Adaptivity for Stochastic Magnetization Dynamics

Dissertation
der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Christian Schellnegger
aus Schorndorf

Tübingen
2018

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Prüfung:	20.06.2018
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Andreas Prohl
2. Berichterstatter:	Prof. Dr. Ľubomír Baňas

Abstract

This thesis is concerned with the adaptive approximation of the stochastic Landau-Lifshitz-Gilbert equation (SLLG), which models the dynamics of a ferromagnetic body at elevated temperatures. The SLLG is a non-linear stochastic partial differential equation which possesses an inherent non-convex side constraint. In a first step, space-time and statistical adaptivity is addressed to a convection-dominated SPDE with linear drift, and the stochastic harmonic map heat flow (HMHF) to the sphere $\mathbb{S}^2$. Secondly, these adaptive concepts are applied to the SLLG. The latter two equations possess a weak martingale solution, rather than a probabilistically strong solution as for the first problem; however our concept of space-time and statistical adaptivity is based on distributions rather than single trajectories, and therefore is applicable also there.

The thesis is splitted in three parts. The first part is concerned with ordinary differential equations (ODEs) in which we focus on local extrapolation methods to adapt the local time step size. In the second part, we repeat in the literature already existing adaptive time stepping methods based on the explicit Monte Carlo Euler method for weak approximation of the SDE. These strategies are limited to lower order systems of SDEs due to the use of Kolmogorov's backward equation. Inspired by this, we perform a local non-parametric high-dimensional density estimation based on Monte-Carlo (MC) sampling.

In the third part we propose a new adaptive time stepping method to numerically solve the SLLG where local step sizes are chosen in regard of the distance between empirical laws of current Euler iterates, and extrapolated data. This histogram-based estimator uses a data-driven partitioning of the high-dimensional state space, and efficient sampling by bootstrapping to save computer resources. Time adaptivity is then complemented by a local refinement/coarsening strategy of the spatial mesh via a stochastic version of the Zienkiewicz-Zhu estimator. Computational experiments compare the efficiency of the proposed adaptive space-time and statistical strategies of already in the literature existing stable discretizations of the SLLG. Of particular interest is the choice of the distance to measure closeness of subsequent laws (in time and space), having an possible impact on the empirical variance of standard estimators; especially in the case of discrete blow-up dynamics of the SLLG.

Zusammenfassung

Ziel dieser Arbeit ist die adaptive Approximation der stochastischen Landau-Lifshitz-Gilbert-Gleichung (SLLG). Die SLLG ist eine nicht-lineare stochastische partielle Differentialgleichung mit einer nicht-konvexen Zwangsbedingung und wird bei der Modellierung ferromagnetischer Dynamiken eingesetzt. Im ersten Schritt werden die adaptiven Strategien für dynamische Konvektions-Diffusionsprobleme im Falle von dominanter Konvektion adressiert sowie auf eine der SLLG zugehörige Problemstellungen angewandt.

Die Dissertation ist in drei wesentliche Teile untergliedert. Im ersten Teil werden adaptive Strategien für Systeme von gewöhnlichen Differentialgleichungen wiederholt. Der Schwerpunkt liegt hierbei auf der Schätzung des lokalen Konsistenzfehlers durch den Vergleich zweier numerischer Lösungen, erzielt durch unterschiedliche Zeitschrittweiten und Extrapolation. Im zweiten Teil wiederholen wir in der Literatur bereits existente adaptive Methoden zur Lösung von stochastischen Differentialgleichungen (SDEs), deren Konstruktion die zugehörige Kolmogorov-Gleichung verwendet, und daher praktisch die Dimension von SDEs auf bis zu drei limitiert.

Die im dritten Kapitel neu vorgeschlagenen, adaptiven Konzepte zur Simulationen der SLLG basieren auf dem Abstand zweier empirischer Verteilungen von Euler-Iterierten, erzielt durch unterschiedliche Zeitschrittweiten und Extrapolation. Hierbei ermöglicht eine daten-abhängige Partitionierung des hoch-dimensionalen Zustandsraumes eine adäquate Histogramm-basierte Abstandsschätzung beider Verteilungen, deren Monte-Carlo Komplexität durch den Einsatz von Bootstrapping drastisch reduziert wird. Ergänzt wird die Zeitadaptivität durch eine lokale Verfeinerungs- bzw. Vergröberungsstrategie des räumlichen Gitters mittels einer stochastischen Version des Zienkiewicz-Zhu Schätzers. Numerische Studien vergleichen die Leistungsfähigkeit der vorgeschlagenen adaptiven Raum-Zeit-und statistischen Strategien hinsichtlich verschiedener in der Literatur existenter Diskretisierungsansätze der SLLG. Von besonderem Interesse ist dabei die Wahl der Abstandsfunktion für Wahrscheinlichkeitsmaße (in Zeit und Ort) und der damit einhergehenden möglichen Varianzreduktion von Standardschätzern, insbesondere im Fall von diskreten Singularitäten der SLLG.

