methods. Here by numerical complexity we mean the computational work needed by a numerical method to solve a problem with a given accuracy. This work first considers adaptive numerical methods for weak approximation of Itô Stochastic Differential Equations (SDEs) and then analyzes Galerkin finite element approximations for Stochastic Partial Differential Equations (SPDEs) that are linear and elliptic.

SDEs are often part of mathematical models that describe the evolution of dynamical systems under uncertainty. A mathematical model establishes mathematical relations between the relevant variables of a given system. For example, a differential equation modeling the temperature of a hot metal surface subject to water cooling, describes the relation between the given initial condition –the initial temperature–, the flux function –which tells how the heat is convected in the system–, and the final value –the final temperature we want to know–. The purpose of a mathematical model is to predict the outcome of events –which can be past, present or future–, for example the result of a certain physical experiment, and possibly take advantage of that knowledge, since an accurate mathematical model may be used as a basic tool to control the outcomes.

Mathematical models can be deterministic or stochastic. The first case arises when the data and the relations described in the model are deterministic, like in the case of an ordinary differential equation with fixed data, whereas in the second either the data or the relations between the variables are stochastic. As an example of this, consider an ordinary differential equation, whose initial value is not deterministic, but follows a given probability distribution. Another example is to consider a perturbation of an ordinary differential equation, where the evolution itself is affected by some “noise”. Setting a formal description of the this intuitive
Chapter 1. Introduction

A notion leads to the concept of Stochastic Differential Equations. The field of applications is quite wide, e.g. it comprises ground water flow and financial markets [KP92] [Øks98].

Uncertainty comes basically from two sources, namely the lack of complete information about the dynamics of the system to model, or the fact that, for fixed data, the system does not always offer the same outcome. As an example of the second case, when rolling a fair dice a sufficiently large number of times, we tend to observe that all the values appear in similar proportions in the outcomes. It is possible then to use this statistical information within a probability model to answer questions related to dice games. This second step is related with the approach pursued here, that is, we shall assume that the stochastic model has been properly identified by some statistical procedure and is given, and then try to compute some related quantities. The need to compute expected values or averages—functionals—depending on the solution of an SDE will guide us towards the notion of weak convergence, as opposite to strong convergence, where good approximation of realization paths is required.

The discretization of an SDE can be more subtle than for ODEs, for example forward and backward differences do not in general converge to the same limit. Therefore, the model must also include information on the discretization to be used.

Numerical methods offer approximate computable solutions to mathematical problems, and are usually applied when the exact solution is either unknown or its computation is costly or involved. In particular, adaptive numerical methods aim for efficient use of computational resources by trying to minimize the degrees of freedom in the numerical discretizations, as well as to provide accurate estimates of the different sources of error present in the computations, like the time discretization error in the solution of an ordinary differential equation. Efficient adaptive numerical methods rely on a posteriori information, i.e. information offered by the computable numerical solution, both to estimate the error present in the numerical solution and to apply a refinement criterion when adding degrees of freedom to a given discretization. On the other hand, a priori information, i.e. information about the unknown and usually non computable exact solution, is of qualitative kind, e.g. provides smoothness properties of the exact solution, and may be used to prove convergence of numerical approximations, to identify the order of such convergence, as well as to select an appropriate numerical method [EEHJ95] [EEHJ96].

Regarding applications, mathematical finance is an area where stochastic modeling with SDEs has obtained a sound success, in particular when dealing with contingent claims pricing theory. A derivative product or a contingent claim is a financial contract whose value depends on a risk factor, also known as the underlying, such as the price of a bond, commodity, currency, share, a yield or rate of interest, an index of prices or yields, etc.

These contracts are also known as ”derivatives”, for short, and are common in financial markets. The application of derivatives is increasing, consider for example the case of energy derivatives [Boh98] [EL00] currently traded in many new regional
markets, arising from deregulation of former national monopolies. A simple example of a derivative, is the so called European Call Option, which gives to its owner the right, but not the obligation, to buy the underlying asset at the previously agreed-upon price on the expiration date \[Hul93\]. The usual valuation method assumes that the financial markets are efficient, that is, that there is no opportunity of making riskless profits, or in other words, that there exist no arbitrage opportunity in the market. This assumption leads to a consistency requirement between the price of the underlying, e.g. a stock price, which can be observed directly in the market, and a fair price for a related derivative product, e.g. the call option introduced above. Mathematically this consistency relation can be expressed by the existence of a probability measure \(Q\) such that given today’s date and today’s stock price, the price of the derivative is the expectation under \(Q\) of its discounted final payoff. The relevant point is that the expectations must be taken under \(Q\), which is known as a martingale measure, and not under the objective probability measure \[BR96, Bjo98\].

After the celebrated work of Black and Scholes \[BS73\], stochastic differential equations have been playing a major role in financial applications. Black and Scholes’ model can be used to fit observed data through implied quantities, and the related valuation formula can be interpreted as a nonlinear interpolation procedure to estimate derivative prices. Even though few of the model’s assumptions are fully respected in practice, e.g. constant volatility and constant riskless interest rate, the model is quite robust, specially for relatively short maturity options. However, when the life of the option becomes larger, extensions of the Black and Scholes model, e.g. allowing stochastic volatility, are of practical use to explain the so called volatility smile effect observed in the market \[FSP00, SP99, WO97\]. On the other hand, since only relatively few stochastic differential equations have explicit solutions, as the financial models get more and more refined the need for deeper understanding and better numerical methods increases.

Thanks to Kolmogorov’s stochastic representation formulae, see \[KSSS8\], numerical methods for weak approximation of SDEs can be based either on the numerical solution of a Kolmogorov backward partial differential equation, see \[Bjo94a, Bjo94b\], using finite differences schemes \[WHD95, Wil98\], or the finite element method \[BS94\], or by the time discretization of the SDE and the computation of sample averages by the Monte Carlo Euler method, see \[KP92\]. The convergence properties of finite difference schemes and the finite element method make them the best tools whenever the dimension of the given system of SDEs is low, say less or equal than four, since their computational cost increases exponentially with such dimension. On the other hand, the computational cost of Monte Carlo methods is only polynomial in the dimension of the SDE system, making them a feasible alternative to compute with large systems of SDEs. Tree methods \[CRR79\] are popular and have pedagogical advantages. They may be thought of as a special case of explicit finite difference schemes, although they are non optimal, i.e. with the same amount of computational work there exist other finite difference schemes with better convergence properties.
This work uses the Euler Monte Carlo method for weak approximation of SDEs, developing a posteriori error approximations proposing and analyzing related adaptive numerical methods for weak approximation that are well suited to solve problems with systems of SDEs.

Stochastic Partial Differential Equations (SPDEs) are also used to describe the behavior of systems under uncertainty. Due to the great development in computational resources and scientific computing techniques, more mathematical models can be solved efficiently. Ideally, these techniques could be used to solve many classical partial differential equations (PDEs) to high accuracy. However, in many cases, the information available to solve a given problem is far from complete. This is the case when solving a partial differential equation whose coefficients depend on material properties that are known to some accuracy. The same may occur with its boundary conditions, and even with the geometry of its domain, see for example the work [BCa, BCb]. Naturally, since the current engineering trends are toward more reliance on computational predictions, the need for assessing the level of accuracy in the results grows accordingly. More than ever, the goal then becomes to represent and propagate the uncertainties from the available data to the desired result through our partial differential equation. By uncertainty we mean either intrinsic variability of physical quantities or simply lack of knowledge about some physical behavior, cf. [Roa98]. If variability is interpreted as randomness then naturally we can apply probability theory. To be fruitful, probability theory requires considerable empirical information about the random quantities in question, generally in the form of probability distributions or their statistical moments. Uncertainties may arise at different levels. They could appear in the mathematical model, e.g. if we are not sure about the linear behavior of some material, or in the variables that describe the model, e.g. if the linear coefficient that describes the material is not completely known. Here we shall discuss the second alternative, and use a probabilistic description for the coefficient variability, leading us to the study of stochastic partial differential equations.

Regarding the approximation of SPDEs, this thesis describes and analyzes two numerical methods for a linear elliptic problem with stochastic coefficients and homogeneous Dirichlet boundary conditions. The first method generates iid approximations of the solution by sampling the coefficients of the equation and using a standard Galerkin finite elements variational formulation. The Monte Carlo method then uses these approximations to compute corresponding sample averages. The second method is based on a finite dimensional approximation of the stochastic coefficients, turning the original stochastic problem into a deterministic parametric elliptic problem. A Galerkin finite element method, of either h or p version, then approximates the corresponding deterministic solution yielding approximations of the desired statistics. We include a comparison of the computational work required by each method to achieve a given accuracy, the numerical complexity, to illustrate their nature and possible use.

The thesis is organized as follows: Section 1.1 describes the problem of weak approximation of Itō stochastic differential equations and the contributions from
papers I, II and III. Finally, Section 1.2 describes a problem from linear elliptic SPDEs and describes the contribution from paper IV.

1.1 Itô Stochastic Differential Equations

1.1.1 Weak Approximation of SDEs

Let $(\Omega, \mathcal{F}, P)$ be a probability space, where $\Omega$ is a set of outcomes, $\mathcal{F}$ is a set of events in $\Omega$, $P : \mathcal{F} \to [0,1]$ is a probability measure; and then let $W : \mathbb{R} \times \Omega \to \mathbb{R}_+$ be a Wiener process on $(\Omega, \mathcal{F}, P)$. On what follows, $\{\mathcal{F}_t^W\}_{t \in [0,T]}$ denotes the natural filtration, i.e. the filter structure of $\sigma$-algebras generated by $W$, or equivalently the filter generated by the random variables $\{W(s) : 0 \leq s \leq t\}$.

Let $a(t,x) \in \mathbb{R}^d$ and $b^\ell(t,x) \in \mathbb{R}^d$, $\ell = 1, \ldots, \ell_0$, be given drift and diffusion fluxes and consider the Itô stochastic differential equation in $\mathbb{R}^d$

\[ dX_k(t) = a_k(t, X(t))dt + \sum_{\ell=1}^{\ell_0} b^\ell_k(t, X(t))dW^\ell(t), \quad k = 1, \ldots, d, \quad t > 0, \]

\[ X_k(0) = X_{0,k}, \quad k = 1, \ldots, d. \]  

A classical reference for SDEs is the book [KS88]. An existence proof for strong solutions of SDEs, based on Piccard iterations and Lipschitz continuity of the drift and diffusion coefficients can be found in e.g. [Øks98], while a description of an alternative proof based on the Euler method can be found in [MSTZ00b].

The weak approximation of SDEs consists in approximating the expectation $E[g(X(T))]$, where $g : \mathbb{R}^d \to \mathbb{R}$ is a given function, $T$ is a given positive number and the stochastic process $X$ is the solution of (1.1) with initial datum $X(0)$. In finance applications the function $g$ can be a discounted payoff function of a $T$-contingent claim, and the fluxes $a, b$ describe the dynamics of the underlying process, e.g. a vector of stock values $X$. Figure 1.1 shows a simple example of such a problem, with $g(x) = \max(X - 0.9, 0)$ corresponding to the payoff diagram of a call option with strike price 0.9, maturity time $T = 1$ and an underlying process that follows a geometric Brownian motion, in this case $dX(t) = \frac{X(t)}{10} dt + \frac{X(t)}{5} dW(t)$, with initial condition $X(0) = 1$. The functional to compute is then

\[ E[g(X(1))] = \int_{\mathbb{R}} \max(x - 0.9, 0)p(1, x; 0, X_0)dx, \]

where $p(1, \cdot, 0, X_0)$ is the probability density of $X(1)$.

A first step towards the development of numerical solutions of the weak approximation problem is the forward Euler method, which is a time discretization of
Chapter 1. Introduction

Consider the time nodes \(0 = t_0 < t_1 < \cdots < t_N = T\) and define the discrete time stochastic process \(\bar{X}\) by

\[
\bar{X}(t_{n+1}) = \bar{X}(t_n) + a(t_n, \bar{X}(t_n))\Delta t_n + \sum_{\ell=1}^{\ell_0} b^\ell(t_n, \bar{X}(t_n))\Delta W^\ell_n, \quad 0 \leq n \leq N - 1
\]

\[
\bar{X}(0) = X_0.
\]

Even though a realization of \(\bar{X}(t_n)\) is computable, the expectation \(E[g(\bar{X}(T))]\) is in general not; however, \(E[g(\bar{X}(T))]\) can be approximated by a sample average of \(M\) independent realizations, \(\frac{1}{M} \sum_{j=1}^{M} g(\bar{X}(T; \omega_j))\), which is the basis of Monte Carlo methods [KP92].

Therefore, the exact computational error, \(\mathcal{E}_C\), naturally separates into the two parts

\[
\mathcal{E}_C \equiv E[g(X(T))] - \frac{1}{M} \sum_{j=1}^{M} g(\bar{X}(T; \omega_j))
\]

\[
= E[g(X(T)) - g(\bar{X}(T))] + [E[g(\bar{X}(T))] - \frac{1}{M} \sum_{j=1}^{M} g(\bar{X}(T; \omega_j))] \equiv \mathcal{E}_T + \mathcal{E}_S,
\]

where the first term, \(\mathcal{E}_T \equiv E[g(X(T)) - g(\bar{X}(T))]\), is the time discretization error, and the second, \(\mathcal{E}_s \equiv [E[g(\bar{X}(T))] - \frac{1}{M} \sum_{j=1}^{M} g(\bar{X}(T; \omega_j))]\), is the statistical error. The time steps for the trajectories \(\bar{X}\) are determined from statistical approximations of the time discretization error \(\mathcal{E}_T\). The number of realizations, \(M\) of \(\bar{X}\), are

\(\text{Figure 1.1. Weak approximation example. Left: Realizations for the Wiener process, } \Delta t = 0.01. \text{ Right: Realizations of the process } X(t), \text{ the function } g \text{ and a final sample density } p(1, x; 0, X_0) \text{ corresponding to } M = 1000 \text{ realizations.}\)
determined from the statistical error $\mathcal{E}_S$. Therefore, the number of realizations can be asymptotically determined by the Central Limit Theorem

$$\sqrt{M}\mathcal{E}_S \to \chi,$$

where the stochastic variable $\chi$ has the normal distribution, with mean zero and variance $\text{var}[g(\bar{X}(T))]$. The objective here is to choose the time nodes, which may be different for different realizations of $W$,

$$0 = t_0 < t_1 < \cdots < t_N = T,$$

and the number of realizations, $M$, so that the absolute value of the computational error is below a given tolerance, $|\mathcal{E}_C| \leq \text{TOL}$, with probability close to one and with as few time steps and realizations as possible.

Other aspects of the use of the Euler method for the weak approximation of SDEs have been addressed before. Milstein [Mil78] proved that the weak order of the Euler method is 1, i.e. that for uniform deterministic time steps $\Delta t = \frac{T}{N}$,

$$E[g(X(T)) - g(\bar{X}(T))] = O\left(\frac{1}{N}\right),$$

where $N$ is the number of time steps. Later, Talay and Tubaro [TT90] proved that for uniform deterministic time steps there is an a priori expansion

$$E[g(X(T)) - g(\bar{X}(T))] = \int_{0}^{T} E[\Psi(s, X(s))] ds + O\left(\frac{1}{N^2}\right),$$

where

$$\Psi(t, x) \equiv \frac{1}{2}(a_k a_n \partial_{kn} u)(t, x) + (a_i d_{jk} \partial_{ijk} u)(t, x) + \frac{1}{2}(d_{ij} d_{kn} \partial_{ijkn} u)(t, x) + \frac{1}{2} \partial_t u(t, x) + (a_i \partial_t \partial_i u)(t, x) + (d_{ij} \partial_t \partial_{ij} u)(t, x),$$

is based on the definition of the conditional expectation

$$u(t, x) \equiv E[g(X(T)) | X(t) = x]$$

and the following notation

$$d_{ij} \equiv \frac{1}{2} b_i b_j,$$

$$\partial_k \equiv \frac{\partial}{\partial x_k},$$

$$\partial_{ki} \equiv \frac{\partial^2}{\partial x_k \partial x_i},$$

$$\vdots$$

with the summation convention, i.e., if the same subscript appears twice in a term, the term denotes the sum over the range of this subscript, e.g.

$$c_{ik} \partial_k b_j \equiv \sum_{k=1}^{d} c_{ik} \partial_k b_j.$$
For a derivative $\partial_\alpha$ the notation $|\alpha|$ is its order. Kloeden and Platen [KP92] extended the results of Talay and Tubaro on the existence of leading order error expansion in a priori form, for first and second order schemes, to general weak approximations of higher order. Bally and Talay [BT95, BT96] extended the proof to the case where the payoff function $g$ is not smooth, using Malliavin calculus [Nua95]. This expansion motivates the use of Richardson’s extrapolation for the development of higher order methods. The case of killed diffusions, e.g. arising in the computation of barrier options, where the distribution of $X$ is not absolutely continuous with respect to the Lebesgue measure, was analyzed by Gobet [Gob00].

An introduction to numerical approximation of SDEs and an extensive review of the literature can be found in the inspiring book by Kloeden and Platen [KP92], including information about the construction and the analysis of the convergence order for higher order methods, either implicit or explicit.

Asymptotical optimal adapted adaptive methods for strong approximation of stochastic differential equations, are analyzed in [HMGR00] and [MG00], which include the hard problem to obtain lower error bounds for any method based on the number of evaluations of $W$ and requires roughly the $L^2$ norm in time of the diffusion $\max_{i}d_{ii}$ to be positive pathwise. The work [GL97] treats a first study on strong adaptive approximation.

### 1.1.2 Overview of Paper 1

The main result is new expansions of the computational error, with computable leading order term in a posteriori form, based on stochastic flows and discrete dual backward problems. The expansions lead to efficient and accurate computation of error estimates. In the first simpler expansion, the size of the time steps $\Delta t_n$ may vary in time but they are deterministic, i.e. the mesh is fixed for all samples. This is useful for solutions with singularities, or approximate singularities, at deterministic times or for problems with small noise. The second error expansion uses time steps which may vary for different realizations of the solution $\bar{X}$. Stochastic time steps are advantageous for problems with singularities at random times. Stochastic time steps use Brownian bridges and require more work for a given number of time steps. The optimal stochastic steps depend on the whole solution $\bar{X}(t)$, $0 < t < T$, and in particular the step $\Delta t(t)$ at time $t$ depends also on $W(\tau)$, $\tau > t$. In stochastic analysis the concept adapted to $W$ means that a process at time $t$ only depends on events generated by $\{W(s), s < t\}$. In numerical analysis an adaptive method means that the approximate solution is used to control the error, e.g. to determine the time steps. Our stochastic steps are in this sense adaptive non adapted, since $\Delta t(t)$ depends slightly on $W(\tau)$, $\tau > t$.

The number of realizations needed to determine the deterministic time steps is asymptotically at most $O(TOL^{-1})$, while the number of realizations for the Monte Carlo method to approximate $E[g(\bar{X}(T))]$ is $O(TOL^{-2})$. Therefore, the additional work to determine optimal deterministic time steps becomes negligible as the error tolerance tends to zero.
Efficient adaptive time stepping methods, with theoretical basis, use a posteriori error information, since the a priori knowledge usually cannot be as precise as the a posteriori. This work develops adaptive time stepping methods by proving in Theorems 2.2 and 3.3 error estimates of $\mathcal{E}_T$ with leading order terms in computable a posteriori form. Theorem 2.2 uses deterministic time steps, while Theorem 3.3 also holds for stochastic time steps, which are not adapted.

The main new idea here is the efficient use of stochastic flows and dual functions to obtain the error expansion with computable leading order term in Theorems 2.2 and 3.3, including also non adapted adaptive time steps. The use of dual functions is standard in optimal control theory and in particular for adaptive mesh control for ordinary and partial differential equations, see [BMV83], [Joh88], [JS95], [EEHJ96], and [BR96]. The authors are not aware of other error expansions in a posteriori form or adaptive algorithms for weak approximation of stochastic differential equations. In particular error estimates with stochastic non adapted time steps seem to not have been studied before.

Theorem 2.2 describes a computable error expansion, with deterministic time steps, to estimate the computational error,

$$E[g(X(T)) - g(\bar{X}(T))] \simeq \sum_n E[\rho(t_n, \omega)](\Delta t_n)^2$$

where $\rho(t_n, \omega)\Delta t_n$ is the corresponding error density function. Section 3 proves in Theorem 3.3 an analogous error expansion

$$E[g(X(T)) - g(\bar{X}(T))] \simeq E[\sum_n \rho(t_n, \omega)(\Delta t_n)^2]$$

which can be used also for stochastic time steps. The leading order terms of the expansion have less variance compared to the expansion in Theorem 2.2, but use upto the third variation, which requires more computational work per realization.

The focus in the paper is on computable error estimates for weak convergence of stochastic differential equations. The technique used here is based on the transition probability density and Kolmogorov’s backward equation, which was developed in [SV69] and [SV79] to analyze uniqueness and dependence on initial conditions for weak solutions of stochastic differential equations. The analogous technique for deterministic equations was introduced in [Grö67] and [Ale61].

### 1.1.3 Overview of Paper 2

Convergence rates of adaptive algorithms for weak approximations of Itô stochastic differential equations are proved for the Monte Carlo Euler method.

Here the focus is on the adaptivity procedures, and we derive convergence rates of two algorithms including dividing and merging of time steps, with either stochastic or deterministic time steps. The difference between the two algorithms is that the stochastic time steps may use different meshes for each realization, while the deterministic time steps use the same mesh for all realizations. The construction and
the analysis of the adaptive algorithms are inspired by the related work [MSTZ00a],
on adaptive algorithms for deterministic ordinary differential equations, and use the
error estimates from [STZ01]. The main step in the extension is the proof of the
almost sure convergence of the error density. Both adaptive algorithms are proven
to stop with optimal number of steps up to a problem independent factor defined
in the algorithm.

There are two main results on efficiency and accuracy of the adaptive algo-
rithms described in Section 3. In view of accuracy with probability close to one,
the approximation errors in (1.4) are asymptotically bounded by the specified error
tolerance times a problem independent factor as the tolerance parameter tends to
zero. In view of efficiency, both the algorithms with stochastic steps and determin-
istic steps stop with the optimal expected number of final time steps and optimal
number of final time steps respectively, up to a problem independent factor. The
number of final time steps is related to the numerical effort needed to compute
the approximation. To be more precise, the total work for deterministic steps is
roughly $M \cdot N$ where $M$ is the final number of realizations and $N$ is the final num-
ber of time steps, since the work to determine the mesh turns out to be negligible.
On the other hand, the total work with stochastic steps is on average bounded by
$M \cdot E[N_{\text{tot}}]$, where the total number, $N_{\text{tot}}$, of steps including all refinement levels
is bounded by $O(N \log N)$ with $N$ steps in the final refinement; for each realization
it is necessary to determine the mesh, which may vary for each realization.

The accuracy and efficiency results are based on the fact that the error den-
sity, $\rho$ which measures the approximation error for each interval following (1.5),(1.6),
converges almost surely or a.s. as the error tolerance tends to zero. This conver-
gence can be understood by the a.s. convergence of the approximate solution, $\bar{X}$,
as the maximal step size tends to zero. Once this convergence is established, the
techniques to develop the accuracy and efficiency results are similar to those from
[MSTZ00a]. Although the time steps are not adapted to the standard filtration gen-
erated by $W$ for the stochastic time stepping algorithm, the work [STZ01] proved
that the corresponding approximate solution converges to the correct adapted so-
lution $X$. This result makes it possible to prove the martingale property of the
approximate error term with respect to a specific filtration, see Lemma 4.2. There-
fore Theorem 4.1 and 4.4 use Doob’s inequality to prove the a.s. convergence of $\bar{X}$.  Similar results of pointwise convergence with constant step sizes, adapted to the
standard filtration, are surveyed by Talay in [Tal95].

This work can be easily modified following [MSTZ02] yielding adaptive algo-
rithms with no merging that have several theoretical advantages.

1.1.4 Weak Approximation of an Infinite Dimensional SDE

The extension of the weak approximation problem described in Section 1.1.1 to
the infinite dimensional case is here motivated by financial applications, in partic-
ular the valuation of contingent claims that have the market interest rate as the
underlaying.
Due to the relatively long life span of these products, the interest rate is modeled as a stochastic process, which may be Markovian, like in the case of spot rate models. In the context of interest rates products, the most elementary one is the zero coupon bond with maturity time $\tau$ that gives to its owner the right to receive one unit of currency on the date $\tau$. The price at time $t < \tau$ of such a contract is denoted by $P(t, \tau)$.

Since we can choose in principle any possible values of $\tau > t$ we have an unlimited number of bonds. However, due to the assumption of absence of arbitrage in the market, the bond prices corresponding to different maturities are not independent. Mathematical models that take into account the joint evolution of zero coupon bonds with different maturities can use the so called forward rate,

$$ f(t, \tau) \equiv -\partial_\tau \log P(t, \tau), \quad \text{for } t \in [0, \tau] \text{ and } \tau \in [0, \tau_{\max}]. $$

See [BR96, Bjö94b, Bjö98, Hul93]. Intuitively, we can think of the instantaneous forward rate $f(t, \tau)$ as the risk-free rate at which we can borrow and lend money over the infinitesimal time interval $[\tau, \tau + d\tau]$, provided that the contract is written at time $t$.

In such models, the absence of arbitrage and friction in the market implies that the drift and the diffusion of the forward rate dynamics must fulfill the Heath-Jarrow-Morton condition [HJM90, HJM92], i.e. that under a risk neutral probability measure, the forward rate $f(t, \tau)$ follows an infinite dimensional Itô stochastic differential equation of the form

$$ \begin{align*}
    df(t, \tau) &= \sum_{j=1}^{J} \sigma_j(t, \tau) \left( \int_{t}^{\tau} \sigma_j(t, s) ds \right) dt + \sum_{j=1}^{J} \sigma_j(t, \tau) dW_j(t), \\
    f(0, \tau) &= f_0(\tau), \quad \tau \in [0, \tau_{\max}].
\end{align*} \tag{1.7} $$

Here $W(t) = (W^1(t), \ldots, W^J(t))$, is a $J$-dimensional Wiener process with independent components, and $\sigma_j(t, T)$, $j = 1, \ldots, J$ are stochastic processes, adapted to the filtration generated by $W$ and that may depend on $f$. Beside this, the initial data $f_0 : [0, \tau_{\max}] \to \mathbb{R}$ is a given deterministic $C^1$ function, obtained from the observable prices $P(0, \tau)$. Figure 1.2 depicts a typical realization of the surface $f(t, \tau)$. Observe that the $t$-sections, $f(t, \cdot)$, are smooth functions of $\tau$, while the $\tau$-sections, $f(\cdot, \cdot)$, are continuous but not smooth.

A basic contract to price is a call option, with exercise time $t_{\max}$ and strike price $K$, on a zero coupon bond with maturity $\tau_{\max}$. The price of this option can be written as

$$ E\left[ \exp \left( -\int_{0}^{t_{\max}} f(s, s) ds \right) \max \left\{ \exp \left( -\int_{t_{\max}}^{\tau_{\max}} f(t_{\max}, \tau) d\tau \right) - K, 0 \right\} \right]. $$
Forward rate realization
time \( t \)
maturity \( \tau \)

**Figure 1.2.** Forward rate modeling: a typical realization of \( f(t, \tau) \).

With this motivation, we consider computation of the functionals

\[
\mathcal{F}(f) \equiv \mathbb{E}
\left[
F\left(\int_0^{t_{\text{max}}} f(s, s) \, ds\right) \ G\left(\int_{\tau_{\text{a}}}^{\tau_{\text{max}}} Q(f(t_{\text{max}}, \tau)) \, d\tau\right)
+ \int_0^{t_{\text{max}}} F\left(\int_0^s f(v, v) \, dv\right) U(f(s, s)) \, ds
\right]
\]

The functions \( F : \mathbb{R} \to \mathbb{R} \), \( G : \mathbb{R} \to \mathbb{R} \), \( Q : \mathbb{R} \to \mathbb{R} \), \( U : \mathbb{R} \to \mathbb{R} \), and their derivatives up to a sufficiently large order \( m_0 \) are assumed to have polynomial growth. Beside this \( 0 < t_{\text{max}} \leq \tau_{\text{a}} < \tau_{\text{max}} \) are given positive numbers.

Here the aim is to provide a computable approximation of the above functional \( \mathcal{F}(f) \). This is accomplished in two steps, namely by a \( t \) and \( \tau \) discretization of (1.7), yielding a numerical solution \( \overline{f} \), and then the computation of sample averages by the Monte Carlo method.

As in the previous Section, an important issue is to estimate the different sources of computational error. Here the new ingredient is the analysis of the \( \tau \) discretization error, which appears together with the time discretization error and statistical error introduced in Section 1.1.1.

### 1.1.5 Overview of Paper 3

This work studies the problem introduced in Section 1.1.4 considering numerical solutions, based on the so called Monte Carlo Euler method, for the price of financial instruments in the bond market, using the Heath Jarrow Morton model for the forward rate [HJM90, HJM92]. The main contribution is to provide rigorous error expansions, with leading error term in computable a posteriori form, offering computational reliability in the use of more complicated HJM multifactor models,
1.2 Stochastic Partial Differential Equations

Let $D$ be a convex bounded polyhedral domain in $\mathbb{R}^d$ and $(\Omega, \mathcal{F}, P)$ be a complete probability space. Here $\Omega$ is the set of outcomes, $\mathcal{F} \subset 2^\Omega$ is the $\sigma$-algebra of events and $P : \mathcal{F} \to [0, 1]$ is a probability measure. Consider the stochastic linear elliptic boundary value problem: find a stochastic process, $u : \Omega \times \overline{D} \to \mathbb{R}$, such that $P$-almost everywhere in $\Omega$, or in other words almost surely (a.s.), we have

$$-\nabla \cdot (a(\omega, \cdot)\nabla u(\omega, \cdot)) = f(\omega, \cdot) \quad \text{on} \quad D,$$

$$u(\omega, \cdot) = 0 \quad \text{on} \quad \partial D. \quad (1.8)$$

Here $a, f : \Omega \times D \to \mathbb{R}$ are stochastic processes that are correlated in space. If we denote by $B(D)$ the Borel $\sigma$-algebra generated by the open subsets of $D$, then $a, f$ where no explicit formula can be found for the pricing of contingent claims. These error estimates can be used to handle simultaneously different sources of error, e.g. time discretization, maturity discretization, and finite sampling. To develop error estimates we use a Kolmogorov backward equation in an extended domain and carry out further the analysis in [STZ01], from general weak approximation of Itô stochastic differential equations in $\mathbb{R}^n$, to weak approximation of the HJM Itô stochastic differential equations in infinite dimensional spaces. Therefore the main new ingredient here is to provide error estimates useful for adaptive refinement not only in time $t$ but also in maturity time $\tau$, see Figure 1.3. Another contribution is the removal of the error in the numerical approximations, produced by the representation of the initial term structure in a finite maturity partition. Finally, the formulae to compute sharp error approximations are simplified by exploiting the structure of the HJM model, reducing the work to compute such error estimates.
are jointly measurable with the $\sigma$-algebra $(\mathcal{F} \otimes B(D))$. It is usual to assume that
the elliptic operator is bounded and uniformly coercive, i.e.

$$\exists a_{\min}, a_{\max} \in (0, +\infty) : a(\omega, x) \in [a_{\min}, a_{\max}], \quad \forall x \in D \quad \text{a.s.}$$

(1.9)

To ensure regularity of the solution $u$ we assume also that $a$ has a uniformly bounded and continuous first derivative, i.e. there exists a real deterministic constant $C$ such that

$$a(\omega, \cdot) \in C^1(\mathcal{D}) \quad \text{and} \quad \max_{\mathcal{D}} |\nabla_x a(\omega, \cdot)| < C \quad \text{a.s.}$$

(1.10)

In addition, the right hand side in (1.8) satisfies

$$\int_{\Omega} \int_{D} f^2(\omega, x) dx \ dP(\omega) < +\infty \quad \text{which implies} \quad \int_{D} f^2(\omega, x) dx < +\infty \quad \text{a.s.}$$

(1.11)

Stochastic Sobolev spaces are used to obtain existence and uniqueness results
for the solution of (1.8). As in the deterministic space, cf. [BS94], the tools are
Hilbert spaces, a suitable weak formulation and then the application of the Lax-
Milgram’s lemma. All the necessary definitions then have to be extended to the
stochastic setting using tensor product of Hilbert spaces, cf. [Lar86] and [ST02],
which is a standard procedure.

**Definition 1.1.** Let $H_1, H_2$ be Hilbert spaces. The tensor space $H_1 \otimes H_2$ is the
completion of formal sums $u(y, x) = \sum_{i=1, \ldots, n} v_i(y) w_i(x)$, $\{v_i\} \subset H_1$, $\{w_i\} \subset H_2$,
with respect to the inner product $(u, \hat{u})_{H_1 \otimes H_2} = \sum_{i,j} (v_i, \hat{v}_j)_{H_1} (w_i, \hat{w}_j)_{H_2}$.

For example, let us consider two domains, $y \in \Gamma$, $x \in D$ and the tensor space
$L^2(\Gamma) \otimes H^1(D)$, with tensor inner product

$$(u, \hat{u})_{L^2(\Gamma) \otimes H^1(D)} =$$

$$\int_{\Gamma} \left( \int_{D} u(y, x) \hat{u}(y, x) dx \right) dy + \int_{\Gamma} \left( \int_{D} \nabla_x u(y, x) \cdot \nabla_x \hat{u}(y, x) dx \right) dy.$$

Thus, if $u \in L^2(\Gamma) \otimes H^k(D)$ then $u(y, \cdot) \in H^k(D)$ a.e. on $\Gamma$ and $u(\cdot, x) \in L^2(\Gamma)$
a.e. on $D$. Moreover, we have the isomorphism

$$L^2(\Gamma) \otimes H^k(D) \simeq L^2(\Gamma; H^k(D)) \simeq H^k(D; L^2(\Gamma))$$

with the definitions

$L^2(\Gamma; H^k(D)) =$

$$\left\{ v : \Gamma \times D \to \mathbb{R} \mid v \text{ is jointly measurable and } \int_{\Gamma} \|v(y, \cdot)\|_{H^k(D)}^2 < +\infty \right\},$$

$H^k(D; L^2(\Gamma)) =$

$$\left\{ v : \Gamma \times D \to \mathbb{R} \mid v \text{ is jointly measurable, } \forall |\alpha| \leq k \quad \exists \quad \partial_\alpha v \in L^2(\Gamma) \otimes L^2(D) \text{ and } \right.$$  

$$\left. \int_{D} \partial_\alpha v(y, x) \phi(x) dx = (-1)^{|\alpha|} \int_{\Gamma} \int_{D} v(y, x) \partial_\alpha \phi(x) dx, \quad \forall \phi \in C_0^\infty(D) \quad \text{a.e. on } \Gamma \right\}$.
1.2. Stochastic Partial Differential Equations

Intuitively, a function \( v(\omega, x) \) that belongs to the stochastic Sobolev space \( \tilde{W}^{s,q}(D) \) will have its realizations accordingly regular, i.e. \( v(\omega, \cdot) \in W^{s,q}(D) \) a.s. We first recall the definition of stochastic weak derivatives. Let \( v \in L^2_p(\Omega) \otimes L^2(D) \), then the \( \alpha \) stochastic weak derivative of \( v \), \( w = \partial_\alpha v \in L^2_p(\Omega) \otimes L^2(D) \), satisfies

\[
\int_D v(\omega, x)\partial^\alpha \phi(x)dx = (-1)^{|\alpha|} \int_D w(\omega, x)\phi(x)dx, \quad \forall \phi \in C_0^\infty(D), \text{ a.s.}
\]

We shall work with stochastic Sobolev spaces \( \tilde{W}^{s,q}(D) = L^q_p(\Omega, W^{s,q}(D)) \) containing stochastic processes, \( v : \Omega \times D \to \mathbb{R} \), that are measurable with respect to the product \( \sigma \)-algebra \( \mathcal{F} \otimes \mathcal{B}(D) \) and equipped with the averaged norms

\[
\|v\|_{\tilde{W}^{s,q}(D)} = E[\|v\|_{W^{s,q}(D)}^q]^{1/q} = E\left[\sum_{|\alpha| \leq s} \int_D |\partial^\alpha v|^q dx\right]^{1/q}, \quad 1 \leq q < +\infty
\]

and

\[
\|v\|_{\tilde{W}^{s,\infty}(D)} = \max_{|\alpha| \leq s} \left(\text{ess sup}_{\Omega \times D} |\partial^\alpha v|\right).
\]

Observe that if \( v \in \tilde{W}^{s,q}(D) \) then \( v(\omega, \cdot) \in W^{s,q}(D) \) a.s. and \( \partial^\alpha v(\cdot, x) \in L^q_p(\Omega) \) a.e. on \( D \) for \( |\alpha| \leq s \). Whenever \( q = 2 \), the above space is a Hilbert space, i.e. \( \tilde{W}^{s,2}(D) = \tilde{H}^s(D) \simeq L^2_p(\Omega) \otimes H^s(D) \).

Now we recall the definition of weak solutions for (1.8). Consider the bilinear form, \( B : \tilde{H}^1_0(D) \times \tilde{H}^1_0(D) \to \mathbb{R} \),

\[
B(v, w) = E\left[\int_D a \nabla v \cdot \nabla w dx\right], \quad \forall v, w \in \tilde{H}^1_0(D).
\]

The standard assumption (1.10) yields both the continuity and the coercivity of \( B \), i.e.

\[
|B(v, w)| \leq a_{\text{max}} \|v\|_{\tilde{H}^1_0(D)} \|w\|_{\tilde{H}^1_0(D)}, \quad \forall v, w \in \tilde{H}^1_0(D), \tag{1.12}
\]

and

\[
a_{\text{min}} \|v\|^2_{\tilde{H}^1_0(D)} \leq B(v, v), \quad \forall v \in \tilde{H}^1_0(D). \tag{1.13}
\]

A direct application of the Lax Milgram’s lemma, see [BS94], implies the existence and uniqueness for the solution to the variational formulation: find \( u \in \tilde{H}^1_0(D) \) such that

\[
B(u, v) = \mathcal{L}(v), \quad \forall v \in \tilde{H}^1_0(D). \tag{1.14}
\]

Here \( \mathcal{L}(v) = E[\int_D f v dx] \), \( \forall v \in \tilde{H}^1_0(D) \) defines a bounded linear functional since the random field \( f \) satisfies (1.11). Moreover, standard arguments from measure theory show that the solution to (1.14) also solves (1.8).

Usually in practical problems the information about the stochastic processes \( a \) and \( f \) is only limited. For example, we may only have approximations for their
expectations and covariance functions to use in the implementation of a numerical method for (1.8).

The Karhunen-Loève expansion, known also as the Proper Orthogonal Decomposition (POD), is a suitable tool for the approximation of stochastic processes. This expansion is used extensively in the fields of detection, estimation, pattern recognition, and image processing as an efficient tool to approximate random processes.

Now we describe the Karhunen-Loève expansion of a stochastic process. Consider a stochastic process $a$ with continuous covariance function, $\text{Cov}[a] : D \times D \rightarrow \mathbb{R}$. Besides this, let $\{(\lambda_i, b_i)\}_{i=1}^{+\infty}$ denote the sequence of eigenpairs associated with the compact self adjoint operator that maps

$$f \in L^2(D) \mapsto \int_D \text{Cov}[a](x, \cdot) f(x) dx \in L^2(D).$$

The real and non-negative eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \ldots$$

satisfy

$$0 \leq \lambda_i \leq \sqrt{\int_D \int_D (\text{Cov}[a](x_1, x_2))^2 dx_1 dx_2, \quad i = 1, \ldots}$$

$$\sum_{i=1}^{+\infty} \lambda_i = \int_D \text{Var}[a](x) dx,$$

and $\lambda_i \to 0$. The corresponding eigenfunctions are orthonormal, i.e.

$$\int_D b_i(x) b_j(x) dx = \delta_{ij}$$

and by Mercer’s theorem, cf. [RSN90] p. 245,

$$\|\text{Cov}[a](x, y) - \sum_{n=1}^{N} \lambda_n b_n(x) b_n(y)\|_{L^\infty(D \times D)} \to 0. \quad (1.15)$$

The Karhunen-Loève expansion of the stochastic process $a$, cf. [Lév92] [Yag87a] [Yag87b], is

$$a_N(\omega, x) = E[a](x) + \sum_{i=1}^{N} \sqrt{\lambda_i} b_i(x) Y_i(\omega)$$

where $\{Y_i\}_{i=1}^{+\infty}$ is a sequence of uncorrelated real random variables, with mean zero and unit variance. These random variables are uniquely determined by

$$Y_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_D (a(\omega, x) - E[a](x)) b_i(x) dx$$
for $\lambda_i > 0$. Then applying (1.15) yields the uniform convergence
\[
\sup_{x \in D} E[(a - a_N)^2](x) = \sup_{x \in D} \left( \text{Var}[a](x) - \sum_{i=1}^{N} \lambda_i b_i^2(x) \right) \to 0, \quad \text{as} \quad N \to \infty.
\]
A standard approach then is to approximate the stochastic coefficients from (1.8) using the Karhunen-Loève expansion and then solve the resulting stochastic partial differential equation.

In this thesis we study the approximation of some statistical moments of the solution from (1.8), e.g. the deterministic function $E[u]$. For example, we are interested in approximating this function using either $L^2(D)$ or $H^1(D)$.

Depending on the structure of the noise that drives an elliptic partial stochastic differential equation, there are different numerical approximations. For example, when the size of the noise is relatively small, a Neumann expansion around the mean value of the equation's operator is a popular alternative. It requires only the solution of standard deterministic partial differential equations, the number of them being equal to the number of terms in the expansion. Equivalently, a Taylor expansion of the solution around its mean value with respect to the noise yields the same result. Similarly, the work [KH92] uses formal Taylor expansions up to second order of the solution but does not study their convergence properties. Recently, the work [BC02] proposed a perturbation method with successive approximations. It also proves that the condition of uniform coercivity of the diffusion is sufficient for the convergence of the perturbation method.

When only the load is stochastic, it is also possible to derive deterministic equations for the moments of the solution. This case was analyzed in [Bab61] and more recently in the work [ST02], where a new method to solve these equations with optimal complexity is presented.

On the other hand, the work by Babuška et al. [Deb00, DB01] and by Ghanem and Spanos [GS91] address the general case where all the coefficients are stochastic. Both approaches transform the original stochastic problem into a deterministic one with higher dimensions, and they differ in the choice of the approximating functional spaces. The work [Deb00] uses finite element to approximate the noise dependence of the solution, while [GRH99, GS91] uses a formal expansion in terms of Hermite polynomials.

Monte Carlo methods are both general and simple to code and they are naturally suited for parallelization. They generate a set of independent identically distributed (iid) approximations of the solution by sampling the coefficients of the equation, using a spatial discretization of the partial differential equation, e.g. by a Galerkin finite element formulation. Then, using these approximations we can compute corresponding sample averages of the desired statistics. The drawback of Monte Carlo methods is their slow rate of convergence. It is worth mentioning that in particular cases their convergence can be accelerated by variance reduction techniques [JCM01] or even the convergence rate improved with Quasi Monte Carlo methods [Caf98, Sob94, Sob98]. Moreover, if the probability density of the
random variable is smooth, the convergence rate of the Monte Carlo method for the approximation of the expected value can be improved, cf. [Nov88, TW98].

Another way to provide a notion of stochastic partial differential equations is based on the Wick product and the Wiener chaos expansion, see [HØUZ96] and [Våg98]. This approach yields solutions in Kondratiev spaces of stochastic distributions which are not the same as those from (1.14). The choice between (1.14) and [HØUZ96] is a modeling decision, based on the physical situation under study. For example, with the Wick product we have $E[a \circ u] = E[a]E[u]$ regardless of the correlation between $a$ and $f$, whereas this is in general not true with the usual product. A numerical approximation for Wick stochastic linear elliptic partial differential equations is studied in [The00], yielding a priori convergence rates.

1.2.1 Overview of Paper 4

We describe and analyze two numerical methods for the linear elliptic problem (1.8). Here the aim of the computations is to approximate the expected value of the solution. Since the approximation of the stochastic coefficients $a$ and $f$ by the Karhunen-Loève expansion is in general not exact, we derive related a priori error estimates. The first method generates iid approximations of the solution by sampling the coefficients of the equation and using a standard Galerkin finite elements variational formulation. The Monte Carlo method then uses these approximations to compute corresponding sample averages. More explicitly, we follow:

1. Give a number of realizations, $M$, and a piecewise linear finite element space on $D$, $X^d_h$.

2. For each $j = 1, \ldots, M$ sample iid realizations of the diffusion $a(\omega_j, \cdot)$ and the load $f(\omega_j, \cdot)$ and find a corresponding approximation $u_h(\omega_j, \cdot) \in X^d_h$ such that

$$
\int_D a(\omega_j, x) \nabla u_h(\omega_j, x) \nabla \chi(x) dx = \int_D f(\omega_j, x) \chi(x) dx, \quad \forall \chi \in X^d_h.
$$

3. Finally use the sample average $\frac{1}{M} \sum_{j=1}^M u_h(\omega_j, \cdot)$ to approximate $E[u]$.

Here we only consider the case where $X^d_h \subseteq H^1_0(D)$ is the same for all realizations, i.e. the spatial triangulation is deterministic.

The second method is based on a finite dimensional approximation of the stochastic coefficients, turning the original stochastic problem into a deterministic parametric elliptic problem. In many problems the source of the randomness can be approximated using just a small number of mutually uncorrelated, sometimes mutually independent, random variables. Take for example the case of a truncated Karhunen-Loève expansion described previously.
Assumption 1.2. Whenever we apply some numerical method to solve (1.8), we assume that the coefficients used in the computations satisfy

\[ a(\omega, x) = a(Y_1(\omega), \ldots, Y_N(\omega), x) \quad \text{and} \quad f(\omega, x) = f(Y_1(\omega), \ldots, Y_N(\omega), x) \] (1.17)

where \( \{Y_j\}_{j=1}^N \) are real random variables with mean value zero, unit variance, are mutually independent, and their images, \( \Gamma_{i,N} \equiv Y_i(\Omega) \) are bounded intervals in \( \mathbb{R} \) for \( i = 1, \ldots, N \). Moreover, we assume that each \( Y_i \) has a density function \( \rho_i : \Gamma_{i,N} \rightarrow \mathbb{R}_+ \) for \( i = 1, \ldots, N \). Use the notations \( \rho(y) = \prod_{i=1}^N \rho_i(y_i) \) \( \forall y \in \Gamma \), for the joint probability density of \( (Y_1, \ldots, Y_N) \), and \( \Gamma = \prod_{i=1}^N \Gamma_{i,N} \subset \mathbb{R}^N \), for the support of such probability density.

After making assumption (1.17), we have by Doob-Dynkin’s lemma, cf. [Øks98], that \( u \), the solution corresponding to the stochastic partial differential equation (1.8) can be described by just a finite, hopefully small, number of random variables, i.e. \( u(\omega, x) = u(Y_1(\omega), \ldots, Y_N(\omega), x) \). Now the goal is to approximate the function \( u(y, x) \). The stochastic variational formulation (1.14) now has a deterministic equivalent in the following: find \( u \in L^2_{\rho}(\Gamma) \otimes H^1_0(D) \) such that

\[
\int_{\Gamma} \rho(y) \int_D a(y, x) \nabla u(y, x) \cdot \nabla v(y, x) dxdy = \int_{\Gamma} \rho(y) \int_D f(y, x) v(y, x) dxdy,
\]

\[ \forall \ v \in L^2_{\rho}(\Gamma) \otimes H^1_0(D). \] (1.18)

A Galerkin finite element method, of either \( h \) or \( p \) version, then approximates the corresponding deterministic solution yielding approximations of the desired statistics, i.e. we seek \( \bar{u}_h \in Z^p_k \otimes X^d_h \), that satisfies

\[
\int_{\Gamma} \rho(y) \int_D a(y, x) \nabla \bar{u}_h(y, x) \cdot \nabla \chi(y, x) dx dy = \int_{\Gamma} \rho(y) \int_D f(y, x) \chi(y, x) dx dy,
\]

\[ \forall \ \chi \in Z^p_k \otimes X^d_h. \] (1.19)

Here \( Z^p_k \subseteq L^2_{\rho}(\Gamma) \) is a finite element space that contains tensor products of polynomials with degree less than or equal to \( p \) on a mesh with size \( k \). Section 7 explains how to use the tensor product structure to efficiently compute \( \bar{u}_h \) from (1.19). Finally, Section 8 uses the a priori convergence rates of the different discretizations to compare the computational work required by (1.16) and (1.19) to achieve a given accuracy in the approximation of \( E[u] \), i.e. their numerical complexity, offering a way to understand the efficiency of each numerical method.
Bibliography


Bibliography


Adaptive Weak Approximation of Stochastic Differential Equations

ANDERS SZEPESSY
RAÚL TEMPONE
Kungl. Tekniska Högskolan
AND
GEORGIOS E. ZOURARIS
University of the Aegean

Abstract
Adaptive time-stepping methods based on the Monte Carlo Euler method for weak approximation of Itô stochastic differential equations are developed. The main result is new expansions of the computational error, with computable leading-order term in a posteriori form, based on stochastic flows and discrete dual backward problems. The expansions lead to efficient and accurate computation of error estimates. Adaptive algorithms for either stochastic time steps or deterministic time steps are described. Numerical examples illustrate when stochastic and deterministic adaptive time steps are superior to constant time steps and when adaptive stochastic steps are superior to adaptive deterministic steps. Stochastic time steps use Brownian bridges and require more work for a given number of time steps. Deterministic time steps may yield more time steps but require less work; for example, in the limit of vanishing error tolerance, the ratio of the computational error and its computable estimate tends to 1 with negligible additional work to determine the adaptive deterministic time steps. © 2001 John Wiley & Sons, Inc.

1 Introduction to the Monte Carlo Euler Method
This work develops adaptive methods and proves a posteriori error expansions, with a computable leading-order term, for weak approximations of Itô stochastic differential equations

\[ dX_k(t) = a_k(t, X(t))dt + \sum_{\ell=1}^{\ell_0} b_\ell^k(t, X(t))dW^\ell(t), \]

\[ k = 1, 2, \ldots, d, \quad t > 0, \]

where \((X(t; \omega))\) is a stochastic process in \(\mathbb{R}^d\), with randomness generated by the independent Wiener processes \(W^\ell(t; \omega), \ell = 1, 2, \ldots, \ell_0\) on \(\mathbb{R}\); cf. [18, 26]. The

© 2001 John Wiley & Sons, Inc.
functions \( a(t, x) \in \mathbb{R}^d \) and \( b^\ell(t, x) \in \mathbb{R}^d, \ell = 1, 2, \ldots, \ell_0 \), are given drift and diffusion fluxes.

The goal is to construct approximations to the expected value \( E[g(X(T))] \) by a Monte Carlo method for a given function \( g \). Examples of such Monte Carlo simulations are to compute option prices in mathematical finance or to simulate stochastic dynamics; cf. [18, 19, 26]. The Monte Carlo method approximates the unknown process \( X \) by the Euler method \( X(t_n) \) (cf. [19, 20]), which is a time discretization based on the nodes \( 0 = t_0 < t_1 < t_2 < \cdots < t_N = T \), with

\[
X(t_{n+1}) - X(t_n) = (t_{n+1} - t_n)a(t_n, X(t_n)) + \sum_{\ell=1}^{\ell_0} (W^\ell(t_{n+1}) - W^\ell(t_n))b^\ell(t_n, X(t_n)).
\]

The aim is to choose the size of the time steps

\[
\Delta t_n \equiv t_{n+1} - t_n
\]

and the number \( M \) of independent identically distributed samples \( X(\cdot; \omega_j), j = 1, 2, \ldots, M \), such that for a given tolerance \( \text{TOL} \),

\[
\left| E[g(X(T))] - \frac{1}{M} \sum_{j=1}^{M} g(X(T; \omega_j)) \right| \leq \text{TOL},
\]

with high probability and as few time steps and realizations as possible.

We study two algorithms for determining the time steps \( \Delta t_n \). In the first simpler algorithm, the size of the time steps \( \Delta t_n \) may vary in time but are deterministic; i.e., the mesh is fixed for all samples. This is useful for solutions with singularities, or approximate singularities, at deterministic times or for problems with small noise.

The second algorithm uses time steps that may vary for different realizations of the solution \( X \). These steps yield optimal time steps in the sense that the expected number of final steps is minimal for this error density and a given global error. Stochastic time steps are advantageous for problems with singularities at random times. The optimal stochastic steps depend on the whole solution \( X(t), 0 < t < T \); in particular, the step \( \Delta t(t) \) at time \( t \) depends also on \( W(\tau), \tau > t \). In stochastic analysis the concept adapted to \( W \) means that a process at time \( t \) only depends on events generated by \( \{W(s) : s < t\} \). In numerical analysis, an adaptive method means that the approximate solution is used to control the error, e.g., to determine the time steps. Our stochastic steps are in this sense adaptive nonadapted, since \( \Delta t(t) \) depends on \( W(\tau), \tau > t \).

For a fixed number of time steps, the algorithm with deterministic steps requires less work than the algorithm with stochastic steps. The number of realizations needed to determine the deterministic time steps is asymptotically at most \( \Theta(\text{TOL}^{-1}) \), while the number of realizations for the Monte Carlo method to approximate \( E[g(X(T))] \) is \( \Theta(\text{TOL}^{-2}) \). Therefore, the additional work to determine
optimal deterministic time steps becomes negligible as the error tolerance tends to zero. Section 4 compares deterministic with stochastic time steps.

The computational error naturally separates into the two parts,

\[ E[g(X(T))] - \frac{1}{M} \sum_{j=1}^{M} g(\overline{X}(T; \omega_j)) \]

\[ = \left( E[g(X(T)) - g(\overline{X}(T))] \right) + \left( E[g(\overline{X}(T))] - \frac{1}{M} \sum_{j=1}^{M} g(\overline{X}(T; \omega_j)) \right) \]

\[ \equiv \mathcal{E}_T + \mathcal{E}_S. \]

The time steps for the trajectories \( \overline{X} \) are determined from statistical approximations of the time discretization error \( \mathcal{E}_T \). The number of realizations, \( M \) of \( \overline{X} \), are determined from the statistical error \( \mathcal{E}_S \). Therefore, the number of realizations can be asymptotically determined by the central limit theorem

\[ \sqrt{M} \mathcal{E}_S \rightarrow \chi, \]

where the stochastic variable \( \chi \) has the normal distribution, with mean zero and variance \( \text{Var}[g(\overline{X}(T))] \).

Efficient adaptive time stepping methods, with theoretical basis, use a posteriori error information, since the a priori knowledge usually cannot be as precise as the a posteriori. This work develops adaptive time-stepping methods by proving in Theorems 2.2 and 3.3 error estimates of \( \mathcal{E}_T \) with leading-order terms in computable a posteriori form. Theorem 2.2 uses deterministic time steps, while Theorem 3.3 also holds for stochastic time steps, which are not adapted.

The main inspiration of Theorems 2.2 and 3.3 is the work by Talay and Tubaro in [32]. They proved that for uniform deterministic time steps \( \Delta t = T/N \),

\[ E[g(X(T))] - g(\overline{X}(T))] = \int_0^T \frac{T}{N} E[\Psi(s, X(s))] ds + O\left( \frac{1}{N^2} \right), \]

where

\[ \Psi(t, x) \equiv \frac{1}{2} (a_k a_n \partial_{kn} u)(t, x) + (a_i d_{jk} \partial_{ij} u)(t, x) \]

\[ + \frac{1}{2} (d_{ij} d_{kn} \partial_{ijkn} u)(t, x) + \frac{1}{2} \partial_t u(t, x) \]

\[ + \left( a_i \frac{\partial}{\partial t} \partial_i u \right)(t, x) + \left( d_{ij} \frac{\partial}{\partial t} \partial_{ij} u \right)(t, x) \]

is based on the definition of the conditional expectation

\[ u(t, x) \equiv E[g(X(T)) \mid X(t) = x] \]

and the notation

\[ d_{ij} \equiv \frac{1}{2} b_i b_j, \quad \partial_k \equiv \frac{\partial}{\partial x_k}, \quad \partial_{ki} \equiv \frac{\partial^2}{\partial x_k \partial x_i}, \ldots, \]
with the summation convention; i.e., if the same subscript appears twice in a term, the term denotes the sum over the range of this subscript, e.g.,

\[ c_{ik} \partial_k b_j \equiv \sum_{k=1}^{d} c_{ik} \partial_k b_j . \]

For a derivative \( \partial_\alpha \), the notation \(|\alpha|\) is its order.

The proof of Talay and Tubaro extends directly to nonuniform deterministic time steps. The difference between the expansion in (2.5) of Theorem 2.2 and (1.4) is in the leading-order terms of the expansion. In (2.5) the leading terms are directly computable known variables and involve computed stochastic flow approximations of only up to second derivatives of \( u(t, x) \). On the other hand, the expansion (1.4) is an a priori estimate based on the unknown exact solution \( X \) and the unknown derivatives \( \partial_\alpha u \) up to fourth order.

An adaptive time-stepping algorithm can, of course, use only computed variables. Suppose that \( X(t) \in \mathbb{R}^d \) and that the Jacobians of the fluxes \( a \) and \( b \) are sparse; i.e., \(#\{j : \partial_j a_i \Delta t + b^j_i \Delta W^k \neq 0\} \) is independent of the dimension \( d \). Then these reductions decrease the work to evaluate the error estimate for one realization and one time step to the order \( \Theta(d^2) \) in Theorem 2.2 compared to the work of order \( \Theta(d^4) \) for evaluating all fourth-order derivatives.

Kloeden and Platen [19] extend the results of Talay and Tubaro on the existence of leading-order error expansion in a priori form, for first- and second-order schemes, to general weak approximations of higher order. Extensions of [32] to approximation of nonsmooth functions \( g \) and the probability density of \( X \) are studied in [3, 4].

The main new idea here is the efficient use of stochastic flows and dual functions to obtain the error expansion with computable leading-order term in Theorems 2.2 and 3.3, also including nonadapted adaptive time steps. The use of dual functions is standard in optimal control theory and in particular for adaptive mesh control for ordinary and partial differential equations; see [2, 5, 7, 16, 17]. The authors are not aware of other error expansions in an a posteriori form or adaptive algorithms for weak approximation of stochastic differential equations. In particular, error estimates with stochastic nonadapted time steps do not seem to have been studied before.

Asymptotical optimal adapted adaptive methods for strong approximation of stochastic differential equations are analyzed in [15, 24], which include the hard problem of obtaining lower error bounds for any method based on the number of evaluations of \( W \) and requires roughly the \( L^2 \)-norm in time of the diffusion max \( d_{ij} \) to be positive pathwise. The work [11] treats a first study on strong adaptive approximation.

Theorem 2.2 describes the basis of an adaptive algorithm, with deterministic time steps, to estimate the computational error; the time steps are then chosen by

\[ (\Delta t_n)^2 |E[\rho(t_n, \omega)]| = \text{const}, \]
where \( \rho(t_n, \omega) \Delta t_n \) is the function defined by the sum of the terms inside of the two square brackets in (2.5) based on the weight functions \( \varphi \) and \( \varphi' \) defined in (2.7) and (2.9). Provided the path \( \overline{X}(t_n), n = 0, 1, \ldots, N, \) is stored, the leading-order error bound can be evaluated by solving, step by step, the two backward problems (2.7) and (2.9).

The backward evolutions (2.7) and (2.9) of the weight functions \( \varphi \) and \( \varphi' \) avoid solving for the two variables \( t \) and \( s \) present in \( \partial \overline{X}(t)/\partial x(s) \), which appears in the forward \( t \)-evolution equation for \( \partial \overline{X}(t)/\partial x(s) \) in the identity \( \varphi_i(t_n) = \partial_j g(\overline{X}(T)) \partial \overline{X}_j(t_n)/\partial x_i(t_n) \); cf. (2.37). A solution with two variables \( s \) and \( t \) would require work of the order \( N^2 \) for each realization instead of the corresponding work of the order \( N \) in Theorems 2.2 and 3.3. Computational aspects of the algorithms are discussed in more detail in Sections 3 and 5. Section 3 proves in Theorem 3.3 an analogous error expansion

\[
E[g(X(T)) - g(\overline{X}(T))] \approx E \left[ \sum_n \rho(t_n, \omega)(\Delta t_n)^2 \right],
\]

which can also be used for stochastic time steps. The leading-order terms of the expansion have less variance compared to the expansion in Theorem 2.2 but use up to the third variation, which requires more computational work per realization. The optimal time steps, which for given error tolerance and error density minimize the expected number of time steps, are then obtained by adapting the step sizes to each individual realization \( X(\cdot, \omega) \) by choosing

\[
(\Delta t_n)^2 |\rho(t_n, \omega)| = \text{const}.
\]

The constant is the same for all realizations. Here the evaluation of a realization of the Wiener process \( W(\cdot, \omega) \) for the Euler method is constructed by successive Brownian bridges; cf. [18] in order to adaptively increase the resolution of \( W \). These optimal steps depend on the whole path \( W \).