Acknowledgments

I thank everyone who supported me during the preparation of this thesis.

There are some people deserving a special note of thanks.

I warmly thank my advisor, Prof. Dr. Andreas Prohl, for the opportunity to write my thesis in the interesting and important field of adaptivity for stochastic partial differential equations. I thank him for his scholastic guidance, his constant inspiration, and his valuable suggestions and patience.
I also thank him, Prof. Dr. Ľubomír Baňas, for being the second evaluator of this thesis.

I am grateful to my colleagues Dr. Thomas Dunst and Dr. Ananta Kumar Majee, who were involved in many helpful mathematical discussions, who gave me programming advice and a constant moral support over the last years.

Much gratitude to Dr. Ananta Kumar Majee for reading parts of this thesis and giving me feedback.

Finally, I would like to thank my parents Hilde and Siegfried, my sister Victoria, and my girlfriend Andrea, who have always been supporting and encouraging me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Magnetism is of fundamental importance for modern civilization: For example, compasses greatly simplified navigation on the sea, loudspeakers are part of every modern audio equipment and of every mobile phone, magnetic recording is the key technology for mass storage, and magnetic resonance imaging is used in medical diagnosis. Magnetization processes are described by ferromagnetic models: their study leads to a deeper understanding of this process and provides information which is experimentally inaccessible or which is linked to unnecessary costs.

One well-accepted physical model describing the magnetization of a ferromagnetic body has been proposed in the pioneering works of Landau and Lifshitz [LL35] and of Gilbert [Gil04]. The time evolution in $[0,T]$ of the magnetization of a ferromagnetic body $D \subset \mathbb{R}^d$, $d \in \{1,2,3\}$ is modeled by the function

$$\boldsymbol{X} : [0,T] \times D \to \{\mathbf{x} \in \mathbb{R}^3;\, |\mathbf{x}| = M_s\}, \tag{1.0.1}$$

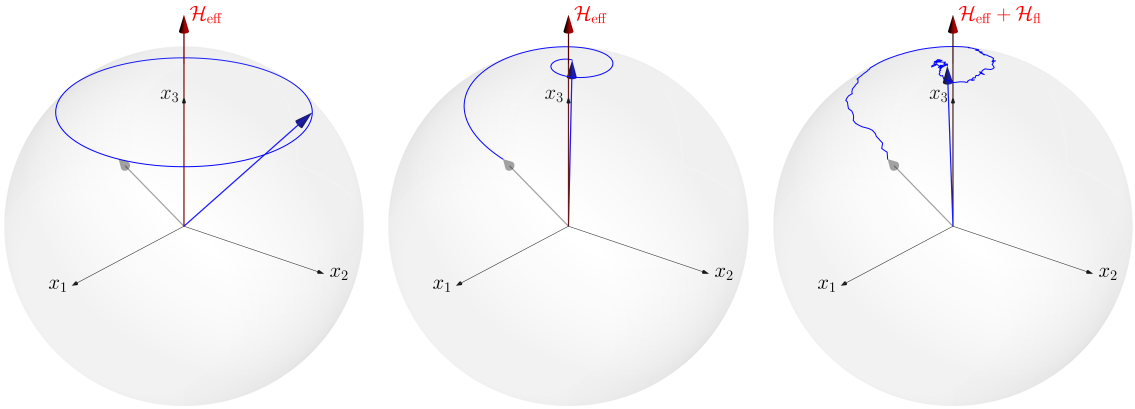
where $M_s > 0$ denotes the saturation magnetization. The vector $\boldsymbol{X}(t,\mathbf{x})$ is the direction of the magnetization at time $t \in [0,T]$ for $T > 0$ and at position $\mathbf{x} \in D$. Since we consider the equation without any physical dimensions, it is natural to set $M_s \equiv 1$. The Landau-Lifshitz-Gilbert equation (LLG)

$$\frac{\partial \boldsymbol{X}_t}{\partial t} = \zeta_1 \boldsymbol{X} \times \boldsymbol{\mathcal{H}}_{\text{eff}} - \zeta_2 \boldsymbol{X} \times \big(\boldsymbol{X} \times \boldsymbol{\mathcal{H}}_{\text{eff}}\big), \tag{1.0.2}$$

where $\zeta_1 \geqslant 0$, and $0 < \zeta_2 \ll 1$, describes the dynamics of the magnetization in the presence of an effective field $\boldsymbol{\mathcal{H}}_{\text{eff}} = \boldsymbol{\mathcal{H}}_{\text{eff}}(\boldsymbol{X}) = -\nabla \mathcal{E}(\boldsymbol{X})$. The first term on the right-hand side of the LLG describes the rotation of the magnetization $\boldsymbol{X}$ around the effective field $\boldsymbol{\mathcal{H}}_{\text{eff}}$. The second term is a phenomenological term which was introduced by Landau and Lifshitz [LL35] and accounts for the damping of the magnetization $\boldsymbol{X}$ towards the effective field $\boldsymbol{\mathcal{H}}_{\text{eff}}$. The dynamics of these terms are visualized in Figure 1.1(a)–(b). The effective field $\boldsymbol{\mathcal{H}}_{\text{eff}}$ consists of several contributions:

$$\boldsymbol{\mathcal{H}}_{\text{eff}} = \boldsymbol{\mathcal{H}}_{\text{exch}} + \boldsymbol{\mathcal{H}}_{\text{ext}} + \boldsymbol{\mathcal{H}}_{\text{anis}} + \boldsymbol{\mathcal{H}}_{\text{magn}}.$$

The first contribution is the exchange energy, which penalizes spatial changes in the magnetization and is usually modeled by $\boldsymbol{\mathcal{H}}_{\text{eff}} = \Delta \boldsymbol{X}$. The second part $\boldsymbol{\mathcal{H}}_{\text{ext}}$ is an external magnetic field (the so-called Zeeman contribution) and favors the alignment with an external field. It is



(a) precession **(b)** precession/damping **(c)** precession/damping

Figure 1.1: Dynamics of the precession and damping term in the LLG, respectively SLLG.

modeled according to $\boldsymbol{\mathcal{H}}_{\text{eff}} = \boldsymbol{u}$, for some $\boldsymbol{u} : [0,T] \times D \to \mathbb{R}^3$. Crystallographic properties of the ferromagnetic material are taken into account by the anisotropy energy $\boldsymbol{\mathcal{H}}_{\text{anis}}$, where the magnetization $\boldsymbol{X}$ prefers to align with the given (crystallographic) easy axis $\boldsymbol{e} \in \mathbb{R}^3$; its contribution is modeled by $\boldsymbol{\mathcal{H}}_{\text{anis}} = -\nabla\phi(\boldsymbol{X})$, where $\phi : \mathbb{S}^2 \to \mathbb{R}$ denotes the anisotropy density. Finally, the fourth contribution is the demagnetization field (also called stray-field), which takes the interaction with a surrounding magnetic field into account and is usually modeled using Maxwell's equation.

The LLG is non-linear and its solutions must satisfy the non-convex side constraint (1.0.1). Existence of a solution has been analyzed in the literature: for $d \in \{2,3\}$, the existence of weak solutions is shown for the prototype case $\boldsymbol{\mathcal{H}}_{\text{eff}} = \Delta\boldsymbol{X}$, see e.g. [AS92, GH93]. Numerical studies carried out in [BP06, BBP08] show that in this setup possible finite time blow-up from smooth initial data may be expected. For $d = 1$ however, existence of a weak solution with improved regularity properties is well known. From a physical viewpoint, the mathematical concept of blow-up translates to defects and spatial energy concentration in ferromagnetic materials. The issue of non-uniqueness is important for micromagnetic simulation. Understanding the source and characterizing nature of singularities affects issues involving the validity of the physical model as well as what allowable features should be present in the definition of weak solution.

Let $D \subset \mathbb{R}^d$ be a bounded smooth domain. According to [BKP08], aspects of finite time finite energy blow-up dynamics are studied with respect to the behavior of the solution $\boldsymbol{X} : D_T \to \mathbb{S}^2$ of the LLG for $\boldsymbol{\mathcal{H}}_{\text{eff}} = \Delta\boldsymbol{X}$,

$$\begin{aligned} \frac{\partial \boldsymbol{X}_t}{\partial t} &= \zeta_1 \boldsymbol{X} \times \Delta\boldsymbol{X} - \zeta_2 \boldsymbol{X} \times \big(\boldsymbol{X} \times \Delta\boldsymbol{X}\big) && \text{on } D_T := (0,T) \times D, \\ \boldsymbol{X}_0 &= \boldsymbol{x}_0, && \text{on } D, \end{aligned} \tag{1.0.3}$$

which is supplemented with Dirichlet or homogeneous Neumann boundary conditions. In this case, problem (1.0.3) linearly combines the Schrödinger flow ($\zeta_2 = 0$), and the heat flow of harmonic maps to the sphere $\mathbb{S}^2$ ($\zeta_1 = 0$). For the latter case, possible blow-up behavior of corresponding gradients caused by the geometry of the target manifold is known ($d \geqslant 2$).

Let $B_1(\mathbf{0}) \subset \mathbb{R}^2$ be the unit ball, and set $\zeta_1 = 0$ in (1.0.3). Consider radially symmetric functions $\boldsymbol{u} : [0,T] \times B_1(\mathbf{0}) \to \mathbb{S}^2$ given in terms of polar coordinates by

$$\boldsymbol{u}(t,r,\varphi) = \begin{pmatrix} \cos(\varphi)\sin\big(\theta(t,r)\big) \\ \sin(\varphi)\sin\big(\theta(t,r)\big) \\ \cos\big(\theta(t,r)\big) \end{pmatrix},$$

where functions of this type solve problem (1.0.3), provided $\theta : (0,T) \times (0,1) \to \mathbb{R}$ satisfies

$$\theta_t = \theta_{rr} + \frac{1}{r}\theta_r - \frac{\sin(2\theta)}{2r^2} \quad \text{on } (0,T) \times (0,1).$$