The focus in the paper is on error estimates for weak convergence of stochastic differential equations. However, the estimates in Theorems 2.2 and 3.3 also work without noise, and they give a contribution to the analysis of a posteriori error estimates and adaptive methods for deterministic ordinary differential equations. A posteriori error estimates for deterministic differential equations are usually based on linearization around the solution path (cf. [5, 8, 14, 16, 17]). Linearization in the classical sense by directly comparing two different solution trajectories is not relevant for weak convergence, i.e., convergence in distribution, since the trajectories might not be close. The technique used here is based on the transition probability density and Kolmogorov’s backward equation, which was developed in [30, 31] to analyze uniqueness and dependence on initial conditions for weak solutions of stochastic differential equations. The analogous technique for deterministic equations was introduced in [1, 12]. The deterministic restriction of Theorem 2.2 is extended to general higher-order methods for ordinary and partial differential equations in [22, 23].
Section 2 proves error estimates for deterministic time steps. Section 3 derives an alternative error expansion with less statistical error but more computational work, which also works for stochastic time steps. Section 4 compares deterministic step size control with stochastic time steps, which can be better when the noise is not small. Section 5 presents implementations of adaptive algorithms and numerical experiments.

2 An Error Estimate with Deterministic Time Steps

This section first proves an error expansion for deterministic time steps in Theorem 2.2. The starting point for the analysis is Lemma 2.1. It uses the fact that the Euler method can be extended, for theoretical purposes only, to

\[ X(t) - X(t_n) = \int_{t_n}^{t} \alpha(s; X)ds + \sum_{\ell=1}^{\ell_0} \int_{t_n}^{t} b^\ell(s; X)dW^\ell(s) \]

where \( \alpha \) and \( b^\ell \) are the piecewise constant approximations

\[ \alpha(s; X) = a(t_n, X(t_n)) \quad \text{and} \quad b^\ell(s; X) = b^\ell(t_n, X(t_n)) \]

for \( s \in [t_n, t_{n+1}) \).

The dependence of the noise \( \omega_j \) is often omitted in the notation. We also use the space \( C^{m_0}([0, T] \times \mathbb{R}^d) \) of functions with bounded continuous derivatives of order up to \( m_0 \) in the supremum norm on \( [0, T] \times \mathbb{R}^d \) and the space \( C^{m_0}_{\text{loc}}([0, T] \times \mathbb{R}^d) \) of functions with continuous derivatives of order up to \( m_0 \) which have bounded maximum norm on each compact set in \( [0, T] \times \mathbb{R}^d \). The space \( C^{m_1, m_2}_{\text{loc}}([0, T] \times \mathbb{R}^d) \) denotes the subset of \( C^{m_1+m_2}_{\text{loc}}([0, T] \times \mathbb{R}^d) \) with time derivatives of up to order \( m_1 \) and space derivatives of up to order \( m_2 \).

**Lemma 2.1** Suppose that, for some \( m_0 > [d/2] + 10 \), there are positive constants \( k \) and \( C \) such that

\[ g \in C^{m_0}_{\text{loc}}(\mathbb{R}^d), \quad |\partial_\alpha g(x)| \leq C(1 + |x|^k), \quad \text{for all } |\alpha| \leq m_0, \]

\[ E[|X(0)|^{2k+d+1} + |X(0)|^{2k+d+1}] \leq C, \]

and

\[ a \text{ and } b \text{ are bounded in } C^{m_0}([0, T] \times \mathbb{R}^d). \]
Then the solution $X$ of the Itô differential equation (1.1) and its Euler approximation $\overline{X}$ in (1.2) and (2.1), based on deterministic time steps, satisfy

$$E[g(X(T)) - g(\overline{X}(T))|F_T] =$$

$$\int_0^T E\left[(a_k(t, \overline{X}(t)) \overline{X}(t)) \overline{\partial}_k u(t, \overline{X}(t)) \right] dt$$

$$+ \int_0^T E\left[(d_{ij}(t, \overline{X}(t)) - \overline{d}_{ij}(t; \overline{X})) \overline{\partial}_{ij} u(t, \overline{X}(t)) \right] dt$$

$$+ E[u(0, X(0)) - u(0, \overline{X}(0))] ,$$

where

$$d_{ij} = \frac{1}{2} b_i^t b_j^f , \quad \overline{d}_{ij} = \frac{1}{2} \overline{b}_i^t \overline{b}_j^f ,$$

and

$$u(t, x) = E[g(X(T)) | X(t) = x] .$$

Lemma 2.1 is combined with stochastic flows to derive the a posteriori error expansion in Theorem 2.2. Theorem 2.2 is based on the variations of the processes $X$ and $\overline{X}$. For a process $\overline{X}$, the first variation of a function $F(X(T))$ with respect to a perturbation in the initial location of the path $\overline{X}$ at time $s$ is denoted by

$$F'(T; s) = \partial_{x(s)} F(\overline{X}(T))$$

$$= \left( \frac{\partial}{\partial x_1} F(\overline{X}(T); \overline{X}(s) = x), \ldots, \frac{\partial}{\partial x_d} F(\overline{X}(T); \overline{X}(s) = x) \right) .$$

The proof of Theorem 2.2 uses mainly that the error in replacing $g(X(T))$ in Lemma 2.1 by $g(\overline{X}(T))$ in the representation (2.4) of $\partial_{x} u$ yields the small deterministic remainder term $\int_0^T O((\Delta t)^2) dt$ in (2.5) of Theorem 2.2, which is analogous to the $O(1/N^2)$ term in (1.4), and needs some a priori estimate to be controlled. Lemma 2.1 can be applied to estimate this error.

The second important ingredient in the proof is the Markov property of $\overline{X}$. Let the standard $\sigma$-algebra generated by $\{W(s) : s \leq t_n\}$ be $\mathcal{F}_{t_n}$. Then the nested expected values

$$E[a_j(t, \overline{X}(t)) \overline{\partial}_{ij}(t) E[g(\overline{X}(T)) | \mathcal{F}_{t}]$$

in (2.3) can, by definition (2.7) of $\varphi$, be decoupled to

$$E[a_j(t, \overline{X}(t)) \varphi_j(t)] ,$$

which reduces the computational complexity substantially; see Lemma 2.5.

**Theorem 2.2** Suppose that $a$, $b$, $g$, $X$, and $\overline{X}$ satisfy the assumptions in Lemma 2.1 and that the initial data $X(0)$ and its approximation $\overline{X}(0)$ have the same
distribution. Then the time discretization error has the expansion
\[
E[g(X(T)) - g(X(T))] = 
\sum_{n=0}^{N-1} \sum_{j=1}^{M} \left[ (a_i(t_{n+1}, X(t_{n+1}; \omega_j)) - a_i(t_n, X(t_n; \omega_j))) \varphi_i(t_{n+1}; \omega_j) \right] \frac{\Delta t_n}{2M}
\]

\[
+ \sum_{n=0}^{N-1} \sum_{j=1}^{M} \left[ (d_{ik}(t_{n+1}, X(t_{n+1}; \omega_j)) \varphi_{ik}^\prime(t_{n+1}; \omega_j) \right] \frac{\Delta t_n}{2M}
\]

\[
+ \sum_{n=0}^{N-1} (\Delta t_n)^2 \left\{ \varnothing(\Delta t_n) + \sum_{m=n}^{N-1} \varnothing((\Delta t_m)^2) \right\} + \int_0^T (I_M + I_{II_M}) dt,
\]

where the two leading-order terms are in computable a posteriori form and based on the discrete dual functions \( \varphi(t_n) \in \mathbb{R}^d \) and \( \varphi'(t_n) \in \mathbb{R}^{d \times d} \), which are determined as follows: For all \( t_n \) and \( x \in \mathbb{R}^d \), let \( \Delta W_n \equiv W(t_{n+1}) - W(t_n) \) and

\[
c_i(t_n, x) \equiv x_i + \Delta t_n a_i(t_n, x) + \Delta W_n^i b_i(t_n, x);
\]

then the function \( \varphi \) is defined by the dual backward problem
\[
\varphi_i(t_n) = \partial_i c_j(t_n, X(t_n)) \varphi_j(t_{n+1}), \quad t_n < T,
\]

and its first variation
\[
\varphi_{ik}^\prime(t_n; \omega) = \partial_{x_k(t_n)} \varphi_i(t_n; \omega) = \frac{\partial \varphi_i(t_n; \overline{X}(t_n) = x)}{\partial x_k}
\]
satisfies for \( t_n < T \)
\[
\varphi_{ik}^\prime(t_n) = \partial_i c_j(t_n, X(t_n)) \partial_k c_p(t_n, X(t_n)) \varphi_{jp}^\prime(t_{n+1})
\]

\[
+ \partial_{ik} c_j(t_n, X(t_n)) \varphi_j(t_{n+1}) ,
\]

\[
\varphi_{ik}^\prime(T) = \partial_{ik} g(X(T)).
\]

The distributions of the statistical errors \( \sqrt{M} I_M \) and \( \sqrt{M} II_M \) tend to the normal distributions with mean zero and variance
\[
\text{Var}[ (a_i(t_{n+1}, X(t_{n+1})) - a_i(t_n, X(t_n))) \varphi_i(t_{n+1}) ] = \varnothing(\Delta t_n)
\]

and
\[
\text{Var}[ (d_{ik}(t_{n+1}, X(t_{n+1})) - d_{ik}(t_n, X(t_n))) \varphi_{ik}^\prime(t_{n+1}; t_{n+1}) ] = \varnothing(\Delta t_n),
\]

respectively.
Proof of Lemma 2.1: A standard energy estimate, using the regularity assumptions on $a$, $b$, and $g$, can be combined with the Sobolev inequality to show that there exists a unique solution $u \in C^1_{0, \text{loc}}([0, T] \times \mathbb{R}^d)$ of the Kolmogorov backward equation

\begin{equation}
Lu := -\frac{\partial}{\partial t} u - a_k \partial_k u - \frac{1}{2} b_k^\ell b_n^\ell \partial_{kn} u = 0, \quad u(T, \cdot) = g,
\end{equation}

satisfying the polynomial growth condition

$$\max_{0 \leq t \leq T} |\partial_\alpha u(t, x)| \leq C (1 + |x|^{k+(d+1)/2}), \quad |\alpha| \leq 6,$$

for some positive $k$ and $C$; cf. [10]. The Feynman-Kac formula without potential (cf. [18]) implies that $u$ can be represented by the expected value

\begin{equation}
u(t, x) = E[g(X(T)) \mid X(t) = x].
\end{equation}

The Itô formula

$$du(t, \overline{X}(t)) = \left(\frac{\partial}{\partial t} u(t, \overline{X}(t)) + \overline{a}_i(t; \overline{X}) \partial_i u(t, \overline{X}(t)) + \overline{d}_{ij}(t; \overline{X}) \partial_{ij} u(t, \overline{X}(t))\right)dt + \overline{b}_i^\ell(t; \overline{X}) \partial_i u(t, \overline{X}(t))dW^\ell(t)$$

combined with the Kolmogorov equation (2.12) to eliminate $\frac{\partial}{\partial t} u$ yields

\begin{equation}
u(0, \overline{X}(0)) - g(\overline{X}(T)) = \int_0^T \left(\alpha_i(t, \overline{X}(t)) - \overline{a}_i(t; \overline{X})\right) \partial_i u(t, \overline{X}(t))dt + \int_0^T \left(d_{ij}(t, \overline{X}(t)) - \overline{d}_{ij}(t; \overline{X})\right) \partial_{ij} u(t, \overline{X}(t))dt + \int_0^T \overline{b}_i^\ell(t; \overline{X}) \partial_i u(t, \overline{X}(t))dW^\ell(t) .
\end{equation}

The expected value of the last integral is zero, by the martingale property of Itô integrals, and the construction of $u$ in (2.13) shows that

$$E[u(0, X(0))] = E[g(X(T))].$$

Therefore, the expected value of equality (2.14) implies the result (2.3). \hfill \Box

Proof of Theorem 2.2: The purpose of Theorem 2.2 is to replace the estimate in Lemma 2.1 by an expansion with computable leading-order term. The expected values $\partial_\alpha u(x, t) = \partial_\alpha E[g(X(T)) \mid X(t) = x]$ will be replaced by the corresponding derivatives of

$$\overline{u}(x, t) = E[g(\overline{X}(T)) \mid \overline{X}(t) = x],$$

based on the known approximate solution $\overline{X}$. This proof is divided into the three steps of Lemmas 2.3 through 2.5: The first lemma estimates the quadrature error; the second lemma shows the error in replacing $\partial_\alpha u$ by $\partial_\alpha \overline{u}$; and the third lemma derives the representation of $\partial_\alpha \overline{u}$ by the dual functions $\varphi$ and $\varphi'$ solving the backward evolution problems (2.7) and (2.9). The lemmas are prepared by first deriving the
representation of $\partial_a u$ and $\partial_a \tilde{u}$ in terms of expected values of stochastic flows, i.e., first and higher variations of $X$ and $\tilde{X}$, respectively.

In the theorem, the derivatives $\partial_a \tilde{u}$ are computed by means of the discrete dual functions $\varphi$ and $\varphi'$, defined in (2.6)–(2.9). In this first step, $\partial_a u$ and its approximation $\partial_a \tilde{u}$ are evaluated by expected values of stochastic flows of $X$ and $\tilde{X}$, respectively. Lemma 2.3 then relates the duals and the stochastic flows, which have the following equations. Recall the definition of the variation of a process $Y$: The first variation of a function $F(Y(T))$ with respect to a perturbation in the initial location of the path $Y$ at time $s$ is denoted by

$$F'(T; s) = \partial_{y(s)} F(Y(T))$$

(2.15)

$$= \left( \frac{\partial}{\partial y_1} F(Y(T); Y(s) = y), \ldots, \frac{\partial}{\partial y_d} F(Y(T); Y(s) = y) \right).$$

Let

$$\delta_{ik} \equiv \begin{cases} 0, & i \neq k, \\ 1, & i = k. \end{cases}$$

The equations for the first variation,

$$dX'_{ij} = \partial_i a_i(s, X(s))X'_{kj}(s)ds + \partial_k b_i^\epsilon(s, X(s))X'_{kj}(s)dW^\epsilon(s), \quad s > t,$$

(2.16)

$$X'_{ij}(t) = \delta_{ij},$$

the second variation,

$$dX''_{ij} = (\partial_i a_i(s, X(s))X''_{kj}(s) + \partial_k a_i(s, X(s))X'_{kj}(s)X''(s))ds$$

$$+ (\partial_k b_i^\epsilon(s, X(s))X'_{kj}(s))dW^\epsilon(s), \quad s > t,$$

(2.17)

$$X''_{ij}(t) = 0,$$

the third variation, for $s > t$,

$$dX'''_{ijmn} =$$

$$\left( \partial_k a_i(s, X(s))X'''_{kjmn}(s) + \partial_k a_i(s, X(s))X'_{kj}(s)X'''_{rmn}(s) \right. \right.$$

$$+ \partial_k a_i(s, X(s))X''_{kn}(s)X'''_{rjm}(s) + \partial_k a_i(s, X(s))X'_{kn}(s)X''(s)$$

$$+ \partial_k a_i(s, X(s))X''_{kn}(s)X'''_{rjm}(s) + \partial_k b_i^\epsilon(s, X(s))X'_{kn}(s)X'''(s)$$

$$+ \partial_k b_i^\epsilon(s, X(s))X''_{kn}(s)X'''(s) + \partial_k b_i^\epsilon(s, X(s))X'_{kn}(s)X''(s)$$

$$+ \partial_k b_i^\epsilon(s, X(s))X''_{kn}(s)X'''(s) + \partial_k b_i^\epsilon(s, X(s))X'_{kn}(s)X''(s)$$

$$+ \partial_k b_i^\epsilon(s, X(s))X''_{kn}(s)X'''(s) + \partial_k b_i^\epsilon(s, X(s))X'_{kn}(s)X''(s)$$

$$+ \partial_k b_i^\epsilon(s, X(s))X''_{kn}(s)X'''(s) + \partial_k b_i^\epsilon(s, X(s))X'_{kn}(s)X''(s)$$

(2.18)

$$X'''_{ijmn}(t) = 0,$$

and similarly the fourth variation,

$$dX''''_{ijmn} = \partial_p \text{(right-hand side of (2.18))},$$

(2.19)

$$X''''_{ijmn}(t) = 0,$$
imply the representation with expected values of stochastic flows, cf. [28, 29].

\[ (2.20) \quad \partial_k u(t, x) = E\left[ \partial_i g(X(T))X'_{ik}(T) \mid X'_{ij}(t) = \delta_{ij}, X(t) = x \right], \]

\[ (2.21) \quad \partial_{kn} u(t, x) = E\left[ \partial_k g(X(T))X''_{ikn}(T) \right. \]
\[ \left. + \partial_{ir} g(X(T))X'_{irk}(T)X''_{rkn}(T) \mid X''_{ikn}(t) = 0, X'_{ij}(t) = \delta_{ij}, X(t) = x \right], \]

\[ (2.22) \quad \partial_{knm} u(t, x) = E\left[ \partial_k g(X(T))X''_{iknm}(T) + \partial_{ir} g(X(T))X'_{irk}(T)X''_{rknm}(T) \right. \]
\[ \left. + \partial_{ir} g(X(T))X'_{im}(T)X''_{rkm}(T) + \partial_{ir} g(X(T))X'_{ikm}(T)X''_{irm}(T) \mid X''_{iknm}(t) = X''_{ikn}(t) = 0, X'_{ij}(t) = \delta_{ij}, X(t) = x \right], \]

and

\[ (2.23) \quad \partial_{kmp} u(t, x) = \partial_p \text{(right-hand side of (2.22))}. \]

Let \( Y = (X, X', X'', X''', X''''\)^T \) and let \( I \) denote the \( d \times d \) identity matrix. Then the system (2.16)–(2.19) can be written

\[ (2.24) \quad dY = A(t, Y)dt + B^\ell(t, Y)dW^\ell(t), \quad t > t_0, \quad Y(t_0) = (x, I, 0, 0, 0)^T. \]

Furthermore, rewrite the representation (2.20)–(2.23) as

\begin{align*}
    f_i(Y) &\equiv \partial_k g(X)X'_{ki}, \\
    f_{ij}(Y) &\equiv \partial_k g(X)X''_{kij} + \partial_{kn} g(X) X'_{ki} X'_{nj}, \\
    f_{ijm}(Y) &\equiv \partial_k g(X)X''''_{kijm} + \partial_{kn} g(X) X'_{ki} X''''_{njm} \\
    &\quad + \partial_{kn} g(X) X'_{kj} X''''_{njm} + \partial_{kn} g(X) X'_{km} X''''_{nji} \\
    &\quad + \partial_{kn} g(X) X'_{ki} X''''_{njm}, \\
    f_{ijmn}(Y) &\equiv \partial_n \text{(right-hand side of } f_{ijm}(Y) \text{ in (2.25))}. 
\end{align*}

The Euler approximation \( \tilde{Y} = (\tilde{Y}^0, \tilde{Y}^1, \tilde{Y}^2, \tilde{Y}^3, \tilde{Y}^4)^T \) of \( Y \) in (2.24) with piecewise constant drift and diffusion fluxes

\[ (2.26) \quad d\tilde{Y} = \tilde{A}(t; \tilde{Y})dt + \tilde{B}^\ell(t; \tilde{Y})dW^\ell(t), \]

defined as in (2.1), yields the following error of the expected value:

\[ E\left[ f_\alpha(Y(T)) - f_\alpha(\tilde{Y}(T)) \mid Y(t) = \tilde{Y}(t) = (x, I, 0, 0, 0)^T \right]. \]

An important consequence of the Euler method is that the first variation of \( \tilde{Y}^0 \equiv \tilde{Y} \) is in fact equal to the function \( \tilde{Y}^1 \) in \( \tilde{Y} \). The expected value has, by Lemma 2.1, the
estimate

\[ E\left[ f_\alpha(Y(T)) - f_\alpha(\overline{Y}(T)) \mid Y(t) = \overline{Y}(t) = (x, I, 0, 0, 0)^T \right] = \]

\begin{equation}
(2.27)
\int_t^T E\left[ (A - \overline{A})_k \partial_k v^\alpha(s, \overline{Y}(s)) \mid \overline{Y}(t) = (x, I, 0, 0, 0)^T \right] ds
+ \int_t^T E\left[ (D - \overline{D})_{kn} \partial_{kn} v^\alpha(s, \overline{Y}(s)) \mid \overline{Y}(t) = (x, I, 0, 0, 0)^T \right] ds
\end{equation}

to be used in Lemmas 2.3 through 2.5, where

\[ D_{kn} = \frac{1}{2} B_k^\ell B_n^\ell, \quad \overline{D}_{kn} = \frac{1}{2} \overline{B}_k^\ell \overline{B}_n^\ell, \]

and

\[- \frac{\partial}{\partial t} v^\alpha \equiv A_k \partial_k v^\alpha \equiv D_{kn} \partial_{kn} v^\alpha = 0, \quad t < T, \quad v^\alpha(T, \cdot) = f_\alpha.\]

The next steps in the proof are the three lemmas.

**Lemma 2.3** Suppose that the assumptions in Lemma 2.1 hold. Then the quadrature error satisfies

\[
\int_{t_n}^{t_{n+1}} E\left[ (a_i(t, \overline{X}(t)) - \overline{a}_i(t; \overline{X})) \partial_i u(t, \overline{X}(t)) \right] dt
- E\left[ (a_i(t_{n+1}, \overline{X}(t_{n+1})) - a_i(t_n, \overline{X}(t_n))) \partial_i u(t_{n+1}, \overline{X}(t_{n+1})) \right] \frac{\Delta t_n}{2} = O((\Delta t_n)^3)
\]

and

\[
\int_{t_n}^{t_{n+1}} E\left[ (d_{ij}(t, \overline{X}(t)) - \overline{d}_{ij}(t; \overline{X})) \partial_{ij} u(t, \overline{X}(t)) \right] dt
- E\left[ (d_{ij}(t_{n+1}, \overline{X}(t_{n+1})) - d_{ij}(t_n, \overline{X}(t_n))) \partial_{ij} u(t_{n+1}, \overline{X}(t_{n+1})) \right] \frac{\Delta t_n}{2} = O((\Delta t_n)^3).
\]

**Proof:** Introduce the notation

\[ f(t, \overline{X}(t)) \equiv (d_{ij}(t, \overline{X}(t)) - \overline{d}_{ij}(t; \overline{X})) \partial_{ij} u(t, \overline{X}(t)), \]

\[ \dot{f}(t) \equiv \frac{t - t_n}{\Delta t_n} (d_{ij}(t_{n+1}, \overline{X}(t_{n+1})) - \overline{d}_{ij}(t_n; \overline{X})) \partial_{ij} u(t_{n+1}, \overline{X}(t_{n+1})). \]

Then the quadrature error satisfies

\[
\int_{t_n}^{t_{n+1}} E[f(t, \overline{X}(t)) - \dot{f}(t)] dt =
\int_{t_n}^{t_{n+1}} E[(d_{ij}(t, \overline{X}(t)) - \overline{d}_{ij}(t; \overline{X})) \partial_{ij} u(t, \overline{X}(t))] dt
- E[(d_{ij}(t_{n+1}, \overline{X}(t_{n+1})) - \overline{d}_{ij}(t_n; \overline{X})) \partial_{ij} u(t_{n+1}, \overline{X}(t_{n+1}))] \frac{\Delta t_n}{2},
\]
and \( E[f(t)] \) is the linear nodal projection of the smooth function \( E[f(t, \bar{X}(t))] \) in the interval \([t_n, t_{n+1}]\). Therefore, a standard interpolation estimate yields
\[
\left| \int_{t_n}^{t_{n+1}} E[f(t, \bar{X}(t)) - \bar{f}(t)] dt \right| \leq \frac{1}{8}(\Delta t_n)^2 \int_{t_n}^{t_{n+1}} \left| \frac{d^2}{dt^2} E[f(t, \bar{X}(t))] \right| dt.
\]

Itô’s formula shows that
\[
\begin{align*}
\frac{d}{dt} E[f(t, \bar{X}(t))] &= E[Lf(t, \bar{X}(t))] \leq C, \\
\frac{d^2}{dt^2} E[f(t, \bar{X}(t))] &= E[L^2 f(t, \bar{X}(t))] \leq C,
\end{align*}
\]
where
\[
Lf \equiv \frac{\partial}{\partial t} f + \bar{a}_i \partial_i f + \frac{1}{2} \bar{d}_{ij} \partial_{ij} f,
\]
which combined with (2.28) proves the estimate of the diffusion terms in the lemma. The estimate of the drift terms follows analogously.

\[\square\]

Lemma 2.4 Let the standard \( \sigma \)-algebra generated by \( \{W(s) : s \leq t_n\} \) be \( \mathcal{F}_{t_n} \), and let the piecewise constant mesh function \( 1_t \) be defined by
\[
1_t = \begin{cases} 
1, & t_n \leq t < t_{n+1}; \\
0, & \text{otherwise}.
\end{cases}
\]
Suppose that the assumptions in Lemma 2.1 hold. Then the discretization errors of the stochastic flows satisfy, for \( t_n \leq t < t_{n+1},
\]
\[
\left| \partial_\alpha (u - \bar{u})(t, \bar{X}(t)) \right| = \int_t^T \mathcal{O}(\Delta t(s)) ds, \quad |\alpha| \leq 4,
\]
and
\[
\begin{align*}
E[(a_i(t, \bar{X}(t)) - \bar{a}_i(t_n; \bar{X}))(\partial_i u(t, \bar{X}(t)) - \partial_i \bar{u}(t, \bar{X}(t)))] &= \\
\Delta t_n \int_t^T \mathcal{O}(\Delta t(s)) ds.
\end{align*}
\]

\[
\begin{align*}
E[(d_{ij}(t, \bar{X}(t)) - \bar{d}_{ij}(t_n; \bar{X}))(\partial_{ij} u(t, \bar{X}(t)) - \partial_{ij} \bar{u}(t, \bar{X}(t)))] &= \\
\Delta t_n \int_t^T \mathcal{O}(\Delta t(s)) ds.
\end{align*}
\]

Proof: Lemma 2.1 and the representation of \( \partial_\alpha u \) and \( \partial_\alpha \bar{u} \) by the stochastic flows (2.20)–(2.21) and (2.25)–(2.26) in (2.27) show that, for \( |\alpha| \leq 4,
\]
\[
\begin{align*}
\partial_\alpha (u - \bar{u})(t, \bar{X}(t)) = \\
\int_t^T E[(A_i - \bar{A}_i)\partial_i v^\alpha(s, \bar{Y}(s)) + (D_{ij} - \bar{D}_{ij})\partial_{ij} v^\alpha(s, \bar{Y}(s)) | \mathcal{F}_s] ds.
\end{align*}
\]

Introduce the notation
\[
f(s, \bar{Y}(s)) \equiv (A_i - \bar{A}_i)\partial_i v^\alpha(s, \bar{Y}(s)) + (D_{ij} - \bar{D}_{ij})\partial_{ij} v^\alpha(s, \bar{Y}(s)).
\]
and let, for \( t_m \leq s < t_{m+1} \),

\[
    L_Y w(s, \bar{Y}(s)) \equiv \left( \frac{\partial}{\partial t} w + \bar{A}_n \partial_n w + \bar{D}_{kn} \partial_{kn} w \right)(s, \bar{Y}(s)).
\]

Then Itô’s formula implies, as in (2.29), that

\[
    E\left[ f(s, \bar{Y}(s)) \mid \mathcal{F}_t \right] = \int_{t_m}^{s} E\left[ L_Y f \mid \mathcal{F}_t \right](\tau) \, d\tau = O(\Delta t_m), \quad t_m \leq s < t_{m+1},
\]

which combined with (2.33) proves (2.30).

The estimate (2.32) follows similarly by defining

\[
    \tilde{f}(t, \bar{X}(t)) \equiv (d_{ij} - \bar{a}_{ij}) \partial_{ij} (u - \bar{u})(t, \bar{X}(t)).
\]

Then the Itô’s formula shows as in (2.28) and (2.29)

\[
    E[\tilde{f}(t, \bar{X}(t))] = \int_{t_n}^{t} E[L \tilde{f}](s) \, ds, \quad t_n \leq t < t_{n+1}.
\]

The final step to prove (2.32) is to establish

\[
    E[L \tilde{f}](s) = \int_s^T O(\Delta t(\tau)) \, d\tau.
\]

The function \( L \tilde{f} \) splits into the two types of terms \( f_1 \equiv (d - \bar{d})v \) and \( f_2 \equiv v \partial_\omega (u - \bar{u}) \), with smooth functions \( v \) of \((s, \bar{X}(s))\). The Itô formula again shows that

\[
    E[f_1](s) = \int_{t_m}^{s} E[Lf_1](\tau) \, d\tau = O(\Delta t_m), \quad t_m \leq s < t_{m+1}.
\]

Moreover, (2.30) implies

\[
    E[f_2](s) = \int_s^T O(\Delta t(\tau)) \, d\tau,
\]

and consequently (2.32) holds. The estimate (2.31) of the drift terms follows analogously. \( \square \)

**Lemma 2.5** Suppose that the assumptions in Lemma 2.1 hold. Then the dual functions \( \varphi \) and \( \varphi' \), defined by (2.6)–(2.9), satisfy

\begin{align*}
    \partial_i \tilde{u}(t_n, \bar{X}(t_n)) &= E[\varphi_i(t_n) \mid \mathcal{F}_{t_n}], \\
    \partial_{ij} \tilde{u}(t_n, \bar{X}(t_n)) &= E[\varphi'_{ij}(t_n) \mid \mathcal{F}_{t_n}],
\end{align*}

and for \( t = t_{n+1} \)

\[
    E\left[ (a_i(t, \bar{X}(t)) - \bar{a}_i(t_{n}; \bar{X})) \varphi_i(t) \mid \mathcal{F}_t \right] = E\left[ (a_i(t, \bar{X}(t)) - \bar{a}_i(t_{n}; \bar{X})) \varphi_i(t) \right],
\]

\[
    E\left[ (d_{ij}(t, \bar{X}(t)) - \bar{d}_{ij}(t_{n}; \bar{X})) \varphi_{ij}(t) \mid \mathcal{F}_t \right] = E\left[ (d_{ij}(t, \bar{X}(t)) - \bar{d}_{ij}(t_{n}; \bar{X})) \varphi'_{ij}(t) \right].
\]
PROOF: Equations (2.1), (2.16), and (2.20) show that the first variation of the Euler approximation $\bar{X}$ is in fact equal to the Euler approximation of the first variation $X'$ and consequently

$$\partial_{ij}\bar{u}(t, \bar{X}(t)) = E\left[\partial_{ij}g(\bar{X}(T))\bar{X}_{ji}(T; t) \mid \mathcal{F}_t\right],$$

where $\bar{X}_{ji}(s; t), \ s > t, \ \text{is the Euler approximation (2.26)} \ \text{of} \ X' \ \text{with initial data} \ \bar{X}_{ji}(t; t) = \delta_{ij}$. Let $\varphi$ be the solution of (2.7). Then (2.26) implies

$$0 = \sum_{n=m}^{N-1} (\varphi(t_n) - \partial_jc_j(t_n, \bar{X}(t_n))\varphi_{i}(t_{n+1}))\bar{X}'_{ik}(t_n; t_m)$$

$$= \sum_{n=m}^{N-1} \varphi(t_{n+1})(\bar{X}'_{ik}(t_{n+1}; t_m) - \partial_jc_i(t_n, \bar{X}(t_n))\bar{X}_{ik}'(t_n; t_m))$$

$$+ \varphi(t_m)\bar{X}'_{ik}(t_m; t_m) - \varphi(T)\bar{X}'_{ik}(T; t_m)$$

$$= \varphi(t_m)\bar{X}'_{ik}(t_m; t_m) - \varphi(T)\bar{X}'_{ik}(T; t_m).$$

Therefore, the initial conditions of (2.7) and (2.26), in (2.24), yield

$$\varphi_k(t_m) = \partial_i g(\bar{X}(T))\bar{X}'_{ik}(T; t_m),$$

which proves (2.34).

Equality (2.37) implies that

$$\partial_{ij}\bar{u}(t_n, \bar{X}(t_n)) = E\left[\partial_{ij}(t_n)\varphi_i(t_n) \mid \mathcal{F}_n\right].$$

The next step is to verify that the first variation of $\varphi$,

$$\varphi'_{ij}(t_n) = \partial_{ij}(t_n)\varphi_i(t_n) = \frac{\partial \varphi_i(t_n; \bar{X}(t_n) = x)}{\partial x_j},$$

satisfies the backward recursive equation (2.9). First, differentiate equation (2.7) for $\varphi$ to obtain

$$\varphi'_{ik}(t_n) = \partial_i c_j(t_n, \bar{X}(t_n))\partial_{s_k(t_n)}\varphi_j(t_{n+1}) + \partial_j \partial_i c_j(t_n, \bar{X}(t_n))\varphi_j(t_{n+1}),$$

$$t_n < T,$$

$$\varphi'_{ik}(T) = \partial_{ik}g(\bar{X}(T)).$$

Then the linear backward equation (2.7) shows that $\varphi(t_{n+1})$ depends only on the point values

$$\{\bar{X}(t_m) : t_{n+1} \leq t_m \leq T\}$$

so that

$$\partial_{s_k(t_n)}\varphi_j(t_{n+1}) = \partial_{s_p(t_{n+1})}\varphi_j(t_{n+1})\partial_{s_k(t_n)}\bar{X}_p(t_{n+1}).$$

Finally, the definitions of $\bar{X}$ and $c$ in (2.1) and (2.6) imply

$$\partial_k c_p(t_n, \bar{X}(t_n)) = \partial_{s_k(t_n)}\bar{X}_p(t_{n+1}).$$
which combined with (2.39) and (2.40) prove that \( \varphi' \) satisfies recursive equation (2.9).

The measurability of \( (a_i(t_{n+1}, \bar{X}(t_{n+1})) - \bar{a}_i(t_n, \bar{X}(t_n))) \in \mathcal{F}_{t_{n+1}} \) proves for \( t = t_{n+1} \) that

\[
E[(a_i(t, \bar{X}(t)) - \bar{a}_i(t_n, \bar{X}(t_n)))] = E[E[(a_i(t, \bar{X}(t)) - \bar{a}_i(t_n, \bar{X}(t_n))) | \mathcal{F}_t]] = E[E((a_i(t, \bar{X}(t)) - \bar{a}_i(t_n, \bar{X}(t_n))) \varphi_i(t) | \mathcal{F}_t)],
\]

and a similar argument for the diffusion term proves (2.36).

The proof of Theorem 2.2 is now concluded by combining Lemmas 2.3–2.5 and the central limit theorem to estimate \( I_M \) and \( II_M \).

This section ends with three remarks on higher-order methods, the variance in the error bound, and general expected values.

**Remark 2.6 (Higher-Order Methods).** Talay and Tubaro point out in [32] that expansion (1.4) justifies Romberg extrapolation to obtain second-order accuracy provided the statistical error is sufficiently small. By Theorems 2.2 and 3.3, Romberg extrapolation also works for variable time steps if every time step is halved. However, the error density for this second-order method is not known in the computation and therefore a choice of time steps would be suboptimal.

Let us compare the computational work for ordinary and stochastic differential equations approximated by \( p \)-th-order accurate methods. Assume that the solution of a deterministic ordinary differential equation is smooth and bounded so that to approximate it with accuracy \( \text{TOL} \) requires the work \( \mathcal{O}(p \text{TOL}^{-1/p}) \) for a \( p \)-th-order method. To approximate the expected value \( E[g(X(T))] \) with accuracy \( \text{TOL} \) by a \( p \)-th-order method requires the additional work of \( \mathcal{O}(\text{TOL}^{-2}) \) realizations; therefore the total work for the Monte Carlo method is \( \mathcal{O}(p \text{TOL}^{-2-1/p}) \). With these assumptions, the optimal choice of the order is \( p = \log \text{TOL}^{-1} \) for both deterministic and stochastic problems.

However, if we compare, for fixed work, the gain in accuracy to do high-order methods instead of first-order, the result is as follows: For a deterministic differential equation, a first-order method with error \( \text{TOL} \) requires the same work as the optimal higher-order method with error \( \mathcal{O}(e^{-\text{TOL}^{-1}}) \), while for a stochastic differential equation a first-order method with error \( \text{TOL} \) requires the same work as the optimal higher-order method with error slightly larger than \( \text{TOL}^{3/2} \). Hence the maximal gain in accuracy to use higher-order methods is much less in the stochastic case and in a sense comparable with the improvement to use a third-order instead of a second-order accurate method in the deterministic case provided \( \text{Var}[g(X(T))] \gg \text{TOL}^2 \). The opposite with negligible noise, satisfying \( \text{Var}[g(X(T))] \ll \text{TOL}^2 \), behaves like the deterministic case. The work [21] assumes small noise to construct new, simplified higher-order methods for weak approximations of stochastic differential equations.
Remark 2.7 (Variance of the Error Bound). The number of realizations to determine a reliable error estimate is in general much smaller than the number of realizations to approximate $E[g(X(T))]$, which is proportional to $\text{TOL}^{-2}$. Let us now study how many realizations $M$ are needed for the error estimate. Write the sum of the leading-order term in (2.5), i.e., the sum of the terms inside of the two square brackets, as

\begin{equation}
\xi \equiv \sum_{n=0}^{N-1} \sum_{j=1}^{M} \frac{\rho_n(\omega_j)}{M} \Delta t_n^2, \tag{2.42}
\end{equation}

where $M$ is the number of realizations to evaluate the error estimate. The stochastic variable $\xi$ has mean $\int_0^T \mathbb{O}(\Delta t) dt$ and, by (2.10) and (2.11), variance $\int_0^T \mathbb{O}(\Delta t) dt / M$. A useful error estimate therefore requires

$$
\sqrt{\int_0^T \mathbb{O}(\Delta t) dt / M} \ll \text{TOL} ,
$$

and consequently

\begin{equation}
M \gg \int_0^T \mathbb{O}(\Delta t) dt / \text{TOL}^2 = \mathcal{O}(\text{TOL}^{-1}). \tag{2.43}
\end{equation}

Estimates (2.10) and (2.11) imply that a good approximation of the density based directly on $\sum_j \rho_n(\omega_j) / M$ would require $M \sim \mathcal{O}((\epsilon^2 \Delta t)^{-1})$, with statistical error $\epsilon$, which can be more demanding than (2.43) for a nonuniform mesh. A remedy for this is to use approximate ergodicity and local averages in time to obtain a statistically good approximation of the density with $M$ based on (2.43); see Section 5 and the appendix.

In Theorem 3.3 an alternative error expansion with leading-order term

$$
\sum_{n=0}^{N-1} \sum_{j=1}^{M} \tilde{\rho}_{jn} \Delta t_n^2 / M
$$

is given, which has the same mean but a smaller variance $(\int_0^T \mathbb{O}(\Delta t) dt)^2 / M$. A consequence for deterministic time steps is that $M \gg \mathcal{O}(1)$ is enough in this case. However, the estimate in Theorem 3.3 requires more computational work per realization.

Remark 2.8 (More General Expected Values). Suppose that $h : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ is sufficiently smooth. Then the error estimates in Lemma 2.1 and Theorems 2.2 and 3.3 include estimates of expected values of the form

$$
E \left[ \int_0^T h(t, X(t)) dt + g(X(T)) \right]
$$
by introducing the additional variable $X_{d+1}$ and equation $dX_{d+1} = h(t, X(t))\, dt$ to (1.1). By eliminating the additional variables in $X$ and $\varphi$, (2.7) is replaced by

$$ \varphi_i(t_n) = \partial_i c_j(t_n, \overline{X}(t_n)) \varphi_j(t_{n+1}) + \partial_i h(t_n, \overline{X}(t_n)) \Delta t_n, \quad t_n < T, $$

$$ \varphi_i(T) = \partial_i g(\overline{X}(T)), $$

and (2.9) by

$$ \varphi_{ik}(t_n) = \partial_i c_j(t_n, \overline{X}(t_n)) \partial_k c_p(t_n, \overline{X}(t_n)) \varphi_{jp}(t_{n+1}) + \partial_{ik} c_j(t_n, \overline{X}(t_n)) \varphi_j(t_{n+1}) $$

$$ + \partial_{ik} h(t_n, \overline{X}(t_n)) \Delta t_n, \quad t_n < T, $$

$$ \varphi_{ik}(T) = \partial_{ik} g(\overline{X}(T)). $$

3 An Error Estimate with Stochastic Time Steps

This section derives error estimates with time steps that are stochastic and determined individually for each realization by the whole solution path $\overline{X}$. The analysis will use the Malliavin derivative, $\partial_{W(t)} Y$, which is the first variation of a process $Y$ with respect to a perturbation $dW(t)$ at time $t$ of the Wiener process; cf. [25]. The Malliavin derivative for a stochastic integral $X$ is related to the first variation $\partial_x(t)\overline{X}$ for a perturbation of the position at time $t$ by

$$ \partial_{W(t)} X(\tau) = \frac{\partial X_k(t)}{\partial W^\epsilon(t)} \partial_{\tau_k(t)} X(\tau) = b^\epsilon_k(X(t)) \partial_{\tau_k(t)} X(\tau), \quad \tau > t, $$

$$ \partial_{W(t)} X(\tau) = 0, \quad \tau < t, $$

if $dX_k = a_k(X(t))\, dt + b_k^\epsilon(X(t))\, dW^\epsilon$; cf. (2.15).

We shall restrict the analysis to time steps that are constructed by the refinement criterion

$$ \Delta t(t) = T 2^{-n} \quad \text{for some natural number } n = n(t, \omega), $$

$$(3.2) \quad |\rho(t, \omega)|(|\Delta t(t)|)^2 < \text{const}, $$

with an approximate error density function $\rho$ satisfying, for $s \in [0, T]$, $t \in [0, T]$, and all outcomes $\omega$, the uniform upper and lower bounds

$$ c(\text{TOL}) \leq |\rho(s, \omega)| \leq C(\text{TOL}), $$

$$(3.3) \quad |\partial_{W(t)} \rho(s, \omega)| \leq C(\text{TOL}), $$

for some positive functions $c$ and $C$, with TOL$/c(\text{TOL}) \rightarrow 0$ as TOL $\rightarrow 0$. For each realization successive subdivisions of the steps yield the largest time steps satisfying (3.2). The corresponding stochastic increments $\Delta W$ will have the correct distribution, with the necessary independence, if the increments $\Delta W$ related to the new steps are generated by Brownian bridges (cf. [18]); i.e., the time steps are generated by conditional expected values of the Wiener process.

Let $\delta$ be a constant approximating $\text{TOL}/E[N]$, where $E[N]$ is the expected number of steps. The analysis in this section with adaptive nonadapted time steps satisfying (3.2) and (3.3) is based on the following:
Stochastic Time Step Algorithm

for $M$ realizations of the Wiener process $W(\omega_j)$, $j = 1, 2 \ldots, M$, do

Step 1: Set $k = 0$. Start with an initial coarse mesh $\Delta t[0]$ and compute $\Delta W[0]$.

Step 2: For the piecewise constant mesh function $\Delta t[k]$ with corresponding noise $\Delta W[k]$, compute $X[k]$ and the weight function $\rho[k]$ defined in (3.12).

Step 3: Define $r(t) \equiv |\rho[k](t)|(\Delta t[k](t))^2$ and let for all $t$

$$\Delta t[k + 1](t) = \begin{cases} 
\Delta t[k](t) & \text{if } r(t) < \delta \quad (\dagger) \\
\Delta t[k](t)/2 & \text{if } r(t) \geq \delta \quad (\ast)
\end{cases}$$

and in the refinement case ($\ast$) construct $\Delta W[k+1]$ by Brownian bridges based on the already known $\Delta W[k]$.

Step 4:

if at least one step of $\Delta t[k]$ is refined by ($\ast$) then

increment $k$ by 1 and goto Step 2.

else

all steps of $\Delta t[k]$ satisfy ($\dagger$) and accept the approximation $g(\overline{X}(T, \omega_j))$ and goto the next realization of $W$.

end if

end for

if the statistical error $E[g(\overline{X}(T))] - \sum_{j=1}^{M} g(\overline{X}(T, \omega_j))/M$ is sufficiently small then

stop

else

restart with a larger $M$.

end if

Section 5 presents a more precise formulation of this algorithm. The adaptive algorithm will, for example, not include discretization by the backward Euler method

$$X(t_{n+1}) = X(t_n) + b(X(t_n+1))\Delta \tilde{W}_n$$

with step size $h$, which is the forward Euler method with time steps

(3.4)

$$\Delta t_n = h \left( \frac{b(X(t_{n+1}))}{b(X(t_n))} \right)^2.$$

For the case $b(x) = x$, the relation (3.4) becomes $\Delta t_n = h/(1 - \Delta \tilde{W}_n)^2$ and $X(t_{n+1}) = X(t_n)(1 + \Delta \tilde{W}_n\sqrt{\Delta t_n/h})$, which violates the requirement that the noise $\Delta \tilde{W}_n\sqrt{\Delta t_n/h}$ is $N(0, \Delta t_n)$. The stochastic time step algorithm generates the noise $\Delta W$ with the correct distribution through Brownian bridges, although the time steps may depend on the future. Here $\Delta \tilde{W}$ are random variables with distribution close to the normal $N(0, h)$ but with compact support, so that the backward Euler method makes sense. We want to exclude such inconsistent time steps (3.4),
since the backward Euler method converges to a different limit than the Itô stochastic integrals generated by limits of the forward Euler method with adapted steps; cf. [19]. Lemma 3.1 and Theorem 3.3 below show that although the steps generated by (3.2) and (3.3) through the algorithm above are not adapted, the method indeed converges to the correct limit as the forward Euler method with adapted time steps.

**Lemma 3.1** Suppose that $a$, $b$, $g$, and $X$ satisfy the assumptions in Lemma 2.1 and that $\overline{X}$ is constructed by the forward Euler method, based on the stochastic time step algorithm above, with step sizes $\Delta t_n$ satisfying (3.2) and (3.3). Also assume that their corresponding $\Delta W_n$ are generated by Brownian bridges. Finally, assume that $\overline{X}(0) = X(0)$. Then

$$\sup_{0 \leq t \leq T} \sqrt{E[(X(t) - \overline{X}(t))^2]} = o\left(\sqrt{\Delta t_{\sup}}\right) = o\left(\frac{T_{\text{TOL}}}{c(T_{\text{TOL}})}\right) \to 0,$$

as $T_{\text{TOL}} \to 0$, where $\Delta t_{\sup} = \sup_{n, \omega} \Delta t_n(\omega)$.

**Proof:** The standard proof of strong convergence of the Euler method with adapted time steps, cf. [20], can be modified to include the adaptive nonadapted steps by first verifying that $\overline{X}(t_n)$ is essentially independent of the increments $W(t + \Delta t_n) - W(t)$, provided $\Delta t_n$ is sufficiently small and $t_n < t$.

Let us now derive this independence claim. The refinement algorithm can be formulated: Refine a time step $\Delta t_n$ if $r_n \geq \delta$, where $r_n = \Delta t_n^2|\rho(t_n, \omega)|$ and the constant $\delta$ is an approximation of $T_{\text{TOL}}/E[N_s]$. Here $E[N_s]$ is the expected number of steps for the method, which has to be guessed at the beginning of the algorithm. Consequently, for each final acceptable time discretization, the minimum $\min_n(\delta - r_n)$ is positive for each realization. The approximate solution $\overline{X}(s)$ depends on $dW(t)$, for $s < t$, only through changes in the mesh. Let us use the notation $\hat{\Delta}t$ for a small but finite time increment, with $\hat{\Delta}t \ll \Delta t$ and its corresponding stochastic increment $\hat{\Delta}W$. We shall show that, provided $\hat{\Delta}t$ is sufficiently small and conditioned on the $\sigma$-algebra $\mathcal{M}(t, \hat{\Delta}t)$ generated by $\{dW(\tau) : \tau < t \text{ or } \tau > t + \hat{\Delta}t\}$, the probability to change the mesh by a change in only $\hat{\Delta}W(t)$ is arbitrarily small; thus $\overline{X}(s)$ will be essentially independent of $\hat{\Delta}W(t)$ for $s < t$. Conditioned on $\mathcal{M}(t, \hat{\Delta}t)$, let $r_n(\hat{\Delta}W) = E[r_n | \mathcal{M}(t, \hat{\Delta}t)]$ denote the dependence of the error indicator $r_n$ on the noise $\hat{\Delta}W$. The Malliavin derivative and Taylor’s formula imply

$$r_n(\hat{\Delta}W) = r_n(0) + \Delta W^\ell \int_0^1 \partial_{W^\ell(t)} r_n(s \hat{\Delta}W) ds.$$

The mesh generated by $r_n(\hat{\Delta}W)$ and $r_n(0)$ is the same provided

$$0 < \delta - r_n(\hat{\Delta}W) = \delta - \left(r_n(0) + \Delta W^\ell \int_0^1 \partial_{W^\ell(t)} r_n(s \hat{\Delta}W) ds\right) \equiv \delta - (r_n(0) + \Delta W \cdot r_n') \text{ for all } n.$$
and
\[(3.5b) \quad \min_n(\delta - r_n(0)) \equiv \epsilon(\omega) > 0.\]

Therefore, (3.5a) and (3.5b) hold if \(\hat{\Delta} W \cdot r_n' < \epsilon(\omega)\). Let
\[\hat{\epsilon} \equiv \frac{\epsilon}{(\sup_n \Delta t_n)^2C(TOL)};\]
then \(|\hat{\Delta} W| < \hat{\epsilon}\) implies \(\hat{\Delta} W \cdot r_n' < \epsilon\), since by (3.3)
\[\hat{\Delta} W \cdot r_n' \leq |\hat{\Delta} W||r_n'| < \hat{\epsilon}|r_n'| = \epsilon \left(\frac{(\Delta t_n)^2}{\sup \Delta t_n} \int_0^1 \frac{\partial W_s(t_n, s \hat{\Delta} W)ds}{(\Delta t_n)^2C(TOL)}\right) \leq \epsilon,\]
where \(\rho(t_n, \hat{\Delta} W) = E[\rho(t_n) | M(t, \hat{\Delta} t)]\). Consequently, the following independence claim holds:
\[(3.6a) \quad \bar{X}(t_n) \text{ is independent of } \hat{\Delta} W(t) \text{ conditioned on } |\hat{\Delta} W(t)| < \hat{\epsilon} \text{ and } M(t, \hat{\Delta} t) \text{ for } t_n < t,\]
and the probability of having different meshes with \(r.(0)\) or \(r.(\hat{\Delta} W(t))\) is, for sufficiently small \(\hat{\Delta} t\), bounded by
\[(3.6b) \quad P(|\hat{\Delta} W| \geq \hat{\epsilon} | M(t, \hat{\Delta} t)) \leq C_{\epsilon_0} \exp\left(-\frac{(\hat{\epsilon})^2}{4(\hat{\Delta} t)}\right) \ll (\hat{\Delta} t)^3.\]

Let \(\hat{X}\) be a forward Euler approximation of \(X\) with uniform time steps \(\hat{\Delta} t\) on a much finer grid than \(\Delta t\) so that \(\{n\hat{\Delta} t : n = 0, 1, \ldots, \hat{N}\}\) includes the time steps \(\{t_m\}\) for \(\bar{X}\) and \(\hat{\Delta} t(t, \omega) \leq \hat{\Delta} t \leq \Delta t(t, \omega)\). The standard proof of strong convergence for the Euler method starts by writing \(t = n\hat{\Delta} t\) and
\[\hat{X}(n\hat{\Delta} t) - \bar{X}(n\hat{\Delta} t) = \sum_{i=0}^{n-1} \left(\Delta_i a \hat{\Delta} t_i + \Delta_i b^\ell \hat{\Delta} W_i^\ell\right),\]
where
\[\Delta_i a = a(i\hat{\Delta} t, \hat{X}(i\hat{\Delta} t)) - a(t_n, \bar{X}(t_n)), \quad t_n \leq i\hat{\Delta} t < t_{n+1},\]
\[\Delta_i b^\ell = b^\ell(i\hat{\Delta} t, \hat{X}(i\hat{\Delta} t)) - b^\ell(t_n, \bar{X}(t_n)), \quad t_n \leq i\hat{\Delta} t < t_{n+1},\]
\[\hat{\Delta} W_i^\ell = W^\ell((i + 1)\hat{\Delta} t) - W^\ell(i\hat{\Delta} t).\]
Therefore,
\[(3.7) \quad E[|\hat{X}(n\hat{\Delta} t) - \bar{X}(n\hat{\Delta} t)|^2] = \sum_{i, i'} E\left[(\Delta_i a_j \hat{\Delta} t_i + \Delta_i b_j^\ell \hat{\Delta} W_i^\ell)(\Delta_{i'} a_j \hat{\Delta} t_{i'} + \Delta_{i'} b_j^\ell \hat{\Delta} W_{i'}^\ell)\right].\]
To prove convergence, we shall use the uniform bound \( E[|\hat{X}(t)|^2] \leq C \). This bound follows by the same steps as the convergence proof below applied to

\[
E[|\hat{X}(\hat{\Delta}t)|^2] = \sum_{i,i'} E[(\hat{a}_j \hat{\Delta}t_i + \hat{b}_j \hat{\Delta}W_i^\ell)(\hat{a}_j \hat{\Delta}t_{i'} + \hat{b}_j \hat{\Delta}W_{i'}^\ell)]
\]

instead of (3.7). Let us first analyze one term in the right-hand side of (3.7),

\[
\sum_{i,i'} E[\Delta_i b_j^\ell \hat{\Delta}W_i^\ell \Delta_{i'} b_{j'}^\ell \hat{\Delta}W_{i'}^\ell] =

2 \sum_{i<i'} E[\Delta_i b_j^\ell \hat{\Delta}W_i^\ell \Delta_{i'} b_{j'}^\ell \hat{\Delta}W_{i'}^\ell] + \sum_i E[\Delta_i b_j^\ell \hat{\Delta}W_i^\ell \Delta_i b_{j'}^\ell \hat{\Delta}W_{i'}^\ell].
\]

The goal for the estimate of the right-hand side above is to use the approximate independence of \( \hat{\Delta}W_{i'}^\ell \) and \( \Delta_i b_j^\ell \hat{\Delta}W_i^\ell \Delta_i b_{j'}^\ell \). Divide the interval \([i' \hat{\Delta}t, (i' + 1) \hat{\Delta}t]\) into the union of disjoint steps \( \hat{\Delta}t_m \) so that \( \hat{\Delta}t = \sum_{m=1}^{\hat{N}} \hat{\Delta}t_m \) with the corresponding Wiener increments

\[
\hat{\Delta}W_m^\ell = W^\ell (i' \hat{\Delta}t + \sum_{n=1}^{m} \hat{\Delta}t_n) - W^\ell (i' \hat{\Delta}t + \sum_{n=1}^{m-1} \hat{\Delta}t_n),
\]

\[
\hat{\Delta}W_m^\ell = W^\ell (i' \hat{\Delta}t + \sum_{n=1}^{m} \hat{\Delta}t_n) - W^\ell (i' \hat{\Delta}t + \sum_{n=1}^{m-1} \hat{\Delta}t_n).
\]

Let us define the \( \sigma \)-algebra \( \mathcal{G}_{i'} \equiv \mathcal{M}(i' \hat{\Delta}t, \hat{\Delta}t_i) \) and introduce the simplifying notation \( \beta \equiv \Delta_i b_j^\ell \hat{\Delta}W_i^\ell \Delta_i b_{j'}^\ell \). The conclusions (3.6a) and (3.6b) give

\[
E[\beta \hat{\Delta}W_{i'}^\ell] = E[E[\beta \hat{\Delta}W_{i'}^\ell | \mathcal{G}_{i'}]]
\]

\[
= E[E[\beta (1_{|\hat{\Delta}W_i|<\delta} + 1_{|\hat{\Delta}W_i|\geq\delta})(\hat{\Delta}W_i^\ell + \hat{\Delta}W_{i'}^\ell) | \mathcal{G}_{i'}]]
\]

\[
\approx E[E[\beta \hat{\Delta}W_i^\ell | \mathcal{G}_{i'}]] + E[E[\beta 1_{|\hat{\Delta}W_i|<\delta} \hat{\Delta}W_i^\ell | \mathcal{G}_{i'}]] + E[\mathcal{O}((\hat{\Delta}t_i)^3)]
\]

\[
= E[E[\beta \hat{\Delta}W_i^\ell | \mathcal{G}_{i'}]] + E[E[\beta | \mathcal{G}_{i'}, {\hat{\Delta}W_i} < \delta]]
\]

\[
\times E[1_{|\hat{\Delta}W_i|<\delta} \hat{\Delta}W_i^\ell | \mathcal{G}_{i'}, {\hat{\Delta}W_i} < \delta] P(|\hat{\Delta}W_i| < \delta | \mathcal{G}_{i'})
\]

\[
= E[\beta \hat{\Delta}W_i^\ell] + E[\mathcal{O}((\hat{\Delta}t_i)^3)]
\]

\[
= E[\beta \hat{\Delta}W_i^\ell] + E[\mathcal{O}((\hat{\Delta}t_i)^3)].
\]
Apply this \( \hat{\Delta}W, \hat{\Delta} \) argument recursively on \( E[\Delta_i b_j^\varepsilon \hat{\Delta}W_t^\varepsilon \Delta_r b_j^\varepsilon \hat{\Delta}W_r^\varepsilon] \), for \( i < i' \), conditioned on \( M(i') \Delta t + \sum_{n=1}^{m} \hat{\Delta} t_n, \hat{\Delta} t_{m+1} \) to get

\[
E[\Delta_i b_j^\varepsilon \hat{\Delta}W_t^\varepsilon \Delta_r b_j^\varepsilon \hat{\Delta}W_r^\varepsilon] = E\left[ \sum_m O((\hat{\Delta} t_m)^3) \right] = E\left[ \int_{(i') \Delta t}^{(i'+1) \Delta t} O((\hat{\Delta} t(t))^2) dt \right] \leq O((\hat{\Delta} t)^3).
\]

A similar argument shows

\[
E[\Delta_i b_j^\varepsilon \hat{\Delta}W_t^\varepsilon \Delta_r b_j^\varepsilon \hat{\Delta}W_r^\varepsilon] = E[\Delta_i b_j^\varepsilon \Delta_r b_j^\varepsilon] \Delta t + O((\hat{\Delta} t)^3).
\]

Summation therefore implies

\[
\sum_{i,i'} E[\Delta_i b_j^\varepsilon \hat{\Delta}W_t^\varepsilon \Delta_r b_j^\varepsilon \hat{\Delta}W_r^\varepsilon] = 2 \sum_{i < i'} E[\Delta_i b_j^\varepsilon \hat{\Delta}W_t^\varepsilon \Delta_r b_j^\varepsilon \hat{\Delta}W_r^\varepsilon] + \sum_i E[\Delta_i b_j^\varepsilon \Delta_r b_j^\varepsilon \hat{\Delta}W_t^\varepsilon \hat{\Delta}W_r^\varepsilon] = O(\hat{\Delta} t) + \sum_i E[\Delta_i b_j^\varepsilon \Delta_r b_j^\varepsilon] \Delta t.
\]

From this step there is no change in the simple standard proof, cf. [20], which does not require Doob’s inequality. Using the global Lipschitz continuity of \( a \) and \( b \), the proof continues by the estimate

\[
E[|\Delta_i b_j^\varepsilon|^2] \leq C(E[|\hat{\Delta}(i \Delta t) - \hat{\Delta}(i \Delta t)|^2] + \Delta t_{sup}).
\]

Similar estimates for the other terms in (3.7) and a final application of Grönwall’s lemma prove the first equality in the theorem.