By adding boundary conditions $\theta(t,0) = 0$ and $\theta(t,1) = \theta^\star > \pi$, the solution $\boldsymbol{u} : D_T \to \mathbb{S}^2$ blows up at a finite time $t^* > 0$, i.e. $\lim_{t \uparrow t^*} \theta_r(t,0) = \infty$, see [Str96], and the solution must look locally (near the blow-up time $t^\star$) like a rescaled harmonic map at the so-called quasi-stationary scale, cf. [vdBW13]. That is, there is a scale on which the solution takes the form $f_c(r) := 2\arctan(cr)$, for $c \in \mathbb{R}$. However, there is an associated Lyapunov functional,

$$\mathcal{E}\big(\theta(t)\big) = \pi \int_0^1 \left[\theta_r^2(t,r) + \frac{\sin^2\big(\theta(t,r)\big)}{r^2}\right] r\mathrm{d}r,$$

whose only stationary points are the family f_c. It is this paradox that leads to blow-up: there is a finite collection of (possibly finite) times at which the solution $\theta(t,0)$ 'jumps' only by $\pm\pi$ by losing 4π of energy; see [Str96, BvdHH11]. This is visualized in Figure 1.2.

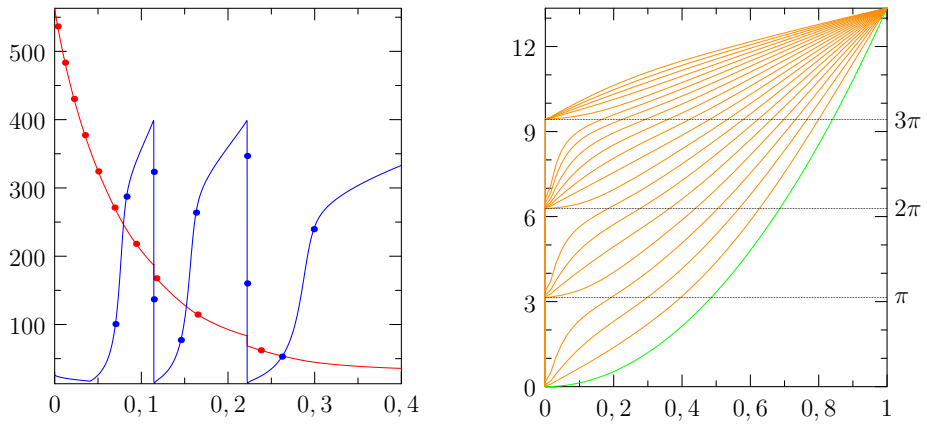


Figure 1.2: Discrete blow-up dynamic: (Left) Behavior of $t \mapsto |\theta_r(t,\cdot)|_\infty$(—●—), as well as $t \mapsto \mathcal{E}\big(\theta(t)\big)$(—●—). (Right) Snapshots of $t \mapsto \theta(t,\cdot)$(——), with initial data $\theta(0,\cdot)$(——).

Finite time blow-up behavior of solutions for the harmonic map heat flow is so far only known for the special case of radial symmetry. The results in [vdBW13] suggest that these blow-up solutions are not generic, i.e. they vanish for small (non-symmetric) perturbations of (initial) data. Solutions of the LLG are not invariant any more under radial symmetry, which is due to the additional precessional motion; however, there is a computational evidence concerning the LLG for a corresponding discrete blow-up behavior [BP06, BKP08]. According to [vdBW13], those singularities are again not generic. In accordance with the results in Figure 1.2, as long as $\|\nabla \boldsymbol{X}\|_{\mathbf{L}^\infty}$ is bounded, the solution remains regular for all time so discrete blow-up dynamics

in this case are indicated by loss of control on $\|\nabla \boldsymbol{X}\|_{\mathbf{L}^\infty}$. In this case, regular, as well as even non-regular, but fixed temporal or spatial meshes are often unable to resolve discrete blow-up features due to the transient character of such behavior. Thus, accurately resolving features with small time and length scales requires to use some adaptive space-time meshes allowing enhanced resolution in certain regions of the time-space domain where non-adaptive methods breakdown.

Currently there are only a few works that deal with the use of spatial adaptation for micromagnetic problems. For micromagnetic problems with or without magnetostriction we refer to [BBNP14b], where spatial improvement of the solution is achieved by local considerations based on the refinement of the mesh by successive refinement of the elements of the mesh with a high error contribution (h-adaption or static regridding methods). Another approach is the so-called r-adaptation with evolving internal structure. This strategy performs re-meshing of the structure without changing the number of degrees of freedom, but can produce highly distorted element geometries. Numerical studies in the case of r-adaption on the LLG have been done in [vdBW13], where the results impose the adaptivity in space and time in order to simulate possible singularities of LLG reliable.

A more complicated situation is to simulate fast switching in magnetic systems [Ber02], modeled by the stochastic version of the Landau-Lifshitz-Gilbert equation (SLLG). In this case, thermal fluctuations $\boldsymbol{\mathcal{H}}_{\text{fl}}$ are included into the LLG by perturbing the effective field $\boldsymbol{\mathcal{H}}_{\text{eff}}$ to describe random changes which occur by the interaction of the ferromagnet with a surrounding heat bath. The study of random fluctuations in the dynamics of magnetism has been proposed by Néel [N46], and has later been studied by W.F. Brown [Bro63] where the focus is on a single nanomagnetic particle. After that, this stochastic equation has been considered by many physicists; see e.g. [BMS09, GPL98] among others. The random thermal fluctuations are usually modeled by $\boldsymbol{\mathcal{H}}_{\text{fl}} = \dot{\boldsymbol{W}}$ which has to be understood in the sense of Stratonovich in order to satisfy the saturation magnetization. Here $\boldsymbol{W}$ is a $\boldsymbol{Q}$-Wiener process; see [BBNP14b, Chapter 2] for a detailed description. The SLLG then takes the form

$$\frac{\partial \boldsymbol{X}_t}{\partial t} = \boldsymbol{X} \times \big(\zeta_1 \boldsymbol{\mathcal{H}}_{\text{eff}} + \boldsymbol{\mathcal{H}}_{\text{fl}}\big) - \zeta_2 \boldsymbol{X} \times \big(\boldsymbol{X} \times \boldsymbol{\mathcal{H}}_{\text{eff}}\big), \tag{1.0.4}$$

where thermal fluctuations in the damping part are neglected due to $0 < \zeta_2 \ll 1$, see [GPL98], and the works [BBNP14b, GPL98, Ber02] for further background of the physical model. The influence of the fluctuations in the equation is visualized in Figure 1.1(c).

The modeling analysis, and numerics of the stochastic Landau-Lifshitz-Gilbert equation is an active field of research: In recent works [AdBH14, BBNP14b, BGJ13] the existence of a weak martingale solution of the SLLG is established. In [BGJ13], where a rigorous mathematical treatment of the SLLG is initiated, the existence of a weak martingale solution in the case of a scalar-valued noise is provided using a Faedo-Galerkin approximation, and compactness arguments. By similar arguments, the existence of a weak martingale solution is shown in [BGJ17] for one-dimensional domains, and moreover, by pathwise uniqueness which holds in one space dimension, the existence of a strong solution is proven.

Stable numerical approximation schemes for the SLLG are studied in [AdBH14, BBNP14b] for $d \in \{2, 3\}$. In [BBNP14b] a finite element method, a midpoint scheme, and a random walk approximation of Wiener increments are combined to obtain a fully practical discretization of the SLLG, whereas in [AdBH14] a semi-discrete scheme which extends the projection algorithm of [AJ06] to the stochastic case is used to construct a weak martingale solution.

Monte-Carlo-based numerical studies in [BBNP14a] for the SLLG motivate discrete (finite time finite energy) blow-up for individual paths, whereas corresponding expectation values have a (expected) smooth evolution but retains some discrete blow-up like characteristic. These numerical studies motivate a pathwise discrete blow-up behavior of solutions in finite time, so that uniqueness is questionable.

Based on the manuscript [PS18] submitted for publication, focuses on adaptive concepts of SPDEs with linear drift, including the convection-dominated case where the streamline diffusion method is adopted to attain a stable discretization, and the stochastic version of the non-linear harmonic map heat flow to $\mathbb{S}^2$ where approximate solutions exhibit discrete blow-up dynamics. Next to this, our particular interest in this thesis is to compare existing stable discretizations of the SLLG [BBNP14b, Hoc15] for the proposed adaptive space-time and statistical concepts in [PS18].

The first part of this thesis is of independent relevancy, though it lays the fundaments to the numerical setup for the second and third part. Here the focus is on adaptive strategies for the time step size in the case of ordinary differential equations (ODEs), based on local extrapolation methods.

In the second part, the deterministic concepts are extended to SDEs ($n, L \in \mathbb{N}$)

$$\mathrm{d}\boldsymbol{X}_t = f\big(t, \boldsymbol{X}_t\big)\mathrm{d}t + \sigma\big(t, \boldsymbol{X}_t\big)\mathrm{d}\boldsymbol{W}_t, \quad t \in (t_0, T], \quad \boldsymbol{X}_{t_0} = \boldsymbol{x}_0 \in \mathbb{R}^L, \tag{1.0.5}$$

where $\boldsymbol{W}$ is a $\mathbb{R}^n$-valued Wiener process, and $f : \mathbb{R}_0^+ \times \mathbb{R}^L \to \mathbb{R}^L$ and $\sigma : \mathbb{R}_0^+ \times \mathbb{R}^L \to \mathbb{R}^{L\times n}$ are given drift and diffusion fluxes. Here we focus on both the strong and weak approximation of the SDE (1.0.5) using the explicit Euler-Maruyama method. In the case of a strong approximation to (1.0.5), we recall the relevancy of Brownian bridges and Brownian trees in integrating the pathwise solution. However, in many applications it is not necessary to generate an almost exact replica of the sample path of the solution of the underlying SDE. This motivates to consider adaptive strategies for the time step size of weak approximations of SDEs, as proposed in [STZ01]. These strategies are limited to lower order systems of SDEs due to the use of Kolmogorov's backward equation. Inspired by this, to overcome this curse of applicability to larger system of SDEs, as it e.g. occurs after a spatial discretization of SPDEs, we perform a local non-parametric high-dimensional density estimation based on Monte-Carlo sampling, which is concerned in the third part of this thesis.