We have \( \inf |\rho|(\Delta t_{sup})^2 \leq TOL/E[N] \leq TOL\Delta t_{sup}/T \), so that the assumption \( \inf |\rho| \geq c(TOL) \) implies \( \Delta t_{sup} \leq TOL/(Tc(TOL)) \), which by assumption tends to zero as TOL tends to zero and verifies the last equality in the theorem.

\[ \square \]

**Lemma 3.1’** Suppose the assumptions of Lemmas 2.1 and 3.1 hold, and let \( p \) be an even integer \( 2 \leq p \leq 2k + d + 1 \equiv k_0 \). Then

\[
\sup_{0 \leq t \leq T} \left( E[|X(t) - \bar{X}(t)|^p] \right)^{1/p} = O(\sqrt{\Delta t_{sup}}) = O\left( \sqrt{\frac{TOL}{c(TOL)}} \right) \to 0,
\]

\[
\sup_{0 \leq t \leq T} \left( E[|\bar{X}(t)|^p] \right)^{1/p} = O(1),
\]

as \( TOL \to 0 \).
The proof of Lemma 3.1 shows

\[
|E[(\bar{X}_i((n+1)\hat{\Delta}t))^p]| = |E[(\bar{X}_i(n\hat{\Delta}t))^p] + E[p(\bar{X}_i(n\hat{\Delta}t))^{p-1}\hat{\Delta}X_i] \hat{\Delta}t + \frac{p}{2}(\bar{X}_i(n\hat{\Delta}t))^{p-2}(\hat{\Delta}X_i)^2 + \cdots + (\hat{\Delta}X_i)^p|.
\]

The proof of Lemma 3.1 shows

\[
|E[(\bar{X}_i((n+1)\hat{\Delta}t))^p]| = |E[(\bar{X}_i(n\hat{\Delta}t))^p] + E[p(\bar{X}_i(n\hat{\Delta}t))^{p-1}\bar{a}_i] \hat{\Delta}t + \frac{p}{2}(\bar{X}_i(n\hat{\Delta}t))^{p-2}(\bar{a}_i\hat{\Delta}t + (\bar{a}_i\hat{\Delta}t)^2) + E[\left(\frac{p}{3}\right)X_i(n\hat{\Delta}t)^{p-3}(\hat{\Delta}X_i)^3 + \cdots + (\hat{\Delta}X_i)^p]|
\]

Let \( f = \max(|\bar{a}|, |\bar{b}|) \). Then

\[
|E[(\bar{X}_i((n+1)\hat{\Delta}t))^p]| \leq |E[(\bar{X}_i(n\hat{\Delta}t))^p]| + C_p E[(\bar{X}_i(n\hat{\Delta}t))^p] + f^p \hat{\Delta}t.
\]

We know from the assumptions in Lemma 2.1 that \( E[f^p] \leq C_p \); therefore an application of Grönwall’s lemma proves the last estimate in Lemma 3.1’. The first estimate in the lemma follows analogously by replacing \( \bar{X} \) above with the error \( \hat{\Delta}X \), which has the increments \( \Delta a \hat{\Delta}t + \Delta b \hat{\Delta}W \) following Lemma 3.1. Then \( f = \max(|\Delta a|, |\Delta b|) \) satisfies \( E[f^p] \leq C(E[|\hat{\Delta}X| - \bar{X}(n\hat{\Delta}t)|^p] + (\Delta t_{\sup})^{p/2}) \), which implies the first estimate in the lemma.

In addition to the dual functions \( \varphi \) and \( \varphi' \) in Theorem 2.2, the new error expansion for stochastic time steps in Theorem 3.3 below also uses the discrete dual variation

\[
\varphi''_{ikm}(t_n) \equiv \partial_{x_m}(t_n)\varphi'_{ik}(t_n) \equiv \frac{\partial \varphi'_{ik}(t_n; \bar{X}(t_n) = x)}{\partial x_m},
\]

which satisfies the following:

**Lemma 3.2** Let \( c, \varphi, \varphi' \), and \( \varphi'' \) be defined by (2.6)–(2.9) and (3.9); then

\[
\begin{align*}
\varphi''_{ikm}(t_n) &= \partial_i c_j(t_n, \bar{X}(t_n))\partial_k c_p(t_n, \bar{X}(t_n))\partial_m c_r(t_n, \bar{X}(t_n))\varphi''_{jpr}(t_{n+1}) \\
&\quad + \partial_{im}\partial_j c_j(t_n, \bar{X}(t_n))\partial_k c_p(t_n, \bar{X}(t_n))\varphi''_{jpr}(t_{n+1}) \\
&\quad + \partial_i c_j(t_n, \bar{X}(t_n))\partial_{km} c_p(t_n, \bar{X}(t_n))\varphi''_{jpr}(t_{n+1}) \\
&\quad + \partial_{ik} c_j(t_n, \bar{X}(t_n))\partial_m c_p(t_n, \bar{X}(t_n))\varphi''_{jpr}(t_{n+1}) \\
&\quad + \partial_{ikm} c_j(t_n, \bar{X}(t_n))\varphi''_{jpr}(t_{n+1}), \quad t_n < T,
\end{align*}
\]

\[
\varphi''_{ikm}(T) = \partial_{ikm} g(\bar{X}(T)).
\]
PROOF: Differentiation of the backward recursive equation (2.9) and the relations (2.37)–(2.41) together with
\( \partial_{x_m(t_n)} \phi_{jp}'(t_{n+1}) = \partial_{x_r(t_{n+1})} \phi_{jp}'(t_{n+1}) \partial_{x_m(t_n)} X_r(t_{n+1}) \)
prove as in (2.38)–(2.41) that \( \phi'' \) satisfies (3.10). Here, (3.11) holds since the linear system of (2.7) and (2.9) for the variable \( (\phi(t_{n+1}), \phi'(t_{n+1})) \) depends only on the point values \( \{X(t_m) : t_{n+1} \leq t_m \leq T\} \).

The following theorem derives an error estimate applicable both to adaptive deterministic time steps and to the stochastic time step algorithm. The assumptions and the proof of the theorem focus on stochastic steps; however, a modification to deterministic time steps is straightforward. The computable error density \( \tilde{\rho} \) of this error estimate can then be cut off for small and large values to satisfy (3.3); see Remark 5.1.

**Theorem 3.3** Suppose that \( a, b, g, \) and \( X \) satisfy the assumptions in Lemma 2.1 and that \( \vec{X} \) is constructed by the forward Euler method with step sizes \( \Delta t_n \) satisfying (3.2)–(3.3) and the corresponding \( \Delta W_n \) are generated by Brownian bridges, following the stochastic time step algorithm in Lemma 3.1. Assume also that \( \vec{X}(0) = X(0) \) and \( E[|X(0)|^k_0] \leq C \) for some \( k_0 \geq 16 \). Then the time discretization error has the following expansion based on both the drift and diffusion fluxes and the discrete dual functions \( \phi, \phi', \) and \( \phi'' \) given in (2.6)–(2.9) and (3.10), with computable leading-order terms

\[
E[g(X(T)) - g(\vec{X}(T))] = E\left[\sum_{n=0}^{N-1} \tilde{\rho}(t_n, \vec{X})(\Delta t_n)^2\right] + \mathcal{O}\left(\sqrt{\frac{\text{TOL}}{c(TOL)}} \left(\frac{C(TOL)}{c(TOL)}\right)^{8/k_0}\right) E\left[\sum_{n=0}^{N-1} (\Delta t_n)^2\right],
\]

(3.12a)

where

\[
\tilde{\rho}(t_n, \vec{X}) = \frac{1}{2} \left( \left( \frac{\partial}{\partial t} a_k + \partial_j a_k a_j + \partial_{ij} a_k d_{ij} \right) \phi_k(t_{n+1}) \right.
\]
\[
+ \left( \frac{\partial}{\partial t} d_{km} + \partial_j d_{km} a_j + \partial_{ij} d_{km} d_{ij} + 2 \partial_j a_k d_{jm} \right) \phi'_m(t_{n+1})
\]
\[
+ \left( 2 \partial_j d_{km} d_{jr} \right) \phi''_{kmr}(t_{n+1}) \bigg),
\]

(3.12b)

and the terms in the sum of (3.12b) are evaluated at the a posteriori known points \( (t_n, \vec{X}(t_n)) \), i.e.,

\[
\partial_a a = \partial_a a(t_n, \vec{X}(t_n)), \quad \partial_a b = \partial_a b(t_n, \vec{X}(t_n)), \quad \partial_a d = \partial_a d(t_n, \vec{X}(t_n)).
\]

**Proof of Theorem 3.3:** Let

\[
\vec{X}(t_{n+1}) = c(\vec{X}(t_n))
\]

(3.13a)
denote one step with the Euler method following (2.6) and similarly write one step with the exact solution

\[(3.13b) \quad X(t_{n+1}) = C(X(t_n)).\]

Introduce the notation \(X^n \equiv X(t_n)\) and \(\bar{X}^n \equiv \bar{X}(t_n)\). The proof has two steps. The first step is to verify the representation

\[(3.14) \quad g(X(T)) - g(\bar{X}(T)) = \sum_{n=0}^{N-1} (C(\bar{X}^n) - c(\bar{X}^n))_j \tilde{\phi}_j(t_{n+1})\]

where the weight functions are defined recursively by

\[(3.15) \quad \tilde{\phi}_i(T) = \int_0^1 \frac{\partial_i g(sX(T) + (1-s)\bar{X}(T))}{(1-s)\bar{X}(T)}ds, \quad \tilde{\phi}_i(t_n) = \left( \int_0^1 \frac{\partial_i C_j(sX(t_n) + (1-s)\bar{X}(t_n))}{(1-s)\bar{X}(t_n)}ds \right) \tilde{\phi}_j(t_{n+1}), \quad n = N - 1, N - 2, \ldots, 0.\]

The next step is to use the Malliavin derivative to analyze the expectation of the representation (3.14) by studying the dependence of \(\bar{X}\) and \(\tilde{\phi}\) on a small increment \(dW\).

To verify (3.14), observe that telescoping cancellation gives

\[(3.16) \quad g(X(T)) - g(\bar{X}(T)) = \sum_{n=0}^{N-1} \left( (X^{n+1} - \bar{X}^{n+1})_j \tilde{\phi}_j(t_{n+1}) - (X^n - \bar{X}^n)_j \tilde{\phi}_j(t_n) \right).\]

Use the definitions (3.13) and split the first term in the sum of (3.16) into

\[(C(X^n) - C(\bar{X}^n))_j \tilde{\phi}_j(t_{n+1}) + (C(\bar{X}^n) - c(\bar{X}^n))_j \tilde{\phi}_j(t_{n+1}).\]

The two first terms above and the last term in the sum of (3.16) combine to zero by (3.15):

\[(C(X^n) - C(\bar{X}^n))_j \tilde{\phi}_j(t_{n+1}) = (X^n - \bar{X}^n)_j \tilde{\phi}_j(t_n),\]

which proves (3.14).

To analyze the expectation of the representation, we use the local exact solution \(\tilde{X}\) satisfying

\[d\tilde{X}_k = a_k(t, \tilde{X}(t))dt + b_k^\ell(t, \tilde{X}(t))dW^\ell, \quad t_n < t < t_{n+1},\]

and

\[\tilde{X}(t_n) = \bar{X}(t_n).\]

Therefore

\[(3.17) \quad C(\bar{X}^n) - c(\bar{X}^n) = \int_{t_n}^{t_{n+1}} (a(t, \tilde{X}(t)) - a(t_n, \bar{X}^n))dt + \int_{t_n}^{t_{n+1}} (b^\ell(t, \tilde{X}(t)) - b^\ell(t_n, \bar{X}^n))dW^\ell.\]
Ito’s formula gives
\[
\int_{t_n}^{t_{n+1}} \left( a_k(t, \tilde{X}(t)) - a_k(t_n, \bar{X}^n) \right) dt
\]
\[
= \int_{t_n}^{t_{n+1}} \int_{t_n}^{t} \left( \frac{\partial a_k}{\partial t} + \partial_i a_k a_i + \frac{1}{2} \partial_{ij} a_k b_i^\ell b_j^\ell \right)(s, \tilde{X}(s)) ds \, dt
\]
\[
+ \int_{t_n}^{t_{n+1}} \int_{t_n}^{t} (\partial_i a_k b_i^\ell)(s, \tilde{X}(s)) dW^\ell(s) dt
\]
and
\[
\int_{t_n}^{t_{n+1}} \left( b_k^\ell(t, \tilde{X}(t)) - b_k^\ell(t_n, \bar{X}^n) \right) dW^\ell(t)
\]
\[
= \int_{t_n}^{t_{n+1}} \int_{t_n}^{t} \left( \frac{\partial b_k^\ell}{\partial t} + \partial_i b_k^\ell a_i + \frac{1}{2} \partial_{ij} b_k^\ell b_i^\ell b_j^\ell \right)(s, \tilde{X}(s)) ds \, dW^\ell(t)
\]
\[
+ \int_{t_n}^{t_{n+1}} \int_{t_n}^{t} (\partial_i b_k^\ell b_i^\ell)(s, \tilde{X}(s)) dW^\ell(s) dW^\ell(t)
\]

Given the partition \( \{ t_n \}_{n=1}^N \), define the function \( \chi : [0, T] \times [0, T] \to \{0, 1\} \) by
\[
\chi_{\Delta t}(s, t) = \begin{cases} 1, & \exists t_n : t_n < s < t < t_{n+1}, \\ 0, & \text{otherwise}. \end{cases}
\]

Let us study the last term in (3.19), inserted into (3.14), separately, and write it in short form as
\[
\sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \int_{t_n}^{t} (\partial_i b_k^\ell b_i^\ell)(s, \tilde{X}(s)) dW^\ell(s) dW^\ell(t) \tilde{\varphi}_k(t_{n+1})
\]
\[
= \int_{0}^{T} \int_{0}^{T} \chi_{\Delta t}(s, t) (\partial_i b_k^\ell b_i^\ell)(s, \tilde{X}(s)) dW^\ell(s) dW^\ell(t) \tilde{\varphi}_k(t_{n+1})
\]
\[
= \int_{0}^{T} \int_{0}^{T} \alpha(s, t) \tilde{\varphi}_k(t_{n+1}) dW^\ell(s) dW^\ell(t)
\]

Let us also use the notation \( \hat{\Delta}t \) for a small but finite time increment, with \( \hat{\Delta}t \ll \Delta t \) and corresponding stochastic increment \( \hat{\Delta}W \), following the proof of strong convergence in Lemma 3.1. Conditioned on the \( \sigma \)-algebra \( \mathcal{M}(\hat{\Delta}t) \) generated by \( \{ dW(s) : s < t \} \), conclusion (3.6) shows that \( \bar{X}(t_n) \) is independent of \( \hat{\Delta}W(t) \) conditioned on \( |\hat{\Delta}W(t)| < \hat{\varepsilon} \) and \( t_n < t \). Let \( \tilde{\varphi}(\hat{\Delta}W(s), \hat{\Delta}W(t)) \) denote the dependence of the weight \( \tilde{\varphi} \) on the noise \( (\hat{\Delta}W(s), \hat{\Delta}W(t)) \). Use the Malliavin
derivative and Taylor’s formula to get

\[
\tilde{\varphi}(\hat{\Delta} W(s), \hat{\Delta} W(t)) =
\]

\[
\tilde{\varphi}(0) + \partial_{W^r(t)} \tilde{\varphi}(0) \hat{\Delta} W^r(t) + \partial_{W^\ell(s)} \tilde{\varphi}(0) \hat{\Delta} W^\ell(s)
\]

\[
+ \partial_{W^r(t)} \partial_{W^\ell(s)} \tilde{\varphi}(0) \hat{\Delta} W^r(s) \hat{\Delta} W^\ell(t) + \frac{1}{2} \partial_{W^r(t)} \partial_{W^\ell(t)} \tilde{\varphi}(0) \hat{\Delta} W^r(t) \hat{\Delta} W^\ell(t)
\]

\[
+ \frac{1}{2} \partial_{W^r(s)} \partial_{W^\ell(s)} \tilde{\varphi}(0) \hat{\Delta} W^r(s) \hat{\Delta} W^\ell(s)
\]

\[
+ O(|\hat{\Delta} W(s)|^3 + |\hat{\Delta} W(s)|^2 |\hat{\Delta} W(t)| + |\hat{\Delta} W(s)||\hat{\Delta} W(t)|^2 + |\hat{\Delta} W(t)|^3).
\]

Following (3.8) in the proof of Lemma 3.1, this Taylor expansion gives

\[
E[\alpha(s, t)\tilde{\varphi}(\hat{\Delta} W(s), \hat{\Delta} W(t)) \hat{\Delta} W^r(s) \hat{\Delta} W^\ell(t)]
\]

\[
= E \left[ \left( \frac{1}{|\hat{\Delta} W(s)| < \varepsilon} + \frac{1}{|\hat{\Delta} W(s)| \geq \varepsilon} \right) \left( \frac{1}{|\hat{\Delta} W(t)| < \varepsilon} + \frac{1}{|\hat{\Delta} W(t)| \geq \varepsilon} \right) \times \alpha(s, t) \tilde{\varphi}(\hat{\Delta} W(s), \hat{\Delta} W(t)) \hat{\Delta} W^r(s) \hat{\Delta} W^\ell(t) \mid \sigma\{\hat{\Delta} W(\tau) : \tau < s, s + \hat{\Delta}s < \tau < t, \tau > t + \hat{\Delta}t \} \right]
\]

(3.21)

\[
= E[\alpha(s, t)\partial_{W^r(s)} \partial_{W^\ell(t)} \tilde{\varphi}(0)] E[(\hat{\Delta} W^r(s))^2] E[(\hat{\Delta} W^\ell(t))^2] + E[O(\hat{\Delta}s \hat{\Delta}t)]
\]

\[
= E[\alpha(s, t)\partial_{W^r(s)} \partial_{W^\ell(t)} \tilde{\varphi}(0)] E[\hat{\Delta}s] E[\hat{\Delta}t] + E[O(\hat{\Delta}s \hat{\Delta}t)].
\]

Therefore, taking the limit in the expected value of the sum of terms from (3.20) when \(\hat{\Delta}t \to 0^+\) and \(\hat{\Delta}s \to 0^+\) yields

\[
E \left[ \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \int_{t_n}^{t} (\partial_t b^r_k b^\ell_i)(s, \tilde{X}(s)) dW^r(s) dW^\ell(t) \tilde{\varphi}_k(t_{n+1}) \right]
\]

\[
\overset{(3.20)}{=} E \left[ \int_0^T \int_0^T \alpha(s, t) \tilde{\varphi}_k(t_{n+1}) dW^r(s) dW^\ell(t) \right]
\]

\[
\overset{(3.21)}{=} \int_0^T \int_0^T E[\alpha(s, t)\partial_{W^r(s)} \partial_{W^\ell(t)} \tilde{\varphi}(0)] ds \ dt
\]

(3.22)

\[
= E \left[ \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \int_{t_n}^{t} (\partial_t b^r_k b^\ell_i)(s, \tilde{X}(s)) dW^r(s) dW^\ell(t) \tilde{\varphi}_k(t_{n+1}) ds \ dt \right]
\]

\[
= O(\Delta t_{\text{sup}}).
\]

The other terms in (3.17)–(3.19) can be estimated similarly to obtain

\[
E[g(X(T)) - g(\tilde{X}(T))] = O(\Delta t_{\text{sup}}).
\]
The final step in the proof is to use the strong convergence of Lemma 3.1 to give an expansion of the integrals in (3.22) and relate the Malliavin derivative to the variations \( \varphi' \) and \( \varphi'' \). Relation (2.37) shows that \( \varphi_k(t_{n+1}) = \partial_i g(\mathbf{X}(T)) \partial_i \mathbf{X}_i(T; t_{n+1}) \), and hence for \( t_n < s < t < t_{n+1} \)

\[
\begin{align*}
\partial W^t(t) \varphi_k(t_{n+1}) &= \partial W^t(t) \mathbf{X}_p(t_{n+1}) \partial_{xp}(t_{n+1}) \varphi_k(t_{n+1}) \\
&= b_p^s(\mathbf{X}(t_n)) \varphi_{kp}(t_{n+1}),
\end{align*}
\]

(3.23)

\[
\begin{align*}
\partial W^r(s) \partial W^t(t) \varphi_k(t_{n+1}) &= b_p^s(\mathbf{X}(t_n)) \partial W^r(s) \varphi_{kp}(t_{n+1}) \\
&= b_p^s(\mathbf{X}(t_n)) \partial_{q\ell}(\mathbf{X}(t_n)) \varphi_{kpq}(t_{n+1}).
\end{align*}
\]

Recall the notation \( Y = (X, X', X'', X''') \) and \( f(Y) \) in (2.25) for the first, second, and third variations of \( g(X(T)) \), with the corresponding equation \( dY = A(Y)dt + B(Y)dW \) following (2.24). Lemmas 3.1 and 3.1' directly show

\[
(E[|f_i(Y) - \varphi_i|^p + |f_{ij}(Y) - \varphi_{ij}|^p + |f_{ijk}(Y) - \varphi_{ijk}''|^p])^{1/p} = O(\sqrt{\Delta t_{\sup}}).
\]

To obtain similar estimates of \( \tilde{\varphi}_i - f_i \) and its variations, it is useful to introduce the function \( \tilde{X}' \) determined by the linear equation

\[
d\tilde{X}'_{ij} = \left( \int_0^1 \partial_k a_i(s, \tau X(s) + (1 - \tau) \mathbf{X}(s)) d\tau \right) \tilde{X}'_{kj}(s) ds + \left( \int_0^1 \partial_k b_i^s(s, \tau X(s) + (1 - \tau) \mathbf{X}(s)) d\tau \right) \tilde{X}'_{kj}(s) dW^s(s), \quad s > t,
\]

(3.25)

\( \tilde{X}'_{ij}(t) = \delta_{ij}, \)

and compare \( \tilde{X}' \) with the analogous linear equation (2.16) for \( X' \), which is linearized around \( X(s) \). Consider (3.25) as a perturbation of the linear equation (2.16). Then the error \( \tilde{X}' - X' \) has an integral representation following Duhamel’s principle; cf. (8.5) in [19] (which follows from a variant of (3.14)). The integral representation combined with Lemmas 3.1 and 3.1’ and their proofs yield \( (E[|\tilde{X}' - X'|^p])^{1/p} = O(\sqrt{\Delta t_{\sup}}) \). Equations (2.17)–(2.18) similarly show \( (E[|\tilde{X}'' - X''|^p + |X''' - X''|^p])^{1/p} = O(\sqrt{\Delta t_{\sup}}) \), which requires \( 8p \leq k_0 \). Consequently, the functions

\[
\tilde{\varphi}_i(t) = \left( \int_0^1 \partial_k g(s X(T) + (1 - s) \mathbf{X}(T)) ds \right) \tilde{X}'_{ki}(T; t)
\]

and

\[
f_i(t) = \partial_k g(X(T)) X'_k(T; t)
\]

and their derivatives can be estimated by

\[
(3.26) \quad (E[|f_i - \tilde{\varphi}_i|^p + |f_{ij} - \tilde{\varphi}_{ij}|^p + |f_{ijk} - \tilde{\varphi}_{ijk}''|^p])^{1/p} = O(\sqrt{\Delta t_{\sup}}).
\]
Lemmas 3.1 and 3.1′ imply \((E[|\hat{X} - \bar{X}|^p])^{1/p} = \Theta(\sqrt{\Delta t_{\text{sup}}})\), which together with the combination (3.24) and (3.26) show that the functions

\[
v(\hat{X}) = (\partial_i b_k^i b_l^j)(s, \hat{X}(s)) \partial_{W_{\ell}(s)} \partial_{W_{\ell}(t)} \bar{q}_k (t_n + 1),
\]

\[
v(\bar{X}) = (\partial_i b_k^i b_l^j)(t_n, \bar{X}(t_n)) \partial_{W_{\ell}(s)} \partial_{W_{\ell}(t)} \bar{q}_k (t_n + 1),
\]

satisfy the estimate \((E[|v(\hat{X}) - v(\bar{X})|^p])^{1/p} = \Theta(\sqrt{\Delta t_{\text{sup}}})\). This pathwise error estimate and (3.23) imply, with \(2 \leq p \leq k_0/8\) and \(1/p + 1/q = 1\),

\[
\left| E \left[ \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \int_{t_n}^{t'} \left( \partial_i b_k^i b_l^j(s, \hat{X}(s)) \partial_{W_{\ell}(s)} \partial_{W_{\ell}(t)} \bar{q}_k (t_n + 1) ds dt 
\right.ight.
\]

\[
- (\partial_i b_k^i b_l^j b_p^r b_q^s)(t_n, \bar{X}(t_n)) \bar{q}_{kp}^{\prime} (t_n + 1) \Delta t_n^2 / 2 \left| \right. \right]
\]

\[
= \left| \int_0^T \int_0^T E \left[ \chi_{\Delta t}(s, t) (v(\hat{X}) - v(\bar{X})) \right] ds dt \right|
\]

\[
\leq \left( \int_0^T \int_0^T \chi_{\Delta t}(s, t) dP ds \right)^{1/q} \left( \int_0^T \int_\Omega \chi_{\Delta t}(s, t) (v(\hat{X}) 
\]

\[
- v(\bar{X}))^p dP ds \right)^{1/p} dt,
\]

where

\[
\left( \int_0^T \int_0^T \chi_{\Delta t}(s, t) dP ds \right)^{1/q} \leq (E[\Delta t_n])^{1/q} \leq \frac{E[\Delta t_n]}{\left( \inf_{n, \omega} \Delta t_n (\omega) \right)^{1/p}},
\]

\[
\left( \int_0^T \int_0^T \chi_{\Delta t}(s, t) (v(\hat{X}) - v(\bar{X}))^p dP ds \right)^{1/p} \leq C(\Delta t_{\text{sup}})^{1/2 + 1/p}.
\]

Consequently,

\[
\left| E \left[ \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \int_{t_n}^{t'} \left( \partial_i b_k^i b_l^j(s, \hat{X}(s)) \partial_{W_{\ell}(s)} \partial_{W_{\ell}(t)} \bar{q}_k (t_n + 1) ds dt 
\right.ight.
\]

\[
- (\partial_i b_k^i b_l^j b_p^r b_q^s)(t_n, \bar{X}(t_n)) \bar{q}_{kp}^{\prime} (t_n + 1) \Delta t_n^2 / 2 \left| \right. \right]
\]

\[
\leq CE \left[ \int_0^T \Delta t_n \Delta t_{\text{sup}}^{1/2} \left( \frac{\Delta t_{\text{sup}}}{\inf_{n, \omega} \Delta t_n (\omega)} \right)^{1/p} dt \right]
\]

\[
\leq CE \left[ \int_0^T \Delta t_n \sqrt{\frac{TOL}{C(TOL)}} \left( \frac{C(TOL)}{c(TOL)} \right)^{1/p} dt \right],
\]

which together with similar estimates of the other terms in the expected value of (3.14) and (3.17)–(3.19) prove the theorem. \qed
4 Stochastic or Deterministic Time Steps?

To simplify the comparison between deterministic and stochastic step size control, we assume in this section that the stochastic error density function \( \tilde{\rho} \equiv \rho_n(\omega) \) in (2.42), (3.12), (5.2), and (5.3) is nonnegative.

Within the class of deterministic time steps, the minimal number of steps is

\[
N_D = \frac{\left(\int_0^T \sqrt{E[\tilde{\rho}(t)]} \, dt\right)^2}{\text{TOL}},
\]

with the constraint \( E[g(X(T)) - g(\overline{X}(T))] = \text{TOL} \), which is achieved by choosing \( E[\tilde{\rho}(t)] \Delta t^2 = \text{const} \); cf. (5.6).

The minimal expected number of time steps in the larger class of stochastic time steps \( \Delta t(t, \omega) \) also depends on stochastic data and on the individual realizations \( \overline{X} \), with the same constraint \( E[g(X(T)) - g(\overline{X}(T))] = \text{TOL} \). This minimal expected number of steps is achieved by determining \( \Delta t(t, \omega) \) individually for each realization \( \overline{X}(\cdot, \omega) \) from

\[
\tilde{\rho}(t, \omega)(\Delta t(t, \omega))^2 = \text{const},
\]

where \( \text{const} = \text{TOL}/E[N_S] \). This yields the minimal expected number of time steps

\[
E[N_S] = \frac{\left(E[\int_0^T \sqrt{\tilde{\rho}(t)} \, dt]\right)^2}{\text{TOL}}.
\]

As a further comparison, a constant time step with the constraint \( E[g(X(T)) - g(\overline{X}(T))] = \text{TOL} \) yields the number of time steps

\[
N_C = \frac{T \int_0^T E[\tilde{\rho}(t)] \, dt}{\text{TOL}}.
\]

The following basic relation between the three time-stepping methods is obtained by applying Jensen’s inequality,

\[
E[N_S] \leq N_D \leq N_C.
\]

It would be interesting to have more precise information on the relation between the work \( E[N_S], N_D, \) and \( N_C \) in order to compare the different methods of time stepping. For instance, if the error density function satisfies

\[
\frac{\sqrt{E[\tilde{\rho}(t)]}}{E[\sqrt{\tilde{\rho}(t)}]} \leq C,
\]

then

\[
N_D \leq C^2 E[N_S];
\]

in particular, \( C \to 1 \) as \( b \to 0 \). Therefore, adaptive deterministic steps seem reasonable for small noise.

To motivate adaptive time steps, consider first the special case of integration of deterministic functions, i.e., \( a(t, x) = a(t) \in \mathbb{R}, b(t, x) = 0 \). Choose the number of steps \( N_a \) for an optimal adaptive Euler method and the number of steps \( N_u \) for
the Euler method with uniform steps such that the global integration error on (0, 1) for both methods is TOL. Then $\text{TOL}N_u \rightarrow \|a^\prime\|_{L^{1/2}(0,1)}$ and $\text{TOL}N_u \rightarrow \|a^\prime\|_{L^1(0,1)}$ as $\text{TOL} \rightarrow 0$. Therefore adaptive integration is more efficient for functions with singularities, such as $a(t) = t^{-1/2}$, where $\|a^\prime\|_{L^{1/2}(0,1)} < \infty$ but $\|a^\prime\|_{L^1(0,1)} = \infty$. Similarly, the example $dX = A(t)Xdt + B(t)XdW$ with possible singularities at deterministic times yields

$$\lim_{\text{TOL} \to 0} \text{TOL} E[N_S] = \lim_{\text{TOL} \to 0} \text{TOL} N_D,$$

so that restricting to deterministic steps is advantageous, since the more complicated stochastic steps do not give a reduction in the number of steps. In particular, Example 2 in Section 5 with $A(t) = |t - 1|^{-1/2}$, $B(t) = 1$, shows $\text{TOL}^{-1} \sim N_D \ll N_C \sim \text{TOL}^{-2}$.

If the singularity occurs at random times, stochastic steps are more efficient. Example 3 in Section 5 shows $\text{TOL}^{-1} \sim E[N_S] \ll N_D \sim \text{TOL}^{-2}$, with $dX = a(t, X(t))dt + X(t)dW$ and

$$a(t, x) \equiv \begin{cases} 0 & \text{if } t \leq \alpha \\ \frac{x}{\sqrt{4|t - \alpha|}} & \text{if } t > \alpha \end{cases}$$

where the random variable $\alpha$ is uniformly distributed on a subset of $(0, T)$ and independent of $W$.

Remark 4.1. Lemma 2.1 also leads to an error representation for stochastic time steps based on the density with $\varphi, \varphi^\prime, \varphi^\prime\prime$ in (3.12) replaced by $u_x, u_{xx}, u_{xxx}$. However, this density requires very expensive computations of $u_x, u_{xx}, u_{xxx}$, and therefore it does not seem useful for adaptive algorithms, although it yields adapted time steps. On the other hand, the alternative density shows that the error density (3.12) is not unique and hence the final number of steps $N_S$, based on (3.12), may not be optimal when compared to all possible error densities.

5 Adaptive Algorithms and Numerical Experiments

This section describes adaptive algorithms and some numerical experiments for the computational problem (1.3). These algorithms adaptively choose the number of realizations and the time steps to bound the approximation error by a given tolerance. The heart of the matter is to estimate, in an a posteriori way, the statistical and the time discretization error.

5.1 Control of the Statistical Error

For $M$ independent samples $\{Y(\omega_j)\}_{j=1}^M$ of a random variable $Y$, with $E[|Y|^6] < \infty$, define the sample average $A(M; Y)$ and the sample standard deviation $S(M; Y)$ of $Y$ by

$$A(M; Y) \equiv \frac{1}{M} \sum_{j=1}^M Y(\omega_j) \quad \text{and} \quad S(M; Y) \equiv [A(M; Y^2) - (A(M; Y))^2]^{1/2}.$$
Let $\sigma \equiv (E[|Y - E[Y]|^2])^{1/2}$ and consider the random variable

$$Z_M \equiv \frac{\sqrt{M}}{\sigma}(A(M; Y) - E[Y])$$

and its cumulative distribution function $F_{Z_M}(x) \equiv P(Z_M \leq x), x \in \mathbb{R}$. Let $\lambda \equiv (E[|Y - E[Y]|^3])^{1/3}/\sigma < \infty$; then the Berry-Esseen theorem (cf. [6, p. 126]) gives the following estimate in the central limit theorem:

$$\sup_{x \in \mathbb{R}} |F_{Z_M}(x) - \Phi(x)| \leq \frac{3}{\sqrt{M}} \lambda^3$$

for the rate of convergence of $F_{Z_M}$ to the cumulative distribution function $\Phi$ of a normally distributed random variable with mean zero and variance one. Since in the examples below $M$ is sufficiently large, i.e., $M \gg 36\lambda^6$, the statistical error

$$E_s(M, Y) \equiv E[Y] - A(M; Y)$$

satisfies, by the Berry-Esseen theorem, the following probability approximations:

$$P\left(\left|E_s(M, Y)\right| \leq c_0 \frac{\sigma}{\sqrt{M}}\right) \simeq 2\Phi(c_0) - 1.$$

In practice, choose some $c_0 \geq 1.65$. Then, $1 > 2\Phi(c_0) - 1 \geq 0.901$, so that

$$|E_s(M, Y)| \leq E_s(M, Y) \equiv c_0 \frac{S(M; Y)}{\sqrt{M}} \tag{5.1}$$

with probability close to 1, which involves the additional step to approximate $\sigma$ by $S(M; Y)$; cf. [9]. Thus, in the computations the deviation $E_s(M, Y)$ is a good approximation of the statistical error $E_s(M, Y)$. For a given $TOL > 0$, the goal is to find $M$ such that $E_s(M, Y) \leq TOL$. The following algorithm adaptively finds $M$ to compute the sample average $A(M; Y)$ as an approximation to $E[Y]$. With large probability, depending on $c_0$, the statistical error in the approximation is then bounded by $TOL$.

**routine Monte-Carlo(TOL, $M_0$; $EY$)**

Set $k = 0$, $M[0] = M_0$, and $E_s[0] = 2TOL$.

**while** $E_s[k] > TOL$ **do**

Compute $M[k]$ new samples of $Y$, along with the sample average $EY = A(M[k]; Y)$, the sample variance $S[k] \equiv S(M[k]; Y)$, and the deviation $E_s[k + 1] \equiv E_s(M[k], Y)$. Compute $M[k + 1]$ by `change_M(M[k], S[k], TOL; $M[k + 1]$). Increment $k$ by 1.

**end while**

**end-of-Monte-Carlo**

**routine change_M($M_{in}$, $S_{in}$, TOL; $M_{out}$)**

$$M_{out} = \min \left\{ \text{integer part} \left(\frac{c_0 S_{in}}{0.95TOL}\right)^2, MCH \times M_{in}\right\}$$

**end-of-change_M**
Here $M_0$ is a given initial value for $M$, and $MCH > 1$ is a positive integer parameter introduced to avoid a large new number of realizations due to a possible inaccurate sample standard deviation. Indeed, $M[k + 1]$ cannot be greater than $MCH \times M[k]$.

5.2 Control of the Time Discretization Error with Deterministic Time Steps

For a given partition $\{t_n\}_{n=0}^N$ of the time interval $[0, T]$, let the piecewise constant mesh function $\Delta t$ be defined by

$$\Delta t(\tau) \equiv \Delta t_n \quad \text{for} \quad \tau \in (t_n, t_{n+1}) \quad \text{and} \quad n = 0, 1, \ldots, N - 1.$$ 

Consider the piecewise constant function $\rho$, which measures the density of the time discretization error, defined by

$$\rho(\tau) \equiv \rho_n \quad \text{for} \quad \tau \in [t_n, t_{n+1}) \quad \text{and} \quad n = 0, 1, \ldots, N - 1,$$

where

$$\rho_n \equiv \frac{1}{\Delta t_n} \left( \frac{1}{2} \left( a_i(t_{n+1}, \overline{X}(t_{n+1})) - a_i(t_n, \overline{X}(t_n)) \right) \phi_i(t_{n+1}) 
+ \frac{1}{2} \left( d_{ik}(t_{n+1}, \overline{X}(t_{n+1})) - d_{ik}(t_n, \overline{X}(t_n)) \right) \phi'_{ik}(t_{n+1}) \right) \tag{5.2}$$

in the case of Theorem 2.2 or

$$\rho_n \equiv \frac{1}{2} \left( \frac{\partial}{\partial t} a_k + \partial_j a_k a_j + \partial_i j a_k d_{ij} \right)(t_n, \overline{X}(t_n)) \phi_k(t_{n+1}) 
+ \left( \frac{\partial}{\partial t} d_{km} + \partial_j d_{km} a_j + \partial_i j d_{km} d_{ij} + 2 \partial_j a_k d_{jm} \right)(t_n, \overline{X}(t_n)) \phi'_{km}(t_{n+1}) 
+ \left( 2 \partial_j d_{km} d_{jr} \right)(t_n, \overline{X}(t_n)) \phi''_{kmr}(t_{n+1}) \right) \tag{5.3}$$

in the case of Theorem 3.3.

Remark 5.1. In order to guarantee that $\Delta t_{\text{sup}} \to 0$ as $\text{TOL} \to 0$, we need to ensure a uniform lower bound of the error density. This requirement is needed both in the deterministic and stochastic time-stepping algorithms. Thus, a modified error density $\hat{\rho}(t) \equiv \text{sign}(\rho(t)) \max(|\rho(t)|, \sqrt{TOL})$ should be used instead of $\rho$. A similar cutoff can be applied to ensure that the error density is uniformly bounded above by, e.g., $\exp(1/TOL)$.

Following Theorems 2.2 and 3.3, the time discretization error

$$E_T(\Delta t) \equiv |E[g(X(T))] - E[g(\overline{X}(T))]|$$
is approximated by sample averages of the expansion $|E\left[\sum_{n=0}^{N-1}(\Delta t_n)^2\rho_n\right]|$ with computable terms. Hence, it follows that

$$\left|E\left[\sum_{n=0}^{N-1}(\Delta t_n)^2\rho_n\right]\right| - E_\tau(M_\tau, \Delta t) \leq E_{TS}(M_\tau, \Delta t),$$

with probability close to 1, where

$$E_\tau(M_\tau, \Delta t) \equiv \left|A\left(M_\tau; \sum_{n=0}^{N-1}(\Delta t_n)^2\rho_n\right)\right|,$$

$$E_{TS}(M_\tau, \Delta t) \equiv E_S\left(M_\tau, \sum_{n=0}^{N-1}(\Delta t_n)^2\rho_n\right).$$

The time discretization error is then approximated a posteriori as the sum of the quantities in (5.4), i.e.,

$$E_\tau(\Delta t) \approx E_\tau(M_\tau, \Delta t) + E_{TS}(M_\tau, \Delta t).$$

For a given $TOL > 0$, the goal is to construct a partition $\Delta t$ of $[0, T]$ with as few time steps and realizations $M_\tau$ as possible such that $E_\tau(M_\tau, \Delta t) \leq TOL$. To do this, first split the tolerance in two positive parts $TOL'$ and $TOL''$ for $E_\tau(M_\tau, \Delta t)$ and $E_{TS}(M_\tau, \Delta t)$, respectively. To this end, the number $k \geq 0$ denotes some integer indexing iterations. Let $\Delta t[k]$ be a partition of $[0, T]$ with $N[k]$ time steps and assume that $M_\tau[k]$ samples of the Euler method (1.2) over $\Delta t[k]$ are given. If $E_{TS}[k] \equiv E_{TS}(M_\tau[k], \Delta t[k]) \leq TOL''$, and $E_\tau[k] \equiv E_\tau(M_\tau[k], \Delta t[k]) > TOL'$, then refine the mesh $\Delta t[k]$ using the procedure refine $\Delta t$ below to decrease the component $E_\tau[k]$ of the time discretization error. If $E_{TS}[k] > TOL''$, then increase $M[k]$ using the procedure change $M$ to decrease the statistical error $E_{TS}[k]$ of the time discretization estimate.

To use the refinement procedure, first compute the approximate error density $\bar{\rho}[k]$ that corresponds to the partition $\Delta t[k]$ by

$$\bar{\rho}[k] \equiv A(M[k]; \rho[k]),$$

where $\rho[k]$ is the error density function on $\Delta t[k]$; see (5.2) or (5.3) and Remark 5.1.

Consider the following minimization problem:

$$(5.5) \min_{\Delta t \in \mathcal{K}[k]} \mathcal{N}(\tilde{\Delta} t)$$

where $\mathcal{K}[k]$ is the feasible set

$$\mathcal{K}[k] \equiv \left\{ \tilde{\Delta} t \in L^2(0, T) : \tilde{\Delta} t \text{ is positive and piecewise constant} \right\}$$

on $\Delta t[k]$ such that $\int_0^T |\bar{\rho}[k](\tau)|\tilde{\Delta} t(\tau)d\tau \leq TOL'$.
and
\[ N(\widetilde{\Delta t}) \equiv \int_0^T \frac{1}{\Delta t(\tau)} d\tau = \sum_{n=0}^{N[k]-1} \frac{\Delta t_n[k]}{\Delta t_n} \]
is the number of steps of the partition \( \widetilde{\Delta t} \). A standard application of a Lagrange multiplier shows that the minimizer of problem (5.5) is
\[
\widetilde{\Delta t}^* = \frac{TOL'}{\sqrt{|\bar{\rho}[k]| \left( \int_0^T \sqrt{|\bar{\rho}[k](s)|} ds \right)}.
\]
To construct the new mesh \( \Delta t[k+1] \), divide, for \( n = 0, 1, \ldots, N[k]-1 \), the interval \([t_n, t_{n+1}]\) of \( \Delta t[k] \) into \( m_n \) uniform subintervals, where
\[
m_n = \min \left\{ \max \left\{ \text{integer part} \left( \frac{\Delta t_n[k]}{\Delta t_n^*} \right), 1 \right\}, \text{NCH} \right\}
\]
and NCH > 1 is an integer parameter, similar to MCH, that bounds the increment of the number of time nodes. The new finer mesh \( \Delta t[k+1] \) is, in general, nonuniform with \( N[k+1] = \sum_{n=0}^{N[k]-1} m_n \) subintervals and, by its construction, \( N[k+1] \leq \text{NCH} \times N[k] \).

Remark 5.2. The strategy above for the construction of a new partition \( \Delta t[k+1] \) is based only on dividing the old time steps \( \Delta t[k] \). It is possible to construct an algorithm that also allows the merging of time steps; see [22]. The dividing and merging algorithm uses more information about the asymptotic behavior of the error density in order to ensure finite termination. The advantage with merging and dividing is that the algorithm can be proven to stop with the optimal number of time steps up to a problem-independent factor.

As pointed out in Remark 2.7, the error density \( \bar{\rho} \) based on (5.2) has a pointwise statistical error, cf. (2.43), which may be larger than the statistical error in the time discretization error estimate (5.4) as the mesh is refined. Unless the number of realizations is based on the pointwise statistical error of the error density, the adaptive mesh specified via (5.6) becomes inaccurate with undesirable oscillations. A remedy to obtain an accurate mesh function, keeping the number of realizations based on the statistical error of the error bound by (2.43), is to filter the sampled error density function. The appendix describes a filtered error density to determine the new mesh by (5.6). This error density is a local \( L^2 \) projection of \( \bar{\rho} \) onto piecewise constant or piecewise linear functions on a mesh determined by minimizing the sum of the pointwise statistical and the approximation errors.

5.3 The Adaptive Algorithm with Deterministic Time Steps

The statistical and time discretization errors are combined in order to bound the computational error
\[
\mathcal{E}_c \equiv E[g(X(T))] - A(M; g(\bar{X}(T))) ,
\]
which is estimated, a posteriori, by the sum of discretization and statistical errors in (5.4) and (5.1), i.e.,

\[(5.7) \quad |\mathcal{E}_c| \approx \mathbb{E}_\tau (M, \Delta t) + \mathbb{E}_{TS} (M, \Delta t) + \mathbb{E}_s (M, g(\overline{X}(T))).\]

To bound the three parts of the computational error by a specified tolerance, the adaptive algorithm, with deterministic time steps, combines the techniques described in the previous subsections related to the control of the statistical and the time discretization error. In particular, in the first loop below it uses the procedures `refine_\Delta t` and `change_M` to determine the mesh with \(N\) steps and the number of realizations \(M_T\) by iteratively increasing \(M_T\) and \(N\) until the time discretization error is below the tolerance. Then, in the second loop, where the mesh is fixed, it increases the number \(M_C\) of realizations of the Euler method (1.2) using only the procedure `change_M` until the statistical error \(\mathbb{E}_S (M_C, g(\overline{X}(T)))\) is below the tolerance. In practice \(M_T \ll M_C\). In this last part of the adaptive algorithm, the computational cost is coming only from the computation of \(M_C\) Euler approximations of \(g(X(T))\), because there is no need to compute the time discretization error component \(\mathbb{E}_{\tau}(M_C, \Delta t)\), which is based on the solution of the discrete backward problems described in Section 2.

**Initialize:** Choose a number of initial realizations \(M[0]\), a coarse mesh function \(\Delta t[0]\) for \([0, T]\), and a tolerance \(TOL > 0\). Split the tolerance into the three parts \(TOL = TOL_S + TOL_T + TOL_{TS}\) of statistical tolerance \(TOL_S\), time discretization tolerance \(TOL_T\), and statistical time discretization tolerance \(TOL_{TS}\), where \(TOL_S = \frac{2}{3}TOL\), \(TOL_T = \frac{2}{9}TOL\), and \(TOL_{TS} = \frac{1}{9}TOL\). Choose \(c_0 \geq 1.65\) and define \(M_{CH} = 50\) and \(N_{CH} = 3\). Set \(k = 0\), \(\mathbb{E}_\tau[0] = 2TOL_T\), and \(\mathbb{E}_{TS}[0] = 2TOL_{TS}\).

**while** \(\mathbb{E}_\tau[k] + \mathbb{E}_{TS}[k] > TOL_T + TOL_{TS}\) **do**

Compute sample averages and error estimates by calling Euler.

**if** \(\mathbb{E}_{TS}[k] > TOL_{TS}\) **then**

Compute \(M_T[k+1] \) by `change_M(M_T[k], \mathbb{V}_{TS}[k], TOL_{TS}; M_T[k+1])`.

\(\Delta t[k+1] = \Delta t[k]\).

Increment \(k\) by 1.

**else if** \(\mathbb{E}_\tau[k] > TOL_T\) **then**

Compute \(\Delta t[k+1]\) by `refine_\Delta t(TOL_T, \Delta t[k], \tilde{\rho}[k]; \Delta t[k+1])`.

\(M_T[k+1] = M_T[k]\).

Increment \(k\) by 1.

**end if**

**end while**

Set \(M_C[k] = M_T[k], \Delta t_c = \Delta t[k]\).

Use Monte-Carlo\((M_C[k], TOL_S; Eg)\) to compute an approximation \(Eg\) for \(E[g(\overline{X}(T))]\).

Accept \(Eg\) as an approximation of \(E[g(X(T))]\) since the computational error is bounded by \(TOL\).
Although values for the parameter $c_0$ have been suggested in the description of the algorithm, the actual value can be set by the user depending on the confidence level required on the computation. The following subroutine completes the description of the algorithm.

**routine Euler**

Compute $M_T[k]$ new realizations of the Euler method (1.2) with the partition $\Delta t[k]$ and update the approximations of $E_T[k] = E_T(M_T[k], \Delta t[k])$ and the statistical time discretization error $E_{TS}[k] = E_{TS}(M_T[k], \Delta t[k])$.

**end-of-Euler**

### 5.4 The Adaptive Algorithm with Stochastic Time Steps

An adaptive, nonadapted time-stepping algorithm based on (3.2)–(3.3) has a structure similar to the Monte-Carlo routine. Its outer loop computes batches of realizations of, for example, $X$, with increasing size until an estimate for the statistical error is below a given tolerance. For each fixed realization, a given initial time partition $1_{t=0}^T U$ is successively refined until an error indicator is below a tolerance for all time steps. This procedure needs to sample the Wiener process $W$ on finer partitions, given its values on coarser ones. Therefore, the use of Brownian bridges is natural to preserve the required independence between the Wiener increments.

**Initialize**: Choose a number of initial realizations $M[0]$, a coarse mesh function $1_{t=0}^T U$ for $T = 1_T$, and a tolerance $TOL > 0$. Split the tolerance into the two parts $TOL = TOL_s + TOL_r$ of statistical tolerance $TOL_s$ and time discretization tolerance $TOL_r$, where $TOL_s = \frac{2}{3}TOL$, $TOL_r = \frac{1}{3}TOL$. Choose $c_0 \geq 1.65$ and define $MCH = 50$ and $NCH = 3$. Set $m = 0$ and $E_s[m] = 2TOL_s$. Let $\delta > 0$ be a constant approximating $TOL_r/E[N]$, where $E[N]$ is the expected number of time steps.

**while** $E_s[m] > TOL_s$ **do**

**for** $j = 1, 2, \ldots, M[m]$ **do**

Start with the initial coarse time partition $\Delta t[0]$, generate $\Delta W[0]$, set $k = 0$ and $r[k] = 2\delta$.

**while** max, $r[k](t) \geq \delta$ **do**

Compute $\overline{X}[k]$ and the weight function $\rho[k]$ defined in (5.3) on $\Delta t[k]$ using the known Wiener increments $\Delta W[k]$. Define $r[k+1] = \rho[k](\Delta t[k])^2$. Let for all $t$

$$\Delta t[k+1](t) = \begin{cases} 
\Delta t[k](t) & \text{if } r[k+1](t) < \delta \\
\Delta t[k](t)/2 & \text{if } r[k+1](t) \geq \delta 
\end{cases}.$$  

Compute $\Delta W[k+1]$ on $\Delta t[k+1]$ using Brownian bridges based on the already known values $\Delta W[k]$.

**if** at least one step of $\Delta t[k]$ is refined by (5.8) **then**

Increment $k$ by 1.
else We have $\Delta t[k + 1] = \Delta t[k]$, so we accept the approximation $g(\overline{X}(T, \omega_j))$.
end if
end while

Compute the sample average $Eg = A(M[m]; g(\overline{X}(T)))$, the sample standard deviation $S[m] = S(M[m]; g(\overline{X}(T)))$, and the a posteriori bound for the statistical error $E_s[m + 1] = E_s(M[m], g(\overline{X}(T)))$.

Compute $M[m + 1]$ by $\text{change}_M(M[m], S[m], \text{TOL}_s; M[m + 1])$.

Update $\delta = \text{TOL}_r/A(M[m]; N)$, where the random variable $N$ is the final number of time steps on each realization, and increase $m$ by 1.
end while

Accept $Eg$ as an approximation of $E[g(X(T))]$ since the computational error is bounded by $\text{TOL}$.

5.5 Numerical Results

This section presents results of numerical examples from the implementation of the adaptive algorithms described above. Example 1 shows the performance of the deterministic time-stepping algorithm, while Examples 2, 3, and 4 illustrate the behavior of the adaptive deterministic and stochastic time-stepping algorithms.

In what follows, the adaptive algorithm with deterministic time steps is called Algorithm $D_1$ when formula (5.2) is applied, and Algorithm $D_2$ when formula (5.3) is used. The computational error $\mathcal{E}_c$ is approximated by means of (5.7). The algorithm with stochastic time steps, based on the error density (5.3), is called Algorithm $S$. Example 1 uses double-precision FORTRAN 77 and simulates the increments of the $\ell_0$-independent Wiener processes by a pseudorandom-number generator. In particular, the program applies a double-precision modification of the functions $\text{ran1}$ and $\text{gasdev}$ proposed in [27], provided with an initial seed $\text{iseed}$, which must be a negative integer. Examples 2, 3, and 4 use MATLAB version 5.3 implementations. In order to diminish the computational effort, we use antithetic variates (see [13]) to reduce the variance in all the computations of Examples 2, 3, and 4.

Example 1

Consider the following Itô stochastic differential equation (1.1) proposed in [32], where $d = 2$, $\ell_0 = 2$,

$$a(t, x) = (-x_2, x_1), \quad b^1(t, x) = \left(0, \frac{\sin(x_1 + x_2)}{\sqrt{1 + t}}\right),$$

$$b^2(t, x) = \left(\frac{\cos(x_1 + x_2)}{\sqrt{1 + t}}, 0\right) \text{ for } t \geq 0, \ x \in \mathbb{R}^2,$$

and the initial condition is $X(0) = (1, 1)$. For $g(x) \equiv (x_1)^2 + (x_2)^2, x \in \mathbb{R}^2$, this problem has the exact solution $E[g(X(T))] = E[g(X(0))] + \log(1 + T)$. 

Algorithm $D_1$ is presented with $\text{iseed} = -1$ and $\text{iseed} = -101$ for Algorithms $D_1$ and $D_2$, respectively. The number of realizations $M$ is sufficiently large so that the total statistical error is small compared to the discretization error. As predicted by Theorems 2.2 and 3.3, Table 5.1 then shows that the ratio of the computational error $\mathcal{E}_c$ and its computable approximation tends to 1 as $N$ increases.

Table 5.2 shows the result of successive iterations of Algorithm $D_1$ with adaptive choice of time steps and number of realizations. The program starts with $M = 100$ initial realizations, and $N = 10$ subintervals as an initial uniform partition of the time interval $[0,1]$. The tolerance $\text{TOL} = 0.02$ is divided into $\text{TOL}_S = 0.01333$, $\text{TOL}_T = 0.00444$, and $\text{TOL}_{TS} = 0.00222$. Here the computation of the errors $\mathcal{E}_T$ and $\mathcal{E}_{TS}$ takes about 11% of the execution time.

Finally, Table 5.3 displays the proportion of computational time used to determine the mesh compared to the total execution time of the program. In this comparison we apply Algorithm $D_1$ for decreasing values of the tolerance. Each entry corresponds to six runs with different values of $\text{iseed}$. The additional work to determine the time steps by computing the backward problems (2.7) and (2.9) becomes negligible as the tolerance decreases.
Example 2

Let us consider a real constant $\alpha \in (0, T)$ and the linear stochastic differential equation

$$
\begin{aligned}
\frac{dX(t)}{dt} &= X(t)\frac{dW(t)}{dt} + \frac{X(t)}{2\sqrt{T-\alpha}} \quad \text{if } t \in [0, \alpha], \\
X(0) &= 1,
\end{aligned}
$$

(5.9)

with the unique solution

$$
X(t) = \begin{cases} 
\exp(W(t) - \frac{1}{2}t), & t \in [0, \alpha], \\
\exp(W(t) - \frac{1}{2}t) \exp(\sqrt{t-\alpha}), & t \in [\alpha, T].
\end{cases}
$$

Here we choose $T = 1$ and $\alpha = T/3$. Our goal is to approximate $E[X(T)] = \exp(\sqrt{T-\alpha})$. To avoid evaluating arbitrarily large values of the drift in (5.9), we modify it to be

$$
a(t, x) = \begin{cases} 
0, & t \in [0, \alpha] \\
\frac{x}{2\sqrt{T-\alpha+TOL^4}}, & t \in (\alpha, T]
\end{cases}
$$

(5.10)

yielding a higher-order perturbation $O(TOL^2)$ in the computed result and not affecting the size of the optimal time steps. Due to the time discontinuity of the drift function and to ensure optimal convergence of the adaptive algorithms, we modify the Euler method by

$$
\bar{X}_{n+1} - \bar{X}_n = a(\hat{t}, \bar{X}_n)\Delta t_n + \bar{X}_n \Delta W_n, \quad n = 0, 1, \ldots,
$$

(5.11)

where we choose the stochastic evaluation time $\hat{t} \in \{t_n, t_{n+1}\}$ by $|a(\hat{t}, \bar{X}_n)| = \max(|a(t_n, \bar{X}_n)|, |a(t_{n+1}, \bar{X}_n)|)$. Observe that the use of $\hat{t}$ does not change the adapted nature of the Euler method. We compare both algorithms, the adaptive adapted deterministic time stepping and the adaptive with stochastic time steps. In this example they both select the time steps in a very similar way, since there is no remarkable influence from the stochastic term $X(t)\frac{dW(t)}{dt}$ in the dynamics of (5.9). However, the smaller number of realizations in the stochastic time-stepping...
Algorithm $D_2$

<table>
<thead>
<tr>
<th>$M$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2.6 \times 10^4$</td>
<td>$4.5 \times 10^2$</td>
</tr>
</tbody>
</table>

Algorithm $S$

<table>
<thead>
<tr>
<th>$M$</th>
<th>$A(M; N)$</th>
<th>$S(M; N)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2.1 \times 10^4$</td>
<td>$4.5 \times 10^2$</td>
<td>$1.0 \times 10^2$</td>
</tr>
</tbody>
</table>

Table 5.4. Example 2: Final $M$ and $\#\Delta t$ using Algorithms $D_2$ and $S$ for $T = 1$ and $\alpha = 1/3$.

Algorithm $D_2$

<table>
<thead>
<tr>
<th>$M$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2.1 \times 10^4$</td>
<td>$5.0 \times 10^3$</td>
</tr>
</tbody>
</table>

Algorithm $S$

<table>
<thead>
<tr>
<th>$M$</th>
<th>$A(M; N)$</th>
<th>$S(M; N)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2.0 \times 10^4$</td>
<td>$3.5 \times 10^2$</td>
<td>$1.0 \times 10^2$</td>
</tr>
</tbody>
</table>

Table 5.5. Example 3: Final $M$ and $\#\Delta t$ of Algorithms $D_2$ and $S$ for $T = 1$ and stochastic $\alpha$.

algorithm is partly explained by a smaller sample variance. The exact variance of $g(X(T))$ is 8.8, while Algorithm $D_2$ gets 2.5 and Algorithm $S$ only 2.1. Finally, Table 5.4 displays results from this example using $TOL = 2.5 \times 10^{-2}$.

Example 3

Now we change (5.9) taking $\alpha$ no longer a constant but a uniformly distributed random variable independent of the Wiener process, i.e., $\alpha \sim U(\tilde{\alpha}, \bar{\alpha})$ with $\alpha = 1/22$ and $\tilde{\alpha} = 1 - 1/22$, for $T = 1$. The conditional expectation $E[X(T) | \alpha] = \exp(\sqrt{T - \tilde{\alpha}})$ can be used to compute

$$E[X(T)] = \frac{1}{\bar{\alpha} - \tilde{\alpha}} \int_{\tilde{\alpha}}^{\bar{\alpha}} \exp(\sqrt{T - \alpha}) d\alpha,$$

which is now the functional to approximate. Since the position of the drift singularity is stochastic, the deterministic time-stepping algorithm $D_2$ gives the approximation error $O(\sqrt{T_{sup}})$, and the stochastic time-stepping algorithm has a clear advantage, using on average much fewer final time steps, as Table 5.5 shows; cf. Section 4. On the other hand, a fairer comparison including the total number of steps and the computer time shows that our recursive refinement algorithms have potential for improvements. Here both algorithms use a similar number of realizations since the sample variances are almost the same. This numerical experiment also uses $TOL = 2.5 \times 10^{-2}$, the regularization (5.10), and the stochastic evaluation time $\hat{t}$ from (5.11).

Example 4

Finally, we include an example where the stochastic time-stepping algorithm uses a different number of time steps for different realizations, the only source of
the noise is the Wiener process, and the adaptivity is not caused by a singularity in time, in contrast to the previous example.

Thus, let us consider the computation of $E[(X(2))^2]$, where $X$ solves

$$dX(t) = 0.2 f(X(t))dt + 0.2 f(X(t))dW(t), \quad X(0) = 1.$$ 

The function $f$ is defined by

$$f(x) = \begin{cases} 
    x, & 0 < x < 2, \\
    p(x), & 2 < x < 2.25, \\
    0, & \text{otherwise},
\end{cases}$$

and the fifth-order polynomial $p$ is chosen such that $f$ is two times continuously differentiable on $\mathbb{R}$. In order to verify the results from the stochastic time-stepping algorithm, we compute a much more accurate numerical approximation based on the Feynman-Kac formula and the Kolmogorov backward partial differential equation. The result of this PDE computation is $E[X^2(2)] \approx 1.34$. After three outer iterations with $\Delta t[0] = T/10$, $TOL = 9 \times 10^{-2}$, $TOL_s = 4 \times 10^{-2}$, and $TOL_r = 5 \times 10^{-2}$, the stochastic time-stepping algorithm stopped with $M = 2.9 \times 10^3$ realizations. The computational result is $\mathcal{A}(M; (\bar{X}(2))^2) \approx 1.31$, yielding a computational error $\mathcal{E}_c \approx 3 \times 10^{-2}$. The time discretization error estimate is $\mathcal{E}_t \approx 1.5 \times 10^{-2}$, while the statistical error estimate, using $c_0 = 1.65$, is $\mathcal{E}_s \approx 3 \times 10^{-2}$. These estimates yield a total error estimate $4.5 \times 10^{-2}$, which is consistent with the observed computational error. With respect to the time partition, the sample mean value of the number of time steps is $\mathcal{A}(M; N) = 2.4 \times 10^2$, while the sample standard deviation is $\mathcal{S}(M; N) = 2.1 \times 10^3$, indicating a large variation of time steps between different realizations.

**Appendix: Filtered Error Density**

The filtering procedure for the density function $\overline{\rho}$ based on (5.2) uses a local average procedure to compute an approximation $\rho^F$, which replaces $\overline{\rho}$ in the construction of the new partition of the time interval. The algorithm to define $\rho^F$ for $n = 0, 1, \ldots, N[k] - 1$ follows the steps:

**FSTEP 1:** Compute first the approximate variance $\overline{\nu}_n \equiv \mathcal{S}(M[k]; \rho_n[k])$. Then specify the length $h_n$ of the filtration width that corresponds to the time node $t_n$ by

$$h_n \equiv \min \left\{ \frac{T}{2}, \left( \frac{\overline{\nu}_n}{1 + \overline{\rho}_A} \right)^{2/5} \right\}$$

where $\overline{\rho}_A = \frac{1}{T} \int_0^T |\overline{\rho}[k]|dt$.

**FSTEP 2:** Compute the length of the window interval $[t_n - h_n, t_n + h_n]$ that is outside of the time interval $[0, T]$ to the left and to the right by $h^L_n \equiv \max\{h_n - t_n, 0\}$ and $h^R_n \equiv \max\{t_{n+1} + h_n - T, 0\}$, respectively. Then find the indexes $j_L$ and
\[ j_R \text{ for nodes of the partition } \Delta t[k] \text{ such that } t_{j_L} = \min\{t_j : t_j \geq t_n - h_n - h_n^R \}, \]
\[ t_{j_R} = \max\{t_j : t_j \leq t_{n+1} + h_n + h_n^L \}. \]

**STEP 3:** If \( h_n^L + h_n^R = 0 \), then set
\[ \rho_n^F \equiv \int_{t_{j_L}}^{t_{j_R}} |\bar{\rho}[k]| dt, \]
which is the \( L^2 \)-projection of \( \bar{\rho}[k] \) onto the functions that are constants over \([t_{j_L}, t_{j_R}]\). If \( h_n^L > 0 \) or \( h_n^R > 0 \), then set
\[ \rho_n^F \equiv \frac{12}{(t_{j_R} - t_{j_L})^2} \left( \int_{t_{j_L}}^{t_{j_R}} |\bar{\rho}[k]| \left( \frac{t - t_{j_L}}{t_{j_R} - t_{j_L}} - \frac{1}{2} \right) dt \right) \left( t_n + \frac{\Delta t_n}{2} \right) + \frac{6}{(t_{j_R} - t_{j_L})} \left( \int_{t_{j_L}}^{t_{j_R}} |\bar{\rho}[k]| \left( \frac{2}{3} - \frac{t - t_{j_L}}{t_{j_R} - t_{j_L}} \right) dt \right), \]
which is the value, at \( t_n + \Delta t_n / 2 \), of the \( L^2 \)-projection of \( \bar{\rho}[k] \) onto the functions that are first-order polynomials over \([t_{j_L}, t_{j_R}]\).

**Acknowledgments.** The work was supported by the EU-TMR Project HCL #ERBFMRXCT960033, the Swedish Research Council for Engineering Science (TFR) Grant #222-148, the Swedish National Network in Applied Mathematics (NTM), and the Commission for Scientific Research (CSIC, UdelaR, Uruguay).

**Bibliography**


**ANDERS SZEPESSY**
Kungl. Tekniska Högskolan
Institutionen för Matematik
S-100 44 Stockholm
SWEDEN
E-mail: szepessy@math.kth.se

**RAÚL TEMPONE**
Kungl. Tekniska Högskolan
Numerisk Analys och Datalogi
S-100 44 Stockholm
SWEDEN
E-mail: rtempone@nada.kth.se

**GEORGIOS E. ZOURARIS**
University of the Aegean
Department of Mathematics
GR-83200 Karlovassi, Samos
GREECE
E-mail: zouraris@aegean.gr

Received December 1999.
CONVERGENCE RATES FOR ADAPTIVE WEAK APPROXIMATION OF STOCHASTIC DIFFERENTIAL EQUATIONS

KYOUNG-SOOK MOON, ANDERS SZEPESSY, RAÚL TEMPONE, AND GEORGIOS E. ZOURARIS

Abstract. Convergence rates of adaptive algorithms for weak approximations of Itô stochastic differential equations are proved for the Monte Carlo Euler method. Two algorithms based either on optimal stochastic time steps or optimal deterministic time steps are studied. The analysis of their computational complexity combines the error expansions with a posteriori leading order term introduced in [A. Szepessy, R. Tempone and G. Zouraris, Comm. Pure and Appl. Math., 54, 1169-1214, 2001] and an extension of the convergence results for adaptive algorithms approximating deterministic ordinary differential equations, derived in [K-S. Moon, A. Szepessy, R. Tempone and G. Zouraris, preprint http://www.nada.kth.se/~szepessy/ode.ps]. The main step in the extension is the proof of the almost sure convergence of the error density. Both adaptive algorithms are proven to stop with the optimal number of steps up to a problem independent factor defined in the algorithm. Numerical examples illustrate the behavior of the adaptive algorithms, motivating when stochastic and deterministic adaptive time step are superior to constant time steps and when adaptive stochastic steps are superior to adaptive deterministic steps.

1. Introduction to Adaptive Algorithms for SDE’s

This work derives convergence rates of adaptive algorithms for weak approximation of Itô stochastic differential equations

\[ dX_k(t) = a_k(t, X(t))dt + \sum_{\ell=1}^{\ell_0} b_{k\ell}(t, X(t))dW_{\ell}(t), \ t > 0 \]

where \( k = 1, \ldots, d \) and \((X(t; \omega))\) is a stochastic process in \( \mathbb{R}^d \), with randomness generated by the independent one dimensional Wiener processes \( W_{\ell}(t; \omega), \ \ell = 1, \ldots, \ell_0 \), on the probability space \((\Omega, \mathcal{F}, P)\), cf. [19], [28]. The functions \( a(t, x) \in \mathbb{R}^d \) and \( b_{\ell}(t, x) \in \mathbb{R}^d, \ \ell = 1, \ldots, \ell_0 \), are given drift and diffusion fluxes.

The goal is to construct approximations to the expected value \( E[g(X(T))] \) by a Monte Carlo method, for a given function \( g : \mathbb{R}^d \rightarrow \mathbb{R} \). A topical example of such an expected value is to compute option prices in mathematical finance, cf. [18]. Other related models based on stochastic dynamics are used, e.g., for stochastic dynamics.

1991 Mathematics Subject Classification. 65C30, 65Y20, 65L50, 60H35.

Key words and phrases. adaptive mesh refinement algorithm, computational complexity, Monte Carlo method, almost sure convergence.

This work is supported by the Swedish Research Council, UdelaR and UdeM in Uruguay, the Swedish Parallel and Scientific Computing Institute (PSCI), the Swedish National Board for Industrial and Technical Development (NUTEK), the EU-TMR project viscosity solutions and their applications.
climate prediction and for wave propagation in random media, cf. [21],[1]. The Monte Carlo Euler method approximates the unknown process \( X \) by the Euler method \( X(t_n) \) (cf. [20, 22]), which is a time discretization based on the nodes 
\[ 0 = t_1 < t_2 < \cdots < t_{N+1} = T \]
where
\[ \Delta t_n \equiv t_{n+1} - t_n, \quad \Delta W_n^\ell \equiv W^\ell(t_{n+1}) - W^\ell(t_n), \quad n = 1, 2, \ldots, N + 1. \]
The aim of the adaptive algorithm is to choose the size of the time steps, \( \Delta t_n \), and the number of independent identically distributed samples \( X(\cdot, \omega_j) \), \( j = 1, 2, \ldots, M \), such that the computational work, \( N \cdot M \), is minimal while the approximation error is bounded by a given error tolerance, \( \text{TOL} \), i.e. the event
\[
\left| \frac{1}{M} \sum_{j=1}^{M} g(X(T; \omega_j)) - E[g(X(T))] \right| \leq \text{TOL}
\]
has a probability close to one. A priori error estimates of the computational error in (1.3) was first derived by Talay and Tubaro in [33]. The work [31] modified Talay’s and Tubaro’s error expansion to an expansion with computable leading order term in a posteriori form, based on computable stochastic flows and discrete dual backward problems.

Here we derive convergence rates of two algorithms including dividing and merging of time steps, with either stochastic or deterministic time steps. The difference between the two algorithms is that the stochastic time steps may use different meshes for each realization, while the deterministic time steps use the same mesh for all realizations. The construction and the analysis of the adaptive algorithms are inspired by the related work [25], on adaptive algorithms for deterministic ordinary differential equations, and use the error estimates from [31].

There are numerous adaptive algorithms for ordinary and partial differential equations, cf. [16], [2], [3], [5, 6], [4], [11], [17], [24], but the theoretical understanding of convergence rates of adaptive algorithms is not as well developed; there are however recent important contributions. The work [14], [15] and [26] prove optimal convergence rates for strong approximation of stochastic differential equations. DeVore studies in [9] the efficiency of adaptive approximation of functions, including wavelet expansions, based on smoothness conditions in Besov spaces. Inspired by this approximation results, Cohen, Dahmen and DeVore prove in [7] that a wavelet-based adaptive \( N \)-term approximation algorithm produces a solution with optimal error \( O(N^{-s}) \) in the energy norm for linear coercive elliptic problems, see also [8]. The work [25] connects DeVore’s smoothness conditions to error densities for adaptive approximation of ordinary differential equations. In particular [25] constructs an algorithm and proves that it stops with the optimal number of time steps, up to a problem independent factor defined in the algorithm; for any \( p \)-th order accurate method, the optimal number of adaptive steps is proportional to the \( p \)-th root of the \( L^{p+1} \) quasi-norm of the error density, while the number of constant steps, with the same error, is proportional to the \( p \)-th root of the larger \( L^1 \)-norm of the error density. This work generalizes [25] to weak approximation of stochastic differential equations.
There are two main results on efficiency and accuracy of the adaptive algorithms described in Section 3. In view of accuracy with probability close to one, the approximation errors in (1.3) are asymptotically bounded by the specified error tolerance times a problem independent factor as the tolerance parameter tends to zero. In view of efficiency, both the algorithms with stochastic steps and deterministic steps stop with the optimal expected number of final time steps and optimal number of final time steps respectively, up to a problem independent factor. The number of final time steps is related to the numerical effort needed to compute the approximation. To be more precise, the total work for deterministic steps is roughly $M \cdot N$ where $M$ is the final number of realizations and $N$ is the final number of time steps, since the work to determine the mesh turns out to be negligible. On the other hand, the total work with stochastic steps is on average bounded by $M \cdot E[N_{\text{tot}}]$, where the total number, $N_{\text{tot}}$, of steps including all refinement levels is bounded by $O(N \log N)$ with $N$ steps in the final refinement; for each realization it is necessary to determine the mesh, which may vary for each realization.

The accuracy and efficiency results are based on the fact that the error density, $\rho$, which measures the approximation error for each interval following (2.1), converges almost surely or a.s. as the error tolerance tends to zero. This convergence can be understood by the a.s. convergence of the approximate solution, $X$, as the maximal step size tends to zero. Although the time steps are not adapted to the standard filtration generated by $W$ for the stochastic time stepping algorithm, the work [31] proved that the corresponding approximate solution converges to the correct adapted solution $X$. This result makes it possible to prove the martingale property of the approximate error term with respect to a specific filtration, see Lemma 4.2. Therefore Theorem 4.1 and 4.4 use Doob’s inequality to prove the a.s. convergence of $X$. Similar results of pointwise convergence with constant step sizes, adapted to the standard filtration, are surveyed by Talay in [32].

The outline of the paper is: Section 2 states the a posteriori error expansion, proved in [31] and used in the adaptive algorithms; Section 3 describes and analyzes the adaptive algorithms with stochastic time steps and deterministic time steps; Section 4 proves a.s. convergence of the error density; and finally Section 5 presents numerical experiments based on the adaptive algorithms.

For simplicity, we introduce the following notation

$$d_{ij} \equiv \frac{1}{2} b_i' b_j', \quad \partial_k \equiv \frac{\partial}{\partial x_k}, \quad \partial_{ki} \equiv \frac{\partial^2}{\partial x_k \partial x_i}, \ldots$$

with the summation convention, i.e., if the same subscript appears twice in a term, the term denotes the sum over the range of this subscript, e.g.

$$c_{ik} \partial_k b_j \equiv \sum_{k=1}^{d} c_{ik} \partial_k b_j.$$

For a derivative $\partial_\alpha$, the notation $|\alpha|$ is its order.

### 2. A Posteriori Error Expansion

The main result of [31] is new expansions of the computational error with computable leading order term in a posteriori form. The result was inspired by a corresponding a priori analysis derived in [33], with the main difference that the
weight for the local error contribution to the global error can be computed efficiently by stochastic flows and discrete dual backward problems in [25]. These a posteriori error expansions can be used in adaptive algorithms, in order to control the approximation error. Although [31] proposed adaptive algorithms, the main focus in that work was on error estimates. Properties regarding the stopping, efficiency and accuracy of the adaptive algorithms, following the ideas in [25] are first studied here. Assume that the process \( X \) satisfies (1.1) and its approximation, \( \tilde{X} \), is given by (1.2), then these error expansions in Theorems 1.2 and 2.2 of [31] have the form

\[
E\left[g(X(T)) - g(\tilde{X}(T))\right] = E\left[\sum_{n=1}^{N} \tilde{\rho}_n \Delta t_n^2\right] + \text{higher order terms},
\]

where \( \rho_n \Delta t_n^2 \) are computable error indicators, i.e. they provide information for further improvement of the time mesh and \( \rho_n \) measures the density of the global error in (2.1). A typical adaptive algorithm does two things iteratively:

1. if the error indicators satisfy an accuracy condition then it stops; otherwise
2. the algorithm chooses where to refine the mesh and then makes an iterative step to (1).

In addition to estimate the global error \( E[g(X(T)) - g(\tilde{X}(T))\] in the sense of (2.1), the indicators \( \rho_n \Delta t_n^2 \) also give simple information on where to refine to reach an optimal mesh, based on the almost sure convergence of the density \( \rho_n \) as we refine the discretization, see Section 4.

In the remaining part of this section we state one error expansion from [31], which can be used with either stochastic or deterministic time steps. The work [31] also proves a second error expansion which requires less computational work per realization, in particular for large \( d \), however this expansion is only valid with deterministic time steps and it has larger statistical error. Although adaptive algorithms based on this second expansion work well in practice, the larger statistical error makes it difficult to analyze precisely. Therefore this work focus on the first error expansion.

The following Lemma 2.1 and Theorem 2.2, derived in [31], describe the error expansion which is used in the adaptive algorithms in Section 3. Assume that for all times \( t \in [t_n, t_{n+1}) \) and all outcomes \( \omega \) the time steps \( \Delta t(t) = \Delta t_n \) are constructed by the refinement criterion

\[
\Delta t(t) = T 2^{-m}, \text{ for some positive integer } m = m(t, \omega),
\]

\[
|\rho(t, \omega) (\Delta t(t))^2 < \text{constant},
\]

with an approximate error density function, \( \rho(t, \omega) = \rho_n(\omega) \), satisfying, for \( s, t \in [0, T] \) and all outcomes \( \omega \), the uniform upper and lower bounds

\[
\frac{c(TOL)}{TOL} \leq |\rho(s, \omega)| \leq C(TOL),
\]

\[
|\partial_{W(t)} \rho(s, \omega)| \leq C(TOL),
\]

for some positive functions \( c \) and \( C \), with \( TOL/c(TOL) \to 0 \) as \( TOL \to 0 \). Here \( \partial_{W(t)} Y \) denotes the Malliavin derivative which is the first variation of a process \( Y \).
with respect to a perturbation $dW(t)$ at time $t$ of the Wiener process, cf. [27] and [31].

**Lemma 2.1.** Suppose there are positive constants $k$ and $C$ and integer $m_0$ with the bounds

$$g \in C_{loc}^{m_0}(\mathbb{R}^d), \quad |\partial_\alpha g(x)| \leq C(1 + |x|^k), \quad \text{for all } |\alpha| \leq m_0,$$

$$E[|X(0)|^{2k+d+1} + |\overline{X}(0)|^{2k+d+1}] \leq C,$$

and

$$a \text{ and } b \text{ are bounded in } C^{m_0}([0, T] \times \mathbb{R}^d).$$

Assume that $\overline{X}$ is constructed by the forward Euler method with step sizes $\Delta t_n$ satisfying (2.2, 2.3) and the corresponding $\Delta W_n = W(t_{n+1}) - W(t_n)$ are generated by Brownian bridges, based on the stochastic time step algorithm in Section 3. Assume also that $\overline{X}(0) = X(0)$. Then there exists a sufficiently large integer $m_0$ such that

$$(2.4) \sup_{0 \leq t \leq T} \sqrt{E[|X(t) - \overline{X}(t)|^2]} = \mathcal{O} \left( \sqrt{\Delta t_{\sup}} \right) = \mathcal{O} \left( \sqrt{\frac{\text{TOL}}{c(TOL)}} \right) \to 0,$$

as $\text{TOL} \to 0$, where $\Delta t_{\sup} \equiv \sup_{n, \omega} \Delta t_n(\omega)$.

**Theorem 2.2.** Suppose that $a, b, g$ and $X$ satisfy the assumptions in Lemma 2.1 and $E[|X(0)|^{k_0}] \leq C$ for some $k_0 \geq 16$. Then the time discretization error has the expansion

$$E[g(X(T)) - g(\overline{X}(T))] = E \left[ \sum_{n=1}^{N} \rho(t_n, \overline{X})(\Delta t_n)^2 \right] + \mathcal{O} \left( \sqrt{\frac{\text{TOL}}{c(TOL)}} \left( \frac{C(TOL)}{c(TOL)} \right)^{8/k_0} \right) E \left[ \sum_{n=1}^{N} (\Delta t_n)^2 \right]$$

with computable leading order terms, where

$$\rho(t_n, \overline{X}) = \frac{1}{2} \left( \frac{\partial}{\partial t} a_k + \partial_j a_k a_j + \partial_i j a_k d_{ij} \right) \varphi_k(t_{n+1})$$

$$+ \frac{1}{2} \left( \frac{\partial}{\partial t} d_{km} + \partial_j d_{km} a_j + \partial_i j d_{km} d_{ij} + 2 \partial_j a_k d_{jm} \right) \varphi_{km}(t_{n+1})$$

$$+ \partial_j d_{km} d_{jr} \varphi_{kmr}(t_{n+1}),$$

and the terms in the sum of (2.6) are evaluated at the a posteriori known points $(t_n, \overline{X}(t_n))$, i.e.

$$\partial_x a \equiv \partial_x a(t_n, \overline{X}(t_n)), \quad \partial_x b \equiv \partial_x b(t_n, \overline{X}(t_n)), \quad \partial_x d \equiv \partial_x d(t_n, \overline{X}(t_n)).$$

Here $\varphi \in \mathbb{R}^d$ is the solution of the discrete dual backward problem

$$\varphi_i(t_n) = \partial_i c_j(t_n, \overline{X}(t_n)) \varphi_j(t_{n+1}), \quad t_n < T,$$

$$\varphi_i(T) = \partial_i g(\overline{X}(T)),$$

with

$$c_i(t_n, x) \equiv x_i + \Delta t_n a_i(t_n, x) + \Delta W_n^i b_i^i(t_n, x)$$
and its first and second variation

\begin{align}
\varphi'_{ikm}(t_n) &= \frac{\partial \varphi_i(t_n; \bar{X}(t_n) = x)}{\partial x_j}, \\
\varphi''_{ikm}(t_n) &= \frac{\partial \varphi'_{ik}(t_n; \bar{X}(t_n) = x)}{\partial x_m},
\end{align}

which satisfy for \( t_n < T \)

\begin{align}
\varphi'_{ik}(t_n) &= \partial_t c_i(t_n, \bar{X}(t_n)) \partial_k c_p(t_n, \bar{X}(t_n)) \varphi'_{jp}(t_{n+1}) \\
\varphi'_{ik}(T) &= \partial_k g(\bar{X}(T)),
\end{align}

and

\begin{align}
\varphi''_{ikm}(t_n) &= -\partial_t c_i(t_n, \bar{X}(t_n)) \partial_k \partial_p c_p(t_n, \bar{X}(t_n)) \partial_{\varphi'} \partial_{\varphi''} \bar{X}(t_n) \partial_{\varphi''} \bar{X}(t_n) \partial_{\varphi'} \bar{X}(t_n) \partial_{\varphi''} \bar{X}(t_n) \\
&\quad + \partial_t c_i(t_n, \bar{X}(t_n)) \partial_k \partial_p c_p(t_n, \bar{X}(t_n)) \partial_{\varphi''} \bar{X}(t_n) \partial_{\varphi'} \bar{X}(t_n) \partial_{\varphi''} \bar{X}(t_n) \\
&\quad + \partial_t c_i(t_n, \bar{X}(t_n)) \partial_{\varphi''} \bar{X}(t_n) \partial_{\varphi'} \bar{X}(t_n) \partial_{\varphi''} \bar{X}(t_n) \\
&\quad + \partial_t c_i(t_n, \bar{X}(t_n)) \partial_k \partial_p c_p(t_n, \bar{X}(t_n)) \partial_{\varphi'} \bar{X}(t_n) \partial_{\varphi''} \bar{X}(t_n) \\
&\quad + \partial_t c_i(t_n, \bar{X}(t_n)) \partial_k \partial_p c_p(t_n, \bar{X}(t_n)) \partial_{\varphi''} \bar{X}(t_n) \partial_{\varphi'} \bar{X}(t_n) \\
\varphi''_{ikm}(T) &= -\partial_{\varphi''} g(\bar{X}(T)),
\end{align}

respectively.

The previous result can be directly applied to the particular case of deterministic time steps. The deterministic time stepping algorithm uses the sample average of \( \rho(t_n, \bar{X}) \) in (2.6), that is the error density becomes

\begin{equation}
\bar{p}(t_n, \bar{X}) = \frac{1}{M_T} \sum_{j=1}^{M_T} \rho(t_n, \bar{X})(\omega_j),
\end{equation}

which has variance \( \text{Var}[\bar{p}_n] = \mathcal{O}(M_T^{-1}) \).

### 3. Adaptive Algorithms for SDE

This section presents two adaptive time stepping algorithms and analyzes the basic properties of the adaptive algorithms for the computational problem (1.1). These adaptive algorithms choose adaptively the number of realizations and the time steps, including dividing and merging, to bound the approximation error by a given error tolerance.

The computational error in (1.3) naturally separates into the time discretization error and the statistical error

\begin{equation}
E[g(X(T))] - \frac{1}{M} \sum_{j=1}^{M} g(\bar{X}(T; \omega_j))
\end{equation}

\begin{align}
= \left( E[g(X(T)) - g(\bar{X}(T))] \right) + \left( E[g(\bar{X}(T))] - \frac{1}{M} \sum_{j=1}^{M} g(\bar{X}(T; \omega_j)) \right)
\end{align}

\begin{align}
= \mathcal{E}_T + \mathcal{E}_S.
\end{align}

The time steps for the realizations of the approximate solution \( \bar{X} \) are determined from statistical approximations of the time discretization error, \( \mathcal{E}_T \), and the number, \( M \), of realizations of \( \bar{X} \) is determined from the statistical error, \( \mathcal{E}_S \). The statistical
error and the time discretization error are combined in order to bound the computational error (3.1). Therefore we split a given error tolerance TOL into a statistical tolerance, TOL\(_S\), and a time discretization tolerance, TOL\(_T\). The computational work is roughly \( O(N \cdot N) = O(TOL_T^{-1} TOL_S^{-2}) \), therefore we use
\[
(3.2) \quad TOL_T = \frac{1}{3} TOL \quad \text{and} \quad TOL_S = \frac{2}{3} TOL,
\]
by minimizing \( TOL_T^{-1} TOL_S^{-2} \) under the constraint \( TOL_T + TOL_S = TOL \).

**Control of the statistical error.** For \( M \) independent samples \( \{Y(\omega_j)\}_{j=1}^M \) of a random variable \( Y \), with \( E[|Y|^6] < \infty \), define the sample average \( A(Y; M) \) and the sample standard deviation \( S(Y; M) \) of \( Y \) by
\[
(3.3) \quad A(Y; M) = \frac{1}{M} \sum_{j=1}^M Y(\omega_j) \quad \text{and} \quad S(Y; M) = \left[ A(Y^2; M) - (A(Y; M))^2 \right]^{1/2}.
\]
Let \( \sigma_Y \equiv (E[|Y - E[Y]|^2])^{1/2} \) and consider the random variable
\[
Z_M \equiv \frac{\sqrt{M}}{\sigma_Y} (A(Y; M) - E[Y])
\]
with cumulative distribution function \( F_{Z_M}(x) \equiv P(Z_M \leq x), x \in \mathbb{R} \). Let \( \lambda \equiv (E[|Y - E[Y]|^3])^{1/3}/\sigma_Y < \infty \), then the Berry-Esseen theorem, cf. [10], gives the following estimate in the central limit theorem
\[
\sup_{x \in \mathbb{R}} |F_{Z_M}(x) - \Phi(x)| \leq \frac{3}{\sqrt{M}} \lambda^3
\]
for the rate of convergence of \( F_{Z_M} \) to the distribution function, \( \Phi \), of a normal random variable with mean zero and variance one, i.e.
\[
(3.4) \quad \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} s^2 \right) ds.
\]
Since in the examples below \( M \) is sufficiently large, i.e. \( M \gg 36 \lambda^6 \), the statistical error
\[
E_S(Y; M) \equiv E[Y] - A(Y; M)
\]
satisfies, by the Berry-Esseen theorem, the following probability approximation
\[
P \left( \left| E_S(Y; M) \right| \leq c_0 \frac{\sigma_Y}{\sqrt{M}} \right) \approx 2\Phi(c_0) - 1.
\]
In practice choose some constant \( c_0 \geq 1.65 \), so the normal distribution satisfies
\[
1 > 2\Phi(c_0) - 1 \geq 0.901
\]
and the event
\[
(3.5) \quad |E_S(Y; M)| \leq E_S(Y; M) \equiv c_0 \frac{S(Y; M)}{\sqrt{M}}
\]
has probability close to one, which involves the additional step to approximate \( \sigma_Y \) by \( S(Y; M) \), cf. [12]. Thus, in the computations \( E_S(Y; M) \) is a good approximation of the statistical error \( E_S(Y; M) \).

For a given \( TOL_S > 0 \), the goal is to find \( M \) such that \( E_S(Y; M) \leq TOL_S \). The following algorithm adaptively finds the number of realizations \( M \) to compute the sample average \( A(Y; M) \) as an approximation to \( E[Y] \). With probability close to one, depending on \( c_0 \), the statistical error in the approximation is then bounded by
TOLs, see Theorems 3.3 and 3.9. Technical reasons to prove a.s. convergence in Lemma 4.4 motivates our choice $M = 2^n$, $n \in \mathbb{N}$.

**Routine Monte-Carlo**($TOL, M_0; EY$)

Set the batch counter $m = 1$, $M[1] = M_0$ and $E_S[1] = +\infty$.

**Do while** ($E_S[m] > TOL$)

Compute $M[m]$ new samples of $Y$, along with the sample average $EY \equiv A(Y; M[m])$, the sample variance $S[m] \equiv S(Y; M[m])$ and the deviation $E_S[m+1] \equiv E_S(Y; M[m])$.

Compute $M[m+1]$ by \texttt{change}_M $M[m]$, $S[m]$, $TOL$; $M[m+1]$. Increase $m$ by 1.

**end-do**

**end of Monte-Carlo**

**Routine change_M** ($M_{in}, S_{in}, TOLs; M_{out}$)

\begin{align}
M^* &= \min \left\{ \text{integer part} \left( \frac{c_0 S_{in}}{TOLs} \right)^2, \ M_{CH} \times M_{in} \right\} \\
N &= \text{integer part} \left( \log_2 M^* \right) + 1 \\
M_{out} &= 2^n.
\end{align}

**end of change_M**

Here, $M_0$ is a given initial value for $M$, and $M_{CH} > 1$ is a positive integer parameter introduced to avoid a large new number of realizations in the next batch due to a possibly inaccurate sample standard deviation $S[m]$. Indeed, $M[m+1]$ cannot be greater than $M_{CH} \times M[m]$.

**Control of the time discretization error.** For given time nodes $0 = t_1 < \ldots < t_{N+1} = T$, let the piecewise constant mesh function $\Delta t$ be determined by

$$\Delta t(\tau) = \Delta t_n \text{ for } \tau \in [t_n, t_{n+1}) \text{ and } n = 1, \ldots, N.$$  

Then the number of time steps that corresponds to a mesh $\Delta t$, for the interval $[0,T]$, is given by

$$N(\Delta t) \equiv \int_0^T \frac{1}{\Delta t(\tau)} d\tau.$$ 

Consider, for $\tau \in [t_n, t_{n+1})$ and $n = 1, \ldots, N$, the piecewise constant function

$$\rho(\tau) \equiv \text{sign}(\rho_n) \max (|\rho_n|, \delta)$$

which measures the density of the time discretization error, where $\rho_n = \rho(t_n, X)$ is defined by (2.6) for the stochastic time stepping algorithm or $\rho_n = \rho(t_n, X)$ from (2.13) for the deterministic time stepping algorithm. Here we have $\text{sign}(x) \equiv x/|x|$, for $x \in \mathbb{R} - \{0\}$, $\text{sign}(0) \equiv 1$, and we use the positive parameter, $\delta$, satisfying $\delta + TOL^2/\delta \to 0$, in order to guarantee that $\Delta t_{\text{sup}} \to 0$ and that $\rho(t)$ and $\rho_n$ have the same limit, as $TOL_T \to 0$. From now on, with a slight abuse of notation, $\rho(t_n) = \rho_n$ denotes the modified density (3.8).

Following the error expansion in Theorem 2.2, the time discretization error is approximated by

$$|E_T| = |E[g(X(T)) - g(\overline{X}(T))]| \lesssim E \left[ \sum_{n=1}^N r_n \right]$$
using the error indicator, $r_n$, defined by
\begin{equation}
    r_n \equiv |\rho(t_n)| \Delta t_n^2
\end{equation}
with the modified error density defined by (3.8). To motivate the adaptivity procedure for the time partition, let us now formulate an optimal choice of the time steps by minimizing the expected computational work subject to the accuracy constraint
\begin{equation}
    E \left[ \sum_{n=1}^{N} r_n \right] \leq \text{TOL}_T.
\end{equation}
More precisely, solve
\begin{equation}
    \min E[N(\Delta t)], \text{ such that } E \left[ \sum_{n=1}^{N} r_n \right] \leq \text{TOL}_T,
\end{equation}
where $\mathcal{K}$ is the feasible set for the mesh function $\Delta t$ and $N(\Delta t)$ is the corresponding number of steps. The optimal choice of time steps in $\mathcal{K}$ is based on the given density $\rho_n[k]$, which is piecewise constant on the mesh $\Delta t[k], k = 1, 2, \ldots$. The choice of $\mathcal{K}$ determines either deterministic time steps or stochastic time steps. For example, if we let
\[ \mathcal{K} \equiv \left\{ \Delta t \in L^2_{\Delta t}((0,T)) : \Delta t \text{ is deterministic, positive and piecewise constant on } \Delta t[k] \right\} \]
then the objective function in (3.12) becomes deterministic and a standard application of a Lagrange multiplier shows that the minimizer of the problem (3.12) satisfies
\begin{equation}
    E[r_n] = \text{constant}, \text{ for all time steps } n,
\end{equation}
which sets the basis for the refinement procedure with deterministic time steps. On the other hand, letting
\[ \mathcal{K} \equiv \left\{ \Delta t \in L^2_{\Delta t \times P}((0,T) \times \Omega)) : \Delta t \text{ is stochastic, positive and piecewise constant on } \Delta t[k](\omega) \right\} \]
leads to
\begin{equation}
    r_n(\omega) = \text{constant}, \text{ for all time steps } n \text{ and for all realizations } \omega,
\end{equation}
which sets the basis for the refinement procedure with stochastic time steps.

Thus, the adaptive algorithm with stochastic time steps uses the optimal conditions (3.11) and (3.14) to construct the mesh which may be different for each realization. On the other hand, the adaptive algorithm with deterministic time steps uses the optimal conditions (3.11) and (3.13) to construct the mesh which is the same for all realizations.

### 3.1. Convergence rates for stochastic time steps.

The optimal conditions (3.11), (3.14) and the restriction (2.2) motivate that the goal of the adaptive algorithm is to construct a time partition $\Delta t$ of $[0,T]$ for each realization such that
\begin{equation}
    s_2 \frac{\text{TOL}_T}{E[N]} \leq |\rho_n| \Delta t_n^2 \leq s_1 \frac{\text{TOL}_T}{E[N]}, \quad \forall \ n = 1,\ldots,N
\end{equation}
where $s_1$ and $s_2$ are given constants satisfying $0 < s_2 < s_1$. Note that in practice the quantity $E[N]$ is not known and we can only estimate it by a sample average.
\( \mathcal{A}(N; M) \) from the previous batch of realizations. The statistical error \( |E[N] - \mathcal{A}(N; M)| \) is then bounded by \( E_S(N; M) \), with probability close to one, by the same argument as in (3.5). The remainder of this section analyzes an adaptive algorithm based on (3.15) with respect to stopping, accuracy and efficiency.

Let \( \overline{N}[j] \equiv \mathcal{A}(N; M[j]) \) be the sample average of the final number of time steps in the \( j \)-th batch of \( M[j] \) numbers of realizations. To achieve (3.15), start with an initial partition \( \Delta t[1] \) and then specify iteratively a new partition \( \Delta t[k + 1] \), from \( \Delta t[k] \), using the following dividing and merging strategy: for each realization in the \( m \)-th batch and for each time step \( n = 1, 2, \ldots, N[k] \),

\[
\text{(3.16) if } r_n[k] \geq s_1 \frac{TOL_T}{N[m - 1]} \text{, then divide } \Delta t_n[k] \text{ into } H \text{ substeps}
\]

\[
\text{(3.17) elseif } \max (r_n[k], r_{n+1}[k]) \leq s_2 \frac{TOL_T}{N[m - 1]} \text{, then merge } \Delta t_n[k] \text{ and } \Delta t_{n+1}[k] \text{ into one step, and increase } n \text{ by 1,}
\]

\[\text{else let the new step be the same as the old one.}\]

\[\text{endif}\]

Here \( H \) is a given integer greater than 1, which bounds the increment of the number of time steps from one iteration to the next. The following analysis, for fixed \( H \), can easily be extended to bounded and varying \( H \).

The dividing and merging strategy (3.16, 3.17) motivates the following stopping criteria: for each realization of the \( m \)-th batch

\[
\text{(3.18) if } \left( r_n[k] < S_1 \frac{TOL_T}{N[m - 1]} , \forall n = 1, \ldots, N[k] \right) \text{ and}
\]

\[
\text{(3.19) } \left( \max (r_n[k], r_{n+1}[k]) > S_2 \frac{TOL_T}{N[m - 1]}, \forall n = 1, \ldots, N[k] - 1 \right) \text{ then stop the dividing-merging process.}
\]

Here \( S_1 \) and \( S_2 \) are given constants satisfying \( 0 < S_2 < s_2 < s_1 < S_1 \). The combination of (3.9), (3.11) and the upper bound (3.18) asymptotically guarantee a given level of accuracy,

\[ |E[g(X(T)) - g(\overline{X}(T))]| < S_1 TOL_T, \]

while the lower bound (3.19) implies efficiency by almost optimal time steps. When almost all \( r_n \) satisfy \( r_n < S_1 TOL_T/N \), the reduction of the error may be slow. Therefore the algorithm stops if \( \max_n r_n < S_1 TOL_T/N \) and \( \min_n (\max(r_n, r_{n+1})) > S_2 TOL_T/N \). What is the right choice for the constants \( S_2 < s_2 < s_1 < S_1 \) of the dividing and merging strategy? Before determining sufficient conditions for the constants, we will describe the algorithm with stochastic time steps in detail.

**Remark 3.1.** In practice, our numerical tests show that

\[ \left| E \left[ g(X(T)) - \mathcal{A}(g(\overline{X}(T)); M) \right] \right| \approx \frac{s_1}{2} < \frac{1}{3} S_1, \]

so we choose \( s_1 = 2 \).

**The adaptive algorithm.** The adaptive, stochastic time stepping algorithm has a structure similar to the Monte-Carlo routine. First we split the specified error tolerance by (3.2) and the outer loop computes the batches of realizations of \( \overline{X} \), until
an estimate for the statistical error (3.5) is below the tolerance, TOLs. In the inner
loop for each realization, we apply our dividing and merging strategy (3.16, 3.17)
to a given initial time mesh iteratively until the approximate solution is sufficiently
resolved, in other words, until the approximate error density and the time steps
satisfy the stopping criteria (3.18, 3.19) with a given time discretization tolerance
TOLT. This procedure needs to sample the Wiener process, W, on finer or coarser
partitions, given its values on coarser or finer ones respectively. Therefore, the use
of Brownian bridges is natural to preserve the required independence between the
Wiener increments.

Now we are ready for the detailed definition of the adaptive algorithm with
stochastic steps:

**Algorithm S**

**Initialization** Choose:
1. an error tolerance, TOL ≡ TOLs + TOLT,
2. a number, N[1], of initial uniform steps Δt[1] for [0, T] and set N = N[1],
3. a number, M[1], of initial realizations,
4. an integer H ≥ 2 for the number of subdivisions of a refined time step, a
   number, s1 = 2 in (3.16) and a rough estimate of c in (3.20) to compute
   s2, S1, S2 using (3.21, 3.22, 3.23), and
5. a constant c0 ≥ 1.65 and an integer MCH ≥ 2 to determine the number of
   realizations in (3.6).

Set the two iteration counters, m for batches and k for time refinement levels,
to 1. Set the stochastic error Es[1] = +∞.

**Do while** ( Es[m] > TOLs )

**For** realizations j = 1, ..., M[m]

Set k = 1 and r[1] = +∞.

**Start** with the initial partition Δt[1] and generate ΔW[1].

**Do while** ( r[k] violates the stopping (3.18, 3.19) )

Compute the approximation X[k] and the error indicator r[k] in
(3.10) using the error density (2.6) on Δt[k] with the known
Wiener increments ΔW[k].

**If** ( r[k] violates the stopping (3.18, 3.19) )

**For** time steps i = 1, ..., N[k]

Do the dividing and merging process (3.16, 3.17) to compute
Δt[k+1] from Δt[k] and compute ΔW[k+1] from ΔW[k]
using Brownian bridges.

**end-for**

**end-if**

Increase k by 1.

**end-do**

Compute the sample average Eg ≡ A (g(X(T)); M[m]), the sample
standard deviation Sm[m] ≡ S(g(X(T)); M[m]) and the a posteriori bound
for the statistical error Es[m] ≡ Es(g(X(T)), M[m]) in (3.5).

**if** ( Es[m] > TOLs )

Compute M[m+1] by change M(M[m], S[m], TOLs; M[m+1]), cf.
(3.6), and update N = A (N; M[m]), where the random variable N is
the final number of time steps on each realization.

\textbf{end-if}

Increase \( m \) by 1.

\textbf{end-do}

Accept \( E_g \) as an approximation of \( E[g(X(T))] \), since the estimate of the computational error is bounded by TOL.

Note that the dividing and merging strategy in (3.16, 3.17) may do infinite loops at some time steps if the constant \( s_2 \) is too close to \( s_1 \). Clearly, we want to avoid the case where the time step \( \Delta t(t)[k] \) is divided but in the next iteration, \( \Delta t(t)[k+1] \) is merged. An important ingredient to avoid such instability is the proof in Section 4 that the error density converges a.s. as the maximal time step tends to zero. Therefore, for sufficiently small initial meshes, which may depend on the realization and small tolerance, there exists a positive constant \( c \) such that for all \( t \in [0, T] \) and for each realization

\[ c \leq \frac{|\rho(t)[k+1]|}{|\rho(t)[k]|} \leq c^{-1}. \]  

\textbf{Stopping of the adaptive algorithm.} The right choice of the parameters \( S_2 < s_2 < s_1 < S_1 \) is explained by

\textbf{Theorem 3.2 (Stopping).} Suppose the assumptions of Theorem 2.2 hold and the adaptive algorithm uses the strategy (3.8), (3.16, 3.17), (3.18, 3.19). Assume that (3.20) holds with \( c \geq 4^{-1} \), and that

\begin{align*}
(3.21) & \quad s_2 < \frac{c}{H^2}s_1, \\
(3.22) & \quad S_2 < cs_2, \\
(3.23) & \quad S_1 > \frac{1}{c}s_1.
\end{align*}

Then for each realization the adaptive dividing-merging process stops by the stopping criteria (3.18, 3.19), after a finite number of operations.

\textbf{Proof.} If \( s_2 \) is too close to \( s_1 \), the algorithm may be unstable in the sense that a time step is first divided and then in the next iteration the step is merged, or similarly a step is first merged and then divided. To analyze this unstable situation assume that (3.20) holds. Merging two steps, where at least one has just been divided requires that

\[ r_i[k] \geq s_1 \frac{\text{TOL}_T}{N}, \quad \text{and} \quad \max (r_i[k+1], r_{i+1}[k+1]) \leq s_2 \frac{\text{TOL}_T}{N} \]

with \( \Delta t_i[k+1] = \Delta t_i[k]/H \), so that by (3.10) and (3.20)

\[ s_2 \geq \frac{N}{\text{TOL}_T} \frac{|\rho[k+1]|}{|\rho[k]|} \left( \frac{\Delta t[k+1]}{\Delta t[k]} \right)^2 r[k] \]

\[ \geq \frac{N}{\text{TOL}_T} \frac{1}{H^2} \left( \frac{1}{s_1 \frac{\text{TOL}_T}{N}} \right) = \frac{c}{H^2}s_1. \]

Similarly, dividing following merging requires \( s_2 \geq cs_1/4 \), which is less demanding than the smaller (3.25). Therefore the condition (3.21) avoids the dividing-merging instability.
The next step is to understand the evolution of
\[ r_{\text{max}}[k] \equiv \max_i r_i[k], \quad \text{and} \quad r_{\text{min}}[k] \equiv \min_i \max(r_i[k], r_{i+1}[k]) \]
for iterations \( k = 1, 2, \ldots \). It is advantageous if \( r_{\text{max}} \) decreases and \( r_{\text{min}} \) increases quickly to levels close to the bounds \( s_1 \text{TOL}_T/N \) and \( s_2 \text{TOL}_T/N \), respectively. Indeed, there holds
\[ r_{\text{max}}[k + 1] < c^{-1} \underbrace{H^2 r_{\text{max}}[k]}_{H^2}, \quad (3.26) \]
provided \( r_{\text{max}}[k] > H^2 s_1 \text{TOL}_T/N \), and
\[ r_{\text{min}}[k + 1] > 4c r_{\text{min}}[k], \quad (3.27) \]
provided \( r_{\text{min}}[k] < s_2 \text{TOL}_T/(4N) \). In other words, the error indicators \( r_{\text{max}} \) and \( r_{\text{min}} \) decrease and increase, respectively, with a constant factor away from the bounds
\[ r_{\text{max}} > H^2 s_1 \text{TOL}_T/N \quad \text{and} \quad r_{\text{min}} < s_2 \text{TOL}_T/(4N). \]
If \( r_{\text{max}} \) is close to the dividing bound, i.e. if
\[ r_{\text{max}}[k] \leq H^2 s_1 \text{TOL}_T/N, \quad (3.28) \]
then in the next iteration \( r_{\text{max}}[k + 1] \leq c^{-1}s_1 \text{TOL}_T/N \), since a divided step cannot be merged by (3.21). Therefore \( r_{\text{max}}[k + 1] \) enters the stopping region, \( r_{\text{max}} < S_1 \text{TOL}_T/N \), provided \( S_1 > s_1/c \) and the condition (3.28) remains satisfied for the later iterations provided \( c^{-1} \leq H^2 \).

Similarly, if \( r_{\text{min}} \) is close to the merging bound, i.e. if
\[ r_{\text{min}}[k] \geq s_2 \text{TOL}_T/(4N), \quad (3.29) \]
then \( r_{\text{min}}[k + 1] \geq cs_2 \text{TOL}_T/N \). Therefore \( r_{\text{min}} \) belongs to the stopping region, \( r_{\text{min}}[k + 1] > S_2 \text{TOL}_T/N \), provided \( S_2 < cs_2 \) and the condition (3.29) remains satisfied for the later iterations provided \( c \geq 4^{-1} \). When both \( r_{\text{max}} \) and \( r_{\text{min}} \) have entered their stopping regions, the dividing-merging process stops, and the algorithm continues with the next realization. \( \square \)

**Accuracy of the adaptive algorithm.** The adaptive algorithm guarantees that the estimated error is bounded by \( S_1 \text{TOL}_T + \text{TOL}_S = (S_1/3 + \frac{2}{3}) \text{TOL} \). The next question is whether the true error is bounded by \( (S_1/3 + \frac{2}{3}) \text{TOL} \) asymptotically. Using the upper bound (3.18) of the error indicators and the a.s. convergence of \( \rho \), proved in Section 4, the approximate error has the estimate

**Theorem 3.3 (Accuracy).** Suppose that the assumptions of Theorem 2.2 hold. Then the adaptive algorithm (3.8), (3.16, 3.17), (3.18, 3.19) satisfies, for any constant \( c_0 > 0 \) defined in (3.6),
\[ \liminf_{\text{TOL} \to 0} \frac{P}{\text{TOL}} \left( \frac{|E[g(X(T))] - A(g(X(T)); M)|}{s_1/3 + 2/3} \right) \geq \int_{-c_0}^{c_0} \frac{e^{-x^2/2}}{\sqrt{2\pi}} \, dx. \]
Proof. To simplify the proof, we split the fraction on the left hand side of (3.30) into a statistical error and a time discretization error, for all $s \in \mathbb{R}_+$,

\begin{equation}
(3.31) \\
\liminf_{\text{TOL} \to 0} P \left( \frac{|E[g(X(T))] - A(g(X(T));M)|}{\text{TOL}} \leq S_1 + \frac{2}{3} \right) \\
\geq \liminf \ P \left( \frac{|E[g(X(T)) - g(X(T))]|}{\text{TOL}} + \frac{|E[g(X(T))] - A(g(X(T));M)|}{\text{TOL}} \leq S_1 + \frac{2}{3} \right) \\
\geq \liminf \ P \left( \frac{|E[g(X(T)) - g(X(T))]|}{\text{TOL}} \leq s \\
\hspace{1cm} \text{and} \quad \frac{|E[g(X(T))] - A(g(X(T));M)|}{\text{TOL}} \leq S_1 + \frac{2}{3} - s \right) \\
= \liminf \ P \left( \frac{|E[g(X(T)) - g(X(T))]|}{\text{TOL}} \leq s \right) \times \\
P \left( \frac{|E[g(X(T))] - A(g(X(T));M)|}{\text{TOL}} \leq S_1 + \frac{2}{3} - s \right).
\end{equation}

The time discretization error. When the adaptive algorithm stops, the error estimate (2.1) and the stopping bound for $\Delta t$ (3.18) imply

\[ |E[g(X(T)) - g(X(T))]| \leq E \left[ \sum_{i=1}^{N} \Delta t_i \int_{t_{i-1}}^{t_i} |\bar{\rho}(\tau)| d\tau \right] \]
\[ < \sqrt{S_1 \text{TOL}_T} \quad E \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} \frac{|\bar{\rho}(\tau)|}{|\rho(\tau)|} d\tau \right]. \]

Apply the dominated convergence theorem to the right hand side of the above equation to get

\[ \limsup_{\text{TOL} \to 0} \frac{|E[g(X(T)) - g(X(T))]|}{\text{TOL}_T} \leq \sqrt{S_1} E \left[ \limsup_{\text{TOL} \to 0} \left( \frac{\int_{0}^{T} |\bar{\rho}(\tau)| / \sqrt{|\rho(\tau)|} d\tau}{\sqrt{\text{TOL}_T N}} \right) \right] \]
\[ \leq \sqrt{S_1} E \left[ \frac{1}{\sqrt{\liminf_{\text{TOL} \to 0} \text{TOL}_T N}} \limsup_{\text{TOL} \to 0} \int_{0}^{T} \frac{|\bar{\rho}(\tau)|}{\sqrt{|\rho(\tau)|}} d\tau \right]. \]

Corollary 4.3 in Section 4 shows that both $\bar{\rho}$ and $\rho$ converge a.s. to the limit $\hat{\rho}$, so that dominated convergence, based on (3.8), establishes

\begin{equation}
(3.32) \\
\lim_{\text{TOL} \to 0} \int_{0}^{T} \frac{|\bar{\rho}(\tau)|}{\sqrt{|\rho(\tau)|}} d\tau = \int_{0}^{T} \sqrt{|\bar{\rho}(\tau)|} d\tau \text{ a.s.}
\end{equation}

since $|\bar{\rho}| / \sqrt{|\rho|}$ is uniformly bounded in $L^1$ using the facts (2.5) and (3.8).

Rewrite the inequality (3.18) as

\[ \sqrt{|\rho|} < \sqrt{S_1 \text{TOL}_T N} \frac{1}{\sqrt{|m-1|} \Delta t_i}, \]
integrate both sides, use the definition (3.7) and take the sample average of $M[m]$ independent samples to obtain

$$
(3.33) \quad \overline{N}[m] \geq \frac{\sqrt{N[m - 1]}}{\sqrt{S_1TOL_T}} A \left( \int_0^T \sqrt{\rho(\tau)} d\tau; M[m] \right).
$$

Take the limit in (3.33) and use that Corollary 4.3 implies

$$
\lim_{TOL \to 0} A \left( \int_0^T \sqrt{\rho(\tau)} d\tau; M[m] \right) = E \left[ \int_0^T \sqrt{\rho(\tau)} d\tau \right]
$$
to get

$$
\sqrt{\liminf_{TOL \to 0} TOL_T N} \geq \frac{1}{\sqrt{S_1}} E \left[ \int_0^T \sqrt{\rho(\tau)} d\tau \right].
$$

and consequently

$$
\limsup_{TOL \to 0} \frac{|E \left[ g(X(T)) - g(X(T)) \right]|}{TOL} \leq S_1.
$$

Since $TOL_T = \frac{1}{TOL}$, this deterministic limit implies that for all $s > S_1/3$

$$
(3.34) \quad \liminf_{TOL \to 0} P \left( \frac{|E \left[ g(X(T)) - g(X(T)) \right]|}{TOL} \leq s \right) = 1.
$$

The statistical error. Use that in (3.6) the number of realizations is

$$
M \geq \left( \frac{c_0 S(g(X(T)); M)}{TOL_s} \right)^2
$$
to rewrite $1/TOL \leq (2/3c_0)\sqrt{M}/S(g(X(T)); M)$ by (3.2), so that

$$
P \left( \frac{|E \left[ g(X(T)) \right] - A(g(X(T)); M)|}{TOL} \leq \frac{S_1}{3} + \frac{2}{3} - s \right) \geq P \left( \frac{|E \left[ g(X(T)) \right] - A(g(X(T)); M)|}{S(g(X(T)); M)} \sqrt{M} \leq \frac{3c_0}{2} \left( \frac{S_1}{3} + \frac{2}{3} - s \right) \right).
$$

Let $\chi \equiv (E[g(X(T)) - A(g(X(T)); M)) / S(g(X(T)); M)$ and $\overline{\chi} \equiv (E[g(X(T)) - A(g(X(T)); M)) / S(g(X(T)); M)$. Then, write

$$
\overline{\chi} = (\chi - \chi) + \chi
$$
and note that $\overline{\chi} - \chi \to 0$, weakly as $\Delta t \to 0$, and that $S(g(X(T)); M) \to \sigma_g(X(T))$, strongly as $M \to \infty$. Therefore the central limit theorem shows that $\lim_{TOL \to 0} \overline{\chi}$ is normally distributed with mean zero and variance one and consequently

$$
(3.35) \quad \sup_{s > \frac{1}{3}} \lim_{TOL \to 0} P \left( \frac{|E \left[ g(X(T)) \right] - A(g(X(T)); M)|}{S(g(X(T)); M)} \sqrt{M} \leq \frac{3c_0}{2} \left( \frac{S_1}{3} + \frac{2}{3} - s \right) \right)
$$

$$
= \sup_{s > \frac{1}{3}} \int_0^{3c_0/2} \left( \frac{s}{\sqrt{2\pi}} \right) e^{-x^2/2} dx = \int_{-c_0}^{c_0} e^{-x^2/2} dx.
$$

Finally the combination of (3.31), (3.34) and (3.35) proves the theorem. $\square$
Efficiency of the adaptive algorithm. The minimal expected number of time steps in the class of stochastic time steps $\Delta t(t, \omega)$ depends on the stochastic data and the individual realizations $X$ through the constraint $E[g(X(T)) - g(\bar{X}(T))] = TOL_T$. The conditions (3.11) and (3.14) imply that the optimal expected number of time steps, $E[N_S]$, satisfies (cf. [31])

$$E[N_S] = \frac{1}{TOL_T} \left( E \left[ \int_0^T \sqrt{\rho(\tau)} d\tau \right] \right)^2. \quad (3.36)$$

On the other hand, for the constant time steps $\Delta t = constant$, the number of steps, $N_C$, to achieve $\sum_{i=1}^{N} |\rho_i| \Delta t_i^2 = TOL_T$, becomes

$$N_C = \frac{T}{TOL_T} \left( \int_0^T E[|\rho(\tau)|] d\tau \right). \quad (3.37)$$

Therefore, Jensen’s inequality shows that the adaptive method with stochastic time steps uses fewer time steps than the method with a constant time step, i.e.,

$$E[N_S] \leq N_C. \quad (3.38)$$

The following theorem uses the lower bound (3.19) of the error indicators to show that the algorithm (3.16 - 3.19) generates a mesh which is optimal, up to a multiplicative constant $C = 4/S_2$.

**Theorem 3.4 (Efficiency).** Suppose that the assumptions of Theorem 2.2 hold. Then there exists a constant $C > 0$, asymptotically bounded by $4/S_2$, such that, using a final batch of $M[m]$ realizations, the final sample average

$$\bar{N}[m] \equiv A(N; M[m])$$

of the number of adaptive steps of the algorithm (3.16, 3.17), (3.18, 3.19) satisfies

$$TOL_T \frac{\bar{N}[m]^2}{\bar{N}[m - 1]} < C \left( A \left( \int_0^T \sqrt{\rho(\tau)} d\tau; M[m] \right) \right)^2, \quad (3.39)$$

and asymptotically

$$\limsup_{TOL_T \to 0} TOL_T E[N] \leq \frac{4}{S_2} \left( E \left[ \int_0^T \sqrt{\rho(\tau)} d\tau \right] \right)^2. \quad (3.40)$$

**Proof.** For each realization, Theorem 3.3 shows that the adaptive dividing-merging stops, by the stopping criteria (3.18, 3.19), and the set of time steps satisfies,

$$\{\Delta t_i : i = 1, \ldots, N\} = D \cup M$$

where

$$D \equiv \{\Delta t_i : S_2 \frac{TOL_T}{\bar{N}[m - 1]} < |\rho_i| \Delta t_i^2 < S_1 \frac{TOL_T}{\bar{N}[m - 1]}, \ i = 1, \ldots, N\}$$

and

$$M \equiv \{\Delta t_i : |\rho_i| \Delta t_i^2 \leq S_2 \frac{TOL_T}{\bar{N}[m - 1]}, \ i = 1, \ldots, N\}. \quad (3.41)$$

Condition (3.19) shows that there is no successive pair of time steps which belongs to the set $M$. This means that the number of steps $N_M$ in $M$ is at most the same as the half of the number of steps, $N$, i.e. $N_M \leq \lceil \frac{N}{2} \rceil$ where $\lceil a \rceil$ rounds the number
a to the nearest integer greater than or equal to a. Then for each realization, the number of steps satisfies

\[(3.43) \quad N[m] = N_D + N_M \leq N_D + \left\lceil \frac{N[m]}{2} \right\rceil \leq 2N_D[m] + 1 \leq 3N_D[m].\]

Take the sample average of both sides of \(M[m]\) independent samples to get

\[(3.44) \quad \frac{1}{3}N[m] \leq N_D[m].\]

The time steps \(\Delta t_i \in D\) satisfy

\[(3.45) \quad \sqrt{S_2 \frac{TOL_T}{N[m-1]}} \frac{1}{\Delta t_i} < \sqrt{|\rho|},\]

and by integrating both sides with the definition (3.7) and taking the sample average, we get

\[(3.46) \quad \sqrt{S_2 \frac{TOL_T}{N[m-1]}} N_D[m] < A \left( \int_0^T \sqrt{|\hat{\rho}(\tau)|} d\tau; M \right).\]

The estimate (3.44) shows

\[(3.47) \quad \sqrt{S_2 \frac{TOL_T}{N[m-1]}} N_D[m] \leq \sqrt{S_2 \frac{TOL_T}{N[m-1]}} N_D[m],\]

which together with (3.46) proves (3.39) where \(C = 9/S_2\). A more precise use of the last inequality in (3.43) shows \(C \rightarrow 4/S_2\), as \(N \rightarrow \infty\).

To prove (3.40), apply dominated convergence theorem to (3.39) and use Corollary 4.3 to get

\[(3.48) \quad \limsup_{TOL_T \to 0} TOL_T N \leq \frac{4}{S_2} \left( E \left( \sqrt{\int_0^T \sqrt{|\hat{\rho}(\tau)|} d\tau} \right) \right)^2.\]

Integrate (3.45) and take the expected value to obtain

\[(3.49) \quad \limsup_{TOL_T \to 0} TOL_T E[N[m]] \leq \sqrt{\frac{4}{S_2}} \limsup_{TOL_T \to 0} E \left( \sqrt{TOL_T N[m-1]} \int_0^T \sqrt{|\hat{\rho}(\tau)|} d\tau \right).\]

Dominated convergence, the a.s. convergence of error density, \(\rho \to \hat{\rho}\), and the bound (3.47) imply

\[(3.50) \quad \limsup_{TOL_T \to 0} TOL_T E[N] \leq \sqrt{\frac{4}{S_2}} \left( \limsup_{TOL_T \to 0} \left( \sqrt{TOL_T N[m-1]} \int_0^T \sqrt{|\hat{\rho}(\tau)|} d\tau \right) \right)^2 \leq \sqrt{\frac{4}{S_2}} \left( \limsup_{TOL_T \to 0} \sqrt{TOL_T N[m-1]} \int_0^T \sqrt{|\hat{\rho}(\tau)|} d\tau \right)^2 \leq \frac{4}{S_2} \left( E \left( \int_0^T \sqrt{|\hat{\rho}(\tau)|} d\tau \right) \right)^2,

which proves (3.40).
Remark 3.5. The error density condition (3.20) also implies constraints on the optimal mesh, for instance $H = 2$ and the assumption $\frac{1}{2}(\rho_i[k] + \rho_{i+1}[k]) = \rho(t_i)[k-1]$ show that

$$2c - 1 \leq \left| \frac{\rho_{i+1}[k]}{\rho_i[k]} \right| \leq 2c^{-1} - 1. \tag{3.49}$$

Remark 3.6. If the number of elements in each refinement iteration increase only very slowly, the total work including all refinement levels becomes proportional to the product of the number of steps in the finest mesh times the number of refinement levels, $J$, which satisfies $\min \Delta t = H^{-J}T/N[1] = O(TOL)$, so that

$$J = O(\log(TOL^{-1})) = O(\log N).$$

Therefore the average of the total number of time steps is essentially bounded by $E[N \log N]$.