Approximation of the error transition probability density in [STZ01] give an indication about the temporal change of the solution $\boldsymbol{X}_t$ on $[t_j, t_{j+1}]$ for some $j \in \mathbb{N}_0$. This is related to the question whether consequent laws on $[t_j, t_{j+1}]$ differ significant or not. From a statistical viewpoint, the error density can be measured in a corresponding manner by comparing subsequent histogram-based empirical laws

$$\widehat{\mu}^{\mathfrak{s},j+1} := \widehat{\mathcal{L}}\big(\boldsymbol{Y}^{\mathfrak{s},j+1}\big) \quad \big(\mathfrak{s} \in \{\mathfrak{1}, \mathfrak{2}\}\big)$$

of *a priori* unknown laws $\mu^{\mathfrak{s},j+1} := \mathcal{L}(\boldsymbol{Y}^{\mathfrak{s},j+1})$ $\big(\mathfrak{s} \in \{\mathfrak{1}, \mathfrak{2}\}\big)$ sampled at time t_{j+1} of related random variables $\boldsymbol{Y}^{\mathfrak{1},j+1}$ and $\boldsymbol{Y}^{\mathfrak{2},j+1}$. Here, the iterate $\boldsymbol{Y}^{\mathfrak{1},j+1}$ solves a discretization by performing one Euler step, while $\boldsymbol{Y}^{\mathfrak{2},j+1}$ results from using two Euler steps, and extrapolation. Since the behavior of the error density is not known *a priori*, adaptation of the local time step size $\tau^{j+1} := t_{j+1} - t_j > 0$ is performed by measuring closeness of laws with respect to some distance $\mathtt{d}$, where we either use the Hellinger $\mathtt{d}_{\mathtt{H}}$, Kullback-Leibler $\mathtt{d}_{\mathtt{KL}}$, or total variation $\mathtt{d}_{\mathtt{TV}}$ distance; see [GS02]. For the prototype problems in [PS18], and the SLLG, computational

studies show that the behavior of the time step size is stable with respect to the choice of $\mathtt{d}$. This is in contrast to the performed adaptivity in space, where, depending on $\mathtt{d}$, the number of degrees of freedom may differ greatly, having an impact on the performance of the algorithm.

Selection of the local time step size $\tau^{j+1} > 0$ is then achieved by checking whether

$$\mathbb{P}\Big[\mathtt{d}\big(\widehat{\mu}^{\mathtt{1},j+1}, \widehat{\mu}^{\mathtt{2},j+1}\big) \geqslant \mathtt{Tol}_\tau\Big] > p_\tau \tag{1.0.6}$$

is satisfied or not, for some given $\mathtt{Tol}_\tau > 0$, and $p_\tau \in [0,1]$. Thus, in order to fix the new time step size $\tau^{j+1} > 0$ at time t_j, several independent M-samples ($M \in \mathbb{N}$) of the subsequent iterates $\boldsymbol{Y}^{\mathtt{1},j+1}$ and $\boldsymbol{Y}^{\mathtt{2},j+1}$ are required to construct a high-dimensional data-dependent partition of the state space $\mathbb{R}^L$ in order to determine $\widehat{\mu}^{\mathfrak{s},j+1}$ $\big(\mathfrak{s} \in \{\mathtt{1},\mathtt{2}\}\big)$ on it, and afterwards to approximate the value of the distribution in (1.0.6) via Monte-Carlo sampling. Thus, since Monte-Carlo sampling is required at different levels of the algorithm, it would be impractical to approximate occurring estimators accurately by ample solving the SPDE.