3.2. Convergence rates for deterministic time steps. The main difference between the stochastic and the deterministic time step algorithm is that the additional work to find the optimal deterministic steps requires a much smaller number, $M_T$, of realizations than total number of realizations $M$. The approximation of the time discretization error in the right hand side of (3.9) can be separated into two parts

$$E \left[ \sum_{n=1}^{N} r_n \right] \leq A \left( \sum_{n=1}^{N} r_n ; M_T \right) + E \left[ \sum_{n=1}^{N} r_n \right] - A \left( \sum_{n=1}^{N} r_n ; M_T \right), \tag{3.50}$$

where the second error term in the right hand side of (3.50) is approximated by

$$\left| E \left[ \sum_{n=1}^{N} r_n \right] - A \left( \sum_{n=1}^{N} r_n ; M_T \right) \right| \lesssim E_{TS} \equiv c_0 \frac{S \left( \sum_{n=1}^{N} r_n ; M_T \right)}{\sqrt{M_T}} \tag{3.51}$$

and the first term defines $E_{TT} \equiv A \left( \sum_{n=1}^{N} r_n ; M_T \right)$. Then for a given $TOL_T > 0$, the goal is to construct a partition $\Delta t$ of $[0, T]$, with as few time steps and realizations $M_T$ as possible, such that

$$E_{TT} + E_{TS} \leq TOL_T.$$

To this end, first split the time discretization tolerance $TOL_T$ in two positive parts $TOL_{TT}$ and $TOL_{TS}$ for $E_{TT}$ and $E_{TS}$ respectively. The statistical error of the time discretization using the density (2.13) is $O(\Delta t_{\text{max}}/\sqrt{M_T})$. Therefore the percentage of the tolerance, $TOL$, devoted to the control of the statistical time discretization error can be arbitrary small as $TOL \to 0$. In practice we choose

$$TOL_{TT} = \frac{2}{3} TOL_T = \frac{2}{9} TOL \text{ and } TOL_{TS} = \frac{1}{3} TOL_T = \frac{1}{9} TOL. \tag{3.52}$$

The control of the statistical time discretization error determines the number of realizations $M_T$ to ensure a reliable choice of the time discretization in the deterministic time stepping algorithm.

Take into account (3.8) and define the error density, $\bar{\rho}$, by

$$E_{TT} \equiv A \left( \sum_{n=1}^{N} r_n ; M_T \right) = \sum_{n=1}^{N} \bar{\rho}_n \Delta t_n^2.$$
Following the optimal conditions (3.11) and (3.13), the goal of the adaptive algorithm described below is to construct a time partition $\Delta t$ of $[0, T]$ such that

$$d_2 \frac{TOL_{TT}}{N} \leq |\bar{\rho}_n| \Delta t_n^2 \leq d_1 \frac{TOL_{TT}}{N}, \quad \forall \ n = 1, \ldots, N$$

(3.53)

where $d_1$ and $d_2$ are given constants satisfying $0 < d_2 < d_1$. This section analyzes an adaptive algorithm based on (3.53), with respect to stopping, accuracy and efficiency.

To achieve (3.53), start as in Section 3.1 with an initial partition $\Delta t[1]$ and then specify iteratively a new partition $\Delta t[k+1]$, from $\Delta t[k]$, using the following dividing and merging strategy: for $n = 1, 2, \ldots, N[k]$ let $\bar{r}_n \equiv |\bar{\rho}_n| \Delta t_n^2$ and

$$\text{if } \bar{r}_n[k] \geq d_1 \frac{TOL_{TT}}{N[k]}, \text{ then divide } \Delta t_n[k] \text{ into } H \text{ substeps}$$

(3.54)

$$\text{elseif } \max(\bar{r}_n[k], \bar{r}_{n+1}[k]) \leq d_2 \frac{TOL_{TT}}{N[k]}, \text{ then merge } \Delta t_n[k] \text{ and } \Delta t_{n+1}[k] \text{ into one step, and increment } n \text{ by } 1,$$

else let the new step be the same as the old one.

endif

until the following stopping criteria is satisfied:

$$\text{if } \left( \bar{r}_n[k] < D_1 \frac{TOL_{TT}}{N[k]}, \forall \ n = 1, \ldots, N[k] \right) \text{ and }$$

(3.56)

$$\left( \max(\bar{r}_n[k], \bar{r}_{n+1}[k]) > D_2 \frac{TOL_{TT}}{N[k]}, \forall \ n = 1, \ldots, N[k] - 1 \right)$$

(3.57)

then the dividing-merging process stops.

Here $D_1$ and $D_2$ are given constants satisfying $0 < D_2 < d_2 < d_1 < D_1$. The combination of (3.50) and (3.56) asymptotically guarantees a given level of accuracy, $E_{TT} < D_1 TOL_{TT}$ and the lower bound (3.57) implies efficiency by almost optimal time steps. The positive numbers $D_1$ and $D_2$ are motivated to avoid slow convergence in case almost all $\bar{r}_n$ satisfy (3.56, 3.57), as in Section 3.1. Before determining the sufficient conditions for the constants $D_2 < d_2 < d_1 < D_1$, we will describe the algorithm with deterministic time steps in detail.

**Remark 3.7.** In practice, our numerical tests show that

$$\frac{|E[g(X(T))] - A(g(X(T)); M)|}{\text{TOL}} \approx \frac{d_1}{2} < \frac{2}{9} D_1,$$

so we choose $d_1 = 2$.

**The adaptive algorithm.** First we split the specified error tolerance into three parts, TOL$_S$, TOL$_{TT}$ and TOL$_{TS}$ by (3.2) and (3.52). The first loop below determines the mesh with $M_T$ realizations by changing iteratively the time steps using our dividing and merging strategy (3.54, 3.55) until $|\bar{\rho}|$ and $\Delta t$ satisfy the stopping criteria (3.56, 3.57) and the statistical error estimate $E_{TS} \leq \text{TOL}_{TS}$. Then, the second loop with fixed mesh chooses the number $M$ of realizations, using (3.6) until $E_S \leq \text{TOL}_S$ holds.

Now we are ready for the detailed definition of the adaptive algorithm with deterministic steps:
Algorithm D

Initialization Choose:

1. an error tolerance, $TOL = TOL_S + TOL_{TT} + TOL_{TS}$,
2. a number, $N[1]$, of initial uniform steps $\Delta t[1]$ for $[0, T]$,
4. an integer $H \geq 2$ for the number of subdivisions of a refined time step, a number, $d_1 = 2$ in (3.54) and a rough estimate of $c$ in (3.58) to compute $d_2, D_1, D_2$ using (3.59, 3.60, 3.61), and
5. a constant $c_0 \geq 1.65$ and an integer $MCH \geq 2$ to determine the number of realizations in (3.6).

Set the iteration counter, $k$, for time refinement levels, to 1 and set the statistical error, $E_{TS} = +\infty$ and $\bar{r}[1] = +\infty$.

Do while ( $\bar{r}[k]$ violates the stopping (3.56, 3.57) or $E_{TS} > TOL_{TS}$ )

Compute, with the same mesh $\Delta t[k]$, sample averages and error estimates by calling Euler.

If ( $\bar{r}[k]$ violates the stopping (3.56, 3.57) )

For all time steps $i = 1, \ldots, N[k]$, do the dividing and merging process (3.54, 3.55) to compute $\Delta t[k+1]$ from $\Delta t[k]$.

elseif ( $E_{TS} > TOL_{TS}$ )

Compute $M_T[k+1]$ by change $M_T[k], \mathcal{V}_{TS}[k], TOL_{TS}; M_T[k+1]$.

Set $\Delta t[k+1] = \Delta t[k]$.

end-if

Increase $k$ by 1.

end-do

Compute an approximation, $E_g$, for $E[g(X(T))]$ with fixed time mesh $\Delta t = \Delta t[k]$ by Monte-Carlo($M_T[k], TOL_S; E_g$).

Accept $E_g$ as an approximation of $E[g(X(T))]$, since the estimate of the computational error is bounded by $TOL$.

routine Euler

Compute $M_T[k]$ new realizations of the Euler method with the same partition $\Delta t[k]$ and update the approximations of the time discretization error indicators $\bar{r}[k]$ and the statistical time discretization error $E_{TS}[k]$.

end-of-Euler

Stopping of the adaptive algorithm. The right choice of the parameters $D_2 < d_2 < d_1 < D_1$ is explained by

Theorem 3.8 (Stopping). Suppose the assumptions of Theorem 2.2 hold and the adaptive algorithm uses the strategy (3.8), (3.54, 3.55), (3.56, 3.57). Assume that there exists a positive constant $c \geq 2^{-1}$ such that for all $t \in [0, T]$ and for each realization

$$c \leq \left| \frac{\bar{r}(t)[k+1]}{\bar{r}(t)[k]} \right| \leq c^{-1}$$
and that
\[(3.59)\]
\[
d_2 < \frac{c}{2H^2}d_1,\]
\[(3.60)\]
\[
D_2 < \frac{c}{H}d_2,\]
\[(3.61)\]
\[
D_1 > \frac{H}{c}d_1.\]

Then the adaptive dividing-merging process stops by the stopping criteria (3.56, 3.57), after a finite number of operations.

Proof. The proof is the same as for Theorem 3.2 in \cite{25}, which is also similar to Theorem 3.2.

Accuracy of the adaptive algorithm. The adaptive algorithm guarantees that the estimated error is bounded by a given error tolerance, \(D_1\text{TOL} + \text{TOL} = (\frac{2}{9}D_1 + \frac{2}{3})\text{TOL}\). By applying the same arguments as in the proof of the Theorem 3.3, the true error for the deterministic time stepping is also bounded by \((\frac{2}{9}D_1 + \frac{2}{3})\text{TOL}\) asymptotically, using the upper bound (3.56) of the error indicators and the a.s. convergence of \(|\bar{\rho}|\) in Section 4.

Theorem 3.9 (Accuracy). Suppose that the assumptions of Theorem 2.2 hold. Then the adaptive algorithm (3.8), (3.54, 3.55), (3.56, 3.57) satisfies, for any constant \(c_0 > 0\) defined in (3.6)
\[
(3.62)\]
\[
\lim \inf_{\text{TOL} \to 0} P \left( \frac{|E[g(X(T))] - A(g(\bar{X}(T)); M)|}{\text{TOL}} \leq \frac{2}{9}D_1 + \frac{2}{3} \right) \geq \int_{-c_0}^{c_0} e^{-x^2/2} \sqrt{2\pi} dx.
\]

Proof. By applying the same arguments as in the proof of Theorem 3.3, we obtain, for the time discretization error,
\[
(3.63)\]
\[
\frac{|E[g(X(T)) - g(\bar{X}(T))]|}{\text{TOL}_{TT}} < D_1 E \left[ \frac{\int_0^T |\hat{\rho}(\tau)|/\sqrt{E[\hat{\rho}(\tau)]} d\tau}{\int_0^T \sqrt{E[\hat{\rho}(\tau)]} d\tau} \right].
\]

Corollary 4.3 and Theorem 2.2 show that \(\hat{\rho}\) converges a.s. to the limit \(\hat{\rho}\) and Theorem 4.5 proves that the sample average, \(\bar{\rho}\), converges to \(E[\hat{\rho}]\). Therefore dominated convergence shows that the fraction in the right hand side of (3.63) converges to 1, i.e.,
\[
\lim_{\text{TOL} \to 0} E \left[ \frac{\int_0^T |\hat{\rho}(\tau)|/\sqrt{E[\hat{\rho}(\tau)]} d\tau}{\int_0^T \sqrt{E[\hat{\rho}(\tau)]} d\tau} \right] = E \left[ \lim_{\text{TOL} \to 0} \frac{\int_0^T |\hat{\rho}(\tau)|/\sqrt{E[\hat{\rho}(\tau)]} d\tau}{\int_0^T \sqrt{E[\hat{\rho}(\tau)]} d\tau} \right] = E \left[ \frac{\int_0^T |\hat{\rho}(\tau)|/\sqrt{E[\hat{\rho}(\tau)]} d\tau}{\int_0^T \sqrt{E[\hat{\rho}(\tau)]} d\tau} \right] = 1.
\]

This deterministic limit implies that for all \(s > \frac{2}{9}D_1\)
\[
(3.64)\]
\[
\lim \inf_{\text{TOL} \to 0} P \left( \frac{|E[g(X(T)) - g(\bar{X}(T))]|}{\text{TOL}} \leq s \right) = 1,
\]
and consequently the combination (3.64) and (3.35) proves (3.62). \(\square\)
**Efficiency of the adaptive algorithm.** Within the class of deterministic time steps, the conditions (3.11) and (3.13) give the minimal number of steps

\[ N_D = \frac{1}{\text{TOL}_T} \left( \int_0^T \sqrt{E[|\bar{\rho}(\tau)|]} d\tau \right)^2. \]  

(3.65)

Using the definitions (3.36) and (3.37) of the minimal expected number, \( E[N_S] \), of stochastic time steps and the number, \( N_C \), of constant time steps respectively, we obtain

\[ E[N_S] \leq N_D \leq N_C \]  

(3.66)

by applying Jensen’s inequality.

Similar to Theorem 3.4 with \( N \) instead of \( \overline{N} \), we prove the following theorem using the lower bound (3.57) of the error indicators to show that the algorithm (3.54 - 3.57) generates a mesh which is optimal, up to a multiplicative constant \( C = 6/D_2 \).

**Theorem 3.10 (Efficiency).** Suppose that the assumptions of Theorem 2.2 hold. Then there exists a constant \( C > 0 \), asymptotically bounded by \( 6/D_2 \), such that the number, \( N \), of adaptive steps of the algorithm (3.8) (3.54, 3.55), (3.56, 3.57) satisfies

\[ \text{TOL}_T N < C \left( \int_0^T \sqrt{|\bar{\rho}(\tau)|} d\tau \right)^2. \]  

(3.67)

4. **Almost sure convergence of the error density**

This subsection proves pathwise a.s. convergence of the error density using a.s. convergence of the approximate solution of the Monte Carlo Euler method based on the adaptive algorithms presented in Section 3. Before presenting the main result, we extend the Euler method, for theoretical purposes only, to \( t \in [0, T] \) by

\[ \overline{X}(t) - \overline{X}(0) = \int_0^t \bar{a}(s; \overline{X}) ds + \sum_{t_n} \int_{t_n}^t \bar{b}^\ell(s; \overline{X}) dW^\ell(s) \]  

(4.1)

where \( \bar{a} \) and \( \bar{b}^\ell \) are the piecewise constant approximations, see Figure 1,

\[ \bar{a}(s; \overline{X}) \equiv a(t_n, \overline{X}(t_n)) \quad \text{and} \quad \bar{b}^\ell(s; \overline{X}) \equiv b^\ell(t_n, \overline{X}(t_n)) \quad \text{for} \quad s \in [t_n, t_{n+1}). \]

4.1. **Stochastic time steps.** Let us now for \( H = 2 \) define the maximum step size for a fixed realization by

\[ \Delta t_{sup}(\eta) \equiv \sup_{1 \leq n \leq N} \Delta t_n(\eta) = T2^{-\eta}, \]  

(4.2)

for a positive integer \( \eta \). Theorem 4.1 and Lemma 4.2 below show that the approximate solution converges a.s. to the correct limit.

**Theorem 4.1.** Suppose that \( a, b, g \) and \( X \) satisfy the assumptions in Lemma 2.1. with the maximum time step size \( \Delta t_{sup}(\eta) \) in (4.2). Then for any \( \alpha < 1/2 \)

\[ \lim_{\eta \to \infty} \Delta t^{-\alpha}_{sup}(\eta) \sup_{t \in [0, T]} |X(t) - \overline{X}(t)| = 0 \quad \text{a.s.}, \]  

(4.3)

where \( \Delta t^{\beta}_{sup}(\eta) \equiv (\Delta t_{sup}(\eta))^{\beta} \) for \( \beta \in \mathbb{R} \).
Proof. To simplify the proof, let us introduce the forward Euler approximation $\hat{X}$ of $X$ with uniform time steps, $\Delta t$, on a much finer grid than $\Delta t$, so that $\{i\Delta t : i = 0, \ldots, \hat{N}\}$ includes all time steps for $X$. We can extend $\hat{X}$ as in (4.1) to $t \in [0, T]$ by

$$\hat{X}(t) - \hat{X}(0) = \int_0^t \hat{a}(s; \hat{X})ds + \sum_{\ell=1}^{\ell_0} \int_0^t \hat{b}^{\ell}(s; \hat{X})dW^\ell(s)$$

where $\hat{a}$ and $\hat{b}^{\ell}$ are the piecewise constant approximations, see Figure 1, $\hat{a}(s; \hat{X}) = a(i\Delta t, \hat{X}(i\Delta t))$ and $\hat{b}^{\ell}(s; \hat{X}) = b^{\ell}(i\Delta t, \hat{X}(i\Delta t))$ for $s \in [i\Delta t, (i+1)\Delta t)$.

Then, the error splits into

$$\Delta t_{\sup}^{-a}(\eta) \sup_t |X(t) - \hat{X}(t)|$$

$$\Delta t_{\sup}^{-a}(\eta) \sup_t |X(t) - \hat{X}(t)| + \Delta t_{\sup}^{-a}(\eta) \sup_t |\hat{X}(t) - \bar{X}(t)|.$$

Let us first study the a.s. convergence of the second term in (4.5). The convergence of the standard Euler method with uniform time steps in the first term follows then by a similar derivation or by [32]. To show a.s convergence of the second term we observe that

$$\hat{X}(t) = \int_0^t \Delta a(s) ds + \int_0^t \Delta b^\ell(s) dW^\ell(s),$$

where for $t \in [t_n, t_{n+1})$,

$$\Delta a(t) = \hat{a}(t; \hat{X}) - \bar{a}(t; \bar{X}),$$

$$\Delta b^\ell(t) = \hat{b}^\ell(t; \hat{X}) - \bar{b}^\ell(t; \bar{X}).$$

Now let us define the second term of the right hand side in (4.6) by

$$Y(t) = \int_0^t \Delta b^\ell(s) dW^\ell(s).$$

We have $\sup_{t} \sqrt{E[|Y(t)|^2]} = O(\Delta t_{\sup}^2)$ by using the mean square strong convergence of the Euler approximation in Lemma 2.1. Lemma 4.2 below verifies that $Y$ is a continuous martingale with respect to a filtration generated by $\{W(s), \Delta t(s) :
\( s \leq t \), so that \( |Y| \) is a submartingale. Therefore Doob’s inequality and Jensen’s inequality give, for any \( \alpha' \in \mathbb{R} \),

\[
P \left( \sup_{0 \leq t \leq T} |Y(t)| \geq \Delta t_{\sup}^{\alpha'}(\eta) \right) \leq \frac{1}{\Delta t_{\sup}^{\alpha'}(\eta)} E[|Y(t)|] \\
\leq \frac{1}{\Delta t_{\sup}^{\alpha'}(\eta)} \sqrt{E[|Y(t)|^2]} \\
= O(\Delta t_{\sup}^{\frac{1}{2} - \alpha'}). 
\]

The definition (4.2) then implies, for a positive constant \( C \),

\[
\sum_{\eta=1}^{\infty} P \left( \sup_{0 \leq t \leq T} |Y(t)| \geq \Delta t_{\sup}^{\alpha'}(\eta) \right) \leq \sum_{\eta=1}^{\infty} O(\Delta t_{\sup}^{\frac{1}{2} - \alpha'}) \\
\leq C \sum_{\eta=1}^{\infty} 2^{-\eta(\frac{1}{2} - \alpha')} < \infty, \tag{4.9}
\]

provided \( \frac{1}{2} - \alpha > 0 \). Therefore the Borel-Cantelli lemma implies

\[
P \left( \sup_{0 \leq t \leq T} |Y(t)| \geq \Delta t_{\sup}^{\alpha'}(\eta) \text{ infinitely often} \right) = 0,
\]

i.e., with \( \alpha < \frac{1}{2} \)

\[
\Delta t_{\sup}^{-\alpha}(\eta) \sup_{0 \leq t \leq T} |Y(t)| \to 0 \text{ a.s., as } \eta \to \infty. \tag{4.10}
\]

The first term in the right hand side of (4.6) satisfies

\[
\sup_{t} |\int_{0}^{t} \Delta a(s)ds| \leq \int_{0}^{T} |\Delta a(s)|ds.
\]

Therefore Chebyshev’s inequality yields

\[
P \left( \sup_{t} |\int_{0}^{t} \Delta a(s)ds| \geq \Delta t_{\sup}^{\alpha'}(\eta) \right) \leq P \left( \int_{0}^{T} |\Delta a(s)|ds \geq \Delta t_{\sup}^{\alpha'}(\eta) \right) \\
\leq \frac{1}{\Delta t_{\sup}^{2\alpha'}(\eta)} \left[ \int_{0}^{T} |\Delta a(s)|^2 ds \right] \\
\leq \frac{T}{\Delta t_{\sup}^{2\alpha'}(\eta)} \left[ \int_{0}^{T} |\Delta a(s)|^2 ds \right] \\
= O \left( \Delta t_{\sup}^{-2\alpha'}(\eta) \right), 
\]

where the last equality follows from Lemma 2.1. The argument in (4.9) and the Borel-Cantelli lemma give similarly \( \Delta t_{\sup}^{-\alpha}(\eta) \sup_{t} |\int_{0}^{t} \Delta a(s)ds| \to 0 \text{ a.s. for } \alpha < \frac{1}{2} \), as \( \eta \to \infty \), and consequently for \( \alpha < \frac{1}{2} \)

\[
(4.11) \quad \Delta t_{\sup}^{-\alpha}(\eta) \sup_{t \in [0,T]} |\hat{X}(t) - \overline{X}(t)| \to 0 \text{ a.s. as } \eta \to \infty.
\]
The same arguments applied to the first term in the right hand side of (4.5) yield for $\alpha < \frac{1}{2}$,

\begin{equation}
\Delta t_n^\alpha(\eta) \sup_{t \in [0,T]} |X(t) - \tilde{X}(t)| \to 0 \text{ a.s. as } \eta \to \infty.
\end{equation}

Here Lemma 4.2 holds directly for $\tilde{X}$, since $\tilde{X}$ is adapted to the standard $\sigma$-algebra generated by $W(s)$. The combination (4.12) and (4.11) proves (4.3). \hfill \Box

The key to our proof of Theorem 4.1 is that $Y_t \equiv Y(t)$ in (4.8) is a martingale with respect to a filtration $\mathcal{F}_t$:

**Lemma 4.2.** Suppose that $a, b, g$ and $X$ satisfy the assumptions in Lemma 2.1. Let the process $Y_t$ be defined by (4.8) and the filtration $\mathcal{F}_t$ be adapted to the standard $\sigma$-algebra generated by \{\text{W(s)}, \Delta t(s) : s \leq t\}. Then $Y_t$ satisfies

(i): $E[|Y_t|] < \infty$, for $0 \leq t \leq T$,

(ii): $Y_t$ is adapted to $\mathcal{F}_t$,

(iii): $E[Y_t | \mathcal{F}_s] = Y_s$ for $t \geq s$,

i.e., $Y_t$ is a martingale with respect to $\mathcal{F}_t$.

**Proof.** To prove the lemma, we observe by Lemma 2.1 that $E[|Y_t|] \leq \sqrt{\text{Vol} Y_t^T} < \infty$ and the construction of $\bar{X}$ implies $Y_t \in \mathcal{F}_t$ for all $t$, so that (i) and (ii) holds. Note that $Y_t$ is not adapted to the standard filtration generated by $\text{W}$ only. Using conditional expectation and the notations in Theorem 4.1

\begin{equation}
E[Y_t | \mathcal{F}_s] = E[Y_s | \mathcal{F}_s] + E \left[ \int_s^t \Delta b^\ell dW^\ell \mid \mathcal{F}_s \right] = Y_s + E \left[ \int_s^t \Delta b^\ell dW^\ell \mid \mathcal{F}_s \right].
\end{equation}

It remains to verify $E \left[ \int_s^t \Delta b^\ell dW^\ell \mid \mathcal{F}_s \right] = 0$. Since $\Delta b^\ell$ is piecewise constant, rewrite $\int_s^t \Delta b^\ell dW^\ell = \sum_{i=n_1}^{n_2} \Delta b^\ell(i\Delta t) \Delta W^\ell_i$, where $\Delta W^\ell_i = W^\ell((i + 1)\Delta t) - W^\ell(i\Delta t)$ with the uniform time steps, $\Delta t$, which is finer than $\Delta t$; at the end point $t$ of the interval $[s, t]$ we let $\Delta W^\ell_{n_2} = W^\ell(t) - W^\ell(n_2\Delta t)$ and similarly for the end point $s$.

Let us now study one term $E \left[ \Delta b^\ell(\hat{n}\Delta t) \Delta W^\ell_{n_2} \right]$. Divide the interval, $[\hat{n}\Delta t, (\hat{n} + 1)\Delta t)$, into the union of disjoint intervals $[\hat{n}\Delta t + \sum_{j=1}^{m-1} \hat{\Delta} t_j, \hat{n}\Delta t + \sum_{j=1}^{m} \hat{\Delta} t_j]$, so that $\hat{\Delta} t = \sum_{m=1}^{\hat{n}} \hat{\Delta} t_m$ with the corresponding Wiener increments

\begin{align*}
\Delta W^\ell_m &= W^\ell(\hat{n}\Delta t + \sum_{j=1}^{m} \hat{\Delta} t_j) - W^\ell(\hat{n}\Delta t + \sum_{j=1}^{m-1} \hat{\Delta} t_j), \\
\Delta W^\ell_{m+1} &= W^\ell((\hat{n} + 1)\Delta t) - W^\ell(\hat{n}\Delta t + \sum_{j=1}^{m} \hat{\Delta} t_j).
\end{align*}

The end intervals $(n_2\Delta t, t)$ and $(s, n_1\Delta t)$ are treated similarly.

We claim that $\bar{X}(t_n)$ is essentially independent of the increments $W(\hat{t} + \hat{\Delta} t) - W(\hat{t})$, provided $\hat{\Delta} t(\hat{t})$ is sufficiently small and $t_n < \hat{t}$. The stopping criteria (3.18, 3.19), for each final acceptable time discretization imply that $\min_n (S_1^{-\text{tol}} r_n - r_n)$
is positive for all realizations. The approximate solution \( \mathbf{X}(\tau) \) depends on \( dW(\hat{t}) \), for \( \tau < \hat{t} \equiv \hat{n}\Delta \hat{t} \), only through changes in the mesh. We shall show that, provided \( \hat{\Delta}t \) is sufficiently small and conditioned on the \( \sigma \)-algebra \( \mathcal{M}(\hat{t}, \hat{\Delta}t) \) generated by \( \{dW(\tau) : \tau < \hat{t}, \text{ or } \tau > \hat{t} + \hat{\Delta}t \} \), the probability to change the mesh by a change in only \( \hat{\Delta}W(\hat{t}) \) is arbitrary small, thus \( \mathbf{X}(\tau) \) will be essentially independent of \( \hat{\Delta}W(\hat{t}) \) for \( \tau < \hat{t} \). Conditioned on \( \mathcal{M}(\hat{t}, \hat{\Delta}t) \), let \( r_n(\hat{\Delta}W) \equiv E[r_n | \mathcal{M}(\hat{t}, \hat{\Delta}t)] \) denote the dependence of the error indicator \( r_n \) on the noise \( \hat{\Delta}W \). The Malliavin derivative, \( \partial_{W(\tau)} r_n \), and Taylor’s formula imply

\[
  r_n(\hat{\Delta}W) = r_n(0) + \hat{\Delta}W^T \int_0^1 \partial_{W(\tau)} r_n(\tau \hat{\Delta}W) d\tau.
\]

The mesh generated by \( r(\hat{\Delta}W) \) and \( r(0) \) is the same provided

\[
  0 < S_1 \frac{\text{TOL}T}{N} - r_n(\hat{\Delta}W) = S_1 \frac{\text{TOL}T}{N} - \left( r_n(0) + \hat{\Delta}W^T \int_0^1 \partial_{W(\tau)} r_n(\tau \hat{\Delta}W) d\tau \right) \equiv S_1 \frac{\text{TOL}T}{N} - (r_n(0) + \hat{\Delta}W \cdot r_n'), \quad \text{for all } n,
\]

and

\[
  \min_n \left( S_1 \frac{\text{TOL}T}{N} - r_n(0) \right) \equiv \epsilon(\omega) > 0.
\]

Therefore, (4.14, 4.15) hold if \( \hat{\Delta}W \cdot r_n' < \epsilon(\omega) \) for all \( n \). Let

\[
  \bar{\epsilon} \equiv \epsilon/((\sup N)\Delta t_n)^2C(\text{TOL}),
\]

then \( |\hat{\Delta}W| < \bar{\epsilon} \) implies \( \hat{\Delta}W \cdot r_n' < \epsilon \), since

\[
  \hat{\Delta}W \cdot r_n' \leq |\hat{\Delta}W||r_n'| < \epsilon|r_n'| = \epsilon(\Delta t_n)^2 \left| \int_0^1 \partial_{W(\tau)} \rho(t_n, \tau \hat{\Delta}W) d\tau \right| \leq \epsilon,
\]

provided \( \rho(t_n, \hat{\Delta}W) \equiv E[\rho(t_n) | \mathcal{M}(\hat{t}, \hat{\Delta}t)] \) and \( \rho \) is an approximate error density function satisfying (2.3).

Consequently, the following independence claim holds:

\[
  \mathbf{X}(t_n) \text{ is independent of } \hat{\Delta}W(\hat{t}) \text{ conditioned on } |\hat{\Delta}W(\hat{t})| < \bar{\epsilon} \text{ and } \mathcal{M}(\hat{t}, \hat{\Delta}t), \text{ for } t_n < \hat{t},
\]

and the conditional probability to have different meshes with \( r(0) \) or \( r(\hat{\Delta}W) \) is, for sufficiently small \( \hat{\Delta}t \), bounded by

\[
  P \left( |\hat{\Delta}W| \geq \bar{\epsilon} \mid \mathcal{M}(\hat{t}, \hat{\Delta}t), \mathcal{F} \right) = 1 - P \left( |\hat{\Delta}W| < \bar{\epsilon} \mid \mathcal{M}(\hat{t}, \hat{\Delta}t), \mathcal{F} \right) \leq (\hat{\Delta}t)^3.
\]
since
\[
P \left( |\hat{\Delta}W| < \epsilon \mid \mathcal{M}(\hat{t}, \hat{\Delta}t), \mathcal{F}_s \right) = P \left( |\hat{\Delta}W| < \epsilon \mid \mathcal{M}(\hat{t}, \hat{\Delta}t) \right) \\
\geq 1 - C t_o \exp \left( - \frac{\epsilon^2}{4 \hat{\Delta}t} \right) > 1 - (\hat{\Delta}t)^3.
\]

Let us define the \(\sigma\)-algebra \(\mathcal{G}\) generated by \(\{dW(t) : \tau < \hat{n}\Delta t \text{ or } \tau > \hat{n}\Delta t + \hat{\Delta}t\}\)
and \(\{\Delta t(t) : \tau < \hat{n}\Delta t\}\). Then using the results (4.17) and (4.16), we get

\[
E \left[ \Delta b^t \hat{\Delta}W^f_h \mid \mathcal{F}_s \right] = E \left[ E \left[ \Delta b^t \hat{\Delta}W^f_h \mid \mathcal{G} \right] \mid \mathcal{F}_s \right] \\
= E \left[ E \left[ \Delta b^t \left( \hat{\Delta}W^f_1 + \frac{1}{\mathbb{1}_{\{\hat{\Delta}W_1 < \epsilon\}}} + \frac{1}{\mathbb{1}_{\{\hat{\Delta}W_1 \geq \epsilon\}}} \hat{\Delta}W^f_1 \right) \mid \mathcal{G} \right] \mid \mathcal{F}_s \right] \\
(4.18) = E \left[ E \left[ \Delta b^t \hat{\Delta}W^f_1 \mid \mathcal{G} \right] \mid \mathcal{F}_s \right] + E \left[ E \left[ \Delta b^t \mathbb{1}_{\{\hat{\Delta}W_1 < \epsilon\}} \hat{\Delta}W^f_1 \mid \mathcal{G} \right] \mid \mathcal{F}_s \right] \\
+ E \left[ E \left[ \Delta b^t \mathbb{1}_{\{\hat{\Delta}W_1 \geq \epsilon\}} \hat{\Delta}W^f_1 \mid \mathcal{G} \right] \mid \mathcal{F}_s \right].
\]

By (4.17), the last term of (4.18) becomes

\[
(4.19) \quad E \left[ E \left[ \Delta b^t \mathbb{1}_{\{\hat{\Delta}W_1 \geq \epsilon\}} \hat{\Delta}W^f_1 \mid \mathcal{G} \right] \mid \mathcal{F}_s \right] = E \left[ O((\hat{\Delta}t_1)^3) \right]
\]
and from (4.16) the second term of (4.18) becomes

\[
(4.20) \quad E \left[ E \left[ \Delta b^t \mathbb{1}_{\{\hat{\Delta}W_1 < \epsilon\}} \hat{\Delta}W^f_1 \mid \mathcal{G} \right] \mid \mathcal{F}_s \right] \\
= E \left[ E \left[ \Delta b^t \mid \mathcal{G}, \{\hat{\Delta}W_1 < \epsilon\}, \mathcal{F}_s \right] \right] \times E \left[ \mathbb{1}_{\{\hat{\Delta}W_1 < \epsilon\}} \hat{\Delta}W^f_1 \mid \mathcal{G}, \{\hat{\Delta}W_1 < \epsilon\}, \mathcal{F}_s \right] P(\hat{\Delta}W_1 \leq \epsilon \mid \mathcal{G}, \mathcal{F}_s) = 0.
\]

Therefore we obtain from (4.19) and (4.20)

\[
(4.21) \quad E \left[ \Delta b^t \hat{\Delta}W^f_h \mid \mathcal{F}_s \right] = E \left[ E \left[ \Delta b^t \hat{\Delta}W^f_1 \mid \mathcal{G} \right] \mid \mathcal{F}_s \right] + E \left[ O((\hat{\Delta}t_1)^3) \right].
\]

Apply this \(\hat{\Delta}W, \hat{\Delta}W\) argument recursively to \(E[\Delta b^t \hat{\Delta}W^f_m] \mid \mathcal{F}_s\) conditioned on \(\mathcal{M}(i\Delta t + \sum_{j=1}^{m} \hat{\Delta}t_j, \hat{\Delta}t_{m+1})\) to get

\[
(4.22) \quad E \left[ \Delta b^t \hat{\Delta}W^f_h \mid \mathcal{F}_s \right] = E \left[ \sum_{m} O((\hat{\Delta}t_m)^3) \right] = E \left[ \int_{(n-1)\Delta t}^{n\Delta t} O((\hat{\Delta}t)^2) dt \right].
\]

Letting \(\hat{\Delta}t \to 0^+\) in (4.22) proves \(E[\Delta b^t \hat{\Delta}W^f_h \mid \mathcal{F}_s] = 0\) and apply the same arguments to all intervals \([\hat{n}\Delta t, (\hat{n} + 1)\Delta t]\), so that by (4.13) \(E[Y_t \mid \mathcal{F}_s] = Y_s\) for \(t \geq s\). \(\Box\)

To verify a.s. convergence of the error density, let us recall the definition of the variation of a process \(Y\): the first variation of a function \(F(Y(T))\) with respect to
a perturbation in the initial location of the path $Y$, at time $s$, is denoted by

$$F'(T; s) = \partial_{(s)} F(Y(T))$$

(4.23)

$$= \left( \frac{\partial}{\partial y_1} F(Y(T); Y(s) = y), \ldots, \frac{\partial}{\partial y_d} F(Y(T); Y(s) = y) \right).$$

The definition (4.23) implies that the first variation, $X'$, of the solution $X$ with respect to a perturbation in the initial location at $t$ satisfies

$$dX'_{ik}(s) = \partial_j a_i(s, X(s))X'_{jk}(s) + \partial_j b_i^k X'_{jk}(s) dW^t(s), \quad s > t,$$

(4.24)

$$X'_{ik}(t) = \delta_{ik} = \left\{ \begin{array}{ll}
0, & i \neq k, \\
1, & i = k,
\end{array} \right.$$ 

and similarly one can derive the equations for the second and the third variation of $X$, cf. [31].

The definition of $c_i$ in (2.8) shows that the forward Euler approximation, $X'$, in (4.24) can be written

$$X'_{ik}(t_{n+1}; t_m) = \partial_j c_i(t_n, X(t_n))X'_{jk}(t_n; t_m)$$

(4.25)

$$X'_{ik}(t_m) = \delta_{ik}.$$ 

Then the equations for $\varphi$ in (2.7) and for $X'$ in (4.25) imply

$$0 = \sum_{n=m}^{N-1} \left( \varphi_i(t_n) - \partial_i c_j(t_n, X(t_n)) \varphi_j(t_{n+1}) \right) X'_{ik}(t_n; t_m)$$

$$= \sum_{n=m}^{N-1} \varphi_i(t_{n+1}) \left( X'_{ik}(t_{n+1}; t_m) - \partial_j c_i(t_n, X(t_n)) X'_{jk}(t_n; t_m) \right)$$

$$+ \varphi_i(t_m) X'_{ik}(t_m; t_m) - \varphi_i(T) X'_{ik}(T; t_m)$$

$$= \varphi_i(t_m) X'_{ik}(t_m; t_m) - \varphi_i(T) X'_{ik}(T; t_m),$$

i.e.,

$$\varphi_k(t_m) = \partial_i g(X(T)) X'_{ik}(T; t_m),$$

(4.26)

by using the initial conditions of $\varphi$ and $X'$. The definitions of the first and second variations of $\varphi$ in (2.9) and (2.10) also yield

$$\varphi'_{kn}(t_m) = \partial_i g(X(T)) X''_{ikn}(T; t_m) + \partial_{ir} g(X(T)) X'_{rn}(T; t_m) X'_{ik}(T; t_m)$$

(4.27)

and

$$\varphi''_{knm}(t_m) = \partial_i g(X(T)) X'''_{iknm}(T; t_m) + \partial_{ir} g(X(T)) X''_{rm}(T; t_m) X''_{ikn}(T; t_m)$$

$$+ \partial_{ir} g(X(T)) X''_{ik}(T; t_m) X''_{rn}(T; t_m) + \partial_{itr} g(X(T)) X''_{in}(T; t_m) X''_{ikm}(T; t_m)$$

(4.28)

$$+ \partial_{itr} g(X(T)) X''_{ik}(T; t_m) X''_{rn}(T; t_m).$$

Define the function $\hat{\rho}$ following the error density $\rho$ in (2.6) by substituting $X$ by the limit $X$ and replacing $\varphi, \varphi', \varphi''$ by the corresponding limits in (4.26-4.28) where $X$ is substituted by $X$.

Now we are ready to prove the a.s. convergence of the error density:
Corollary 4.3. Suppose the assumptions of Theorem 4.1. Then the error density $\rho$ in (2.6) converges a.s. to the limit $\hat{\rho}$, defined above, as the specified error tolerance tends to zero.

Proof. To prove the corollary, it is necessary to understand the convergence of the discrete dual solutions, $\varphi$, in (2.7) and its first and second variation, $\varphi', \varphi''$. Using the definitions of the variations of $X$, let us consider an augmented system for $Z = (X, X', X'', X''')^T$ and let $I$ denote the $d \times d$ identity matrix. Then (4.24) and the equations for $X''$ and $X'''$ can be written

$$dZ = A(t, Z)dt + B^\ell(t, Z)dW^\ell(t), \quad t > t_0$$

$$Z(t_0) = (x, I, 0, 0, 0)^T.$$ 

The Euler approximation $\tilde{Z} = (\tilde{X}, \tilde{X}', \tilde{X}'', \tilde{X}''')^T$ of $Z$ in (4.29) with piecewise constant drift and diffusion fluxes satisfies

$$d\tilde{Z} = \tilde{A}(t; \tilde{Z})dt + \tilde{B}^\ell(t; \tilde{Z})dW^\ell$$

defined as in (4.1).

Applying Theorem 4.1 to $\tilde{Z}$ and the augmented system (4.29) shows that the approximation $\tilde{Z}$ converges a.s. to the process $Z$. Therefore, the representations (4.26-4.28) and the regularity assumptions on the functions $g$, $a$ and $b$ imply that $\varphi$, $\varphi'$ and $\varphi''$ converge a.s.. Consequently, the error density $\rho$ converges a.s. to the true error density $\hat{\rho}$ as the specified error tolerance, TOL, tends to zero, i.e., as the maximum step size tends to zero. □

4.2. Deterministic time steps.

Lemma 4.4. Suppose that $a, b, g$ and $X$ satisfy the assumptions in Lemma 2.1. Then for any $\alpha \in (0, 1/2)$

$$\lim_{M \to \infty} M^\alpha \sup_{t \in [0, T]} |E[\tilde{X}(t)] - A(\tilde{X}(t); M)| = 0 \text{ a.s.}$$

Proof. The idea is to split the difference in (4.30) into a sum of a martingale and a bounded integral and use Doob’s inequality for the martingale. Thus, representation (4.1) shows

$$E[\tilde{X}(t)] - A(\tilde{X}(t); M) = E\left[\int_0^t \tilde{b}^\ell(s; \tilde{X}(s))dW^\ell(s; M)\right]$$

$$= \int_0^t \left(E[\tilde{a}(s; \tilde{X}(s))] - A(\tilde{a}(s; \tilde{X}(s)); M)\right)ds.$$
is then a $\mathcal{F}_{M,t}$ martingale and Doob’s inequality yields

$$P\left( M^\alpha \sup_{t \in [0,T]} |Y_M(t)| \geq \epsilon \right) \leq M^{2\alpha} \frac{\text{Var}[Y_M(T)]}{\epsilon^2}$$

$$= M^{2\alpha} \frac{\text{Var}[E[\int_0^T b^\ell(s, \overline{X}(s))dW^\ell(s); M]|\sigma(\Delta t)]]}{\epsilon^2}$$

$$= \frac{\text{Var}[E[\int_0^T b^\ell(s, \overline{X}(s))dW^\ell(s)|\sigma(\Delta t)]]}{M^{1-2\alpha} \epsilon^2}$$

$$\leq \frac{C_T}{M^{1-2\alpha} \epsilon^2}.$$  

Here $C_T$ is a constant that, due to the smoothness assumptions in the drift and diffusion coefficients, bounds uniformly the variance of $\int_0^T b^\ell(s, \overline{X}(s))dW^\ell(s)$. The construction (3.6) shows that all $M$ belong to the set $\{2^k | k \in \mathbb{N}\}$. Let $M_k \equiv 2^k$, then

$$\sum_{k=1}^{\infty} P\left( (M_k)^\alpha \sup_{t \in [0,T]} |Y_{M_k}(t)| \geq \epsilon \right) \leq \frac{C_T}{\epsilon^2} \sum_{k=1}^{\infty} \frac{1}{(M_k)^{1-2\alpha}} < \infty$$

and Borel-Cantelli’s lemma implies

$$P\left( (M_k)^\alpha \sup_{t \in [0,T]} |Y_{M_k}(t)| \geq \epsilon \text{ infinitely often} \right) = 0,$$

i.e. $(M_k)^\alpha \sup_{t \in [0,T]} |Y_{M_k}(t)| \to 0$ a.s., as $k \to \infty$ for $0 < \alpha < 1/2$. Now consider the second term in (4.31), namely $\int_0^t (E[\alpha(s; \overline{X}(s))] - \mathcal{A}(\alpha(s; \overline{X}(s)); M_k)) ds$. First note that we have

$$\sup_{t \in [0,T]} \left| \int_0^t E[\alpha(s; \overline{X}(s))] - \mathcal{A}(\alpha(s; \overline{X}(s)); M_k) ds \right| \leq$$

$$\int_0^T \left| E[\alpha(s; \overline{X}(s))] - \mathcal{A}(\alpha(s; \overline{X}(s)); M_k) \right| ds$$

so that a combination of Chebychev’s inequality and Cauchy-Schwartz inequality yield

$$P\left( (M_k)^\alpha \sup_{t \in [0,T]} \left| \int_0^t (E[\alpha(s; \overline{X}(s))] - \mathcal{A}(\alpha(s; \overline{X}(s)); M_k)) ds \right| \geq \epsilon \right)$$

$$\leq \frac{M_k^{2\alpha}}{\epsilon^2} E \left[ \left( \int_0^T \left| E[\alpha(s; \overline{X}(s))] - \mathcal{A}(\alpha(s; \overline{X}(s)); M_k) \right| ds \right)^2 \right]$$

$$\leq \frac{M_k^{2\alpha}}{\epsilon^2} \frac{T}{M_k^{1-2\alpha}} \int_0^T E \left[ \left( E[\alpha(s; \overline{X}(s))] - \mathcal{A}(\alpha(s; \overline{X}(s)); M_k) \right)^2 \right] ds$$

$$= \frac{T}{\epsilon^2 M_k^{1-2\alpha}} \int_0^T \text{Var}[\alpha(s; \overline{X}(s))] ds \leq \frac{C_T}{\epsilon^2 M_k^{1-2\alpha}}.$$ 

The analogous Borel-Cantelli arguments (4.32-4.33) therefore imply for $\alpha < 1/2$

$$\sum_{k=1}^{\infty} P\left( (M_k)^\alpha \sup_{t \in [0,T]} \left| \int_0^t E[\alpha(s; \overline{X}(s))] - \mathcal{A}(\alpha(s; \overline{X}(s)); M_k) ds \right| \to 0 \text{ a.s.} \right)$$

$$= 1.$$
which combined with (4.33) proves the lemma.

\[ \square \]

**Theorem 4.5** (Convergence). Suppose the assumptions of Lemma 4.4 hold. Then the error density \( \tilde{\rho} \) in (2.13) converges a.s. to \( E[\tilde{\rho}] \), defined in Corollary 4.3, as the specified error tolerance tends to zero.

\[ \text{Proof.}\] First split \( E[\tilde{\rho}] - A(\tilde{\rho}; M) = E[\tilde{\rho} - \bar{\rho}] + (E[\bar{\rho}] - A(\bar{\rho}; M)) \). The first term is bounded by \( O(\Delta t_{\text{sup}}) \) using, for the augmented system in (4.29), the weak convergence of order one, proved for both stochastic and deterministic time steps in Theorem 2.2. The a.s. convergence of the second term follows from applying Lemma 4.4 to the augmented system in (4.29). \[ \square \]

5. Numerical Experiments

This section presents numerical results from the implementation of the adaptive algorithms described in Section 3, namely the adaptive algorithms with deterministic time steps (Algorithm D) and with stochastic time steps (Algorithm S).

**Algorithm D** uses a Matlab version 6 implementation and simulates the \( \ell_0 \) independent Wiener processes with a pseudo-random number generator, based on either a linear congruential recursion,

\[ u_{k+1} = (au_k + c) \mod m \]

with \( a = 16807, c = 0 \) and \( m = 2^{31} - 1 \), or on a more advanced random number generator that uses several seeds proposed by Marsaglia, see [23].

**Algorithm S** uses a double precision FORTRAN 77 implementation, with the same linear congruential pseudo-random number generator (5.1). In particular, the program applies a double precision modification of the functions ran1 and gasdev proposed in [30], provided with an initial seed.

In all computations the number of subdivisions of a refined step is \( H = 2 \), and the constants to determine the number of realizations in (3.6) are \( c_0 = 1.65 \) and \( MCH = 16 \).

We prescribe the error tolerance \( TOL = 0.025 \) with the splitting \( TOL_S = \frac{2}{3}TOL \) and \( TOL_T = \frac{1}{3}TOL \) and Algorithm D uses \( TOL_T = TOL_{TS} + TOL_{TT} \), where \( TOL_{TS} = \frac{1}{3}TOL \) and \( TOL_{TT} = \frac{2}{9}TOL \). Observe that the choice \( TOL_{TT} = \frac{2}{9}TOL \) instead of \( TOL_{TT} = \frac{1}{3}TOL \) will in general give a final number of time steps in Algorithm D that is around 50\% larger than the expected value of the number of time steps in Algorithm S. However, this apparent disadvantage is largely overcome by the fact that in Algorithm S we have to compute the time error approximations for all realizations, while in Algorithm D this is done just in the first loop, which usually has a smaller number of realizations.

Unless otherwise stated, the time adaptivity procedures in Algorithm D and Algorithm S are applied with initial uniform partitions of \([0, T]\) having \( N[1] = 20 \) subintervals and the initial number of realizations is \( M[1] = 64 \).

The constants for the merging and dividing procedure are taken in accordance with Remarks 3.1 and 3.7 and the corresponding stopping results in Theorems 3.2 and 3.8. Indeed, we use \( d_1 = s_1 = 2 \) and \( d_2 = s_2 = 1/40 \), while the corresponding constants for the stopping criteria are \( D_1 = S_1 = 8 \) and \( D_2 = S_2 = 1/160 \), respectively.

In order to reduce the computational effort, we use antithetic variates, see [13], to reduce the variance in some computations.
To see the practical effect of the merging procedure, both algorithms are tested with and without it. The results with merging correspond to the table values Merge = 1 and similarly, the results with the antithetic variance reduction correspond to Anti = 1. Due to the construction of our algorithms, the number of realizations, $M$, is a power of 2, see (3.6). In fact, the final number of realizations is either $M = 2^{15} = 32768$ or $M = 2^{17} = 131072$, depending on the use of variance reduction and the accuracy level, $TOL = 0.025$.

In Algorithm S the total amount of computational work is proportional to the total number of time steps performed in all realizations. The latter is shown for all the computations under the table entry $\sum MN$. Similarly, in Algorithm D we show for each computation the table entry $\sum MN$, which corresponds to the total number of time steps in all realizations of the first loop that determines the mesh, and $2^{nd}\sum MN$, which corresponds to the total number of time steps in all realizations of the second loop with fixed mesh. We make a distinction between the amount of work in the first and the second loop since only the first entails the computation of time discretization error approximations, and therefore the first loop uses much more computational effort per time step.

For each computation with Algorithm D we show the final number of time steps, $N$, as well as an estimate of the optimal number of time steps, $N^*$, that would be required to achieve the desired accuracy. The optimal number of time steps is based on the approximate error density defined in (2.13) and (3.8), and it is given by $N^* = (\int_0^T \sqrt{\rho(\tau)}d\tau)^2/TOLTT$.

Similarly, for each of the computations with Algorithm S we show the sample average of the final number of time steps, $A(N;M)$ and its sample standard deviation, $S(N;M)$. We use the estimate $\left(\frac{A(\int_0^T \sqrt{\rho(\tau)}d\tau;M)}{\sqrt{TOLTT}}\right)^2$ for the optimal expected number of time steps that would be required to achieve the desired accuracy, based on the approximate error density defined in (2.6) and (3.8).

**Example 5.1.**

Let us consider a real constant $\alpha \in (0, T)$ and the linear stochastic differential equation

$$
\begin{align*}
\frac{dX(t)}{dt} & = \begin{cases} 
X(t)dW(t), & t \in [0, \alpha] \\
\frac{X(t)dt}{2\sqrt{T-\alpha}} + X(t)dW(t), & t \in (\alpha, T] 
\end{cases} \\
X(0) & = 1
\end{align*}
$$

with the unique solution

$$
X(t) = \begin{cases} 
\exp(W(t) - \frac{1}{2}t), & t \in [0, \alpha] \\
\exp(W(t) - \frac{1}{2}t)\exp(\sqrt{T-\alpha}), & t \in [\alpha, T].
\end{cases}
$$

Here we choose $T = 1$ and $\alpha = T/3$. Our goal is to approximate $E[X(T)] = \exp(\sqrt{T-\alpha})$. To avoid evaluating arbitrarily large values of the drift in (5.2) we modify the drift to be

$$
a(t, x) = \begin{cases} 
0, & t \in [0, \alpha] \\
\frac{x}{2\sqrt{T-\alpha + TOL^2}}, & t \in (\alpha, T]
\end{cases}
$$
yielding a higher order perturbation $O(TOL^2)$ in the computed result and in the
size of the optimal time steps. Due to the time discontinuity of the drift function
and to ensure optimal convergence of the adaptive algorithms, we modify the Euler
method by
\[ X_{n+1} - X_n = a(\hat{t}, X_n) \Delta t_n + X_n \Delta W_n, \quad n = 0, \ldots \]
where we choose the stochastic evaluation time $\hat{t} \in \{t_n, t_{n+1}\}$ by
\[ |a(\hat{t}, X_n)| = \max(|a(t_n, X_n)|, |a(t_{n+1}, X_n)|). \]
Observe that the use of $\hat{t}$ does not change the adapted nature of the Euler method.

We compare the results of Algorithm D and Algorithm S in Tables 1, 2, 3 and 4. First, Tables 1 and 2 show results from Algorithm D, the only difference being
the choice of the pseudo random number generator. Results in Table 1 use [23] and
those in Table 2 are based on the linear congruential random number generator
(5.1). As expected, the level of difference between these results is comparable to
the effect of taking different initial seeds in the computations.

Next, Tables 3 and 4 show results from Algorithm S. Table 3 shows results when
the time adaptivity procedure is started with the same mesh for each realization,
namely a uniform mesh with 20 time steps, whereas the results of Table 4 use that
the time adaptivity procedure is started with the optimal mesh from the previous
realization, yielding less computational work since the drift singularity occurs at
a deterministic time. Table 4 shows a situation where merging is needed, namely
when using the stochastic time stepping algorithm and the optimal mesh from the
previous realization as a starting guess for the next realization. In this case the
merging procedure avoids an excessive increment of the average number of time
steps.

In this example Algorithm D and Algorithm S select the final number of time
steps in a similar way, since there is no remarkable influence from the stochastic
term $X(t)dW(t)$ in the dynamics of (5.2). Besides this, Figure 3 shows that both
Algorithm D and Algorithm S have a remarkable advantage over a computation
with uniform time steps which needs $1.7 \times 10^5$ time steps to achieve the given
level of accuracy $TOL_{FR} \approx 5.6 \times 10^{-3}$, compared to on average less than 600 final
adaptive steps. However, there is a noticeable difference in the total number of operations in all levels, and the deterministic time stepping algorithm has a clear
advantage. Table 1 shows that variance reduction is useful in the second loop of
the deterministic time stepping algorithm where the statistical error to control has
an associated large variance. Finally, there is no significant difference between the
results with the merging procedure or without merging and in addition, the final
number of time steps is close to the estimated optimal number of time steps.

<table>
<thead>
<tr>
<th>Merge</th>
<th>Anti</th>
<th>$M$</th>
<th>$N$</th>
<th>$1^{st} \sum MN$</th>
<th>$2^{nd} \sum MN$</th>
<th>$N^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>$2^{13}$</td>
<td>$4.8 \times 10^2$</td>
<td>$0.6 \times 10^6$</td>
<td>$48 \times 10^6$</td>
<td>$4.2 \times 10^2$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>$2^{17}$</td>
<td>$5.2 \times 10^2$</td>
<td>$0.3 \times 10^6$</td>
<td>$77 \times 10^6$</td>
<td>$4.7 \times 10^2$</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>$2^{15}$</td>
<td>$4.6 \times 10^2$</td>
<td>$0.6 \times 10^6$</td>
<td>$46 \times 10^6$</td>
<td>$4.4 \times 10^2$</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>$2^{17}$</td>
<td>$5.0 \times 10^2$</td>
<td>$0.3 \times 10^6$</td>
<td>$74 \times 10^6$</td>
<td>$4.4 \times 10^2$</td>
</tr>
</tbody>
</table>

Table 1. (Example 5.1) Numerical results using the Matlab 6
pseudo random number generator, see [23].
Algorithm D

\[ \sum_{M}^{1} 2^{15} \times 10^2 \quad 0.6 \times 10^6 \quad 50 \times 10^6 \quad 4.1 \times 10^2 \]

\[ \sum_{M}^{2} 2^{17} \times 10^2 \quad 0.35 \times 10^6 \quad 95 \times 10^6 \quad 5.8 \times 10^2 \]

\[ \sum_{M}^{N} 2^{15} \times 10^2 \quad 0.6 \times 10^6 \quad 50 \times 10^6 \quad 4.9 \times 10^2 \]

\[ \sum_{M}^{N^*} 2^{17} \times 10^2 \quad 0.35 \times 10^6 \quad 85 \times 10^6 \quad 4.5 \times 10^2 \]

Table 2. (Example 5.1) Similar to Table 1, but using the linear congruential pseudo random number generator from (5.1) to simulate the Wiener increments.

Algorithm S

\[ \sum_{M}^{1} 2^{15} \times 10^2 \quad 2.2 \times 10^2 \quad 4 \times 10 \quad 170 \times 10^6 \]

\[ \sum_{M}^{2} 2^{15} \times 10^2 \quad 2.3 \times 10^2 \quad 4 \times 10 \quad 175 \times 10^6 \]

Table 3. (Example 5.1) For each realization the time adaptivity procedure is started with the same uniform mesh that has 20 time steps. This computation uses antithetic variates and the estimate for the expected value of the optimal number of time steps is in this case \(2.1 \times 10^2\).

Algorithm S

\[ \sum_{M}^{1} 2^{15} \times 10^2 \quad 5.8 \times 10^2 \quad 1.5 \times 10^2 \quad 56 \times 10^6 \]

Table 4. (Example 5.1) For each realization the time adaptivity procedure is started with the optimal mesh from the previous realization. This computation uses both antithetic variates and the merging procedure. The estimate for the expected value of the optimal number of time steps is in this case \(2.1 \times 10^2\).

Example 5.2.

Now we change (5.2) taking \(\alpha\) no longer a constant but a uniformly distributed random variable independent of the Wiener process, i.e. \(\alpha \sim U(\underline{\alpha}, \bar{\alpha})\) with \(\underline{\alpha} = 1/22\) and \(\bar{\alpha} = 1 - 1/22\), for \(T = 1\). The conditional expectation \(E[X(T)|\alpha]\) can be used to compute

\[ E[X(T)] = \frac{1}{\bar{\alpha} - \underline{\alpha}} \int_{\underline{\alpha}}^{\bar{\alpha}} \exp(\sqrt{T - \alpha}) d\alpha, \]

which is now the functional to approximate. Since the position of the drift singularity is stochastic, the deterministic time stepping Algorithm D gives the approximation error \(O(\sqrt{\Delta t_{\text{sup}}})\), so that \(N_D \sim \text{TOL}^2\); and Algorithm D behaves like a uniform time stepping algorithm which needs \(1.3 \times 10^3\) time steps to achieve the given level of accuracy \(\text{TOL}_T \approx 5.6 \times 10^{-3}\). Thus, Algorithm S has a clear advantage in this example, with asymptotic accuracy \(O(\Delta t_{\text{sup}})\), so that \(E[N_S] \sim \text{TOL}^{-1}\) using on average less than 200 final time steps, as Table 5 and Figure 4 show.
Figure 2. (Example 5.1). Computation with variance reduction and the merging procedure. Final mesh function for Algorithm D and TOL = 0.025.

Figure 3. (Example 5.1). Computations with uniform time steps and variance reduction. The computational order of convergence for the time discretization is 0.49 and the estimated number of time steps to achieve the accuracy $TOL_{TT} \approx 5.6 \times 10^{-3}$ becomes $1.7 \times 10^5$. The computational time discretization error is given by $A(g(X(T)) - g(\bar{X}(T)))/2^{15}$.

References
Algorithm S

<table>
<thead>
<tr>
<th>Merge</th>
<th>$M$</th>
<th>$A(N; M)$</th>
<th>$S(N; M)$</th>
<th>$\sum MN$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$2^{15}$</td>
<td>$1.8 \times 10^2$</td>
<td>$5 \times 10$</td>
<td>$140 \times 10^9$</td>
</tr>
<tr>
<td>0</td>
<td>$2^{15}$</td>
<td>$1.9 \times 10^2$</td>
<td>$5 \times 10$</td>
<td>$146 \times 10^9$</td>
</tr>
</tbody>
</table>

Table 5. (Example 5.2) Both computations use antithetic variates and in this case there is no significant advantage in using the merging procedure. The estimate for the expected value of the optimal number of time steps is in this case $1.7 \times 10^2$.

Figure 4. (Example 5.2). Computations with uniform time steps and variance reduction. The order of convergence for the time discretization is 0.49 and the estimated number of time steps to achieve the accuracy $TOL_{TT} \approx 5.6 \times 10^{-3}$ becomes $1.3 \times 10^5$. The computational time discretization error is given by $A(g(X(T)) - g(X(T)))$.

MONTE CARLO EULER APPROXIMATION OF HJM TERM STRUCTURE
FINANCIAL MODELS

T. BJÖRK†, A. SZEPESSY†, R. TEMPONE§, AND G. ZOURARIS‡

Abstract. We present two Monte Carlo Euler methods for a weak approximation problem of the Heath Jarrow Morton (HJM) term structure model, based on Itô stochastic differential equations in infinite dimensional spaces, and prove error estimates useful for adaptive algorithms. These error estimates are based on stochastic flows and discrete dual backward problems, and they can be used to identify different error contributions arising from time and maturity discretization as well as the classical statistical error due to finite sampling. Explicit formulas for efficient computation of sharp error approximation are included. Due to the structure of the HJM models considered here, the computational effort devoted to the error estimates is low compared to the work to compute Monte Carlo solutions to the HJM model. Numerical examples with known exact solution are included in order to show the behavior of the estimates.

1. The HJM model

When valuing derivatives in the bond market it is important to use models that are consistent with the initial term structure observed in the market. The Heath-Jarrow-Morton model for the forward rate has this property and in addition there is freedom to choose the volatility structure, for example to be able to fit other derivative prices quoted in the market, see [6, 7, 15, 19]. This Heath Jarrow Morton model approach is particularly suitable for Monte Carlo computations, since in general the alternative of tree methods leads, for the multifactor case, to non recombining trees with higher computational cost.

This work studies numerical solutions for the price of financial instruments in the bond market, using the Heath Jarrow Morton model of forward rates. The main contribution is to provide rigorous error expansions, with leading error term in computable a posteriori form, offering computational reliability in the use of more complicated HJM multifactor models, where no explicit formula can be found, or such a formula is just too complicated to use, for the pricing of contingent claims. These error estimates can be used in adaptive algorithms to handle simultaneously different sources of error, e.g. time discretization, maturity discretization, and finite sampling, see [21]. To develop error estimates we use a Kolmogorov backward equation in an extended domain and carry out further the analysis in [21], from general weak approximation of Itô stochastic differential equations in \( \mathbb{R}^n \), to weak approximation of the HJM Itô stochastic differential equations in infinite dimensional spaces. Therefore, the main new ingredient here is to provide error estimates useful for adaptive refinement not only in time \( t \) but also in maturity time \( \tau \). In addition, using the structure of the HJM model studied here, the application of a simple transformation removes the error

1991 Mathematics Subject Classification. Primary 65C05, 65C30, 65C20 Secondary 91B28, 91B70.

Key words and phrases. A posteriori error estimates, stochastic differential equations, Monte Carlo methods, HJM model, option price, bond market.

†Institutionen för finansiell ekonomi, Handelshögskolan, Box 6501, S-113 83 Stockholm.
†Matematiska Institutionen, Kungl. Tekniska Högskolan, S-100 44 Stockholm.
§Department of Numerical Analysis and Computing Science (NADA), Royal Institute of Technology (KTH), S-100 44 Stockholm, Sweden.
‡Department of Mathematics, University of the Aegean, GR-832 00 Karlovassi, Samos, Greece, and Institute of Applied and Computational Mathematics, FO.R.T.H., Heraklion, Crete, Greece.

1
caused by the representation of the initial term structure in a finite maturity partition. Finally, the formulas to compute sharp error approximations are simplified by exploiting the structure of the HJM model, reducing the work to compute such error estimates. The use of the error estimates proposed here is compatible with the application of variance reduction techniques, allowing for faster Monte Carlo computations, see [4].

The bond market is assumed to be efficient and without friction, i.e. there is no arbitrage opportunity, and there exists a martingale probability measure, under which bond contracts can be priced as expected values of properly discounted cash flows, see [1, 3, 9]. On what follows, all the equations are assumed to be under such a probability measure.

The HJM model is based on the so called forward rate, $f(t, \tau)$, which relates to the price of the most simple type of bond, the zero coupon bond, with contracting time $t$ and maturity time $\tau$, by

$$p(t, \tau) = \exp\left(-\int_t^\tau f(t, \eta)d\eta\right).$$

The equations are assumed to be under such a probability measure.

In particular, the non arbitrage assumption in the HJM formulation, see [13, 14], yields an Itô stochastic differential equation, for $\tau \in [0, \tau_{\text{max}}]$,

$$df(t, \tau) = \sum_{j=1}^J \sigma^j(t, \tau)\left(\int_t^\tau \sigma^j(t, s)ds\right)dt + \sum_{j=1}^J \sigma^j(t, \tau)dW^j(t), \quad t \in [0, t_{\text{max}}]$$

$$f(0, \tau) = f_0(\tau).$$

Here $W^j$, $j = 1, \ldots, J$ are independent Wiener processes, and $\sigma^j(t, \tau)$, $j = 1, \ldots, J$ are stochastic processes, adapted to the filter structure generated by the Wiener processes. Furthermore, the initial datum for the term structure, $f_0 : [0, \tau_{\text{max}}] \to \mathbb{R}$, is a deterministic function in $C^1([0, \tau_{\text{max}}])$. In this setting, the short rate, $r(t)$, is defined as $r(t) \equiv f(t, t)$.

On what follows the volatility function $\sigma = (\sigma^1, \ldots, \sigma^J)$ is assumed to be of the form

$$\sigma(t, \tau) = \xi(r(t)) \lambda(t, \tau)$$

$$= \xi(f(t, t)) \lambda(t, \tau),$$

where $\xi : \mathbb{R} \to \mathbb{R}$ and $\lambda : [0, t_{\text{max}}] \times [0, \tau_{\text{max}}] \to \mathbb{R}^J$ are given bounded functions on $C^m(\mathbb{R})$ and $C^m([0, t_{\text{max}}] \times [0, \tau_{\text{max}}])$, respectively, for $m_0$ a sufficiently large integer. Then, problem (1.1) reads: for $t \in [0, t_{\text{max}}]$ and $\tau \in [0, \tau_{\text{max}}]$, find $f(t, \tau)$ such that

$$df(t, \tau) = \xi^2(f(t, t)) \tilde{\lambda}(t, \tau)dt + \xi(f(t, t)) \lambda(t, \tau) \cdot dW(t), \quad t \leq \tau$$

$$f(0, \tau) = f_0(\tau), \quad \tau \in [0, \tau_{\text{max}}]$$

where

$$\tilde{\lambda}(t, \tau) \equiv \lambda(t, \tau) \cdot \int_t^\tau \lambda(\tau, z)dz \quad \text{for} \quad t \in [0, t_{\text{max}}], \text{and} \quad \tau \in [0, t_{\text{max}}]$$

Here the notation $a \cdot b$ denotes the standard inner product in $\mathbb{R}^J$, i.e. $a \cdot b \equiv \sum_{j=1}^J a_j b_j$.

In many models used in practice, the function $\lambda$ has the form $\lambda(t, \tau) = \lambda_0(\tau - t)$, and then

$$\tilde{\lambda}(t, \tau) = \lambda_0(\tau - t) \cdot \int_0^{\tau-t} \lambda_0(\tau - z)dz \equiv \tilde{\lambda}_0(\tau - t).$$

Observe that to solve for $f$ it is enough to have $\lambda_0 : \mathbb{R}^+ \to \mathbb{R}$. This work extends the usual domain of definition of $\lambda$ and $f$, $D \equiv \{ (t, \tau) \in [0, t_{\text{max}}] \times [0, \tau_{\text{max}}] : t \leq \tau \}$, to the set $[0, t_{\text{max}}] \times [0, \tau_{\text{max}}]$, leaving $f|_D$ unchanged. The extension of $D$ helps to develop a posteriori approximations for the
time and maturity discretization errors, depending on a linear backward problem, cf. Theorem 3.1.

A typical contract to price is a call option, with exercise time \( t_{\text{max}} \) and strike price \( K \), on a zero coupon bond. Its price can be written in terms of the forward rate as
\[
E \left[ \exp \left( - \int_0^{t_{\text{max}}} f(s,s) \, ds \right) \max \left\{ \exp \left( - \int_0^{t_{\text{max}}} f(t_{\text{max}}, \tau) \, d\tau \right) - K, 0 \right\} \right].
\]

Another basic contract is a continuous cap, with price
\[
E \left[ \int_0^{t_{\text{max}}} \exp \left( - \int_0^s f(u,u) \, du \right) (f(s,s) - r_c)^+ \, ds \right].
\]

With this motivation, and bearing in mind other possible contracts, we consider the approximation of functionals
\[
E[\mathcal{F}(f)] \equiv E \left[ F \left( \int_0^{t_{\text{max}}} f(s,s) \, ds \right) \, G \left( \int_{t_{\text{max}}}^{t_{\text{max}}} Q(f(t_{\text{max}}, \tau)) \, d\tau \right) + \int_0^{t_{\text{max}}} F(Y(s)) U(f(s,s)) \, ds \right]
\equiv E \left[ F \left( Y(t_{\text{max}}) \right) \, G \left( \mathcal{I}_Q \right) + Z(t_{\text{max}}) \right]
\] (1.4)

where
\[
\mathcal{I}_Q \equiv \int_{t_{\text{max}}}^{t_{\text{max}}} Q(f(t_{\text{max}}, \tau)) \, d\tau, \quad Y(t) \equiv \int_0^t f(s,s) \, ds, \quad \text{and} \quad Z(t) \equiv \int_0^t F(Y(s)) U(f(s,s)) \, ds.
\] (1.5)

On what follows, we say that a function \( S : \mathbb{R} \to \mathbb{R} \) has polynomial growth if there exist positive constants \( k, C \) such that \( |S(x)| \leq C(1 + |x|^k) \) for all \( x \in \mathbb{R} \).

The functions \( F : \mathbb{R} \to \mathbb{R}, G : \mathbb{R} \to \mathbb{R}, Q : \mathbb{R} \to \mathbb{R}, U : \mathbb{R} \to \mathbb{R}, \) and their derivatives up to a sufficiently large order \( m_0 \) are assumed to have polynomial growth. Beside this \( 0 < t_{\text{max}} \leq \tau_a < \tau_{\text{max}} \) are given positive numbers.

Although we have considered contracts with continuous time cash flows given by the function \( U \), the analysis in this work can be easily suited to discrete time cash flows, cf. Remark 3.4.

In addition, the error estimates included in this work can be modified to accomodate different contingent claims, like e.g. swaptions.

**Remark 1.1.** Let us consider the system of differential equations describing the dynamics for the forward rates (1.2-1.3), \( Y(t) \) and \( Z(t) \), \( i.e. \), let, for \( t \in [0,t_{\text{max}}] \), \( \tau(t) \equiv f(t,t) \) and
\[
df(t,\tau) = \xi(t) \, dt + \eta(t) \, dW(t), \quad \tau(t) \in [0,\tau_{\text{max}}],
\]
\[
dY(t) = \nu(t) \, dt,
\]
\[
dZ(t) = F(Y(t))U(\nu(t)) \, dt,
\]
with the initial conditions
\[
f(0,\tau) = f_0(\tau), \quad \tau \in [0,\tau_{\text{max}}]
\]
\[
Y(0) = 0,
\]
\[
Z(0) = 0.
\] (1.7)

In the discretizations given below, \( Y(t) \) and \( Z(t) \) will be always considered to be the last components of the approximate solution vector.

**Remark 1.2** (Removing the error from initial datum). As we shall see below, the maturity discretization error has two sources, namely the error in the \( \tau \) discretization of the dynamics (1.2) and
the error arising from representing the initial data, \( f_0 \), in the \( \tau \)-partition. Due to the special structure of (1.2-1.3), the initial error can be avoided by using the difference \( g(t, \tau) = f(t, \tau) - f_0(\tau) \), which implies \( r(t) \equiv g(t, t) + f_0(t) \), and

\[
\begin{align*}
\mathcal{d}g(t, \tau) &= \xi^2(r(t)) \tilde{\lambda}(t, \tau) dt + \xi(r(t)) \lambda(t, \tau) \cdot dW(t), \quad \tau \in [0, \tau_{\text{max}}], \\
\mathcal{d}Y(t) &= r(t) dt, \\
\mathcal{d}Z(t) &= F(Y(t)) U(r(t)) dt,
\end{align*}
\]

with homogeneous initial conditions

\[
\begin{align*}
g(0, \tau) &= 0, \quad \tau \in [0, \tau_{\text{max}}] \\
Y(0) &= 0, \\
Z(0) &= 0.
\end{align*}
\]

Here as before

\[
\tilde{\lambda}(t, \tau) \equiv \lambda(t, \tau) \cdot \int_{t}^{\tau} \lambda(t, z) dz.
\]

The computations below use different discretizations of (1.8-1.9), so the \( \tau \) approximation error does not have a contribution from the initial datum. The work is organized as follows: Section 2 first presents two Monte Carlo Euler methods: a stochastic finite difference method, the Euler Finite Difference (EFD), and a more accurate stochastic finite element method, the Euler Finite Element (EFE), for the Heath Jarrow Morton model (1.8-1.9). Both the EFD and the EFE methods approximate the solution process \( g \) in (1.8-1.9) with processes that are piecewise constant in maturity, but they differ in the maturity discretization of (1.8). Then, Section 2 provides a numerical approximation for the functional \( F(f) \) defined in (1.4-1.5) for both the EFD and the EFE methods. Section 3 states and proves error estimates for the EFD method, giving explicit formulas for efficient computation of the discrete duals. Finally, Section 4 presents numerical examples, giving computational evidence to the error analysis developed in Section 3.

2. Two Monte Carlo-Euler methods

The purpose of this section is to introduce two time and maturity discretizations of (1.8-1.9): the Euler-Finite Difference (EFD) method and the Euler-Finite Element (EFE) method. The discretizations include a numerical quadrature of the functional \( F(f) \) defined in (1.4), see (2.8-2.9).

2.1. Time and maturity discretization. Given extreme points \( 0 < t_{\text{max}} \leq \tau_{\text{a}} < \tau_{\text{max}} \) introduced in (1.4-1.5), let \( N \) and \( L \) denote the number of subintervals on \([0, t_{\text{max}}]\) and \([0, \tau_{\text{max}}]\), respectively. Then, consider partitions

\[
0 = t_0 < \cdots < t_N = t_{\text{max}} \quad \text{and} \quad 0 = \tau_0 < \cdots < \tau_L = \tau_{\text{max}}
\]

of the \( t \)-interval \([0, t_{\text{max}}]\) and of the \( \tau \)-interval \([0, \tau_{\text{max}}]\), respectively. For technical reasons, these partitions are assumed to satisfy the following condition: every \( \tau \)-node in the interval \([0, t_{\text{max}}]\) is also a \( t \)-node, i.e.

\[
\text{there exists an one-to-one index map } \rho, \text{ such that, } \tau_\ell = t_{\rho(\ell)} \text{ for } \tau_\ell \leq t_{\text{max}}. \quad (2.1)
\]

In addition, assume that

\[
\text{there exists an index } \ell_\star \text{ such that } t_{\text{max}} = \tau_{\ell_\star}. \quad (2.2)
\]
Denote by \( \ell_n \) the index corresponding to \( \tau_n \), and define the auxiliary index function, \( \ell_n \), by
\[
\ell_n \equiv \max \left\{ \ell \in \mathbb{Z} : 0 \leq \ell \leq L, \quad \text{s.t.} \quad \tau_n \leq \ell_n \right\}.
\] (2.3)

Use the notation
\[
\Delta t_n \equiv t_{n+1} - t_n, \quad \Delta W_n \equiv W(t_{n+1}) - W(t_n) \quad \text{for} \quad 0 \leq n \leq N - 1,
\]
and
\[
\Delta \tau_n \equiv \tau_{n+1} - \tau_n, \quad \text{for} \quad 0 \leq \ell \leq L - 1.
\]

Finally, introduce the space of piecewise constant and right continuous functions on a \( \tau \)-partition, \( \{\tau \}_{\ell=0}^L \), of the interval \([0, \tau_{\text{max}}]\), by
\[
S_{\Delta \tau} \equiv \left\{ \chi \in L^\infty(0, \tau_{\text{max}}) : \text{there are constants} \quad \{c_\ell\}_{\ell=0}^{L-1} \text{such that} \quad \chi\|_{[\tau_\ell, \tau_{\ell+1}]} = c_\ell \quad \text{for} \quad 0 \leq \ell \leq L-1 \right\}.
\]

Define the standard \( L^2 \)-projection \( \Pi : L^2(0, \tau_{\text{max}}) \to S_{\Delta \tau} \) by
\[
(\Pi w, \chi) = (w, \chi), \quad \forall \chi \in S_{\Delta \tau},
\]
which satisfies, for \( 0 \leq \ell \leq L - 1 \),
\[
\Pi w \big|_{[\tau_\ell, \tau_{\ell+1}]} = \frac{1}{\Delta \tau_\ell} \int_{\tau_\ell}^{\tau_{\ell+1}} w(\tau) \, d\tau.
\]

For \( \chi \in S_{\Delta \tau} \), and \( 0 \leq \ell \leq L - 1 \), denote by \( \chi_\ell \) the constant value of \( \chi \) in \([\tau_\ell, \tau_{\ell+1})\). When considering a function, \( w = w(t, \tau) \), depending on two variables, the \( L^2 \) projection is always with respect to \( \tau \), i.e. for \( \tau \in [\tau_\ell, \tau_{\ell+1}) \), \( \Pi w(t, \tau) = \frac{1}{\Delta \tau_\ell} \int_{\tau_\ell}^{\tau_{\ell+1}} w(t, s) \, ds \).

### 2.2. The Euler-Finite Difference (EFD) method.

For each time level the EFD method approximates \( g(t_n, \cdot) \) by a piecewise constant function, \( \overline{g}(t_n, \cdot) \in S_{\Delta \tau} \).

In particular, it finds the approximate values \( \overline{g}_{n,K} \approx g(t_n, \tau_k) \), \( \overline{g}_{n,L} \approx Y(t_n) \), \( \overline{g}_{n,L+1} \approx Z(t_n) \) by setting first
\[
\overline{g}_{0,\ell} \equiv 0, \quad 0 \leq \ell \leq L + 1;
\] (2.4)
and, then recursively, for \( 0 \leq n \leq N - 1 \), define \( r_n \equiv \overline{g}_{n,\ell_n} + f_0(t_n) \), with \( \ell_n \) as in (2.3), and
\[
\begin{align*}
\overline{g}_{n+1,\ell} &= \overline{g}_{n,\ell} + \xi^2(r_n) \lambda(t_n, \tau_\ell) \Delta t_n + \xi(r_n) \lambda(t_n, \tau_\ell) \cdot \Delta W_n, \quad &\text{for} \quad 0 \leq \ell \leq L - 1, \\
\overline{g}_{n+1,L} &= \overline{g}_{n,L} + r_n \Delta t_n, \\
\overline{g}_{n+1,L+1} &= \overline{g}_{n,L+1} + F(\overline{g}_{n,L}) U(r_n) \Delta t_n.
\end{align*}
\] (2.5)

### 2.3. The Euler-Finite Element (EFE) method.

The EFE method also approximates the \( \tau \)-function \( g(t_n, \cdot) \), \( \tau \in [0, \tau_{\text{max}}) \) by a piecewise constant function \( \overline{g}(t_n, \cdot) \in S_{\Delta \tau} \), but is based in a variational formulation of (1.8-1.9) with \( S_{\Delta \tau} \), being the space of trial and test functions. Let \( r_n \equiv \overline{g}_{n,\ell_n} + f_0(t_n) \), with \( \ell_n \) as in (2.3). The EFE is defined by the initial datum
\[
\overline{g}_{0,\ell} \equiv 0, \quad 0 \leq \ell \leq L + 1
\] (2.6)
and, for \( 0 \leq n \leq N - 1 \), the recursion
\[
\begin{align*}
\overline{g}_{n+1,\ell} &= \overline{g}_{n,\ell} + \xi^2(r_n) \Pi \lambda(t_n, \tau_\ell) \Delta t_n + \xi(r_n) \Pi \lambda(t_n, \tau_\ell) \cdot \Delta W_n, \quad &\text{for} \quad 0 \leq \ell \leq L - 1, \\
\overline{g}_{n+1,L} &= \overline{g}_{n,L} + r_n \Delta t_n, \\
\overline{g}_{n+1,L+1} &= \overline{g}_{n,L+1} + F(\overline{g}_{n,L}) U(r_n) \Delta t_n.
\end{align*}
\] (2.7)
2.4. Approximation of the functional (1.4-1.5). The numerical approximation of \( \mathcal{F}(f) = \mathcal{F}(g + f_0) \) involves both an approximation of the processes \( g, Y, Z \), by computable quantities, and an approximation of the \( \tau \) integral \( \mathcal{I}_Q \) in (1.5) by some quadrature rule of order \( p \). For a fixed realization of \( \overline{g} \) and a smooth function \( f_0 \) this quadrature takes the form
\[
\int_{\tau_n}^{\tau_{\text{max}}} Q(\overline{g}(t_{\text{max}}, \tau))d\tau = \int_{\tau_n}^{\tau_{\text{max}}} Q(\overline{g}(t_{\text{max}}, \tau) + f_0(\tau))d\tau \approx \mathcal{I}_Q
\]
where
\[
\mathcal{I}_Q = \sum_{\ell=1}^{L-1}\{\mathcal{I}_Q, \Delta \tau_{\ell}\}, \quad \text{and} \quad \hat{\mathcal{I}}_{Q,\ell} \equiv \sum_{i=1}^{N_{\text{quad}}} \{ v_i Q(\overline{g}_{N,\ell} + f_0(\tau_i + s_i \Delta \tau_{\ell})) \}.
\]
Note that \( \overline{g}(t, \cdot) \) is piecewise constant for fixed \( t \) and the quadrature error is caused only by approximation of the initial datum \( f_0 \). In particular, if the initial datum for the term structure, \( f_0 \), is a piecewise constant function on the maturity partition, then there is no quadrature error. Here \( s_i, v_i, \ i = 1, \ldots, N_{\text{quad}} \) are the nodes and the weights of a numerical quadrature rule of order \( p \), i.e. a numerical quadrature that integrates exactly all polynomials up to degree \( p \). For example, the Simpson’s rule has \( s = (0, 1, 2, 1) \) and \( v = (1, 4, 1, 4) \), with \( p = 3 \). If instead the Gaussian quadrature with two nodes is applied then \( s = (\frac{1}{2} - \frac{\sqrt{3}}{2}, \frac{1}{2} + \frac{\sqrt{3}}{2}) \) and \( v = (1, 2) \), with \( p = 4 \). Now, using the definition (2.9) of \( \hat{\mathcal{I}}_Q \), we compute the approximation \( \mathcal{F}(\overline{g} + f_0) \) of \( \mathcal{F}(f) \) by
\[
\mathcal{F}(\overline{g} + f_0) \equiv \mathcal{F}(\overline{g}_{N,L}) \ G(\mathcal{I}_Q) + \overline{g}_{N,L+1}
\]
for both the stochastic finite difference and the stochastic finite element methods.

The Monte Carlo method, [16] approximates the expectation of a given random variable \( X \) by a sample average of \( M \) independent realizations of \( X \), i.e. \( E[X] \approx A(M; X) \equiv \frac{1}{M} \sum_{j=1}^{M} X(\omega_j) \). In particular, here we approximate \( E[\mathcal{F}(f)] \) by a sample average of \( \mathcal{F}(\overline{g} + f_0) \),
\[
A(M; \mathcal{F}(\overline{g}_{N,L}) \ G(\mathcal{I}_Q) + \overline{g}_{N,L+1}) \equiv \frac{1}{M} \sum_{h=1}^{M} \left( \mathcal{F}(\overline{g}_{N,L}(\omega_h)) \ G(\hat{\mathcal{I}}_Q(\omega_h)) + \overline{g}_{N,L+1}(\omega_h) \right).
\]

Therefore, the exact computational error, \( \mathcal{E}_c \), naturally separates into three error contributions as follows:
\[
\mathcal{E}_c = E[\mathcal{F}(g + f_0)] - A(M; \mathcal{F}(\overline{g} + f_0))
\]
\[
= E\left[ \mathcal{F}(g + f_0) - \mathcal{F}(\overline{g} + f_0) \right]
\]
\[
+ E\left[ \mathcal{F}(\overline{g} + f_0) - \mathcal{F}(\overline{g} + f_0) \right]
\]
\[
+ \left[ \mathcal{F}(\overline{g} + f_0) - A(M; \mathcal{F}(\overline{g} + f_0)) \right]
\]
\[
\equiv E_{\tau,t} + E_Q + E_S = \begin{cases} \tau, t \text{ Discretization} \quad \text{Quadrature} \quad \text{Statistical} \end{cases}
\]
where \( E_{\tau,t} \) is the error contribution from \( t \) and \( \tau \) discretization, \( E_Q \) is the quadrature error in (2.8), and \( E_S \) is the statistical error. Theorem 3.1 below gives an estimate of \( E_{\tau,t} \) which, as the step size of both the time and maturity partitions goes to zero and the number of realizations goes to infinity, is asymptotically correct for the EFD method. On the other hand, the statistical error \( E_S \) can be analyzed by the Central Limit Theorem, a standard procedure in Monte Carlo methods, cf. Section 4.

Then Lemma 3.1 estimates the quadrature error. Provided that the order of the numerical quadrature in (2.9) is sufficiently large, the quadrature error, \( E_Q \), is a higher order term in the
expansion of the computational error. Here we choose to give explicit computable expressions only for the leading order terms in an expansion of \((2.12)\). In particular, to keep \(E_Q\) as a higher order term the order \(p\) of the quadrature rule needs to be greater than 2 for the EFD method and greater than 3 for the EFE method.

3. A posteriori error approximation

The goal of this section is to develop a computable approximation for the error \(E\), defined in \((2.12)\). This section presents a thorough analysis for the stochastic finite difference method, including an approximation of the time and maturity discretization given in Theorem 3.1 and an a priori estimate of the higher order quadrature based on Theorem 3.2.

3.1. Error approximation for the EFD method. Let the process \(g\) solve
\[
dg(t, \tau) = a(t, \tau, g(t, t))dt + b(t, \tau, g(t, t)) \cdot dW(t),
\]
\[
g(0, \tau) = 0, \quad \tau \in [0, \tau_{\text{max}}]
\]
with \(a(t, \tau, x) \equiv \xi^2(x + f_0(t)) \cdot \overline{\lambda}(t, \tau)\), and \(b(t, \tau, x) \equiv \xi(x + f_0(t)) \cdot \lambda(t, \tau)\). Here as before, \(\xi : \mathbb{R} \to \mathbb{R}\) and \(\lambda : [0, \tau_{\text{max}}] \times [0, \tau_{\text{max}}] \to \mathbb{R}^J\) are given bounded functions on \(C^m_0(\mathbb{R})\) and \(C^m_0([0, \tau_{\text{max}}] \times [0, \tau_{\text{max}}])\), respectively, for \(m_0\) sufficiently large. The Monte Carlo method approximates the unknown process \(g(t, \tau)\) by a time and maturity discretization \(\overline{g}(t, \tau)\), with time nodes \(0 = t_0 < t_1 < \cdots < t_N = \tau_{\text{max}}\) and maturity nodes \(0 = \tau_0 < \tau_1 < \cdots < \tau_L = \tau_{\text{max}}\), based on the Euler method, which for \(n = 0, \ldots, N - 1\) reads
\[
\overline{g}(t_{n+1}, \tau_{\ell}) - \overline{g}(t_n, \tau_{\ell}) = a(t_n, \tau_{\ell}, \overline{g}(t_n, t_n)) \Delta t_n + b(t_n, \tau_{\ell}, \overline{g}(t_n, t_n)) \cdot \Delta W_n
\]
\[
\overline{g}(0, \tau_{\ell}) = 0, \quad \text{for } \ell = 0, \ldots, L
\]
In the analysis of the EFD method, it is useful to extend its definition for all times \(t\) and all maturities \(\tau\) as follows:
\[
\overline{g}(t, \tau) = a(t_n, \tau_{\ell}, \overline{g}(t_n, t_n)) (t - t_n) + b(t_n, \tau_{\ell}, \overline{g}(t_n, t_n)) \cdot (W(t) - W(t_n))
\]
\[
= \int_{t_n}^t \overline{a}(s, \tau, \overline{g}(t_n, t_n)) ds + \int_{t_n}^t \overline{b}(s, \tau, \overline{g}(t_n, t_n)) \cdot dW(s), \quad \text{for } t \in [t_n, t_{n+1}]
\]
\[
\overline{g}(0, \tau) = 0, \quad \text{for } \tau_{\ell} \leq \tau < \tau_{\ell+1} \text{ and for } \ell = 0, \ldots, L,
\]
where for \(t_n \leq t < t_{n+1}\) and \(\tau_{\ell} \leq \tau < \tau_{\ell+1}\), \(\overline{a}\) and \(\overline{b}\) are the piecewise constant approximations
\[
\overline{a}(t, \tau, x) \equiv a(t_n, \tau_{\ell}, x) = \xi^2(x + f_0(t_n)) \cdot \overline{\lambda}(t_n, \tau_{\ell}),
\]
\[
\overline{b}(t, \tau, x) \equiv b(t_n, \tau_{\ell}, x) = \xi(x + f_0(t_n)) \cdot \lambda(t_n, \tau_{\ell}).
\]
Observe that \(\overline{g}(\cdot, \cdot) \in S_{\Delta \tau}\) for any time \(t \in [0, \tau_{\text{max}}]\). Also, for \(t_n \leq t < t_{n+1}\) and \(\tau_{\ell} \leq \tau < \tau_{\ell+1}\), use the notation
\[
d(t, \tau, \tilde{\tau}, x) \equiv \xi^2(x + f_0(t))\frac{\lambda(t, \tilde{\tau}) \cdot \lambda(t, \tau)}{2}
\]
in the following theorem.

**Theorem 3.1** (Time and maturity error approximation). Suppose that \(a, b, F, G, U, g\) and \(\overline{g}\), satisfy the above assumptions and that the initial datum is \(g(0, \cdot) = \overline{g}(0, \cdot) = 0\). Then the computational error of the EFD method (3.2) has the expansion
\[
E_{\tau} = E[|F(g + f_0) - \overline{F}(g + f_0)|] = E_{\tau} + E_{\tau} + O((\Delta t)^2 + (\Delta \tau)^2)
\]
where

\[
E_r = \sum_{n=0}^{N-1} \left\{ \sum_{\ell=0}^{L-1} E \left[ \frac{a(t_n, \tau_{\ell+1}, \overline{f}(t_n, t_n)) - a(t_n, \tau_0, \overline{f}(t_n, t_n))}{2} \overline{v}_{n, \ell'} \right] \Delta \tau_\ell \right\} \Delta t_n
+ \sum_{n=0}^{N-1} \left( \sum_{\ell=0}^{L-1} \sum_{\ell'=0}^{L-1} E \left[ \frac{d(t_n, \tau_{\ell+1}, \tau_{\ell'+1}, \overline{f}(t_n, t_n)) - d(t_n, \tau_0, \tau_0, \overline{f}(t_n, t_n))}{2} \overline{v}_{n, \ell', \ell'} \right] \Delta \tau_\ell \Delta \tau_{\ell'} \right) \Delta t_n, \quad (3.7)
\]

and, recalling that \( r_n \equiv \overline{f}(t_n, t_n) + f_0(t_n) \),

\[
E_l = \sum_{n=0}^{N-1} \left\{ E \left[ \left( F(\overline{y}_{n+1, \ell}) U(r_{n+1}) - F(\overline{y}_{n, \ell'}) U(r_n) \right) \right] + E \left[ \left( r_{n+1} - r_n \right) \overline{v}_{n+1, \ell} \right] \right\} \Delta t_n
+ \sum_{n=0}^{L-1} E \left[ \left( a(t_n, \tau_0, \overline{f}(t_n, t_n)) - a(t_n, \tau_0, \overline{f}(t_n, t_n)) \right) \overline{v}_{n, \ell} \right] \Delta \tau_0 \frac{\Delta t_n}{2}
+ \sum_{n=0}^{N-1} \left( \sum_{\ell=0}^{L-1} \sum_{\ell'=0}^{L-1} E \left[ \left( d(t_n, \tau_0, \overline{f}(t_n, t_n)) - d(t_n, \tau_0, \overline{f}(t_n, t_n)) \right) \overline{v}_{n, \ell', \ell} \right] \Delta \tau_0 \Delta \tau_{\ell'} \right) \Delta t_n. \quad (3.8)
\]

The two leading order terms \( E_r + E_l \) in the right hand side of (3.6) are in a posteriori form and based on the discrete duals \( \overline{v}_{n, \ell} \in \mathbb{R}^{L+1} \) and \( \overline{v}_{n, \ell'} \in \mathbb{R}^{(L+1) \times (L+1)} \) which are determined as follows. Use the notation

\[
\overline{\tau}_{Q, \ell} \equiv \sum_{i=1}^{N_{quad}} w_i Q(\overline{y}_{N, \ell} + f_0(\tau_0 + s_i \Delta \tau_0)), \quad \overline{\tau}_Q \equiv \sum_{\ell=0}^{L-1} \overline{\tau}_{Q, \ell} \Delta \tau_0
\]

\[
\overline{\tau}'_{Q, \ell} \equiv \sum_{i=1}^{N_{quad}} w_i Q'(\overline{y}_{N, \ell} + f_0(\tau_0 + s_i \Delta \tau_0)), \quad \overline{\tau}'_Q \equiv \sum_{i=1}^{N_{quad}} w_i Q'(\overline{y}_{N, \ell} + f_0(\tau_0 + s_i \Delta \tau_0)),
\]

and

\[
c(t_n, \tau_0, x) \equiv \partial_x a(t_n, \tau_0, x) \Delta t_n + \partial_x b(t_n, \tau_0, x) \cdot \Delta W_n, \quad x \in \mathbb{R}.
\]

Then, the function \( \overline{f} \) is defined by the dual backward problem with final datum

\[
\overline{f}_{N, \ell} = \begin{cases} 
0, & 0 \leq \ell \leq \ell_n - 1 \\
F(\overline{y}_{N, L}) G(\overline{I}_Q) \overline{\tau}'_{Q, \ell}, & \ell_n \leq \ell < L - 1 \\
F'(\overline{y}_{N, L}) G(\overline{I}_Q), & \ell = L
\end{cases} \quad (3.9)
\]

and for \( 0 \leq n \leq N - 1 \)

\[
\overline{v}_{n, \ell} = \begin{cases} 
\overline{v}_{n+1, \ell}, & 0 \leq \ell \leq L - 1, \ell \neq \ell_n \\
(\overline{v}_{n+1, L} + F(\overline{y}_{n, L}) U(r_n)) \Delta t_n + \sum_{\ell=\ell_n+1}^{L-1} c(t_n, \tau_0, \overline{y}_{n, \ell}, \overline{v}_{n+1, \ell}) \overline{v}_{n+1, \ell}, & \ell = \ell_n \\
\overline{v}_{n+1, L} + F(\overline{y}_{n, L}) U(r_n) \Delta t_n, & \ell = L.
\end{cases} \quad (3.10)
\]
The second dual, \( \bar{\varphi}_{n,\ell,\tilde{\ell}} \), has final datum

\[
\bar{\varphi}_{n,\ell,\tilde{\ell}} = \begin{cases} 
F(\bar{g}_{n,L})G''(\bar{T}_Q) \bar{T}', \bar{T}'', & \text{if } \ell_n \leq \ell \neq \tilde{\ell} < L - 1, \\
F(\bar{g}_{n,L}) \left\{ G''(\bar{T}_N) (\bar{T}'_Q)^2 + G'(\bar{T}_Q) \bar{T}'_Q \right\}, & \text{if } \ell_n \leq \ell = \tilde{\ell} < L - 1, \\
0, & \text{otherwise};
\end{cases}
\]

and solves the recursion

\[
\bar{\varphi}_{n,\ell,\tilde{\ell}} = \begin{cases} 
(1 + c(t_n, \tau_n, \bar{g}_{n,n}, \bar{\varphi}))^2 \bar{\varphi}_{n+1,\ell,\tilde{\ell}}, & \text{if } \ell \neq \ell_n \text{ and } \tilde{\ell} \neq \ell_n \\
+ \sum_{\ell=0}^{L-1} \frac{\partial_c}{\partial t}(t_n, \tau_n, \bar{g}_{n,n}) \bar{\varphi}_{n+1,\ell} + F(\bar{g}_{n,L}) U'(\tau_n) \Delta t_n, & \text{if } \ell = \tilde{\ell} = \ell_n, \\
c(t_n, \tau_n, \bar{g}_{n,n}) \bar{\varphi}_{n,\ell,\tilde{\ell}}, & \text{if } \ell \neq \ell_n \text{ and } \tilde{\ell} = \ell_n. 
\end{cases}
\]

Proof of Theorem 3.1

The proof is an application of Theorem 2.2 in [21]. To be able to split the time and maturity discretization errors, introduce the semidiscretized fluxes \( \bar{u} \) and \( \bar{b} \) that, for \( \tau_\ell \leq \tau < \tau_{\ell+1} \), are

\[
\bar{u}(t, \tau, x) = \xi^2(x + f_0(t)) \tilde{\lambda}(t, \tau), \\
\bar{b}(t, \tau, x) = \xi(x + f_0(t)) \lambda(t, \tau).
\]

As a first step, replace the exact solution of (3.1), \( g \), by a finite dimensional approximation: a piecewise constant \( g^*(t, \cdot) \), which is an Euler approximation with a much finer discretization, both in time \( t \) and maturity time \( \tau \), than \( \bar{g} \). Thus, \( g^* \) uses a time grid \( \{t_n\}_{n=0}^N \) much finer than \( \{\tilde{t}_n\}_{\ell=0}^\ell \) and a maturity time grid, \( \{\tilde{\tau}_\ell\}_{\ell=0}^M \) much finer than \( \{\tau_\ell\}_{\ell=0}^\ell \). Consequently, the number of time steps satisfy \( P >> N, M >> L \) respectively, and

\[
\Delta t \equiv \max_{p=0, \ldots, P-1} \Delta t_{p+1} - \Delta t_p \ll \max_{n=0, \ldots, N-1} \Delta t_{n+1} - t_n \equiv \Delta t, \\
\Delta \tilde{t} \equiv \max_{m=0, \ldots, M-1} \Delta \tilde{t}_{m+1} - \Delta \tilde{t}_m \ll \max_{\ell=0, \ldots, L-1} \Delta \tilde{t}_{\ell+1} - \Delta \tilde{t}_\ell \equiv \Delta \tilde{t}.
\]

In the application of the mentioned Theorem, include the \( \tau \)-discretization error terms \( a - \bar{u} \), \( b - \bar{b} \) as well as the \( t \)-discretization terms \( \bar{u} - \bar{g} \), \( \bar{b} - \bar{g} \) in the error expansion, following Lemmata 2.1-2.5 in [21], to obtain (3.6-3.8) for \( g \) replaced by the piecewise constant process \( g^* \). For this purpose, observe that \( \bar{g} \) can be also thought of as a piecewise constant function on the finer \( \tau \)-partition that defines \( g^* \).

The second step is to let \( M, P \to \infty \) and \( \Delta \tilde{t}, \Delta \tilde{t} \to 0 \), using

\[
\sup_{0 < t \leq t_{\max}} \sqrt{E[|g(t, \cdot) - g^*(t, \cdot)|_{H}^2]} = O(\Delta \tilde{t} + \sqrt{\Delta t}).
\]

and similar estimates for the corresponding dual functions \( \bar{\varphi}, \bar{\varphi}', \bar{\varphi}'' \), to control the higher order terms in the error expansion. Here \( H \) is a discrete \( H^1 \) norm in \( \tau \), i.e.

\[
\|x\|_{H}^2 \equiv \sum_{m=0}^{M-1} \left( \frac{\chi(\tilde{t}_{m+1}) - \chi(\tilde{t}_m)}{\tilde{t}_{m+1} - \tilde{t}_m} \right)^2 (\tilde{t}_{m+1} - \tilde{t}_m) + \sum_{m=0}^{M} (\chi(\tilde{t}_m))^2 (\tilde{t}_{m+1} - \tilde{t}_m).
\]
The strong convergence (3.13) of $g^*$ to $g$ follows from the standard convergence proof of the Euler method, since from the regularity of $a$ and $b$ and the application of a discrete Sobolev inequality there exist strictly positive constants $C_1, C_2$ such that
\[ \|a(t, \cdot, \chi(t)) - a(t, \cdot, \psi(t))\|_H + \|b(t, \cdot, \chi(t)) - b(t, \cdot, \psi(t))\|_H \leq C_1 |\chi(t) - \psi(t)| \leq C_2 \|\chi - \psi\|_H, \]
for all functions $\chi, \psi$, that are piecewise constant on the $\tau$-grid $\{\hat{t}_n\}_{n=0}^M$.

\begin{remark}
In the EFD method the $\tau$-discretization error of (3.6) and (3.7) can, by (3.4), (3.5) and the notation $r_n = \bar{\pi}_{n, t_n} + f_0(t_n)$, be expressed by
\begin{equation}
E_{r} = \sum_{n=0}^{N-1} \left\{ \sum_{\ell=0}^{L-1} E \left[ \xi^2(r_n) \bar{\pi}_{n, \ell} \right] \frac{\bar{\lambda}(t_n, \tau_{\ell+1}) - \bar{\lambda}(t_n, \tau_{\ell})}{2} \Delta\tau_{\ell} \right. \\
+ \frac{1}{4} \sum_{\ell=0}^{L-1} \sum_{\tilde{\ell}=0}^{L-1} E \left[ \xi^2(r_n) \bar{\pi}'_{n, \ell, \tilde{\ell}} \right] \left( \lambda(t_n, \tau_{\ell+1}) \cdot \lambda(t_n, \tau_{\tilde{\ell}+1}) - \lambda(t_n, \tau_{\ell}) \cdot \lambda(t_n, \tau_{\tilde{\ell}}) \right) \Delta\tau_{\ell} \Delta\tau_{\tilde{\ell}} \left. \right\} \Delta t_n, \tag{3.14} \end{equation}
and the time discretization is
\begin{equation}
E_t = \sum_{n=0}^{N-1} \left\{ E \left[ (F(\bar{\pi}_{n+1, L} U(r_{n+1}) - F(\bar{\pi}_{n, L} U(r_n)) \right] + E \left[ (r_{n+1} - r_n) \bar{\pi}_{n+1, L} \right] \\
+ \sum_{\ell=0}^{L-1} E \left[ \xi^2(r_{n+1}) \bar{\lambda}(t_{n+1}, \tau_{\ell}) \right] \frac{\lambda(t_{n+1}, \tau_{\ell+1}) \cdot \lambda(t_{n+1}, \tau_{\ell})}{2} \Delta\tau_{\ell} \right. \\
+ \sum_{\ell=0}^{L-1} \sum_{\tilde{\ell}=0}^{L-1} E \left[ \xi^2(r_{n+1}) \bar{\lambda}(t_{n+1}, \tau_{\ell}) \right] \frac{\lambda(t_{n+1}, \tau_{\ell+1}) \cdot \lambda(t_{n+1}, \tau_{\tilde{\ell}})}{2} \Delta\tau_{\ell} \Delta\tau_{\tilde{\ell}} \left. \right\} \Delta t_n. \tag{3.15} \end{equation}
\end{remark}

\begin{remark} (Computational cost of the error estimates). The work to approximate $F(f) = E[X]$ within an accuracy $TOL$ is $O(\text{Var}[X])$, provided we use the Monte Carlo version of the EFD method as in (2.11). It is therefore important to try to use both variance reduction techniques and adaptive methods to save computational effort. On the other hand, the work needed to compute sufficiently sharp error estimates as described in Theorem 3.1 is only $O(TOL^{-3})$. The number of realizations needed to have a statistical error in the error bound much smaller than $TOL$ is only $O(TOL^{-1})$ instead of $O(\text{Var}[X]/TOL^2)$ realizations we need to compute an approximation of $F(f)$ using (2.11), while the work to compute the error estimate for each realization is still $O(TOL^{-2})$, including the computation of the duals $\bar{\varphi}$ and $\varphi'$. This surprising reduction of work for $\bar{\varphi}$ and $\varphi'$ is special for the HJM model studied here. For general SDEs the corresponding work would be $O(TOL^{-4})$ instead of $O(TOL^{-2})$. Thus, cheap and sharp error bounds are obtained by the use of the a posteriori error estimates in Theorem 3.1. Observe that if variance reduction techniques are applied to the approximation of $F(f)$, it is natural to try to use them also to reduce the variance in the error estimators.
\end{remark}
Remark 3.3 (Stochastic or Deterministic time steps?). The works [21, 17] study approximations of Itô Stochastic differential equations with either deterministic or stochastic time steps. The difference is that the stochastic time steps may use different meshes for each realization, while the deterministic time steps use the same mesh for all realizations. This work presents only the approximation of the HJM model with deterministic time steps, and, if needed, following closely [21, 17] analogous results can be derived for the case of stochastic time steps. Stochastic time steps use Brownian bridges and require more work for a given number of time steps. Deterministic time steps may yield more time steps, but require less work; for example in the limit of vanishing error tolerance, the ratio of the computational error and its computable estimate tends to one with negligible additional work to determine the adaptive deterministic time steps, cf. Remark 3.2. Stochastic time steps are better suited when there is a remarkable influence from the Wiener process in the dynamics of the solution.

Remark 3.4 (More general expected values). In financial computations it is often necessary to price contracts that have intermediate cash flows, like in the case of caps, floors and swaps, see [3, 15]. In that case we need to evaluate a sum of functionals like the one defined in (1.4-1.5), where each of the functionals corresponds to a different cash flow occurrence, i.e. it has different values of $t_{\text{max}}$, $\tau_{\text{max}}$, etc. The analysis carried out in this section can then be applied provided we include the intermediate cash flow dates in the time partition and, by the linearity of the dual problems (3.9-3.12), and use the dual equations that correspond to the sum of the cash flows.

3.2. Error estimates for the EFE method. The analysis of the EFE method follows a similar line as the estimates of the EFD method. The difference lies in the $\tau$-discretization error, which by virtue of the orthogonality of both $\hat{\lambda} - \Pi \hat{\lambda}$ and $\lambda - \Pi \lambda$ to the subspace of piecewise constant functions $S_{\Delta \tau}$, becomes second order accurate. Therefore, more careful expansions, including interpolation estimates, need to be carried out in order to capture the second order contributions from the $\tau$-discretization. However, unless the function $\xi$ is either a constant or a linear function the authors have not been able to develop an a posteriori error estimate which is asymptotically exact like in the EFD case.

3.3. Analysis of the quadrature error $E_Q$. Here the aim is to justify that the quadrature error $E_Q$ is a higher order term in the expansion of the computational error (2.12), provided that the quadrature rule is sufficiently accurate. Lemma 3.1 provides a strong local estimate for the quadrature error, which implies an estimate for $E_Q$.

**Lemma 3.1.** Let $X$ be a real valued random variable in a probability space $(\Omega, \mathcal{G}, P)$. Assume that for each function $S : \mathbb{R} \to \mathbb{R}$ with polynomial growth there exist a constant $C_S$ such that

$$E[|S(X)|] < C_S < \infty. \quad (3.16)$$

Let $Q : \mathbb{R} \to \mathbb{R}$ be a smooth function such that all its derivatives up to order $p+1$ have polynomial growth, and let $f_0 : [0, \tau] \to \mathbb{R}$ be in $C^{p+1}([0, \tau])$. Then, there exist a constant $C$ such that a quadrature rule of order $p$ yields the approximation

$$E\left[\left(\int_0^\tau Q(X+f_0(v))dv - \sum_{i=1}^{N_{\text{quad}}} \{v_i Q(X+f_0(s_i \tau))\} \tau^2\right)^\frac{1}{2}\right] \leq C \tau^{p+1}. \quad (3.17)$$

*Proof of Lemma 3.1*:

Introduce the auxiliary function $h(v, X) \equiv Q(X+f_0(v))$ and observe that $h(\cdot, X) \in C^{p+1}_{[0,\tau]}$. Thus, for each realization of the random variable $X$, a standard quadrature error estimate for a $p$-th order
quadrature rule gives, cf. [2],
\[
\int_0^\tau Q(X(\omega) + f_0(v))\,dv - \sum_{i=1}^{N_{quad}} \{ v_i \, Q(X(\omega) + f_0(s_i, \tau))\} \tau < K_p \| \partial^{p+1} h(\cdot, X(\omega)) \|_{L_r(\tau, \infty)} \, \tau^{p+1},
\]
where \( K_p \) is a constant depending only on the choice of the quadrature rule. Observe that the function \( S(x) \equiv \left( \| \partial^{p+1} h(\cdot, x) \|_{L_r(\tau, \infty)} \right)^2 \) has polynomial growth with respect to \( x \) and therefore (3.17) follows from a direct application of (3.16) and (3.18). □

The following argument avoids the use of higher order \( \tau \) partial derivatives of the process \( g \) for the analysis of the quadrature error \( E_Q \). Note that the proof of Theorem 3.1 for each function \( S \) of polynomial growth provides the uniform bound
\[
E[|S(\bar{\omega}(t, \tau))|] < C_S, \quad \text{for all } \tau \in [0, \tau_{\text{max}}], \ t \in [0, t_{\text{max}}]
\]
and \( C_S \) is independent from \( \Delta \tau \) and \( \Delta t \). A direct application of Lemma 3.1 therefore yields
\[
E[(\mathcal{I}_Q - \bar{\mathcal{I}}_Q)^2]^{\frac{1}{2}} < O((\Delta \tau)^p) \quad \text{and} \quad E[(G(\mathcal{I}_Q) - G(\bar{\mathcal{I}}_Q))^2]^{\frac{1}{2}} < O((\Delta \tau)^p)
\]
with \( G, \mathcal{I}_Q \) as in (1.4-1.5) and \( \bar{\mathcal{I}}_Q \) defined in (2.9). Then (2.10-2.11) and (3.19) imply

**Theorem 3.2** (A priori estimate of the quadrature error). There holds \( E_Q = O((\Delta \tau)^p) \).

4. **Numerical experiments**

This section provides numerical evidence for the computational error (2.12) of the numerical methods defined in (2.4-2.5) and (2.6-2.7) and the approximation of the functional \( F(f) \) in (1.4-1.5), by computing with examples that have known exact solution. Thus, the computations presented here are related to much simpler examples than the general case of (1.2-1.3), to permit a straightforward derivation of an exact solution to compare with. The implementation uses double precision fortran 77 and simulates the increments of the \( J \) independent Wiener processes by a double precision modification of the functions ran1 and gasdev proposed in [18]. For the particular case of the EFD method, the estimates developed in Theorem 3.1 are compared with the exact computational error: the numerical results are in agreement with the theory and the work to compute the estimates is small.

**Control of the statistical error.** For \( M \) independent samples \( \{ Y(\omega_j) \}_{j=1}^M \) of a random variable \( Y \), with \( E[|Y|^6] < \infty \), define the sample average \( \mathcal{A}(Y; M) \) and the sample standard deviation \( \mathcal{S}(Y; M) \) of \( Y \) by
\[
\mathcal{A}(Y; M) \equiv \frac{1}{M} \sum_{j=1}^M Y(\omega_j) \quad \text{and} \quad \mathcal{S}(Y; M) \equiv \left[ \mathcal{A}(Y^2; M) - (\mathcal{A}(Y; M))^2 \right]^{1/2}.
\]
Let \( \sigma \equiv (E[|Y - E[Y]|^2])^{1/2} \) and consider the random variable
\[
Z_M \equiv \frac{\sqrt{M}}{\sigma} (\mathcal{A}(Y; M) - E[Y])
\]
with cumulative distribution function \( F_{Z_M}(x) \equiv P(Z_M \leq x), \ x \in \mathbb{R} \). Let \( \lambda \equiv (E[|Y - E[Y]|^3])^{1/3} / \sigma < \infty \), then the Berry-Esseen theorem, cf. [10] p. 126, gives the following estimate in the central limit theorem
\[
\sup_{x \in \mathbb{R}} |F_{Z_M}(x) - \Phi(x)| \leq \frac{3}{\sqrt{M}} \lambda^3
\]
for the rate of convergence of \( F_{\tau,M} \) to the distribution function, \( \Phi \), of a normal random variable with mean zero and variance one, i.e.

\[
\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} s^2\right) ds.
\]

Since in the examples below \( M \) is sufficiently large, i.e. \( M \gg 36\lambda^6 \), the statistical error

\[
\mathcal{E}_S(Y; M) \equiv E[Y] - \mathcal{A}(Y; M)
\]
satisfies, by the Berry-Esseen theorem, the following probability approximation

\[
P \left( \left| \mathcal{E}_S(Y; M) \right| \leq c_o \frac{\sigma}{\sqrt{M}} \right) \simeq 2\Phi(c_o) - 1.
\]

In practice choose some constant \( c_o \geq 1.65 \), so the normal distribution satisfies \( 1 > 2\Phi(c_o) - 1 \geq 0.901 \) and the event

\[
\left| \mathcal{E}_S(Y; M) \right| \leq \mathcal{E}_a(Y; M) \equiv c_o \frac{\mathcal{S}(Y; M)}{\sqrt{M}}
\]

has probability close to one, which involves the additional step to approximate \( \sigma \) by \( \mathcal{S}(Y; M) \), cf. [11]. Thus, in the computations \( \mathcal{E}_a(Y; M) \) is a good approximation of the statistical error \( \mathcal{E}_S(Y; M) \).

For a given \( \text{TOL} > 0 \), the goal is to find \( M \) such that \( \mathcal{E}_a(Y; M) \leq \text{TOL} \). The algorithm described in [21] adaptively finds the number of realizations \( M \) to compute the sample average \( \mathcal{A}(Y; M) \) as an approximation to \( E[Y] \). With large probability, depending on \( c_o \), the statistical error in the approximation is then bounded by \( \text{TOL} \). For more details on the implementation of an adaptive algorithm to control the statistical error, see [21].

Now let us introduce some notation to be used in the numerical examples. \( \mathcal{E}_{mn} \) denotes the sample average approximating the \( \tau \)-discretization error (3.13) and \( \mathcal{E}_{tm} \) denotes the sample average approximation to the \( t \)-discretization error (3.14). Beside this, denote by \( \mathcal{E}_a \) the approximation (4.1) to the statistical error introduced in (2.12) and by \( \mathcal{E}_{mn,t} \) the approximation (4.1) to the statistical error in the estimation of the \( \tau \)-discretization error (3.13) by sample averages. Similarly, \( \mathcal{E}_{tm,t} \) denotes the corresponding approximation to the statistical error in the estimation of the expected values in \( t \)-discretization error (3.14).

**Remark 4.1 (Variance reduction techniques).** The use of variance reduction techniques can decrease substantially the statistical errors. In particular the so called antithetic variates technique introduced in [12] reduces the variance in a sample estimator \( \mathcal{A}(M; Y) \) by using another estimator \( \mathcal{A}(M; Y') \) with the same expectation as the first one, but which is negatively correlated with the first. Then, the improved estimator is \( \mathcal{A}(M; \frac{1}{2}(Y + Y')) \). Here, the choice of \( Y \) and \( Y' \) relates to the Wiener process \( W \) and its reflection along the time axis, \( -W \), which is also a Wiener process. If a realization of the Wiener process, \( W(\cdot, \omega_j) \), yields, using one of the numerical discretizations (2.4-2.7), a realization \( \bar{\mathcal{F}}(\cdot, \omega_j) \) and \( -W(\cdot, \omega_j) \) yields \( \bar{\mathcal{F}}(\cdot, \omega_j) \) respectively, then we choose

\[
\frac{1}{M} \sum_{h=1}^{M} \frac{F(\bar{\mathcal{F}}_{N,L}(\omega_h))}{G(\mathcal{I}_Q(\bar{\mathcal{F}}(\omega_h)))} + \frac{\bar{\mathcal{F}}_{N,L+1}(\omega_h)}{G(\mathcal{I}_Q(\mathcal{F}(\omega_h)))} \frac{F(\bar{\mathcal{F}}_{N,L}(\omega_h))}{2}
\]

as a better estimate. All the numerical results presented below use antithetic variates.

In general, the use of control variates, see [5], can be also combined with other variance reduction methods. For example, the control variates technique is based on the knowledge of an estimator \( Y'' \), positively correlated with \( Y \), whose expected value \( E[Y''] \) is known and relatively close to the desired \( E[Y] \), yielding \( Y - Y'' + E[Y''] \) as an improved estimator. The estimates presented in this
work do not preclude the use of control variates, and even though it is not applied here, it can be a valuable tool in practical computations.

4.1. Ho-Lee model. The Ho-Lee model has \( \xi(x) = \sigma \) and \( \lambda_o(x) = 1 \) so (1.2-1.3) takes the form

\[
\begin{align*}
\frac{df(t, \tau)}{dt} &= \sigma^2 (\tau - t) dt + \sigma \, dW(t), \quad \tau \in [t, \tau_{\text{max}}] \\
 f(0, \cdot) &= f_0.
\end{align*}
\]

(4.2)

In this example the initial condition is \( f_0(\tau) = r_0 - \frac{\sigma^2}{2} (\tau-t)^2 + \int_0^\tau \vartheta(s) ds, \) where \( r_0 \) and \( \sigma \) are real positive constants and \( \vartheta : \mathbb{R}^+ \to \mathbb{R} \) is a given function. Then, the exact solution of (4.2) is

\[
f(t, \tau) = r_0 - \frac{\sigma^2}{2} (\tau-t)^2 + \int_0^\tau \vartheta(s) ds + \sigma W(t), \quad \tau \geq t,
\]

which follows the normal distribution and therefore, yields bond prices which are log-normal distributed, allowing the use of Black and Scholes formulas for the pricing of call and put options on bonds. Setting \( \tau_0 = \tau_{\text{max}} \), \( F(x) = 1-x \), \( G(x) = x \), \( Q(x) = x \) and \( U(x) = 0 \) in (1.4-1.5), the functional to be computed has the form

\[
\mathcal{F}(f) = E \left[ \left( 1 - \int_0^{\tau_{\text{max}}} f(s, s) ds \right) \left( \int_{\tau_{\text{max}}}^{t_{\text{max}}} f(t_{\text{max}}, \eta) d\eta \right) \right]. 
\]

(4.3)

In the numerical experiments we choose \( r_0 = 0.05, \sigma = 0.01, \vartheta(s) = \frac{1}{10} \exp(-s) \). Then \( \mathcal{F}(f) \) is a known function of \( t_{\text{max}} \) and \( \tau_{\text{max}} \).

The first experiment sets \( t_{\text{max}} = 1.0 \) and \( \tau_{\text{max}} = 2.0 \), comparing the efficiency of the EFD and EFE methods. Table 1 shows the computational error for both methods and compares the a posteriori approximation of the error with the true computational error for the EFD method.

Observe that the ratio \( A \pm B \) of the a posteriori approximation of the error over the computational error becomes closer and closer to one as we refine the time and maturity partitions, provided that the statistical error is small compared to the \( \lambda \)-discretization error and the \( \tau \)-discretization error. Here a confidence interval for the ratio between the error approximation and the exact computational error, \( \mathcal{E}_c \), introduced in (2.12) is \([A - B, \ A + B]\), with \( A \equiv \frac{\mathcal{E}_{\text{approx}} + \mathcal{E}_{\text{exact}}}{|\mathcal{E}_c|} \) and \( B \equiv \frac{\mathcal{E}_{\text{approx}} - \mathcal{E}_{\text{exact}}}{|\mathcal{E}_c|} \). Whenever we use the EFD method we call \( E_{\text{EFD}} \equiv \mathcal{E}_c \) and if we use the EFE method we call \( E_{\text{EFE}} \equiv \mathcal{E}_c \).