In order to compute empirical laws $\widehat{\mu}^{\mathfrak{s},j+1}$ $\big(\mathfrak{s} \in \{\mathtt{1},\mathtt{2}\}\big)$, the most basic method for estimation is the histogram. A natural means of estimating a high-dimensional density/distribution is to partition the underlying state space $\mathbb{R}^L$ into a collection of uniform cells $\{\widehat{C}_r\}_{r=1}^R$, $R \in \mathbb{N}$ such that $\mathbb{R}^L = \bigcup_{r=1}^R \widehat{C}_r =: \widehat{\mathcal{P}}_{M;R}$, and then form standard estimates locally, within each cell, based on the relative frequency. However, uniform cells are not able to adapt to spatially varying anisotropic smoothness. Further developments of the histogram method allow the cells to depend on the data set, and substantial improvement can be obtained by the so-called 'data-dependent' histograms, i.e., one has to choose cells to be thinner in the direction where the density varies more; see e.g. [DGL96]. In [GO84], the authors introduced CART (Classification and Regression Trees) as a method for estimating classification and regression functions with piecewise constant estimates. For density estimation with CART-type methods, the breakthrough work [LN96] present L^1-consistency results on density estimators based on data-dependent partitions, i.e.,

$$\lim_{M\to\infty} \sup_{t>0} \mathbb{P}\left[\sum_{r=1}^R \Big|\mu\big[\widehat{C}_r\big] - \widehat{\mu}\big[\widehat{C}_r\big]\Big| > t \,\middle|\, \widehat{\mathcal{P}}_{M;R}\right] = 0.$$

Most of these data-dependent partitions are stored as binary trees, which are computationally attractive, since they allow fast range queries as e.g. needed to calculate the relative frequencies to determine $\{\widehat{\mu}^{\mathfrak{s},j+1}\}_{\mathfrak{s}}$, and low storage effort because each node stores only the decision criterions for sub-dividing the state space.

Providing these data sets to build $\widehat{\mathcal{P}}_{M;R}$ and $\{\widehat{\mu}^{\mathfrak{s},j+1}\}_{\mathfrak{s}}$ on it, as well as to approximate $\mathbb{P}$, requires long computational times in case we solve the SPDE; this effort for testing relation (1.0.6) may be drastically reduced by the re-sampling strategy bootstrap [Efr79], where independent identically distributed sub-samples are drawn with replacement from one sample for $\boldsymbol{Y}^{\mathtt{1},j+1}$ and $\boldsymbol{Y}^{\mathtt{2},j+1}$. Bootstrapping a given sample with equal selection probabilities is commonly used in practice; there, the sample is used to build $\widehat{\mathcal{P}}_{M;R}$, and $\widehat{\mu}^{\mathfrak{s},j+1}$ $\big(\mathfrak{s} \in \{\mathtt{1},\mathtt{2}\}\big)$ on it via cross-validation [Sto74]. In this thesis, in the present case of 'grouped' data according to a discrete distribution, we perform a weighted bootstrapping, based on the 'Alias method' [Wal77], in order to reduce the variability of the variance estimate, and to avoid solving the PDEs.

Time adaptivity is then complemented by a local refinement/coarsening strategy of the spatial mesh $\mathcal{T}_h^j$ via a stochastic version of the Zienkiewicz-Zhu (ZZ) estimator [ZZ92]. The ZZ-estimator $G_h(\nabla \boldsymbol{Y}^{j+1})$ is based on post-processing of the computed solution gradient $\nabla \boldsymbol{Y}^{j+1}$ in order to get a better approximation. Since no assumptions are needed about the regularity of the solution, this estimator may be flexible applied to different problems. In accordance with (1.0.6), for fixed $\tau^{j+1} > 0$, local refining/coarsening of an element $K \in \mathcal{T}_h^j$ is then achieved by checking whether

$$\mathbb{P}\Big[\mathtt{d}\big(\widehat{\mu}_K^{\mathtt{1},j+1}, \widehat{\mu}_K^{\mathtt{2},j+1}\big) \geqslant \mathtt{Tol}_h\Big] > p_h \tag{1.0.7}$$

is satisfied or not, for some given $\mathtt{Tol}_h > 0$, $p_h \in [0,1]$, with corresponding empirical laws

$$\begin{aligned}\widehat{\mu}_K^{\mathtt{1},j+1} &:= \widehat{\mathcal{L}}\big(|\nabla \boldsymbol{Y}^{j+1}|_K\big),\\ \widehat{\mu}_K^{\mathtt{2},j+1} &:= \widehat{\mathcal{L}}\big(|G_h(\nabla \boldsymbol{Y}^{j+1})|_K\big).\end{aligned}$$

This leads to a family of parametric empirical probability measures, indexed by K. In this case, an accurate partitioning of the underlying low-dimensional state space $\mathbb{R}_0^+$ requires by far less cells compared to $\mathbb{R}^L$ (temporal adaptivity), however, to build partitions and probability measures $\big\{\widehat{\mu}_K^{\mathfrak{s},j+1};\, K \in \mathcal{T}_h^j\big\}_{\mathfrak{s}}$ on it is quite time consuming. This deficiency may be avoided by the use of parallelization software [DM98] which is enabled by the performed h-adaption.