In this example the \( \lambda \)-discretization gives the largest contribution to the computational error, and there is no practical advantage in the use of the EFE method.

<table>
<thead>
<tr>
<th>iseed = −1</th>
<th>EFE</th>
<th>E_{\text{EFD}}</th>
<th>[A - B, A + B]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N = L )</td>
<td>( E_{\text{EFE}} )</td>
<td>( E_{\text{EFD}} )</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>( -8.40 \times 10^{-4} )</td>
<td>( -8.25 \times 10^{-4} )</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>( -4.16 \times 10^{-4} )</td>
<td>( -4.08 \times 10^{-4} )</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>( -2.07 \times 10^{-4} )</td>
<td>( -2.03 \times 10^{-4} )</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Functional (4.3): Comparing the EFD and EFE methods in the Ho-Lee model with \( M = 5000, c_0 = 1.65 \).

\[ \theta \]
4.2. **Vasicek model.** The Vasicek model has \( \xi(x) = \sigma \) and \( \lambda_\alpha(x) = \exp(-\alpha x) \), so \( \tilde{\lambda}_\alpha(x) = \frac{1}{\alpha}(1 - \exp(-\alpha x))\exp(-\alpha x) \) and the forward rate equation (1.2-1.3) becomes

\[
\begin{align*}
 df(t, \tau) &= \sigma^2 \frac{1}{\alpha} (1 - \exp(-\alpha(\tau - t))) \exp(-\alpha(\tau - t)) \, dt + \sigma \exp(-\alpha(\tau - t)) \, dW(t), \quad \tau \geq t \\
 f(0, \cdot) &= f_0.
\end{align*}
\]

In this example the initial condition is

\[
f_0(\tau) = (r_0 - \frac{\theta}{\alpha}) \exp(-\alpha \tau) + \frac{\theta}{\alpha} - \frac{\sigma^2}{2\alpha^2}(1 - \exp(-\alpha \tau))^2, \quad \tau \in [t, \tau_{\text{max}}]
\]

where \( r_0, \sigma, \alpha \) and \( \theta \) are given positive constants. The solution of (4.4) is then

\[
f(t, \tau) = \left( \exp(-\alpha t) \left( r_0 - \frac{\theta}{\alpha} \right) + \sigma \int_0^t \exp(-\alpha(t - s)) \, dW(s) \right) \exp(-\alpha(\tau - t))
\]

\[
+ \frac{\theta}{\alpha} - \frac{\sigma^2}{2\alpha^2}(1 - \exp(-\alpha(\tau - t)))^2, \quad \tau \geq t
\]

which is normally distributed and yields bond prices that are lognormal, as in the Ho-Lee model. Here we set \( r_0 = \tau_{\text{max}} = 0.3, \tau_{\text{max}} = 6.0 \), and use again the functional defined in (4.3). In addition, we take \( r_0 = 0.03, \alpha = 1.0, \sigma = 0.01 \) and \( \theta = 0.05 \).

Table 2 displays the computational errors for the EFD and EFE methods and compares the a posteriori approximation of the error with the true error for the EFD method. Observe that the ratio \( A \pm B \) of the a posteriori approximation of the error over the computational error becomes closer and closer to 1 as we refine the time and maturity partitions, provided that the statistical error is small compared to the time and \( \tau \)-discretization error.

<table>
<thead>
<tr>
<th>iseed = -1</th>
<th>EFE</th>
<th>EFD</th>
<th>( A - B ), ( A + B )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N = L )</td>
<td>( E_{\text{EFF}} )</td>
<td>( E_{\text{EFF}} )</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>(-2.30 \times 10^{-8})</td>
<td>(-2.07 \times 10^{-8})</td>
<td>(1.92, 1.95)</td>
</tr>
<tr>
<td>10</td>
<td>(-2.05 \times 10^{-8})</td>
<td>(-1.95 \times 10^{-8})</td>
<td>(1.03, 1.05)</td>
</tr>
<tr>
<td>20</td>
<td>(-1.06 \times 10^{-8})</td>
<td>(-1.00 \times 10^{-8})</td>
<td>(0.99, 1.02)</td>
</tr>
</tbody>
</table>

**Table 2.** Functional (4.3): Comparing the EFD and EFE methods in the Vasicek model with \( M = 5000, c_0 = 1.65 \).

4.3. **The Cox-Ingersoll-Ross (CIR) model.** Consider the following CIR short rate model

\[
r(t) = r_0 + \int_0^t (\vartheta - \alpha \, r(s)) \, ds + \int_0^t \sigma \sqrt{r(s)} \, dW(s), \quad t \geq 0,
\]

where \( \vartheta, \alpha \) and \( \sigma \) are real constants. To connect the solution \( r(t) \) of (4.5) to the diagonal value \( f(t, t) \) of the solution of a HJM problem, consider, first, the solution \( B = B(t; \tau) \) of the following Riccati differential equation (see [1]):

\[
\begin{align*}
\partial_t B(t; \tau) &= \frac{1}{2} \sigma^2 B^2(t; \tau) + \alpha B(t; \tau) - 1, \quad t \in [0, \tau], \quad \tau \geq 0, \\
B(\tau; \tau) &= 0,
\end{align*}
\]

which has the form \( B(t; \tau) = \psi(\tau - t) \) where

\[
\psi(x) = -\frac{\alpha}{\sigma^2} + \frac{2}{\sigma^2 \tau_0} \sinh(\gamma_0 x) + \frac{\alpha}{\sigma^2} \cosh(\gamma_0 x) + \frac{\alpha}{2 \tau_0} \sinh(\gamma_0 x) + \frac{\alpha}{2 \tau_0} \sinh(\gamma_0 x) \text{ and } \gamma_0 := \sqrt{(2\sigma^2 + \alpha^2)^{1/2}}.
\]
Provided $\xi(x) = \sigma \sqrt{\max\{x,0\}}$ and $\lambda_e(x) = \psi'(x)$, then $\hat{\lambda}_e(x) = \psi'(x)\psi(x)$ and the stochastic function

$$f(t,\tau) = \varphi(t) \psi'(\tau - t) + \vartheta \psi(\tau - t)$$

solves (1.2-1.3) with the initial condition $f_0(\tau) = r_0 \psi'(\tau) + \vartheta \psi(\tau)$. Taking into account that $\psi'(0) = 1$ and $\psi(0) = 0$, it follows that $f(t,t) = r(t)$. Setting $\tau_0 = \tau_{\text{max}}$, $F(x) = \exp(-x)$, $G(x) = \max\{\exp(-x) - K_\alpha,0\}$, $Q(x) = x$ and $U(x) = 0$ in (1.4-1.5), the functional to compute in this example takes the form

$$\mathcal{F}(f) = E\left[\exp\left(-\int_0^{\tau_{\text{max}}} f(s,s)ds\right) \max\left\{\exp\left(-\int_0^{\tau_{\text{max}}} f(t_{\text{max}},\eta)d\eta\right) - K_\alpha,0\right\}\right].$$

(4.6)

In the numerical experiments we choose $r_0 = 0.15$, $\alpha = 1.0$, $\sigma = 0.1$, $\vartheta = 0.05$, $\tau_{\text{max}} = 5.0$, $\tau_{\text{max}} = 8.0$ and $K_\alpha = 0.5$. Table 3 shows the computational errors for the EFD and EFE methods and the ratio between the error estimate and the exact computational error. There is no practical difference in this case between the EFD and the EFE method since the computational error is mainly time discretization error and the $\tau$ discretization error is relatively unimportant.

In order to have smooth coefficients in the HJM model (1.2-1.3) we approximate the function $\lambda_e(x)$ given by (4.6) via the Feynman Kac representation formula, using a numerical solution of the following backward PDE, cf. [20] p. 313,

$$v_t + (\vartheta - \alpha r)v_r + \frac{1}{2}\sigma^2 rv_{rr} - rv = 0, \quad t \in [0,\tau_{\text{max}}], \quad r \in [0,r_{\text{max}}],$$

with final datum $v(t_{\text{max}},r) = (B(r,t_{\text{max}},\tau_{\text{max}}) - K_\alpha)^+$, where $B(r,t_{\text{max}},\tau_{\text{max}})$ denotes the CIR value for a bond with contracting time $t_{\text{max}}$, maturity time $\tau_{\text{max}}$ and short rate at $t_{\text{max}}$ equal to $r$. We also use the boundary conditions

$$v_t(t,0) + \alpha v_r(t,0) = 0,$$
$$v(t,r_{\text{max}}) = 0,$$

for $t \in [0,t_{\text{max}}]$. The value of $r_{\text{max}} >> \frac{2\sigma}{\vartheta}$ is taken sufficiently large so that the homogeneous Dirichlet boundary at $r = r_{\text{max}}$ has a negligible effect on the numerical approximation for $v(0,0.15) = \mathcal{F}(f)$. The spatial discretization is a centered finite differences scheme and the time stepping is done by a diagonally implicit Runge Kutta method, namely the DIRK2 method, see [8].

<table>
<thead>
<tr>
<th>iseed = -1</th>
<th>EFE</th>
<th>EFD</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = L</td>
<td>$E_{\text{EFE}}$</td>
<td>$E_{\text{EFD}}$</td>
</tr>
<tr>
<td>5</td>
<td>$1.23 \times 10^{-2}$</td>
<td>$1.21 \times 10^{-2}$</td>
</tr>
<tr>
<td>10</td>
<td>$5.83 \times 10^{-3}$</td>
<td>$5.39 \times 10^{-3}$</td>
</tr>
<tr>
<td>20</td>
<td>$2.76 \times 10^{-3}$</td>
<td>$2.79 \times 10^{-3}$</td>
</tr>
</tbody>
</table>

Table 3. Functional (4.6): Comparing the EFD and EFE methods in the CIR model with $M = 2000$, $c_0 = 1.65$. Compute an accurate numerical approximation of the exact $\mathcal{F}(f)$ from (4.6), via the Feynman Kac representation formula, using a numerical solution of the following backward PDE, cf. [20] p. 313,
Another way to estimate the exact solution with high accuracy is to use a formula based on the $\chi^2$ distribution, see [19] pp. 187-193 for details.

5. A two-factor Gaussian model

A two-factor model has randomness introduced by two scalar independent Wiener processes $W_1, W_2$. In particular, for a two-factor Gaussian model we have $\xi(x) = 1$, $\lambda_0(x) = \sigma_1$ and $\lambda_{0,2}(x) = \sigma_2 \exp(-\frac{a_2 x}{2})$ with $\sigma_1$, $\sigma_2$ and $a_2$ real positive constants. Thus (1.2-1.3) takes the form

$$
\begin{aligned}
\frac{df(t, \tau)}{dt} &= ((\sigma_1)^2 (\tau - t) + \frac{2(a_2)^2 \exp(-\frac{a_2(\tau-t)}{2})}{a_2} (1 - \exp(-\frac{a_2(\tau-t)}{2})) dt
+ \sigma_1 dW_1(t) + \sigma_2 \exp(-\frac{a_2(\tau-t)}{2}) dW_2(t), \quad \tau \in [t, \tau_{\text{max}}]
\end{aligned}
$$

(5.1)

Here the initial condition is $f_0(\tau) = b_0 + b_1 \exp(-k\tau)$ where $b_0$, $b_1$ and $k$ are real constants. Then, the exact solution of (5.1) is normal distributed as in the Ho-Lee and Vasicek models, so explicit formulas are available for the pricing of put and call options with bonds as underlyings. In the numerical experiment we take $\sigma_1 = 0.02$, $\sigma_2 = 0.01$, $a_2 = 0.5$, and compute with the functional defined in (4.6) with strike $K_{\lambda} = 0.5$, $t_{\text{max}} = 1$ and $\tau_{\text{max}} = 3$. For the initial condition we set $b_0 = 0.0759$, $b_1 = -0.0439$ and $k = 0.4454$. Table 4 shows the computational errors for the EFD and EFE methods and the ratio between the error estimate and the exact computational error.

<table>
<thead>
<tr>
<th>$is\text{seed} = -1$</th>
<th>$E_{\text{EFE}}$</th>
<th>$E_{\text{EFD}}$</th>
<th>$E_{\text{EFD}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N = L$</td>
<td>$-5.15 \times 10^{-4}$</td>
<td>$-6.90 \times 10^{-4}$</td>
<td>(0.98, 1.02)</td>
</tr>
<tr>
<td>$10$</td>
<td>$-2.78 \times 10^{-4}$</td>
<td>$-3.50 \times 10^{-4}$</td>
<td>(0.96, 1.05)</td>
</tr>
</tbody>
</table>

Table 4. Functional (4.6): Comparing the EFD and EFE methods in the two-factor Gaussian model with $M = 40000$, $c_0 = 1.65$.

Acknowledgements The work was completed while the fourth author was a post-doc at NADA, KTH, supported by HCL and NTU.

References

September 5, 2002

GALERKIN FINITE ELEMENT APPROXIMATIONS OF
STOCHASTIC ELLIPTIC PARTIAL DIFFERENTIAL EQUATIONS

IVO BABUŠKA†, RAÚL TEMPONE§ AND GEORGIOS E. ZOURARIS‡

Abstract. We describe and analyze two numerical methods for a linear elliptic problem with stochastic coefficients and homogeneous Dirichlet boundary conditions. Here the aim of the computations is to approximate the expected value of the solution. Since the approximation of the stochastic coefficients from the elliptic problem is in general not exact, we derive related a priori error estimates. The first method generates iid approximations of the solution by sampling the coefficients of the equation and using a standard Galerkin finite elements variational formulation. The Monte Carlo method then uses these approximations to compute corresponding sample averages. The second method is based on a finite dimensional approximation of the stochastic coefficients, turning the original stochastic problem into a deterministic parametric elliptic problem. A Galerkin finite element method, of either $h$ or $p$ version, then approximates the corresponding deterministic solution yielding approximations of the desired statistics. We include a comparison of the computational work required by each method to achieve a given accuracy. This comparison suggests intuitive conditions for an optimal selection of these methods.

Contents

1. Introduction 2
2. Theoretical Aspects of the Continuous Problem 5
2.1. Notation and function spaces 5
2.2. Existence and uniqueness for the solution of a linear stochastic elliptic problem 6
2.3. Continuity with respect to the coefficients $a$ and $f$ 7
2.4. Finite dimensional noise case 9
3. The finite element spaces 11
3.1. Finite element spaces on the spatial set $D \subset \mathbb{R}^d$: h-version 11
3.2. Tensor product finite element spaces on the outcomes set $\Gamma \subset \mathbb{R}^N$: k-version 11
3.3. Tensor product finite element spaces on $\Gamma \times D$: $k \times h$-version 12
3.4. Tensor product finite element spaces on $\Gamma \times D$: $p \times h$-version 13
4. The Monte Carlo Galerkin Finite Element Method 13
5. The Stochastic Galerkin Finite Element Method: $k \times h$-version 15
6. The Stochastic Galerkin Finite Element Method: $p \times h$-version 18
6.1. Error estimates 19
7. Double orthogonal polynomials 23
8. Asymptotical efficiency comparisons 25

1991 Mathematics Subject Classification. 65N30, 65N15.
Key words and phrases. stochastic elliptic equation, perturbation estimates, Karhunen-Loève expansion, finite elements, Monte Carlo method, $k \times h$-version, $p \times h$-version, expected value, error estimates.
† Texas Institute for Computational and Applied Mathematics (TICAM), ACES, University of Texas at Austin, Austin, TX 78759, USA.
§ Department of Numerical Analysis and Computing Science (NADA), Royal Institute of Technology (KTH), S–100 44 Stockholm, Sweden.
‡ Department of Mathematics, University of the Aegean, GR–832 00 Karlovassi, Samos, Greece, and Institute of Applied and Computational Mathematics, FO.R.T.H., GR–711 10 Heraklion, Crete, Greece.
1. Introduction

The use of computational models has increasingly affected the historical development of engineering products, based on first testing prototypes and then building a final version. Computational models aim to reduce the cost of the experimental tests as well as to discard faulty designs before they are built.

Due to the great development in computational resources and scientific computing techniques, more mathematical models can be solved efficiently. Ideally, this artillery could be used to solve many classical partial differential equations, the mathematical models we shall focus on here, to high accuracy. However, in many cases, the information available to solve a given problem is far from complete and is in general very limited. This is the case when solving a partial differential equation whose coefficients depend on material properties that are known to some accuracy. The same may occur with its boundary conditions, and even with the geometry of its domain, see for example the works [6, 7]. Naturally, since the current engineering trends are toward more reliance on computational predictions, the need for assessing the level of accuracy in the results grows accordingly. More than ever, the goal then becomes to represent and propagate the uncertainties from the available data to the desired result through our partial differential equation. By uncertainty we mean either intrinsic variability of physical quantities or simply lack of knowledge about some physical behavior, cf. [46]. If variability is interpreted as randomness then naturally we can apply probability theory. To be fruitful, probability theory requires considerable empirical information about the random quantities in question, generally in the form of probability distributions or their statistical moments. On the other hand, if the only available information comes in the form of some bounds for the uncertain variables, the description and analysis of uncertainty may be based on other methods, like e.g. convexity methods cf. [10, 19]. Uncertainties may arise at different levels. They could appear in the mathematical model, e.g. if we are not sure about the linear behavior of some material, or in the variables that describe the model, e.g. if the linear coefficient that describes the material is not completely known. Here we shall discuss the second alternative, and use a probabilistic description for the coefficient variability, leading us to the study of stochastic differential equations. Although the results presented on this paper can be generalized to linear elliptic stochastic partial differential equations we now focus our study on the standard model problem, a second order linear elliptic equation with homogeneous Dirichlet boundary conditions.

Let $D$ be a convex bounded polygonal domain in $\mathbb{R}^d$ and $(\Omega, \mathcal{F}, P)$ be a complete probability space. Here $\Omega$ is the set of outcomes, $\mathcal{F} \subset 2^\Omega$ is the $\sigma$-algebra of events and $P : \mathcal{F} \to [0, 1]$ is a probability measure. Consider the stochastic linear elliptic boundary value problem: find a stochastic function, $u : \Omega \times \overline{D} \to \mathbb{R}$, such that $P$-almost everywhere in $\Omega$, or in other words almost surely (a.s.), we have

\begin{align}
-\nabla \cdot (a(\omega, \cdot) \nabla u(\omega, \cdot)) &= f(\omega, \cdot) \quad \text{on } D,

u(\omega, \cdot) &= 0, \quad \text{on } \partial D.
\end{align}

Here $a, f : \Omega \times D \to \mathbb{R}$ are stochastic functions. If we denote by $B(D)$ the Borel $\sigma$-algebra generated by the open subsets of $D$, then $a, f$ are assumed measurable with the $\sigma$-algebras $(\mathcal{F} \otimes B(D))$ and $B(\mathbb{R})$. Here we work with the natural $\sigma$-algebra $\mathcal{F} = \sigma(a, f)$, which is the smallest one that makes both $a$ and $f$ measurable functions. On what follows we shall assume that $a$ is bounded and
uniformly coercive, i.e.

\[ \exists a_{\min}, a_{\max} \in (0, +\infty) : \quad P\left( \omega \in \Omega : \ a(\omega, x) \in [a_{\min}, a_{\max}], \ \forall x \in D \right) = 1 \]

To ensure regularity of the solution \( u \) we assume also that \( a \) has a uniformly bounded and continuous first derivative, i.e. there exists a real deterministic constant \( C \) such that

\[ P\left( \omega \in \Omega : \ a(\omega, \cdot) \in C^1(D) \quad \text{and} \quad \max_D |\nabla_x a(\omega, \cdot)| < C \right) = 1. \]

In addition, the right hand side in (1.1) satisfies

\[ \int_\Omega \int_D f^2(\omega, x) \, dx \, dP(\omega) < +\infty \quad \text{which implies} \quad \int_D f^2(\omega, x) \, dx < +\infty \quad \text{a.s.} \]

Stochastic differential equations driven by the Wiener process have been widely applied in Mathematical Finance, see [32, 34, 42]. On the other hand, stochastic partial differential equations have been more popular to model physical phenomena, e.g. random vibrations, seismic activity, oil reservoir management, composite materials, etc, see [4, 18, 20, 24, 30, 31, 33, 47, 51] and the references therein. Solving a stochastic partial differential equation entails finding the joint probability distribution of the solution, which is a hard problem. In practice we shall usually be satisfied with much less, namely the computation of some moments, e.g. the expected value of the solution, or some probability related to a given event, e.g. the probability of some eventual failure, cf. [29, 38].

Whenever we approximate a given stochastic partial differential equation, a corresponding discretization error exists. The data uncertainties, the discretization error and the error from solving the discretized equations inaccurately add to the uncertainty in the result. Thus, it seems of little use to make one of them small when the others are relatively large, e.g. to solve the equations with high accuracy if the data is largely inaccurate. Moreover, quantifying the uncertainty in computer based simulations is also important for the eventual validation of the results. Without such quantification, validation becomes arbitrary, cf. [2, 41, 44, 46].

Depending on the structure of the noise that drives an elliptic partial stochastic differential equation, there are different numerical approximations. For example, when the size of the noise is relatively small, a Neumann expansion around the mean value of the equation’s operator is a popular alternative. It requires only the solution of standard deterministic partial differential equations, the number of them being equal to the number of terms in the expansion. Equivalently, a Taylor expansion of the solution around its mean value with respect to the noise yields the same result. Similarly, the work [33] uses formal Taylor expansions up to second order of the solution but does not study their convergence properties. Recently, the work [3] and the work [5] proposed a perturbation method with successive approximations. It also proves that uniform coercivity of the diffusion is sufficient for the convergence of the perturbation method.

When only the load is stochastic, it is also possible to derive deterministic equations for the moments of the solution. This case was analyzed in [2, 24] and more recently in the work [48], where a new method to solve these equations with optimal complexity is presented.

On the other hand, the work by Babuška et al. [16, 17] and by Ghanem and Spanos [24] address the general case where all the coefficients are stochastic. Both approaches transform the original stochastic problem into a deterministic one with higher dimensions, and they differ in the choice of the approximating functional spaces. The work [16] uses finite elements to approximate the noise dependence of the solution, while [22, 24] uses a formal expansion in terms of Hermite polynomials.

Monte Carlo methods are both general and simple to code and they are naturally suited for parallelization. They generate a set of independent identically distributed (iid) approximations of the solution by sampling the coefficients of the equation, using a spatial discretization of the partial differential equation, e.g. by a Galerkin Finite Elements formulation. Then, using these
approximations we can compute corresponding sample averages of the desired statistics. The drawback of Monte Carlo methods is their slow rate of convergence. It is worth mentioning that in particular cases their convergence can be accelerated by variance reduction techniques [32] or even the convergence rate improved with Quasi Monte Carlo methods [14, 49, 50]. Moreover, if the probability density of the random variable is smooth, the convergence rate of the Monte Carlo method for the approximation of the expected value can be improved, cf. [40, 54].

Another way to provide a notion of stochastic partial differential equations is based on the Wick product and the Wiener chaos expansion, see [30] and [55]. This approach yields solutions in Kondratiev spaces of stochastic distributions which are not the same as those from (2.3). The choice between (2.3) and [30] is a modeling decision, based on the physical situation under study. For example, with the Wick product we have \( E[a \odot u] = E[a]E[u] \) regardless on the correlation between \( a \) and \( f \), whereas this is in general not true with the usual product. A numerical approximation for Wick stochastic linear elliptic partial differential equations is studied in [53], yielding a priori convergence rates.

This work studies the case of stochastic linear elliptic partial differential equations with random diffusion and load coefficients stating and proving conditions for existence and uniqueness of solutions. For example, to obtain a meaningful numerical solution for (1.1) its diffusion coefficient should be uniformly coercive. This work compares a Monte Carlo Galerkin method with the Stochastic Galerkin Finite Element method introduced in [16] and introduces a related \( p \)-version, providing theoretical a priori convergence analysis in each case. A priori estimates are useful to characterize the convergence, and ultimately they provide information to compare the number of operations required by numerical methods. The conclusion for now is that if the noise is described by a small number of random parameters or if the accuracy requirement is sufficiently strict, then a Stochastic Galerkin method is to be preferred, otherwise a Monte Carlo Galerkin method seems still to be the best choice. It is worth mentioning that the development of numerical methods for stochastic differential equations is still very much ongoing, and better numerical methods are expected to appear.

The rest of the paper is organized as follows.

Section 2 introduces the basic notation and proves the existence and uniqueness for solutions, as well as perturbation estimates in energy norm with respect to perturbations in the stochastic coefficients of the equation, i.e. modeling error estimates. The approximation of the coefficients in (1.1) is done by means of the Karhunen-Loève expansion, see [37]. We assume that a finite number of terms of such an expansion have been identified by a statistical procedure, see [39, 56, 57], and base the numerical approximations on such information. Once the coefficient approximation step is done, the stochastic problem is transformed into a parametric elliptic problem and it is natural to apply finite element techniques to construct related numerical methods.

Section 3 defines some deterministic finite element subspaces and gives their approximation properties. Tensor product of deterministic finite element subspaces are used to construct the Stochastic Galerkin finite element approximation.

Sections 5 and 6 present two versions of Galerkin finite element approximations for the stochastic partial differential equation under study. These versions differ only in the choice of the approximation subspaces, namely in the direction of the parameter related with the noise representation. Section 5 uses polynomial approximation with fixed degree, while Section 6 uses polynomial approximation with increasingly higher degree, proving the standard exponential convergence with respect to degree \( p \), cf. [26]. Beside this, precise implementation details are included in Section 7 to ensure minimal computational cost in both cases.

On the other hand, the Monte Carlo Galerkin finite element approximation is based on first sampling the coefficients in the stochastic partial differential equation and then approximating
the corresponding realization of the solution. Section 8 studies convergence of the Monte Carlo Galerkin finite element method for the approximation of the expected value of the solution.

To characterize the convergence of the numerical approximations, a priori error estimates are developed for each of the methods described in this work. Using this information, Section 8 compares the asymptotical computational work required by each of the approximations, giving some intuition on their possible application.

2. Theoretical Aspects of the Continuous Problem

2.1. Notation and function spaces. Let \( m \in \mathbb{N} \), \( D \) be an open, connected and convex subset of \( \mathbb{R}^m \) with polygonal boundary \( \partial D \). Denote the volume of \( D \) by \( |D| \equiv \int_D 1 \, dx \). For \( s \in \mathbb{N} \) and \( 1 \leq p \leq +\infty \), let \( W^{s,q}(D) \) be the Sobolev space of functions having generalized derivatives up to order \( s \) in the space \( L^p(D) \). Using the standard multi-index notation, \( \alpha = (\alpha_1, \ldots, \alpha_d) \) is a \( d \)-tuple of non-negative integers and the length of \( \alpha \) is given by \( |\alpha| = \sum_{i=1}^d \alpha_i \). The Sobolev norm of \( v \in W^{s,q}(D) \) will be denoted by

\[
\|v\|_{W^{s,q}(D)} = \left\{ \sum_{|\alpha| \leq s} \int_D |\partial^\alpha v|^q \, dx \right\}^{1/q}, \quad 1 \leq q < +\infty,
\]

and

\[
\|v\|_{W^{s,\infty}(D)} = \max_{|\alpha| \leq s} \left( \text{ess sup}_D |\partial^\alpha v| \right).
\]

We shall write \( H^{\alpha}(D) \equiv W^{2,2}(D) \) and omit the index 2 from the symbol of its norm, i.e., \( \|\cdot\|_{H^{\alpha}(D)} = \|\cdot\|_{W^{2,2}(D)} \). As usual, the function space \( H^1_0(D) \) is the subspace of \( H^1(D) \) consisting of functions which vanish at the boundary of \( D \) in the sense of trace.

To characterize the convergence of the numerical approximations, a priori error estimates are developed for each of the methods described in this work. Using this information, Section 8 compares the asymptotical computational work required by each of the approximations, giving some intuition on their possible application.
case. Here the Banach space $C(\Gamma; H)$ comprises all continuous functions $u : \Gamma \rightarrow H$ with the norm $\|u\|_{C(\Gamma; H)} = \sup_{y \in \Gamma} \|u(y)\|_H$. Similar definitions apply to the spaces $C^k(\Gamma; H)$, $k = 1, \ldots$; cf. [21] p. 285.

Let $Y$ be an $\mathbb{R}^N$-valued random variable in $(\Omega, \mathcal{F}, P)$. If $Y \in L^2_P(\Omega)$ we denote its expected value by

$$E[Y] = \int_{\Omega} Y(\omega) dP(\omega) = \int_{\mathbb{R}^N} y \, d\mu_Y(y).$$

where $\mu_Y$ is the distribution measure for $Y$, defined for the Borel sets $B(\mathbb{R}^N)$, by $\mu_Y(B) = P(Y^{-1}(B))$. If $\mu_Y$ is absolute continuous with respect to the Lebesgue measure then there exists a density function $\rho_Y : \mathbb{R} \rightarrow [0, +\infty)$, such that

$$E[Y] = \int_{\mathbb{R}^N} y \, \rho_Y(y) \, dx.$$  

Analogously, whenever $Y_i \in L^2_P(\Omega)$ for $i = 1, \ldots, d$ the covariance matrix of $Y$, $\text{Cov}[Y] \in \mathbb{R}^{d \times d}$, is defined by $\text{Cov}[Y](i, j) = \text{Cov}(Y_i, Y_j) = E[(Y_i - E[Y_i])(Y_j - E[Y_j])]$, $i, j = 1, \ldots, d$. Beside this, whenever $u(\omega, \cdot)$ is a stochastic process the positive semi definite function $\text{Cov}[u](x_1, x_2) = \text{Cov}[u(x_1), u(x_2)] = \text{Cov}[u(x_2), u(x_1)]$ is the covariance function of the stochastic process $u$.

To introduce the notion of stochastic Sobolev spaces we first recall the definition of stochastic weak derivatives. Let $u \in L^2_P(\Omega) \otimes L^2(D)$, then the $\alpha$ stochastic weak derivative of $u$, $w = \partial_\alpha u \in L^2_P(\Omega) \otimes L^2(D)$, satisfies

$$\int_D u(\omega, x) \partial^\alpha \phi(x) \, dx = (-1)^{|\alpha|} \int_D w(\omega, x) \phi(x) \, dx, \quad \forall \phi \in C_0^\infty(D), \ a.s.$$  

We shall work with stochastic Sobolev spaces $\tilde{W}^{s,q}(D) = L^q_P(\Omega, W^{s,q}(D))$ containing stochastic functions, $v : \Omega \times D \rightarrow \mathbb{R}$, that are measurable with respect to the product $\sigma$-algebra $\mathcal{F} \otimes B(D)$ and equipped with the averaged norms

$$\|v\|_{\tilde{W}^{s,q}(D)} = E[\|v\|^q_{W^{s,q}(D)}]^{1/q} = E[\sum_{|\alpha| \leq s} \int_D |\partial^\alpha v|^q \, dx]^{1/q}, \quad 1 \leq q < +\infty$$

and

$$\|v\|_{\tilde{W}^{s,\infty}(D)} = \max_{|\alpha| \leq s} (\text{ess sup}_{\Omega \times D} |\partial^\alpha v|).$$

Observe that if $v \in \tilde{W}^{s,q}(D)$ then $v(\omega, \cdot) \in W^{s,q}(D)$ a.s. and $\partial^\alpha v(\cdot, x) \in L^q_P(\Omega)$ a.e. on $D$ for $|\alpha| \leq s$. Whenever $q = 2$, the above space is a Hilbert space, i.e. $\tilde{W}^{s,2}(D) = \tilde{H}^s(D) \simeq L^2_P(\Omega) \otimes H^s(D)$.

2.2. Existence and uniqueness for the solution of a linear stochastic elliptic problem.

In this section, we discuss some issues related to the formulation of the elliptic boundary value problem (1.1) and its solution. We also develop perturbation estimates corresponding to a change in the coefficients of (1.1). These perturbation estimates are both useful to analyze modeling errors and to develop discretization error estimates. Let us consider the tensor product Hilbert space $H = \tilde{H}^1(\Omega) \simeq L^2_P(\Omega; H^1(\Omega))$ endowed with the inner product $(v, w)_H = E[\int_D \nabla v \cdot \nabla w \, dx]$. Define the bilinear form, $\mathcal{B} : H \times H \rightarrow \mathbb{R}$, by $\mathcal{B}(v, w) = E[\int_D a \nabla v \cdot \nabla w \, dx]$, $\forall v, w \in H$. The standard assumption \[\text{[1.3]}\] yields both the continuity and the coercivity of $\mathcal{B}$, i.e.

$$|\mathcal{B}(v, w)| \leq a_{\text{max}} \|v\|_H \|w\|_H, \quad \forall v, w \in H,$$

and

$$a_{\text{min}} \|v\|_H^2 \leq \mathcal{B}(v, v), \quad \forall v \in H.$$
A direct application of the Lax–Milgram’s Lemma, cf. [13], implies the existence and uniqueness for the solution to the variational formulation: find \( u \in H \) such that

\[
B(u,v) = \mathcal{L}(v), \quad \forall v \in H.
\]

Here \( \mathcal{L}(v) \equiv E[\int_D f v dx] \), \( \forall v \in H \) defines a bounded linear functional since the random field \( f \) satisfies \( (1.4) \). Moreover, standard arguments from measure theory show that the solution to \( (2.3) \) also solves \( (1.4) \). The formulation \( (2.3) \) together with assumption \( (2.10) \) on finite dimensional noise give the basis for the Stochastic Galerkin Finite element (SGFEM) method introduced in Sections 5 and 6, while formulation \( (1.1) \) is the basis for the Monte Carlo Galerkin Finite element (MCGFEM) method, discussed in Section 4. Since the domain \( D \) is convex and bounded and assumption \( (1.3) \) on the diffusion \( a \) holds, the theory of elliptic regularity, cf. [1, 21], implies that the solution of \( (1.1) \) satisfies \( u(\omega, \cdot) \in H^2(D) \cap H_0^1(D) \) a.s.

### 2.3. Continuity with respect to the coefficients \( a \) and \( f \)

Usually in practical problems the information about the stochastic processes \( a \) and \( f \) is only limited. For example, we may only have approximations for their expectations and covariance functions to use in the implementation of a numerical method for \( (1.1) \). Moreover, due to some efficiency considerations, sometimes it may be even useful to use an approximation of \( a \) and \( f \). Therefore, an additional approximation error, of modeling type, appears in the computations together with the usual discretization error. For implementation details, see [56, 57, 23, 31, 8]. In the next proposition we consider a weak formulation with perturbed bilinear form and functional, and estimate the size of the corresponding perturbation in the solution with the energy norm.

**Proposition 2.1.** Let \( (H, (\cdot, \cdot)_H) \) be a Hilbert space. Consider two symmetric bilinear forms \( B, \tilde{B} : H \times H \rightarrow \mathbb{R} \) that are \( H \)-coercive and bounded, i.e. there exist real constants \( 0 < a_{\min} \leq a_{\max} \) such that

\[
a_{\min} \|v\|^2_H \leq \min\{B(v,v), \tilde{B}(v,v)\}, \quad \forall v \in H,
\]

and

\[
\max\{\|B(v,w)\|, \|\tilde{B}(v,w)\|\} \leq a_{\max} \|v\|_H \|w\|_H, \quad \forall v, w \in H.
\]

We assume that the above bilinear forms are comparable, i.e. there exist a constant \( \gamma \) such that

\[
|(B - \tilde{B})(v,w)| \leq \gamma \|v\|_H \|w\|_H, \quad \forall v, w \in H.
\]

Consider two bounded linear functionals, \( \mathcal{L}, \tilde{\mathcal{L}} \in H' \) and let \( u, \tilde{u} \in H \) be the solutions of the problems

\[
B(u,v) = \mathcal{L}(v), \quad \forall v \in H,
\]

\[
\tilde{B}(\tilde{u},v) = \tilde{\mathcal{L}}(v), \quad \forall v \in H.
\]

Then there holds

\[
\|u - \tilde{u}\|_H \leq \frac{1}{a_{\min}} \|\mathcal{L} - \tilde{\mathcal{L}}\|_{H'} + \frac{\gamma}{a_{\min}} \|\tilde{\mathcal{L}}\|_{H'}
\]

**Proof.** Since by Lax-Milgram’s lemma \( u \) and \( \tilde{u} \) are well defined we can consider their difference \( e \equiv u - \tilde{u} \). Then

\[
B(e,e) = B(e,u) - B(e,\tilde{u})
\]

\[
= \mathcal{L}(e) - \tilde{B}(e,\tilde{u})
\]

\[
= \mathcal{L}(e) + (\tilde{B} - B)(e,\tilde{u}) - \tilde{B}(e,\tilde{u})
\]

\[
= (\mathcal{L} - \tilde{\mathcal{L}})(e) + (\tilde{B} - B)(e,\tilde{u})
\]

\[
\leq (\|\mathcal{L} - \tilde{\mathcal{L}}\|_{H'} + \gamma \|\tilde{u}\|_H)\|e\|_H.
\]
Since we have \( \|\hat{u}\|_{H} \leq \frac{1}{a_{\text{min}}} \|\hat{L}\|_{L^{1}} \) and \( \|e\|_{L^{2}} \leq \frac{1}{a_{\text{min}}} \mathcal{B}(e,e) \) then (2.7) follows.

**Remark 2.1.** The proof of the previous lemma gives

\[
\|u - \hat{u}\|_{H}^{2} \leq \frac{1}{a_{\text{min}}} (\|L - \hat{L}\|_{H} \|u - \hat{u}\|_{H} + |(\hat{B} - B)(u - \hat{u})|).
\]

If in addition we know that there exist Banach spaces, \( V_{1}, V_{2} \), and positive constants, \( C, \gamma' \) such that

\[
\|v\|_{V_{1}} \leq C \|v\|_{H}
\]

(2.8)

\[
|(\hat{B} - B)(u, v)| \leq \gamma' \|u\|_{V_{1}} \|v\|_{V_{2}}, \ \forall u \in V_{1}, \ v \in V_{2}
\]

then

(2.9)

\[
\|u - \hat{u}\|_{H} \leq \frac{1}{a_{\text{min}}} (\|L - \hat{L}\|_{H} + C \gamma' \|u\|_{V_{2}}).
\]

We shall see that there are cases where it is not possible to apply (2.7) but (2.9) is still valid. Next we consider a sequence of problems like (2.3) with approximate coefficients. A direct application of Lemma 2.1 yields the following convergence result.

**Corollary 2.1 (Convergence result).** Consider the Hilbert space \( H = \tilde{H}^{1\frac{1}{2}}(D) \), two convergent sequences of stochastic processes, \( \{a_{n}\}, \{f_{n}\} \), satisfying

\[
0 < a_{\text{min}} \leq a_{n} \leq a_{\text{max}} < \infty, \ (P \otimes dx) \text{ a.e. on } D \times \Omega, \ \|a_{n} - a\|_{L^{\infty}(D)} \to 0
\]

and

\[
\|f_{n} - f\|_{L^{2}(D)} \to 0.
\]

Then the stochastic processes \( u \) and \( u_{n} \), defined by

\[
E[\int_{D} a_{n} \nabla u_{n} \cdot \nabla v dx] = E[\int_{D} f_{n} v dx], \ \forall v \in H
\]

\[
\int_{D} a \nabla u \cdot \nabla v dx = \int_{D} f v dx, \ \forall v \in H,
\]

satisfy

\[
\|u - u_{n}\|_{\tilde{H}^{1\frac{1}{2}}(D)} \leq \frac{1}{a_{\text{min}}} (\|f_{n} - f\|_{L^{2}(D)} + \frac{1}{a_{\text{min}}} \|a - a_{n}\|_{L^{\infty}(D)} \|f\|_{L^{2}(D)}) \to 0.
\]

Following a similar line as in the previous Corollary, we now apply (2.9) to produce an alternative convergence estimate. This result requires a weaker approximation of the diffusion coefficient, namely in \( L^{2p}(D) \), \( 1 \leq p < +\infty \), but needs more regularity from the solution \( u \).

**Corollary 2.2 (Second convergence result).** Let \( 1 < p < +\infty \) with \( 1/p + 1/q = 1 \). Assume all the hypothesis from the previous corollary except that related with the convergence of the sequence \( \{a_{n}\} \), which satisfies instead

\[
\|a_{n} - a\|_{L^{2p}(D)} \to 0.
\]

Besides this, assume that the solution \( u \) belongs to the stochastic Sobolev space \( \tilde{W}^{1,2q}(D) \). Then

\[
\|u - u_{n}\|_{\tilde{H}^{1\frac{1}{2}}(D)} \leq \frac{1}{a_{\text{min}}} (\|f_{n} - f\|_{L^{2}(D)} + \|a - a_{n}\|_{L^{2p}(D)} \|u\|_{\tilde{W}^{1,2q}(D)}) \to 0.
\]
Proof. Let \( V_1 = H \) then in order to apply (2.9) it is enough to bound the difference of bilinear forms
\[
\int_D E[(a - \hat{a}) \nabla u \cdot \nabla v] dx \leq \left( \int_D E[(a - \hat{a})^2] \nabla u^2 dx \right)^{1/2} \left( \int_D E[\nabla v^2] dx \right)^{1/2} \leq \left( \int_D E[(a - \hat{a})^2] dx \right)^{1/2p} \left( \int_D E[\nabla u^2] dx \right)^{1/2q} \left( \int_D E[\nabla v^2] dx \right)^{1/2}.
\]

Remark 2.2 (Sufficient conditions for the hypothesis of the previous corollaries). Here we recall the Karhunen-Loève expansion, a suitable tool for the approximation of stochastic processes. Consider a stochastic process \( a \) with continuous covariance function, \( \text{Cov}[a] : D \times D \to \mathbb{R} \). Besides this, let \( \{\lambda_i, b_i\}_{i=1}^\infty \) denote the sequence of eigenpairs associated with the compact self adjoint operator that maps
\[
f \in L^2(D) \mapsto \int_D \text{Cov}[a](x, \cdot)f(x) dx \in L^2(D).
\]
Its non-negative eigenvalues, \( \sqrt{\int_{D \times D} (\text{Cov}[a](x_1, x_2))^2 dx_1 dx_2} \geq \lambda_1 \geq \lambda_2 \geq \ldots \geq 0 \) satisfy
\[
\sum_{i=1}^\infty \lambda_i = \int_D \text{Var}[a](x) dx.
\]
The corresponding eigenfunctions are orthonormal, i.e. \( \int_D b_i(x)b_j(x) dx = \delta_{ij} \). The Karhunen-Loève expansion of the stochastic process \( a \), cf. \([37], [56], [57]\), is
\[
a_N(\omega, x) = E[a](\omega) + \sum_{i=1}^N \sqrt{\lambda_i} b_i(x) Y_i(\omega)
\]
where \( \{Y_i\}_{i=1}^\infty \) is a sequence of uncorrelated real random variables, with mean zero and unit variance. These random variables are uniquely determined by
\[
Y_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_D (a(\omega, x) - E[a](\omega))b_i(x) dx.
\]
for \( \lambda_i > 0 \). Then, by Mercer’s theorem cf. \([35]\) p. 245, we have
\[
\sup_{x \in D} E[(a - a_N)^2](x) = \sup_{x \in D} \sum_{i=N+1}^{+\infty} \lambda_i b_i^2(x) \to 0, \quad \text{as} \quad N \to \infty.
\]
If in addition,
- the images \( Y_i(\Omega) \), \( i = 1, \ldots \), are uniformly bounded in \( \mathbb{R} \)
- the eigenfunctions \( b_i \) are smooth, which is the case when the covariance function is smooth, and uniformly bounded functions, and
- the eigenvalues have the decay \( \lambda_i = O((1 + \cdot)^{-s}) \) for some \( s > 1 \),
then \( \|a - a_N\|_{L^\infty(D)} \to 0 \). Notice that for larger values of \( s \) we can also obtain the convergence of higher spatial derivatives of \( a_N \) in \( L^\infty(D) \). The last two conditions can be readily verified once the covariance function of \( a \) is known. However, observe that it is also necessary to verify the uniform coercivity of \( a_N \), which depends on the probability distributions of \( Y_i \), \( i = 1, \ldots \).

2.4. Finite dimensional noise case. In many problems the source of the randomness can be approximated using just a small number of mutually uncorrelated, sometimes mutually independent, random variables. Take for example the case of a truncated Karhunen-Loève expansion described previously.

Assumption 2.1 (Finite dimensional and independent noise). Whenever we apply some numerical method to solve (1.1) we assume that the coefficients used in the computations satisfy
\[
a(\omega, x) = a(Y_1(\omega), \ldots, Y_N(\omega), x) \quad \text{and} \quad f(\omega, x) = f(Y_1(\omega), \ldots, Y_N(\omega), x)
\]
where \( \{Y_i\}_{i=1}^N \) are real random variables with mean value zero, unit variance, are mutually independent, and their images, \( \Gamma_{i,N} = Y_i(\Omega) \) are bounded intervals in \( \mathbb{R} \) for \( i = 1, \ldots, N \). Moreover, we assume that each \( Y_i \) has a density function \( \rho_i : \Gamma_{i,N} \rightarrow \mathbb{R}^+ \) for \( i = 1, \ldots, N \).

In the sequel we use the notations \( \rho(y) = \Pi_{i=1}^N \rho_i(y) \) \( \forall y \in \Gamma \) for the joint probability density of \((Y_1, \ldots, Y_N)\) and \( \Gamma \equiv \Pi_{i=1}^N \Gamma_{i,N} \subset \mathbb{R}^N \) for the support of such probability density.

After making assumption (2.10), we have by Doob-Dynkin’s lemma, cf. [42], that
\[
\int_I \rho(y) \int_D a(y,x) \nabla u(y,x) \nabla v(y,x) dx dy = \int_I \rho(y) \int_D f(y,x) v(y,x) dx dy, \quad \forall v \in L^2_p(\Gamma) \otimes H^1_0(D).
\]
In this work the gradient notation, \( \nabla \), always means differentiation with respect to \( x \in D \) only unless otherwise stated. The corresponding strong formulation for the problem now becomes an elliptic partial differential equation with an \( N \)-dimensional parameter, i.e.
\[
\begin{align*}
- \nabla \cdot (a(y,x) \nabla u(y,x)) &= f(y,x) \quad \forall (y,x) \in \Gamma \times D, \\
u(y,x) &= 0 \quad \forall (y,x) \in \Gamma \times \partial D.
\end{align*}
\]
Making Assumption (2.10) is a crucial step, turning the original stochastic elliptic equation (1.1) into a deterministic parametric elliptic one and allowing the use of finite element and finite difference techniques to approximate the solution of the resulting deterministic problem.

**Truncation of the outcomes set, \( \Gamma \)**

For computational reasons, it may be useful to compute the solution of (2.12) in a subdomain with strictly positive probability, \( \Gamma_0 \subset \Gamma \), e.g. to save computational work. Beside this, for the sake of the definition and analysis of the numerical methods, we assume the probability density of \( Y \) to be strictly positive in \( \Gamma_0 \). In that case, we shall approximate the function
\[
E[u(Y, \cdot) \chi_{\{Y \in \Gamma_0\}}] = E[u(Y, \cdot) | Y \in \Gamma_0] \text{ } P(Y \in \Gamma_0)
\]
instead of the original \( E[u] \). If \( \bar{u} \) is an approximation of \( u \) in \( \Gamma_0 \) then we have the splitting
\[
\|E[u(Y, \cdot)] - E[\bar{u}(Y, \cdot) \chi_{\{Y \in \Gamma_0\}}]\| \leq \|E[u(Y, \cdot)] - E[u(Y, \cdot) \chi_{\{Y \in \Gamma_0\}}]\| + \|E[u(Y, \cdot)] - \bar{u}(Y, \cdot) | Y \in \Gamma_0]\| P(Y \in \Gamma_0)
\]
Property 2.2 below gives an estimate for the first error contribution, related to the truncation of \( \Gamma \). The second error contribution in (2.13) is the discretization error and it will be analyzed for each numerical approximation, see Sections 3, 5 and 6. In those Sections we shall simplify the notation by writing \( \Gamma = \Gamma_0 \) and work with the corresponding conditional probability space.

**Property 2.2.** Let \( u \) be the solution of the problem (2.12), then there exists a constant \( C \) such that
\[
\|E[u(Y, \cdot)] - E[u(Y, \cdot) \chi_{\{Y \in \Gamma\}}]\|_{L^2_p(D)} \leq C \sqrt{P(Y \notin \Gamma_0)} \|f\|_{L^2_p(\Gamma_0; \otimes L^2(D))}
\]

**Proof.** To derive (2.14), it is enough to estimate
\[
\int_D \int_{\partial \Gamma \cap \Gamma_0} \rho(y) \nabla u(y,x) dy^2 dx \leq P(Y \in \Gamma \backslash \Gamma_0) \int_D \int_{\partial \Gamma \cap \Gamma_0} \rho(y)|\nabla u(y,x)|^2 dy dx \\
\leq P(Y \in \Gamma \backslash \Gamma_0) \frac{C_D}{a_{\min}} \int_{\partial \Gamma \cap \Gamma_0} \rho(y) \int_D |f(y,x)|^2 dx dy.
\]
Remark 2.3. Let us assume that the component \( \{Y_j\}_{j=1}^N \) of \( Y \) are random variables with bounded \( m_0 \)-th moment. Also, we assume that \( \Gamma_0 = \prod_{j=1}^N [-\bar{y}_j, \bar{y}_j] \) where \( \{\bar{y}_j\}_{j=1}^N \) are positive real numbers. Using Markov’s inequality (cf. [11]) we have

\[
P(Y \notin \Gamma_0) = 1 - \prod_{j=1}^N P(|Y_j| \leq \bar{y}_j) \leq 1 - \prod_{j=1}^N \left(1 - \frac{1}{(\bar{y}_j)^{m_0}} E[|Y_j|^{m_0}]\right).
\]

Remark 2.4. If each of the independent random variables \( Y_i, i = 1, \ldots, N \) has a continuous probability density, \( \rho_i \), let \( F_i(x) = \int_{-\infty}^x \rho_i(y)dy \) and define the iid random variables \( Z_i(\omega) \equiv F_i(Y_i(\omega)) \sim U(0,1), i = 1, \ldots, N \). Consider for \( i = 1, \ldots, N \), \( F_i^{-1} : [0,1] \rightarrow \Gamma_i \) and let \( y(z) \equiv (F_i^{-1}(z_i), \ldots, F_N^{-1}(z_N)) \), \( \hat{u}(z,x) \equiv u(y(z),x) \) and \( \hat{f}(z,x) \equiv f(y(z),x) \). Now we can equivalently reformulate (2.11) to: find \( \hat{u} \in L^2([0,1]^N) \otimes H^1_0(D) \) such that

\[
\int_{[0,1]^N} \int_D \hat{a}(z,x) \nabla \hat{u}(z,x) \cdot \nabla v(z,x) dxdz = \int_{[0,1]^N} \int_D \hat{f}(z,x)v(z,x) dxdz, \forall v \in L^2([0,1]^N) \otimes H^1_0(D).
\]

Thus, depending on the features of a particular problem, one may discretize (2.11) or (2.15), and apply the numerical methods and the analysis included in this work.

3. The finite element spaces

In this section, first we consider standard finite element spaces on the spatial set \( D \subset \mathbb{R}^d \) and the outcomes set \( \Gamma \subset \mathbb{R}^N \), separately. Then we define tensor product finite element spaces on the set \( \Gamma \times D \) which we will use to construct approximations of the solution of the parametric boundary value problem (2.12). The section concludes with some useful approximation properties of the tensor finite element spaces.

3.1. Finite element spaces on the spatial set \( D \subset \mathbb{R}^d \): \( h \)-version. Consider a family of piecewise linear (continuous) finite element approximation spaces, \( X_h^d \subset H^1_0(D) \), based on conforming triangulations (of simplices), \( T_h^d \), of the convex polyhedral domain, \( D \subset \mathbb{R}^d \), with a maximum mesh spacing parameter \( h > 0 \). We shall always assume that the triangulations are nondegenerate (sometimes also called regular), cf. [13] p.106. Then, (cf. [13], [15], [33]) the finite element spaces \( X_h^d \) satisfy the standard approximation estimate, namely that for all \( v \in H^2(D) \cap H^1_0(D) \)

\[
(3.1) \quad \min_{\chi \in X_h^d} \|v - \chi\|_{H^1_0(D)} \leq C h \|v\|_{H^2(D)},
\]

where \( C > 0 \) is a constant independent from \( v \) and \( h \).

3.2. Tensor product finite element spaces on the outcomes set \( \Gamma \subset \mathbb{R}^N \): \( k \)-version. Let \( \Gamma = \prod_{j=1}^N \Gamma_j \) be as in Subsection 2.2. Consider a partition of \( \Gamma \) consisting of a finite number of disjoint \( \mathbb{R}^N \)-boxes, \( \gamma = \prod_{j=1}^N (a_j^0, b_j^0), \) with \( (a_j^0, b_j^0) \subseteq \Gamma_j \) for \( j = 1, \ldots, N \). The mesh spacing parameters, \( k_j > 0 \), are defined by \( k_j \equiv \max_j |b_j - a_j| \), for \( 1 \leq j \leq N \). For every non-negative integer \( q \in \mathbb{N} \) consider the finite element approximation space of (discontinuous) piecewise polynomials with degree at most \( q \) on each direction, \( Y_k^{N,q} \subset L^2(\Gamma) \). Thus, if \( \phi \in Y_k^{N,q} \) its restriction to each of the partition boxes satisfies \( \phi \mid_{\gamma_j} \in \text{span}\left( \prod_{j=1}^N y_a^j : \alpha_j \in \mathbb{N} \text{ and } \alpha_j \leq q, \ j = 1, \ldots, N \right) \).

The finite element spaces \( Y_k^{N,q} \) have (cf. Section 4.6 in [13]) the following approximation property: for all \( v \in H^{q+1}(\Gamma) \)

\[
(3.2) \quad \min_{\phi \in Y_k^{N,q}} \|v - \phi\|_{L^2(\Gamma)} \leq C \sum_{j=1}^N (k_j)^{q+1} \|\partial_{\gamma_j}^{q+1} v\|_{L^2(\Gamma)}
\]
where $C > 0$ is a constant independent of $v$ and $k_j > 0$.

### 3.3. Tensor product finite element spaces on $\Gamma \times D$: $k \times h$–version.

Here, we shall discuss some approximation properties of the following tensor product finite element spaces

\[
Y^{N,q}_k \otimes X^d_k \equiv \{ \psi = \psi(y,x) \in L^2(\Gamma \times D) : \psi \in \text{span} \{ \varphi(y) \chi(x) : \varphi \in Y^{N,q}_k, \chi \in X^d_k \} \}.
\]

with $X^d_k$ and $Y^{N,q}_k$ as in Subsections 3.1 and 3.2.

For later use we recall the standard $L^2$–projection operators $\Pi^d_k : L^2(D) \to X^d_k$, $\Pi^{N,q}_k : L^2(\Gamma) \to Y^{N,q}_k$ by

\[
\begin{align*}
(\Pi^d_k v - v, \chi)_{L^2(D)} &= 0, \quad \forall \chi \in X^d_k, \quad \forall v \in L^2(D), \\
(\Pi^{N,q}_k w, \varphi)_{L^2(\Gamma)} &= 0, \quad \forall \varphi \in Y^{N,q}_k, \quad \forall w \in L^2(\Gamma),
\end{align*}
\]

and the $H^1_0$ projection operator $\mathcal{R}^d_k : H^1_0(D) \to X^d_k$ by

\[
(\nabla \mathcal{R}^d_k v - v, \nabla \chi)_{L^2(D)} = 0, \quad \forall \chi \in X^d_k, \quad \forall v \in H^1_0(D).
\]

Estimates (3.1) and (3.2) imply

\[
\|v - \Pi^d_k v\|_{L^2(D)} + h \|v - \mathcal{R}^d_k v\|_{H^1_0(D)} \leq C h^2 \|v\|_{H^2(D)},
\]

\[
\|w - \Pi^{N,q}_k w\|_{L^2(\Gamma)} \leq C \sum_{j=1}^N (k_j)^q + 1 \|\partial_{y_j}^{q+1} w\|_{L^2(\Gamma)},
\]

for all $v \in H^2(D) \cap H^1_0(D)$ and $w \in H^{q+1}(\Gamma)$.

We now present an approximation property for the tensor product finite element spaces defined in (3.3) which is a direct implication of the approximation properties of the spaces $Y^{N,q}_k$ and $X^d_k$.

**Proposition 3.1.** Let $q$ be a nonnegative integer. Then, there exists a constant $C > 0$ depending only on $d$, $N$ and $q$, such that

\[
\inf_{\psi \in Y^{N,q}_k \otimes X^d_k} \|v - \psi\|_{L^2(\Gamma; H^1_0(D))} \leq C \left\{ h \|v\|_{L^2(\Gamma; H^2(D))} + \sum_{j=1}^N (k_j)^q + 1 \||\partial_{y_j}^{q+1} v\||_{L^2(\Gamma; H^1_0(D))} \right\}
\]

for all $v \in C^{q+1}(\Gamma; H^2(D) \cap H^1_0(D))$.

**Proof.** Since $\Pi^d_k(\mathcal{R}^d_k v) \in Y^{N,q}_k \otimes X^d_k$, by (3.6) we obtain

\[
\inf_{\psi \in Y^{N,q}_k \otimes X^d_k} \|v - \psi\|_{L^2(\Gamma; H^1_0(D))} \leq \|v - \Pi^d_k(\mathcal{R}^d_k v)\|_{L^2(\Gamma; H^1_0(D))}
\]

\[
\leq \|v - \mathcal{R}^d_k v\|_{L^2(\Gamma; H^1_0(D))} + \|\mathcal{R}^d_k v - \Pi^d_k(\mathcal{R}^d_k v)\|_{L^2(\Gamma; H^1_0(D))}
\]

\[
\leq C h \|v\|_{L^2(\Gamma; H^2(D))} + \|\mathcal{R}^d_k v - \Pi^d_k(\mathcal{R}^d_k v)\|_{L^2(\Gamma; H^1_0(D))},
\]

Applying the estimate (3.2) and using the boundedness of $\mathcal{R}^d_k$ in $H^1_0(D)$ yields

\[
\|\mathcal{R}^d_k v - \Pi^d_k(\mathcal{R}^d_k v)\|_{L^2(\Gamma; H^1_0(D))} \leq \|v - \Pi^d_k(\mathcal{R}^d_k v)\|_{L^2(\Gamma; H^1_0(D))} \leq \sum_{j=1}^N (k_j)^q + 1 \||\partial_{y_j}^{q+1} v\||_{L^2(\Gamma; H^1_0(D))},
\]

The estimate (3.7) follows combining (3.8) with the last estimate. \(\square\)
3.4. Tensor product finite element spaces on $\Gamma \times D$: $p \times h$--version. Consider again the set $\Gamma = \Gamma_1 \times \ldots \times \Gamma_N \subset \mathbb{R}^N$, a multi-index $p = (p_1, \ldots, p_N)$ and the tensor finite element space

$$Z^p = \bigotimes_{i=1}^N Z_i^{p_i}$$

where the one dimensional global polynomial subspaces, $Z_i^{p_i}$, are defined by

$$Z_i^{p_i} = \{ v : \Gamma_i \rightarrow \mathbb{R} : v \in \text{span}(y^s, s = 0, \ldots, p_i) \}, \quad i = 1, \ldots, N.$$  

Here, the basic tool for approximation are the following tensor product finite element spaces

$$(3.9) \quad Z^p \otimes X^d_h \equiv \{ \psi = \psi(y, x) \in L^2(\Gamma \times D) : \psi \in \text{span}(\varphi(y) \chi(x) : \varphi \in Z^p, \chi \in X^d_h) \}. $$

Similarly to Subsection 3.3 the $p \times h$--version has the following spatial approximation property

**Proposition 3.2.** Let $v \in L^2(\Gamma; H^2(D) \cap H^1_0(D))$ then there exists a constant $C > 0$ independent of $v$ and $h$ such that

$$\inf_{\psi \in L^2(\Gamma) \otimes X^d_h} \| v - \psi \|_{L^2(\Gamma; H^2(D))} \leq C h \| v \|_{L^2(\Gamma; H^2(D))} $$

The approximation properties in the $y$--direction will be studied in Section 6.

4. The Monte Carlo Galerkin Finite Element Method

In this section we describe the use of the standard Monte Carlo Galerkin Finite Element Method (MCGFEM) to construct approximations to some solution statistics. These approximations are based on sample averages of independent, identically distributed realizations corresponding to sample coefficient functions. For each realization of the coefficients a realization of the approximate solution is computed with standard finite element spaces. Our discussion will be based on the approximation properties in the $y$--direction will be studied in Section 6.

Formulation of the Monte Carlo Galerkin Finite Element Method (MCGFEM):  
- Give a number of realizations, $M$, a piecewise linear finite element space on $D$, $X^d_h$, defined in subsection 3.1.  
- For each $j = 1, \ldots, M$ sample iid realizations of the diffusion $a(\omega_j, \cdot)$, the load $f(\omega_j, \cdot)$ and find a corresponding approximation $u_h(\omega_j, \cdot) \in X^d_h$ such that

$$(4.1) \quad B_{\omega_j}(u_h(\omega_j, \cdot), \chi) \equiv \int_D a(\omega_j, \cdot) \nabla u_h(\omega_j, \cdot) \nabla \chi dx = (f(\omega_j, \cdot), \chi), \quad \forall \chi \in X^d_h.$$  

- Finally use the sample average $\frac{1}{M} \sum_{j=1}^M u_h(\omega_j, \cdot)$ to approximate $E[u]$.  

Here we only consider the case where $X^d_h$ is the same for all realizations, i.e. the spatial triangulation is deterministic. The computational error naturally separates into the two parts

$$(4.2) \quad E[u] - \frac{1}{M} \sum_{j=1}^M u_h(\omega_j, \cdot) = \left( E[u] - E[u_h] \right) + \left( E[u_h] - \frac{1}{M} \sum_{j=1}^M u_h(\omega_j, \cdot) \right) \equiv \mathcal{E}_h + \mathcal{E}_S.$$  

The size of the spatial triangulation controls the space discretization error $\mathcal{E}_h$, cf. Proposition 4.2. while the number of realizations, $M$ of $u_h$, controls the statistical error $\mathcal{E}_S$.

To study the behavior of the statistical error, let us first consider the random variable $\| \mathcal{E}_S \|_{L^2(D)}$. Its variance satisfies the estimate

$$(4.3) \quad M E[\| \mathcal{E}_S \|_{H^1_0(D)}^2] \leq \| u_h \|_{H^1_0(D)}^2 \leq \left( \frac{C_D}{\min} \right)^2 \| f \|_{L^2(D)}^2.$$
and a similar result also holds in $L^2(D)$. Then, thanks to (4.3) we have that, for either $H = L^2(D)$ or $H = H^1_0(D)$, the statistical error $\mathcal{E}_S$ tends a.s. to zero as we increase the number of realizations, i.e.

**Proposition 4.1.** Suppose that there exist a constant $C > 0$ independent from $M$ and $h$ such that the statistical error in $H$ norm satisfies

$$M \mathbb{E}[\|\mathcal{E}_S\|_H^2] \leq C, \quad \forall M, h.$$  \hspace{1cm} (4.4)

Then, taking number of realizations increasingly from the set $\{2^k : k \in \mathbb{N}\}$ we have for any $\alpha \in (0, 1/2)$ and any choice of mesh size $h$

$$\lim_{M \to \infty} M^\alpha \|\mathcal{E}_S\|_H = 0 \text{ a.s.}$$

**Proof.** Let $\epsilon > 0$. Then (4.4) and Markov’s inequality give

$$P(M^\alpha \|\mathcal{E}_S\|_H > \epsilon) \leq \frac{\mathbb{E}[M^{2\alpha} \|\mathcal{E}_S\|_H^2]}{\epsilon^2} \leq \frac{C}{\epsilon^2 M^{1-2\alpha}}.$$ 

Thus

$$\sum_{k=1}^{\infty} P(M_k^\alpha \|\mathcal{E}_S\|_H > \epsilon) \leq \frac{C}{\epsilon^2} \sum_{k=1}^{\infty} \frac{1}{M_k^{1-2\alpha}} \leq \frac{C}{\epsilon^2} \sum_{k=1}^{\infty} \frac{1}{(2^{1-2\alpha})^k} < \infty$$

and Borel-Cantelli’s lemma implies that for any given $\epsilon > 0$ and $\alpha \in (0, 1/2)$

$$P(M^\alpha \|\mathcal{E}_S\|_H > \epsilon \text{ infinitely often}) = 0$$

which finishes the proof. 

Under the same assumptions of Proposition 4.1 we have that for any given $\epsilon > 0$ there exist a constant $C > 0$ independent from $\epsilon$, $M$ and $h$ such that

$$P\left(\|\mathcal{E}_H\| \geq \frac{\epsilon}{\sqrt{M}}\right) \leq \frac{C}{\epsilon^2}.$$  \hspace{1cm} (4.5)

Thus, within a given confidence level we have the usual convergence rate for the Monte Carlo method, which is independent from the mesh size $h$. Next we present some error estimates for the space discretization error, $\|E[u - u_h]\|_{H^1_0(D)}$ and $\|E[u - u_h]\|_{L^2(D)}$. The first two lemmata estimate strong errors, which yield in Remark 4.2 a bound for the spatial discretization error in $E[u]$. The proofs are done assuming piecewise linear finite element spaces on regular meshes and follow the standard arguments for the deterministic case, cf. [13].

**Lemma 4.1.** There holds

$$\|\sqrt{a} \nabla (u - u_h)\|_{L^2(D)} \leq C h \|u\|_{H^2(D)}, \quad \text{a.s.}$$ \hspace{1cm} (4.6)

$$\|u - u_h\|_{L^2(D)} \leq C h^2 \|u\|_{H^2(D)}, \quad \text{a.s.}$$ \hspace{1cm} (4.7)

**Proof.** Let $\omega \in \Omega$. Then

$$B_\omega(e, e) = B_\omega(e, u - \mathcal{R}^d_h u)$$

\leq a_{\max} \|e\|_{H^1_0(D)} \|\nabla (u - \mathcal{R}^d_h u)\|_{L^2(D)}

\leq C h \|u\|_{H^2(D)} \|u\|_{H^2(D)}$$

and the use of the uniform coercivity assumption [1.2] yields (4.6). To prove (4.7) we follow a standard duality argument. Let $\omega \in \Omega$, then consider the continuous dual problem

$$B_\omega(\varphi, v) = (e, v), \quad \forall v \in H^1_0(D)$$

$14$
and then we use elliptic regularity, \( \| \varphi \|_{H^2(D)} \leq C \| e \|_{L^2(D)} \). Notice that the constant of elliptic regularity depends in general on \( \omega \) and we have used assumption (1.3) to control it uniformly.

Therefore,
\[
\| e \|_{L^2(D)}^2 = B_\omega(\varphi, e) = B_\omega(\varphi - R_h^d \varphi, e) \\
\leq a_{\text{max}} \| e \|_{H^1_0(D)} \| \varphi - R_h^d \varphi \|_{H^1_0(D)} \\
\leq \| e \|_{H^1_0(D)} C h \| \varphi \|_{H^2(D)} \\
\leq Ch^2 \| f \|_{L^2(D)} \| e \|_{L^2(D)}.
\]

□

A direct application of Lemma 4.1 gives

**Proposition 4.2** (Spatial discretization error estimates). There holds
\[
\| E[u] - E[u_h] \|_{H^1_0(D)} \leq E[\| u - u_h \|_{H^1_0(D)}^2]^{1/2} \leq C h \| f \|_{L^2(D)}
\]
and similarly
\[
\| E[u] - E[u_h] \|_{L^2(D)} \leq E[\| u - u_h \|_{L^2(D)}^2]^{1/2} \leq C h^2 \| f \|_{L^2(D)}.
\]

The results from Proposition 4.2 and estimate (4.5) will be used in Section 8 to compare the SGFEM with other discretizations for (1.1).

5. **The Stochastic Galerkin Finite Element Method: \( k \times h \)-version**

This section defines and analyzes the \( k \times h \)-version of the stochastic Galerkin finite element method (\( k \times h \)-SGFEM) which, via a Galerkin variational formulation, yields approximations, \( u_{kh} \in \mathcal{Y}_k^r \otimes X_h^d \), of the solution of the parametric elliptic boundary value problem (2.12). Then, the section ends by showing how to use \( u_{kh} \) to construct approximations of the expected value of \( u \), analyzing the corresponding approximation error.

**Formulation of the \( k \times h \)-SGFEM:** Let \( q \in \mathbb{N} \) and \( \Gamma \) be a bounded box in \( \mathbb{R}^N \). The \( k \times h \)-SGFEM approximation is the tensor product, \( u_{kh} \in \mathcal{Y}_k^r \otimes X_h^d \), such that
\[
\int_\Gamma \rho (a \nabla_x u_{kh}, \nabla_x \psi)_{L^2(D)} \, dy = \int_\Gamma \rho (f, \psi)_{L^2(D)} \, dy, \quad \forall \psi \in \mathcal{Y}_k^r \otimes X_h^d.
\]

Recall that \( N \in \mathbb{N} \), \( \rho : \Gamma \to (0, +\infty) \) is the density function of the vector-valued random variable \( Y : \Omega \to \Gamma \subset \mathbb{R}^N \) which has mutually independent components. Hence, the assumption (1.2) on the random function \( a(\omega, \cdot) \equiv a(Y(\omega), \cdot) \) reads
\[
a(y, \cdot) \in [a_{\text{min}}, a_{\text{max}}], \quad \forall y, \cdot) \in \Gamma \times \bar{D}.
\]

For the case where \( a \) is a truncated Karhunen Loève expansion Section 7 discusses how to compute efficiently \( u_{kh} \), the solution of (5.1), by a double-orthogonal polynomials technique. To carry out the analysis of the \( k \times h \)-version, assume that
\[
a \in C^{q+1,1}(\bar{\Gamma} \times \bar{D}) \quad \text{and} \quad f \in C^{q+1}(\Gamma; L^2(D)),
\]
which is trivially satisfied whenever \( a \) and \( f \) have the form of a truncated Karhunen Loève expansion (cf. Section 6). Hence, by Lemma 4.1 in [35], the solution \( u \) of (2.12) satisfies
\[
u \in C^{q+1}(\Gamma; H^2(D) \cap H^1_0(D)).
\]

Use (5.2) and (2.12) to obtain
\[
\| u(y, \cdot) \|_{H^1_0(D)} \leq \frac{C_{\text{sp}}}{a_{\text{min}}} \| f(y, \cdot) \|_{L^2(D)}, \quad \forall y \in \Gamma,
\]
where $C_\omega$ is the constant of the Poincaré–Friedrichs inequality on $D$. Also, elliptic regularity yields

$$ \|u(y, \cdot)\|_{H^2(D)} \leq C_{0,\omega} \|f(y, \cdot)\|_{L^2(D)}, \quad \forall y \in \Gamma, \tag{5.5} $$

where $C_{0,\omega}$ is a constant which depends on $D$ and $\|a\|_{L^\infty(\Gamma; W^{1,\infty}(D))}$. Finally, take derivatives with respect to $y_j$ in (2.12), proceed as in the derivation of (5.4), and follow an induction argument arriving at

$$ \|\partial_{y_j}^{q+1}u(y, \cdot)\|_{H^q_{\Gamma}(D)} \leq C_{q+1,\omega} \sum_{0 \leq \beta \leq q+1} \|\partial_{y_j}^{\beta}f(y, \cdot)\|_{L^2(D)}, \quad \forall y \in \Gamma, \quad 1 \leq j \leq N \tag{5.6} $$

where $C_{q+1,\omega}$ is a constant which depends on $q, C_\omega, a_{\min}$ and $\|a\|_{L^\infty(\Gamma; W^{1,\infty}(\Gamma))}$. We now derive an a priori error estimate for the $k \times h$–SGFEM in the energy norm, that will be later used to derive an error estimate for $E[u - u_{kh}]$.

**Proposition 5.1.** Let $u$ be the solution of the problem (2.12) and $u_{kh} \in Y^q_k \otimes X^d_h$ be the $k \times h$–SGFEM approximations of $u$ defined in (5.1). If (5.3) holds and $\rho \in L^\infty(\Gamma)$, then

$$ \left( \int_{\Gamma} \rho \|\sqrt{\alpha} \nabla_x (u - u_{kh})\|_{L^2(D)}^2 dy \right)^{\frac{1}{2}} \leq C \sqrt{\|\rho\|_{L^\infty(\Gamma \times D)}} \left( h \|f\|_{L^2(\Gamma; L^2(D))} + \sum_{j=1}^N (k_j)^{q+1} \sum_{0 \leq \beta \leq q+1} \|\partial_{y_j}^{\beta}f\|_{L^2(\Gamma; L^2(D))} \right) \tag{5.7} $$

where the constant $C$ depends on $q, D, \Gamma$ and $a$, and is independent of $k, h$ and $u$.

**Proof.** Let $V_{kh} = Y^q_k \otimes X^d_h$ and $e = u - u_{kh}$. Combining (5.1) with (2.12) gives the standard Galerkin orthogonality

$$ \int_{\Gamma} \rho \langle a \nabla x e, \nabla x \psi \rangle_{L^2(D)} \ dy = 0, \quad \forall \psi \in V_{kh}, \tag{5.8} $$

and thus

$$ \left( \int_{\Gamma} \rho \|\sqrt{\alpha} \nabla_x e\|_{L^2(D)}^2 dy \right)^{\frac{1}{2}} \leq \inf_{\psi \in V_{kh}} \left( \int_{\Gamma} \rho \|\sqrt{\alpha} \nabla_x (u - \psi)\|_{L^2(D)}^2 dy \right)^{\frac{1}{2}} \leq \|a \rho\|_{L^\infty(\Gamma \times D)}^{\frac{1}{2}} \inf_{\psi \in V_{kh}} \|u - \psi\|_{L^2(\Gamma; H^1(D))}. \tag{5.9} $$

Using (5.9) and the estimate (3.7) implies

$$ \left( \int_{\Gamma} \rho \|\sqrt{\alpha} \nabla_x e\|_{L^2(D)}^2 dy \right)^{\frac{1}{2}} \leq C \|a \rho\|_{L^\infty(\Gamma \times D)}^{\frac{1}{2}} \left( h \|u\|_{L^2(\Gamma; H^2(D))} + \sum_{j=1}^N (k_j)^{q+1} \|\partial_{y_j}^{q+1}u\|_{L^2(\Gamma; H^2_{\Gamma}(D))} \right). $$

The last estimate combined with (5.5) and (5.6) yield (5.7). \( \square \)

As a direct result from (5.7) we obtain

**Corollary 5.1 (Convergence result).** Under the assumptions of Proposition 5.1 we have

$$ \|E[u(Y, \cdot)] - E[u_{kh}(Y, \cdot)]\|_{H^2_{\Gamma}(D)} \leq C \left( h + \sum_{j=1}^N (k_j)^{q+1} \right) $$

with $C > 0$ independent from $u, h$ and $k_j$.

The next step is to use Proposition 5.1 to estimate the $L^2(D)$ error in the approximation of the expected value of $u(Y, \cdot)$. 

16
Theorem 5.2. Let \( u \) be the solution of the problem (2.12) and \( u_{kh} \in Y_k^r \otimes X_h^d \) be the \( k \times h \)-SGFEM approximations of \( u \) defined in (5.1). If (5.3) holds and \( \rho \in L^\infty(\Gamma) \), then

\[
\| E[u(Y, \cdot) - E[u_{kh}(Y, \cdot)] \|_{L^2(D)} \leq C \left( h^2 + \sum_{j=1}^N (k_j)^{2(q+1)} \right) \| a \rho \|_{L^\infty(\Gamma \times D)} \| f \|_{L^2(D; H^{q+1}(\Gamma))}.
\]

The constant \( C \) depends on \( q, D, \Gamma \) and \( a \), and it is independent from \( k, h \) and \( u \).

Proof. Let \( V_{kh} = Y_k^r \otimes X_h^d \), \( e \equiv \hat{u} - u_{kh} \), and \( g \equiv E[e(Y, \cdot)] \in H^1_0(D) \). Then, we consider \( \hat{u} \in C^{q+1}(\Gamma; H^2(D) \cap H^1_0(D)) \) defined, for \( y \in \Gamma \), as the solution of the following elliptic boundary value problem

\[
-\nabla_x^t (a(y, \cdot) \nabla_x \hat{u}(y, \cdot)) = g(\cdot) \text{ in } D,
\]

\[
\hat{u}(y, \cdot) = 0 \text{ on } \partial D.
\]

Then (5.5) reads

\[
\| \hat{u}(y, \cdot) \|_{H^2(D)} \leq C_{0, a} \| g \|_{L^2(D)}, \quad \forall y \in \Gamma,
\]

and since \( g \) is independent from \( y \), the estimate (5.6) for the problem (5.11) reads

\[
\sum_{|\beta| \leq q+1} \| \partial^\beta_y \hat{u}(y, \cdot) \|_{H^1(D)}^2 \leq C_{q+1, a} \| g \|_{L^2(D)}^2, \quad \forall y \in \Gamma.
\]

Use (5.11) and the orthogonality property (5.8) to obtain

\[
\int_D \rho (g, e)_{L^2(D)} \, dy = \int_D \rho (a \nabla_x e, \nabla_x (\hat{u} - \psi))_{L^2(D)} \, dy, \quad \forall \psi \in V_{kh},
\]

which yields, by Cauchy Schwartz inequality,

\[
\| g \|_{L^2(D)}^2 \leq \tilde{B}_1 \tilde{B}_2
\]

with

\[
\tilde{B}_1 \equiv \left( \int_D \rho \| \sqrt{a} \nabla_x e \|_{L^2(D)}^2 \, dy \right)^{\frac{1}{2}}
\]

and

\[
\tilde{B}_2 \equiv \inf_{\psi \in V_{kh}} \left( \int_D \rho \| \sqrt{a} \nabla_x (\hat{u} - \psi) \|_{H^1(D)}^2 \, dy \right)^{\frac{1}{2}}.
\]

Next, observe that the energy estimate (5.7) gives

\[
\tilde{B}_1 \leq C \left( h + \sum_{j=1}^N (k_j)^{q+1} \right) \| a \rho \|_{L^\infty(\Gamma \times D)} \| f \|_{L^2(D; H^{q+1}(\Gamma))}.
\]

Finally, use (3.7), (5.12) and (5.13), to bound \( \tilde{B}_2 \) as follows

\[
\tilde{B}_2 \leq C \| a \rho \|_{L^\infty(\Gamma \times D)} \left\{ h \left( \int_D \| \hat{u} \|_{H^2(D)}^2 \, dy \right)^{1/2}
\right\}
\]

\[
+ \sum_{j=1}^N (k_j)^{q+1} \left( \int_D \| \partial^q_y \hat{u} \|_{H^1(D)}^2 \, dy \right)^{\frac{1}{2}}
\]

\[
\leq C \| a \rho \|_{L^\infty(\Gamma \times L^\infty(D))} \left( h + \sum_{j=1}^N (k_j)^{q+1} \right) \| g \|_{L^2(D)}.
\]

Combining (5.14), (5.15) and (5.16), the estimate (5.10) follows. \( \square \)
Remark 5.1. Let \( k \equiv \max_{1 \leq j \leq N} k_j \) and assume that \( u \in C^{q+1}(\Gamma; H^{s+1}(D) \cap H_0^1(D)) \) for some integer \( s \geq 1 \), and that \( X_h^d \) are standard finite element spaces consisting of continuous piecewise polynomial functions of degree less or equal to \( s \) (cf., e.g., [13, 15, 43]). Then, proceeding in a similar way we get the estimate \( \| E[(u(Y, \cdot) - u_{kh}(Y, \cdot))] \|_{L^2(D)} \leq C \left( h^{s+1} + k^{\frac{(q+1)(s+1)}{2}} \right) \) which is a superconvergence result with respect to \( k \) that generalizes the estimate (5.10).

6. The Stochastic Galerkin Finite Element Method: \( p \times h \)–version

The goal of this section is to introduce and analyze the \( p \times h \)–version of the SGFEM method, that uses global polynomials in the \( y \) direction instead of piecewise discontinuous ones. This method yields, cf. Theorem 6.2, an exponential rate of convergence with respect to \( p \), the degree of the polynomials used for approximation. The work by Gui and Babuška [20] on the \( h - p \) versions of the Finite Element in one dimension inspired the current analysis of the \( p \)-version tensor product approximation in the \( y \) direction. The application of the \( p \)-version in the \( y \) direction is motivated by the fact that \( u \) is infinitely differentiable with respect to \( y \in \Gamma \), cf. Lemma 6.1 and the comparison of computational work in Section 8. As the \( k \times h \)–version studied in Section 5 the \( p \)-version also gives a super-convergent approximation of the expected value of the solution, cf. Theorem 6.3.

The basic assumptions for this section are summarized as follows:

Assumption 6.1. Assume that the density function, \( \rho : \Gamma \rightarrow \mathbb{R} \) of the \( N \) dimensional random vector \( Y \), is bounded and that the components of \( Y \) are mutually independent. Beside this, the functions \( a, f : \Gamma \times D \rightarrow \mathbb{R} \) are finite Karhunen-Loève expansions, i.e., \( a(y, x) = E[a](x) + \sum_{i=1}^N b_i(x)y_i \) and \( f(y, x) = E[f](x) + \sum_{i=1}^N \hat{b}_i(x)y_i \). Moreover, the uniform coercivity assumption \( (1.2) \) and the right hand side assumption \( (1.4) \) yield \( b_i \in C^1(D) \), \( \hat{b}_i \in L^2(D) \) for \( i = 1, \ldots, N \). Use the notation

\[
\hat{\Gamma}_i = \prod_{1 \leq j \leq N, j \neq i} \Gamma_j
\]

and let \( \hat{y}_i \) be an arbitrary element of \( \hat{\Gamma}_i \). Then, for each \( \hat{y}_i \in \hat{\Gamma}_i \) let

\[
\tilde{a}_i(\hat{y}_i) \equiv \min_{x \in D} \left\{ E[a](x) + \sum_{1 \leq j \leq N, j \neq i} b_j(x)y_j \right\}
\]

and assume a slightly stronger uniform coercivity requirement, i.e. there exists a constant \( \bar{\nu} > 0 \) such that

\[
\nu_i(\hat{y}_i) \equiv \tilde{a}_i(\hat{y}_i) - \| b_i \|_{L^\infty(D)} \max_{y \in \Gamma_i} |y| \geq \bar{\nu} > 0, \quad \forall \hat{y}_i \in \hat{\Gamma}_i, \quad i = 1, \ldots, N.
\]

Observe that with the above construction we have \( 0 < \bar{\nu} \leq a_{\min} \). 

\( p \times h \)–version of the SGFEM method: Let and \( \Gamma \) be a bounded box in \( \mathbb{R}^N \). Then for \( h \in (0, h_0) \) the \( p \times h \)–version SGFEM approximation is the tensor product \( u_h^p \in Z^p \otimes X_h^d \), recall (3.9), that satisfies

\[
\int_{\Gamma} \rho \left( a \nabla_x u_h^p, \nabla_x \chi \right)_D \, dy = \int_{\Gamma} \rho \left( f, \chi \right)_D \, dy \quad \forall \chi \in Z^p \otimes X_h^d.
\]
6.1. Error estimates. To illustrate the convergence of the $p$-version here we study for the particular case of the energy error, i.e. consider

$$
\|u - u_h^p\|_H = \sqrt{\int_{\Gamma} \rho(y) \int_D a(y, x)|\nabla(u - u_h^p)(y, x)|^2dxdy}
\leq \sqrt{\|\rho a\|_{L^\infty(\Gamma \times D)} \min_{v \in Z^p \otimes X_h^2} \|u - v\|_{L^2(\Gamma) \otimes H^1_0(D)}}
\leq \sqrt{\|\rho a\|_{L^\infty(\Gamma \times D)} \left\{ \min_{v \in Z^p \otimes H^1_0(D)} \|u - v\|_{L^2(\Gamma) \otimes H^1_0(D)} + \min_{v \in L^2(\Gamma) \otimes X_h^2} \|u - v\|_{L^2(\Gamma) \otimes H^1_0(D)} \right\}}.
$$

This bound splits the error into an $L^2(\Gamma)$ approximation error and a standard $H^1_0(D)$ FEM approximation error. The rest of this section studies the first one, since for the second we can apply the results from Proposition 3.2. The minimizer

$$
\|u - u^p\|_{L^2(\Gamma) \otimes H^1_0(D)} = \min_{v \in Z^p \otimes H^1_0(D)} \|u - v\|_{L^2(\Gamma) \otimes H^1_0(D)}
$$

is the projection

$$
u^p = (\Pi_1 \ldots \Pi_N)u
$$

with $\Pi_i : L^2(\Gamma) \otimes H^1_0(D) \rightarrow L^2(\Gamma) \otimes H^1_0(D)$ being the natural extension of the $L^2$ projection $\Pi_i : L^2(\Gamma_i) \rightarrow Z^p_i$, so the difference $u - u^p$ splits into

$$
u - u^p = (1 - \Pi_1)u + \ldots + (\Pi_1 \ldots \Pi_{N-1})(1 - \Pi_N)u.
$$

In addition, the boundedness of the projections $\Pi_i$ yields

$$
(6.2) \quad \|u - u^p\|_{L^2(\Gamma) \otimes H^1_0(D)} \leq \sum_{i=1}^N \|(1 - \Pi_i)u\|_{L^2(\Gamma) \otimes H^1_0(D)}.
$$

Without loose of generality we now consider the first term on the right hand side of (6.2), since the other terms have a completely similar behavior. Moreover, since

$$
\|(1 - \Pi_1)u\|^2_{L^2(\Gamma) \otimes H^1_0(D)} = \int_{\Gamma_2 \times \ldots \times \Gamma_N} \left( \int_{\Gamma_1} \|(1 - \Pi_1)u(y_1, y_2, \ldots, y_N, \cdot)\|^2_{H^1_0(D)}dy_1 \right) dy_2 \ldots dy_N
$$

it is enough to estimate

$$
(6.3) \quad (E_1)^2(y_2, \ldots, y_N) \equiv \int_{\Gamma_1} \|(1 - \Pi_1)u(y_1, y_2, \ldots, y_N, \cdot)\|^2_{H^1_0(D)}dy_1
$$

and thus our analysis requires only one dimensional arguments in the $y$-direction. Let $\Gamma_1 = (y_{\min}, y_{\max})$ and consider the map $\Psi : (-1, 1) \rightarrow H^1_0(D)$ defined by

$$
\Psi(t) = u(y_1(t), y_2, \ldots, y_N, \cdot) \in H^1_0(D)
$$

with the affine transformation, $y_1 : [-1, 1] \rightarrow \Gamma_1$, $y_1(t) \equiv \left( \frac{y_{\max} + y_{\min}}{2} + \frac{y_{\max} - y_{\min}}{2} \right) t$. In the upcoming estimate of the quantities $\|d_n\|_{H^1_0(D)}$, to be proved in Lemma 6.2 we need to consider a continuation of $\Psi$ to the complex plane, namely that

**Lemma 6.1** (Complex continuation). The function $\Psi : [-1, 1] \rightarrow H^1_0(D)$ can be analytically continued to the complex domain.

**Proof.** Let $t_0 \in (-1, 1)$. We shall prove that the real function $\Psi$ can be represented as a power series for $|t - t_0| < r_{t_0}$, for some $r_{t_0} > 0$. Since $\Psi$ depends linearly on $f$, let us assume that
\[ f(y, x) = f(x) \text{ only, without loss of generality. Let } y(t) = (y_1(t), y_2, \ldots, y_N) \text{ and consider the formal series} \]
\[ u_F(t) \equiv \sum_{j=0}^{+\infty} \left( \frac{|\Gamma_1|(t - t_0)}{2} \right)^j u_j \]
with \( u_j \in H^j_0(D) \) satisfying
\[ -\nabla \cdot (a(y(t_0), x) \nabla u_0(x)) = f(x), \quad \forall x \in D, \]
\[ u_0(x) = 0 \quad \forall x \in \partial D, \]
and for \( j \geq 0 \)
\[ -\nabla \cdot (a(y(t_0), x) \nabla u_{j+1}(x)) = \nabla \cdot (b_1(x) \nabla u_j(x)) \quad \forall x \in D, \]
\[ u_{j+1}(x) = 0 \quad \forall x \in \partial D. \]
This construction implies
\[ \|u_{j+1}\|_{H^j_0(D)} \leq \frac{C_D \|b_1\|_{L^\infty(D)}}{a_{\text{min}}} \|u_j\|_{H^j_0(D)}, \quad j \geq 1 \]
and then
\[ \|u_F\|_{H^1_0(D)} \leq \frac{C_D \|f\|_{L^2(D)}}{a_{\text{min}}} \frac{1}{1 - q} < \infty \]
for \( q \equiv \frac{|t-t_0|\|\Gamma_1\|_{C(D)}\|b_1\|_{L^\infty(D)}}{2a_{\text{min}}} < 1 \). Thus for any \( t_0 \in (-1, 1) \) and \( |t - t_0| < r_{t_0} \equiv \frac{2a_{\text{min}}}{\|\Gamma_1\|_{C(D)}\|b_1\|_{L^\infty(D)}} \), the function \( u_F \) can be represented as a power series in \( t - t_0 \). At the same time, we have the equality \( u_F(t) = \Psi(t) \) for \( t \in (-1, 1) \) since both functions solve the linear elliptic equation
\[ -\nabla \cdot (a(y(t), x) \nabla u(y(t), x)) = f(x), \quad \forall x \in D, \]
\[ u(y(t), x) = 0 \quad \forall x \in \partial D, \]
which has a unique solution. Then \( u_F \) is the analytic continuation of \( \Psi \) and the proof is complete.

\[ \square \]

**Remark 6.1.** Consider the natural extension of the variable \( t \) to the complex \( \eta \). Observe that \( \Psi(\eta) \) from Lemma 6.1 solves
\[ -\nabla \cdot (a(y(\eta), x) \nabla \Psi(\eta, x)) = f(x), \quad \forall x \in D, \]
\[ \Psi(\eta, x) = 0 \quad \forall x \in \partial D. \]
Following [20], we use the Legendre polynomials to prove approximation estimates for the \( p \times h \)-version of the SGFEM. Since the Legendre polynomials,
\[ p_n(t) \equiv \frac{1}{2^n n!} \frac{d^n}{dt^n}((t^2 - 1)^n), \quad n = 0, 1, \ldots \]
are orthogonal with respect to the \( L^2(-1, 1) \) inner product we have the error representation
\[ (E_1)^2(y_2, \ldots, y_N) = \frac{|\Gamma_1|}{2} \sum_{n=p_1+1}^{+\infty} \frac{2}{2n+1} \|d_n\|_{H^1_0(D)}^2 \]
with the corresponding Fourier coefficients
\[ d_n \equiv \frac{2n+1}{2} \int_{-1}^{1} \Psi(t)p_n(t)dt \in H^1_0(D). \]
Therefore, to obtain an estimate for \( E_1 \) we shall study the convergence of the tail series in (6.7).
Notation 6.1. Consider the natural extension of the variable \( t \) to the complex \( \eta \) and introduce the real function, \( \tilde{a} : \mathbb{C} \to \mathbb{R} \),

\[
\tilde{a}(\eta) \equiv \min_{x \in D} \Re \{ a(y_1(\eta), y_2, \ldots, y_N, x) \} = \min_{x \in D} a(\Re \{ y_1(\eta) \}, y_2, \ldots, y_N, x) = \min_{x \in D} [a(0, y_2, \ldots, y_N, x) + \Re \{ y_1(\eta) \} b_1(x)],
\]

with \( \Re \{ c \} \) being the real part of \( c \in \mathbb{C} \). Whenever \( \tilde{a}(\eta) \neq 0 \) the continued function \( \Psi \), the solution of (6.6), satisfies the bound

\[
\| \Psi(\eta) \|_{H^1_0(D)} = \| u(y_1(\eta), y_2, \ldots, y_N, \cdot) \|_{H^1_0(D)} \leq C_D \frac{\| f(y_1(\eta), y_2, \ldots, y_N, \cdot) \|_{L^2(D)}}{\tilde{a}(\eta)}
\]

with \( C_D \) being the Poincaré constant for the domain \( D \). Beside this, define

\[
\tilde{a}_1(y_2, \ldots, y_N) \equiv \min_{x \in D} a(0, y_2, \ldots, y_N, x)
\]

and observe that

\[
\tilde{a}(\eta) \geq \tilde{a}_1 - |y_1(\eta)| \| b_1 \|_{L^\infty(D)}.
\]

We are now ready to estimate the Fourier coefficients in (6.7).

Lemma 6.2. Let \( \tau \in (0,1) \). Then there exists positive constants \( C > 0 \) and \( \theta_f(\hat{y}_1, \tau) > 0 \) such that

\[
\| d_n \|_{H^1_0(D)} \leq C_D \frac{\theta_f(2n+1)}{\tau \tilde{a}_1 2^n} \int_{-1}^{1} \left( \frac{1-t^2}{t+1+C(1-\tau)} \right)^n dt
\]

Proof. Consider

\[
d_n = \frac{2n+1}{2} \int_{-1}^{1} \Psi(t) p_n(t) dt = \frac{(2n+1)(-1)^n}{n!} 2^{n+1} \int_{-1}^{1} \frac{d^n}{dt^n} \Psi(t)(1-t^2)^n dt.
\]

Use the analytic continuation of the real function \( \Psi \) to the complex domain as in Lemma 6.1. The application of Cauchy’s formula gives

\[
\frac{d^n}{dt^n} \Psi(t) = \frac{n!(-1)^n}{2\pi i} \int_{\gamma_t} \frac{\Psi(\eta)}{(\eta-t)^{n+1}} d\eta,
\]

where \( \gamma_t \) is a positively oriented closed circumference with center at the real point \( t \in (-1,1) \), radius \( R(t) \), and such that all singularities from \( \Psi \) are exterior to \( \gamma_t \). Denote the complex closed ball \( B(t, R(t)) = \{ z \in \mathbb{C} : |z-t| \leq R(t) \} \). Estimate (6.8) implies

\[
\| \frac{d^n}{dt^n} \Psi(t) \|_{H^1_0(D)} \leq C_D \frac{n!}{2\pi} \int_{\gamma_t} \frac{\| f(y_1(\eta), \hat{y}_1, \cdot) \|_{L^2(D)} d\eta}{\tilde{a}(\eta)|\eta-t|^{n+1}}
\]

\[
\leq C_D \frac{n!}{2\pi} \left( \sup_{\eta \in \gamma_t} \| f(y_1(\eta), \hat{y}_1, \cdot) \|_{L^2(D)} \right) \int_{\gamma_t} \frac{|d\eta|}{\tilde{a}(\eta)|\eta-t|^{n+1}}
\]

\[
\leq C_D \frac{n!}{(R(t))^n} \left( \sup_{\eta \in \gamma_t} \| f(y_1(\eta), \hat{y}_1, \cdot) \|_{L^2(D)} \right) \sup_{n \in \gamma_t} \frac{1}{\tilde{a}(\eta)}.
\]
Let
\[ \theta_f \equiv \sup_{t \in [-1, 1]} \sup_{\eta \in \gamma} \|f(y_1(\eta), \tilde{y}_1, \cdot)\|_{L^2(D)} \]
then estimate (6.10) implies
\[ \|d_n\|_{H^s(D)} \leq \frac{(2n + 1)C_D\theta_f}{2^{n+1}} \mathcal{J}_{-1}^{1} \left( \frac{1}{\inf_{\eta \in \gamma} \tilde{\alpha}(\eta)} \right) \left( \frac{1 - t^2}{R(t)} \right) n dt. \]

Now choose \( R(t) \) to control the above integral as follows. Let \( \tau \in (0, 1) \) and consider the interval \( I = \{ y \in \mathbb{R} : |y| \leq (1 - \tau)\|\gamma\|_{L^\infty(D)} \} \). By construction, we have the strict inclusion \( \Gamma_1 \subset I \) with \( dist(\Gamma_1, I) > \frac{(1-\tau)\rho}{c_1} > 0 \).

Then, the corresponding preimages satisfy the inclusion \((y_1)^{-1}(\Gamma_1) = [-1, 1] \subset (y_1)^{-1}(I) \) and \((y_1)^{-1}(I) = (-1 - \delta_1, 1 + \delta_2)\) with \( \delta_1, \delta_2 > 0 \). Choose the radius
\[ R(t) \equiv \min\{t + 1 + \delta_1, -t + 1 + \delta_2\} > 0 \]
which implies, by virtue of (6.9),
\[ \inf_{\eta \in \gamma} \tilde{\alpha}(\eta) \geq \tau \tilde{\alpha}_1. \]

Let
\[ \delta(\tau, y_2, \ldots, y_N) \equiv \min\{\delta_1, \delta_2\} \geq \frac{(1 - \tau)\tilde{\nu}}{C_1} \frac{2}{y_{\min} - y_{\max}} \]
and observe that (6.11, 6.14) imply
\[ \|d_n\|_{H^s(D)} \leq \frac{(2n + 1)C_D\theta_f}{\tau \tilde{\alpha}_1 2^{n+1}} \mathcal{J}_{-1}^{1} \left( \frac{1 - t^2}{t + 1 + \delta} \right) n dt. \]
which is what we wanted to prove.

Now we use a result from [26], namely that we have

**Lemma 6.3** (Integral estimate). Let \( \xi < -1 \) and define
\[ r \equiv \frac{1}{|\xi| + \sqrt{\xi^2 - 1}}, \quad 0 < r < 1. \]
Then there holds
\[ (-1)^n \mathcal{J}_{-1}^{1} \left( \frac{t^2 - 1}{t + \xi} \right) n dt = (2r)^n 2^{n+1} \frac{n!}{(2n+1)!!} \Phi_{n,0}(r^2) \]
where \( \Phi_{n,0}(r^2) \) is the Gauss hypergeometric function. Moreover, we have
\[ \Phi_{n,0}(r^2) = \sqrt{1 - r^2} + O\left( \frac{1}{n^{1/3}} \right) \]
uniformly with respect to \( 0 < r < 1 \).

**Lemma 6.4.** Let \( \tau \in (0, 1) \). Under Assumption 6.1 there exist positive constants \( C, \theta_f > 0 \) such that
\[ \|d_n\|_{H^s(D)} \leq \frac{2C_D\theta_f}{\tau \tilde{\alpha}_1} \sqrt{\frac{\pi n}{2}} \left( \sqrt{1 - r^2} + O\left( \frac{1}{n^{1/3}} \right) \right) r^n \]
with
\[ r \equiv \frac{1}{|\xi| + \sqrt{\xi^2 - 1}}, \quad 0 < r < 1 \]
and \( \xi < -1 - C(1 - \tau) < -1 \).
Proof. The result follows from the last two lemmata and the asymptotic equivalence
\[
\frac{(2n)!!}{(2n-1)!!} \sim \sqrt{\frac{n}{2}} \quad n \to \infty.
\]
\[\square\]

Finally, we can state the estimate for the size of the series in (6.7).

**Lemma 6.5.** Let \( \tau \in (0,1) \). Under assumption 6.4 there exist positive constants \( C, \theta_f > 0 \) such that
\[
(E_1)(y_1, \ldots, y_N) \leq \frac{C D \theta_f \sqrt{|\Gamma|}}{\tau a_1} \left( \sqrt{1 - \tau^2} + O \left( \frac{1}{(p_1)^{1/3}} \right) \right) \sqrt{\pi} \frac{r_{p_1+1}}{\sqrt{1 - \tau^2}}
\]

**Proof.** Use Lemma 6.4 to estimate the tail of the series
\[
(E_1)^2(y_2, \ldots, y_N) = \frac{|\Gamma|}{2} \sum_{n=p_1+1}^{+\infty} \frac{2}{2n+1} \|d_n\|^2_{H^{1\sigma}(D)}.
\]

\[\square\]

The main result of this section, namely the exponential convergence with respect to the multi-index \( p \) as in [20], follows from the above lemmata, i.e.

**Theorem 6.2.** Let \( \tau \in (0,1) \) and \( u \) be the solution of (2.11), \( u \in L^2(\Gamma) \otimes H^1_0(D) \), which is analytic with respect to \( y \), onto the subspace \( Z^p \otimes H^1_0(D) \). Then there exist positive constants, \( 0 < C, C_f \), such that
\[
E_p := \min_{v \in Z^p \otimes H^1_0(D)} \|u - v\|_{L^2(\Gamma) \otimes H^1_0(D)}
\]
\[
\leq \frac{C_D C_f \sqrt{|\Gamma|}}{\tau} \left( \sum_{i=1}^{N} |\Gamma_i| \int_{\Gamma_i} \left( 1 + \frac{1}{\sqrt{1 - r_i^2}} O \left( \frac{1}{(p_i)^{1/3}} \right) \right) \left( \frac{r_i}{a_i} \right)^{p_i+1} d\gamma_i \right)^2 \tau_i(\hat{y}_i) \equiv \frac{1}{|\xi_i| + \sqrt{\xi_i^2 - 1}} \quad 0 < r_i < 1, i = 1, \ldots, N
\]

with
\[
\xi_i(\hat{y}_i) < -1 - C(1 - \tau) < -1.
\]

Similarly as the \( k \times h \)-version, cf. (5.10), the \( p \)-version has a superconvergence result for the approximation of expected value of the solution.

**Theorem 6.3** (Superconvergence of the \( p \)-version with piecewise linear FEM in space.). There holds
\[
\|E[u - u_h^p]\|_{L^2(D)} \leq C \left( h^2 + \frac{1}{\tau} \sum_{i=1}^{N} (\|r_i\|_{L^\infty}(\Gamma_i))^{2p_i+2} \right)
\]
with \( 0 < r_i(\hat{y}_i) < 1 \) as in Theorem 6.2 and \( C > 0 \) is independent from \( u, h, p_i \) and \( r_i \).

The proof of the previous Theorem uses Theorem 6.2 and is completely similar to the proof of Theorem 5.2.

7. Double orthogonal polynomials

Here we explain the use of double orthogonal polynomials to compute efficiently the solution of the \( k \times h \)-version and the \( p \times h \)-version studied in Sections 5 and 6 respectively. The idea is to use a special basis to decouple the system in the \( y \)-direction, yielding just a number of uncoupled systems, each one with the size and structure of one Monte Carlo realization of (4.1).

Without any loose of generality we focus on the \( p \)-version, i.e. find \( u_h^p \in Z^p \otimes X_h^p \) such that
\[
\int_{\Gamma} \rho(y) \int_D a(y, x) \nabla u_h^p(y, x) \cdot \nabla v(y, x) dxdy = \int_{\Gamma} \rho(y) \int_D f(y, x)v(y, x) dxdy \quad \forall v \in Z^p \otimes X_h^d.
\]
Let \( \{ \psi_j(y) \} \) be a basis of the subspace \( Z^p \subset L^2(\Gamma) \) and \( \{ \varphi_i(x) \} \) be a basis of the subspace \( X^d_k \subset H^1_0(D) \). Write the approximate solution as
\[
(7.2) \quad u^n_{ij}(y, x) = \sum_{j,i} u_{ij} \psi_j(y) \varphi_i(x)
\]
and use test functions \( \nu(y, x) = \psi_k(y) \varphi_\ell(x) \) to find the coefficients \( u_{ij} \). Then (7.1) gives
\[
\sum_{j,i} \left( \int_\Gamma \rho(y) \psi_k(y) \psi_j(y) \int_D a(y, x) \nabla \varphi_i(x) \cdot \nabla \varphi_\ell(x) dxdy \right) u_{ij} = \int_\Gamma \rho(y) \psi_k(y) \int_D f(y, x) \varphi_\ell(x) dxdy, \quad \forall k, \ell
\]
which can be rewritten as \( \sum_{j,i} \left( \int_\Gamma \rho(y) \psi_k(y) \psi_j(y) K_{i,\ell}(y) dy \right) u_{ij} = \int_\Gamma \rho(y) \psi_k(y) f_\ell(y) dy \quad \forall k, \ell \) with \( K_{i,\ell}(y) = \int_D a(y, x) \nabla \varphi_i(x) \cdot \nabla \varphi_\ell(x) dx \) and \( f_\ell(y) \equiv \int_D f(y, x) \varphi_\ell(x) dx \). Now, if we have that the diffusion coefficient is a truncated Karhunen-Loève expansion, \( a(y, x) = E[a](x) + \sum_{n=1}^N b_n(x) y_n \), and by the independence of the \( Y \) components its joint probability density decouples into the product \( \rho(y) = \prod_{m=1}^N \rho_m(y_m) \) then we have a corresponding “Karhunen-Loève ” like expression for the stiffness matrix
\[
K_{i,\ell}(y) = \int_D (E[a](x) + \sum_{n=1}^N b_n(x) y_n) \nabla \varphi_i(x) \cdot \nabla \varphi_\ell(x) dx = K^0_{i,\ell} + \sum_{n=1}^N y_n K^n_{i,\ell}
\]
with deterministic coefficients
\[
K^0_{i,\ell} = \int_D E[a](x) \nabla \varphi_i(x) \cdot \nabla \varphi_\ell(x) dx
\]
and
\[
K^n_{i,\ell} = \int_D b_n(x) \nabla \varphi_i(x) \cdot \nabla \varphi_\ell(x) dx.
\]
By the same token we have,
\[
\int_\Gamma \rho(y) \psi_k(y) \psi_j(y) K_{i,\ell}(y) dy = K^0_{i,\ell} \int_\Gamma \rho(y) \psi_k(y) \psi_j(y) dy + \sum_{n=1}^N K^n_{i,\ell} \int_\Gamma y_n \rho(y) \psi_k(y) \psi_j(y) dy.
\]
Since \( \psi_k \in Z^p \), with multiindex \( p = (p_1, \ldots, p_N) \), it is enough to take it as the product
\[
\psi_k(y) = \prod_{r=1}^N \psi_{kr}(y_r)
\]
where \( \psi_{kr} : \Gamma_r \to \mathbb{R} \) is a basis function of the subspace
\[
Z^{pr} = \text{span}[\psi_{h} : h = 1, \ldots, p_r + 1].
\]
Keeping this choice of \( \psi_k \) in mind,
\[
\int_\Gamma \rho(y) \psi_k(y) \psi_j(y) K_{i,\ell}(y) dy = K^0_{i,\ell} \int_\Gamma \prod_{m=1}^N \rho_m(y_m) \psi_{km}(y_m) \psi_{jm}(y_m) dy + \sum_{n=1}^N K^n_{i,\ell} \int_\Gamma y_n \prod_{m=1}^N \rho_m(y_m) \psi_{km}(y_m) \psi_{jm}(y_m) dy
\]
Now, for every set $\Gamma_n, n = 1, \ldots, N$ choose the polynomials, $\psi_j(y) = \prod_{n=1}^{N} \psi_{j_n}(y_n)$, to be biorthogonal, i.e. for $n = 1, \ldots, N$ they must satisfy

$$\int_{\Gamma_n} \rho_n(z) \psi_{k_n}(z) \psi_{j_n}(z) dz = \delta_{kj}$$

$$\int_{\Gamma_n} z \rho_n(z) \psi_{k_n}(z) \psi_{j_n}(z) dz = c_{kn}\delta_{kj}.$$  

To find the polynomials $\psi$ we have to solve $N$ eigenproblems, each of them with size $(1 + p_n)$. The computational work required by these eigenproblems is negligible with respect to the one required to solve for $u_{ij}$, cf. [25], section 8.7.2. The orthogonality properties (7.3) for $\psi$ imply the decoupling

$$\int_{\Gamma} \rho_k \psi_j dy = \delta_{kj}, \quad \int_{\Gamma} y_n \rho_k \psi_j dy = c_{kn}\delta_{kj}.$$  

By means of this decoupling we now conclude that

$$\sum_{j,i} \left( \int_{\Gamma} \rho(y) \psi_k(y) \psi_j(y) K_{i,\ell}(y) dy \right) = K_{i,\ell}^0 \int_{\Gamma} \rho(y) \psi_k(y) \psi_j(y) dy$$

$$+ \sum_{n=1}^{N} K_{i,\ell}^n \int_{\Gamma} y_n \rho(y) \psi_k(y) \psi_j(y) dy$$

$$= \left( K_{i,\ell}^0 + \sum_{n=1}^{N} c_{kn} K_{i,\ell}^n \right) \delta_{kj}.$$  

The structure of the linear system that determines $u_{ij}$ now becomes block diagonal, which each block being coercive and with the sparsity structure identical to one deterministic FEM stiffness matrix, i.e.

$$\begin{bmatrix}
K^0 + \sum_{n=1}^{N} c_{1n} K^n & 0 & \ldots & 0 \\
0 & K^0 + \sum_{n=1}^{N} c_{2n} K^n & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \ldots & 0 & K^0 + \sum_{n=1}^{N} c_{Nn} K^n
\end{bmatrix}$$

The conclusion is that the work to find the coefficients $u_{ij}$ in (7.2) is the same as the one needed to compute $\prod_{i=1}^{N} (1 + p_i)$ Monte Carlo realizations of $u_k$ defined in (4.1). Beside this, observe that as a consequence of the uniform coercivity assumption, each of the diagonal blocks in the system above is symmetric and strictly positive definite.

8. ASYMPTOTICAL EFFICIENCY COMPARISONS

In this section we compare the asymptotical numerical complexity for the Monte Carlo Galerkin finite element method, cf. Section 4, with the Stochastic Galerkin finite element method introduced in Sections 5 and 6. In all the cases, the spatial discretization is done by piecewise linear finite elements on globally quasuniform meshes. For the $k \times h$–SGFEM the $\Gamma$ partitions are also assumed to be globally quasuniform. Beside this, the diffusion function $a$ is assumed to be a truncated Karhunen-Loève expansion.

8.1. MCGFEM versus $k \times h$–SGFEM. Here we consider the computational work to achieve a given accuracy bounded by a positive constant TOL, for both the MCGFEM and the $k \times h$–SGFEM methods. This optimal computational work indicates under which circumstances one method may be best suited. When using the MCGFEM method to approximate the solution of (1.1) in energy
norm the error becomes, applying Proposition 4.2 together with Proposition 4.5 that given a confidence level, 0 < c_0 < 1, there exists a constant C > 0 depending only on c_0 such that

\[
P \left( \| E[u] - \frac{1}{M} \sum_{j=1}^{M} u_h(\cdot; \omega_j) \|_{H^1_0(D)} \leq C(h + \frac{1}{\sqrt{M}}) \right) \geq c_0
\]

Then, in the sense of (8.1) we write \( E_{\text{MCGFEM}} = O(h) + O(1/\sqrt{M}) \). The corresponding computational work for the MCGFEM method is \( \text{Work}_{\text{MCGFEM}} = O((1/h^d)^r + 1/h^d)M \), where the parameter \( 1 \leq r \leq 3 \) relates to the computational effort devoted to solve one linear system with \( n \) unknowns, \( O(n^r) \). From now on we continue the discussion with the optimal \( r = 1 \) that can be achieved by means of the Multigrid method, cf. [12, 13, 28]. Thus, choosing \( h \) and \( M \) to minimize the computational work for a given desired level of accuracy \( \text{TOL} \) > 0 yields the optimal work

\[
\text{Work}_{\text{MCGFEM}}^* = O \left( \text{TOL}^{-(2+d)} \right).
\]

Similarly, if we apply a \( k \times h \)–SGFEM with piecewise polynomials of order \( q \) in the \( y \)--direction the computational error in \( H^1_0(D) \) norm is, cf. Proposition 5.1

\[
E_{\text{SGFEM}} = O(h) + NO(k^{q+1})
\]

and the corresponding computational work for the \( k \times h \)--version is

\[
\text{Work}_{\text{SGFEM}}^*\text{SGFEM} = O \left( h^{-d}(1 + q)^N k^{-N} \right).
\]

Here \( N \) is the number of terms in the truncated Karhunen-Loève expansion of the coefficients \( a \) and \( f \) and \( k \) is the discretization parameter in the \( y \) direction. Similarly as before, we can compute the optimal work for the \( k \times h \)--SGFEM method, yielding

\[
\text{Work}_{\text{SGFEM}}^*\text{SGFEM} = O((1 + q)^N \left( \frac{\text{TOL}}{N} \right)^{-\frac{N}{q+d}} \text{TOL}^{-d}).
\]

Therefore, a \( k \times h \)--version SGFEM is likely to be preferred whenever \( \text{TOL} \) is sufficiently small and \( N/2 < 1 + q \), i.e. if the number of terms in the Karhunen-Loève expansion of \( a \) is large then the degree of approximation in the \( y \) direction, \( q \), has to become correspondingly large. We summarize the comparison results in Table 1, where we also include corresponding results from the \( p \times h \)--version, discussed in Subsection 8.2.

<table>
<thead>
<tr>
<th>Work</th>
<th>MCGFEM</th>
<th>( k \times h )--version SGFEM</th>
<th>( p \times h )--version SGFEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0^1(D) ) Error</td>
<td>( M/h^d )</td>
<td>( h + \frac{1}{\sqrt{M}} )</td>
<td>( h + \frac{1}{\sqrt{M}} )</td>
</tr>
<tr>
<td>( H_0^1(D) ) Work*</td>
<td>( (1+q)^{-\frac{N}{q+d}} \text{TOL}^{-d} )</td>
<td>( \frac{(1+p)^{-\frac{N}{p+d}}}{h^d} \text{TOL}^{-d} )</td>
<td>( (\log_r(\text{TOL}))^N \text{TOL}^{-d} )</td>
</tr>
</tbody>
</table>

Table 1. Approximation of the function \( E[u] \) in \( H_0^1(D) \). Asymptotical numerical complexity for the MCGFEM and SGFEM methods.

Similarly, if we are interested to control the difference \( \| E[u] - E[u_h] \|_{L^2(D)} \) the application of Proposition 4.2 and Properties 4.1 and 4.2 for the MCGFEM method and Theorem 5.2 on the superconvergence of the \( k \times h \)--SGFEM method imply the results shown in Table 2. In this case \( k \times h \)--SGFEM is likely to be preferred whenever \( N/4 < (q + 1) \) and \( \text{TOL} \) is sufficiently small. In addition, the comparison tells us that to be able to be competitive with the Monte Carlo method when the number of relevant terms in the Karhunen-Loève expansion is not so small, an optimal method should have a high order of approximation and should avoid as much as possible the coupling between the different
components of the numerical solution to preserve computational efficiency. The approach proposed by Ghanem and Spanos [24] based on orthogonal polynomials has, whenever the approximate diffusion satisfies \(1.2\), a high order of approximation but introduces coupling between the different components of the numerical solution. The uncoupling can be achieved for linear equations using double orthogonal polynomials, see the description in Section 7. With this motivation, Section 6 studies the convergence of the \(p\times h\)–SGFEM.

### 8.2. MCGFEM versus \(p\times h\)–SGFEM

Here we consider the computational work to achieve a given accuracy, for both the \(p\times h\)–version of SGFEM defined in \((8.1)\) and the MCGFEM method for the approximation of \(E[u]\) defined in Section \(6\). i.e. we are interested to control the difference \(|E[u] - E[u^h]|_{L^2(D)}\) or \(|E[u] - \frac{1}{N} \sum_{j=1}^{M} u_h(\cdot; \omega_j)|_{L^2(D)}\), respectively. This computational work indicates under which circumstances one method may be best suited than the other. Besides this, let us assume that we use in our computations a piecewise linear FEM space in \(D\). When using the MCGFEM method to approximate the expected value of the solution of \((1.1)\), we have the optimal work required to achieve a given desired level of accuracy \(\text{TOL} > 0\), cf. \((8.2)\),

\[\text{Work}_{\text{MCGFEM}}^k = O(1/\text{TOL}^{2+d/2}).\]

On the other hand, if we apply a \(p\times h\)–version of the SGFEM, with \(p_i = p, i = 1, \ldots, N\), the computational error is, cf. Theorem \(6.3\),

\[E_{\text{SGFEM}} = O(h^2) + O(r^{2(p+1)}), \quad 0 < r < 1,\]

and the corresponding computational work is, cf. Section \(7\),

\[\text{Work}_{\text{SGFEM}}^k = O \left( \frac{(1 + p)^N}{h^d} \right).\]

Recall that \(N\) is the number of terms in the truncated Karhunen-Loève expansion of the coefficients \(a\) and \(f\) and \(k\) is the discretization parameter in the \(y\) direction. As before, we can compute the optimal work for the SGFEM method, yielding

\[\text{Work}_{\text{SGFEM}}^k \leq O \left( \frac{1}{\log_r(\text{TOL})} \right) \text{N}^{\text{TOL}^{-d/2}} \]

and the asymptotical comparison

\[\lim_{\text{TOL} \to 0} \frac{\text{Work}_{\text{SGFEM}}^k}{\text{Work}_{\text{MCGFEM}}^k} = \lim_{\text{TOL} \to 0} \frac{(1 + p)^N}{h^d} \frac{\text{TOL}^2}{\text{TOL}^{-d/2}} = 0.\]

Therefore, for sufficiently strict accuracy requirements, i.e. sufficiently small TOL, in the computation of \(E[u]\), SGFEM requires less computational effort than MCGFEM. The work of Bahvalov and its subsequent extensions, cf. [9, 27, 57, 40] generalizes the standard Monte Carlo method, taking advantage of the available integrand’s smoothness and yielding a faster order of convergence. The optimal work of such a method is for our case, i.e. the approximation of \(E[u]\) in \(L^2(D)\), \(O(C(N)\text{TOL}^{-d/2} + r^{d(p+1)})\), where it is assumed that the integrand \(u\) has bounded derivatives up to order \(q\) with respect to \(y\) and the integral is done in the \(N\)-dimensional unit cube.
The result on the computational work of the $p\times h$-version of the SGFEM presented in this work is then related to the case $q = \infty$, since $u$ is analytic with respect to $y$. This analyticity allows the exponential convergence with respect to $p$, cf. Theorem 6.3.

Notice that we only discussed the optimal asymptotical computational work required by both methods, but in practice the constants involved in the asymptotic approximations makes these comparisons just indicative, and not conclusive. In addition, we have only studied the case where the integrals \( \int_{\Gamma} \rho y^k dy \) can be computed exactly for $k = 0, 1, \ldots$, and not considered the more general case where quadrature rules are needed to approximate such integrals.

**Remark 8.1** (Use of higher order FEM approximations on $D$). Based on Remark 5.1 and following the approach from Sections 8.1 and 8.2 we can discuss the possible use of higher order FEM approximations on $D$. This use seems a priori always useful for the $p\times h$-version of the SGFEM while for the $k\times h$-version higher order FEM on $D$ are attractive provided $d \gg N/(q+1)$. On the other hand a MCGFEM piecewise linear approximation on $D$ with error $TOL$ requires the same work as the optimal higher order method with error slightly larger than $TOL^{1+d/4}$. See [52], p. 1884, for a similar discussion on the weak approximation of ordinary SDEs.

**Remark 8.2** (Combination of MCGFEM with SGFEM). It is possible to combine in a natural way the Monte Carlo method with an SGFEM version, partitioning $\Gamma$ in order to take advantage of their different convergence rates. Split the domain $\Gamma$ into $\Gamma = \Gamma_{MC} \cup \Gamma_G$, $\Gamma_{MC} \cap \Gamma_G = \emptyset$ and approximate

\[
E[u(Y, \cdot)] = E[u(Y, \cdot) | Y \in \Gamma_{MC}] P(Y \in \Gamma_{MC}) + E[u(Y, \cdot) | Y \in \Gamma_G] P(Y \in \Gamma_G).
\]

using SGFEM in $\Gamma_{G}$ and MCGFEM in $\Gamma_{MC}$.

**Remark 8.3** (Successive approximation method). The work [5] proposes a successive approximation for the solution of (1.1). As any Neumann expansion with $K$ terms, whenever it converges it has an error of the form $\text{Error}_{BC} \leq C(h + \xi^{K+1})$ for some $0 < \xi < 1$. The computational work to achieve this error is $W_{BC} \leq C \frac{N^{K+1}}{(N-1)^3}$. Then, a comparison between this method and the MCGFEM yields that if $\xi < 1/\sqrt{N}$ then the method from [5] is likely to be preferred. On the other hand, for sufficiently small tolerances the $p\times h$-version of the SGFEM is more efficient than the successive approximation.

Aknowledgements The first author was partially supported by the National Science Fundation (grant DMS-9802367) and the Office of Naval Research (grant N00014-99-1-0724). The second author was partially supported by the Swedish Council for Engineering Science grant # 222-148, the National Science Fundation grant DMS-9802367, and UdelaR and UdeM in Uruguay. The third author was partially supported by the European Union TMR grant No ERBFMRX-CT98-0234 (Viscosity Solutions and their Applications).

The authors would like to thank Prof. Anders Szepessy for many valuable discussions regarding to this work.

**References**