Next to an enhanced resolution of multiple scales of the solution, spatial adaptivity also serves as a variance reduction technique, enabling to attain the desired accuracy of our estimator of the distance $\mathtt{d}$ for a reduced number of Monte-Carlo simulations. In fact, adaptive meshes $\{\mathcal{T}_h^j\}_j$ are supported to favor certain scalings of continuous piecewise affine functions shared by the solution of the considered SPDE and thus serves as a strategy for importance sampling. Conversely, a decrease of the variance of the estimator along a sequence of adaptively refined mesh may be seen, as an indication of the performance of adaptive re-meshing; see [AG07, p. 134]. An immediate consequence of the variance reduction of the computed samples are the more accurate resolution of the state space $\mathbb{R}^L$ (respectively $\mathbb{R}_0^+$), leading to (possibly) increased Monte-Carlo approximations of the required histogram-based distributions, which is one of the crucial part of the proposed adaptive concepts.

Chapter 2

Adaptivity for ordinary differential equations

This chapter is about the numerical solution of ordinary differential equations (ODEs), which is to determine a vector-valued function $\boldsymbol{X} \in \mathcal{C}^1\big([t_0, T]; \mathbb{R}^L\big)$ such that ($L \in \mathbb{N}$)

$$\dot{\boldsymbol{X}}_t = f\big(t, \boldsymbol{X}_t\big) \quad \forall\, t \in (t_0, T], \quad \boldsymbol{X}_{t_0} = \boldsymbol{x}_0 \in \mathbb{R}^L. \tag{2.0.1}$$

Since the function $f : \mathbb{R}_0^+ \times \mathbb{R}^L \to \mathbb{R}^L$ is assumed to be continuously differentiable, there exists a unique solution at least locally on some open interval containing t_0.

In [DB02, Chapter 4], several one-step methods were discussed that allow to approximate the solution of (2.0.1) on uniform time meshes $I_\tau := \{t_j\}_{j=0}^J$ covering $[t_0, T]$, with equi-distant time step size $\tau := t_{j+1} - t_j > 0$, where the algorithms yield more precise results as $\tau \downarrow 0$. However, in practice it is often needed to find the approximations of $\boldsymbol{X}_t$ with a given tolerance $\texttt{Tol}_\tau > 0$, which inherits the problem of choosing an adequate (problem adapted) value for the time step size $\tau > 0$. For the construction of adaptive methods, the following aspects ought to be taken into account:

1. Efficiency/Accuracy: Minimization of the required number of time steps $J \in \mathbb{N}$ while maintaining the required accuracy $\texttt{Tol}_\tau$ of the approximations of $\boldsymbol{X}_t$.

2. Robustness/Reliability: In order to obtain a reliable numerical solution, the time step size must be adapted to the behavior of the solution, in particular, in cases for which the solution rapidly changes; see Figure 2.1.

In the case of adaptive one-step methods, a time step size $\tau^{j+1} > 0$ is determined for each time step t_j (and thus the new time step $t_{j+1} := t_j + \tau^{j+1}$), where $t_{j+1} = t_0 + \sum_{i=0}^{j} \tau^{i+1}$ for some given initial time $t_0 \geqslant 0$. Typically, for the user-specified tolerance $\texttt{Tol}_\tau > 0$, a quality measure to determine τ^{j+1} at time t_j may be the local consistency error Θ^{j+1} at time t_{j+1} which can be realized by one of the following two strategies:

(i) Error per step (EPS) to ensure $\big|\Theta^{j+1}\big(t_{j+1}, \boldsymbol{Y}^{j+1}; \tau^{j+1}\big)\big| \simeq \texttt{Tol}_\tau$. If the consistency error Θ^{j+1} does not satisfy the foregoing estimate, then estimate Θ^{j+1} with a smaller τ^{j+1} until Θ^{j+1} is smaller than tolerance $\texttt{Tol}_\tau$. If Θ^{j+1} is unnecessarily small, then one should increase τ^{j+1} on the next time step.

(ii) Error per unit step (EPUS) to ensure $\left|\Theta^{j+1}\big(t_{j+1}, \boldsymbol{Y}^{j+1}; \tau^{j+1}\big)\right| \simeq \tau^{j+1}\mathtt{Tol}_\tau$. If this criterion holds at time steps $\{t_j\}_j$, it is expected that the (accumulated) error at terminal time $t_J = T$ is of order $\mathcal{O}(\mathtt{Tol}_\tau)$. This conclusion is not true for EPS.

The most important reason for estimating local consistency errors is to gain some confidence in the approximations. There is some cost associated with estimating error and adapting τ^{j+1}, but generally this is a bargain because (2.0.1) is solved more efficiently. Indeed, if a solution changes significantly on a scale that is very small compared to $T - t_0$, it may then be impractical to solve the problem with a constant time step size that is small enough to resolve the fastest changes in the solution $\boldsymbol{X}_t$. And, if the time step size τ^{j+1} is too big, the computation may become instable. Controlling the error by adjusting τ^{j+1} can often stabilize the integrator.

The usual way to estimate the local consistency error via extrapolation can be described as taking each time step with two different formulas. There is a basic approximation $\boldsymbol{Y}^{1,j+1}$ of order $p_1 \in \mathbb{N}$, and a 'prior' approximation $\boldsymbol{Y}^{2,j+1}$ of order $\mathbb{N} \ni p_2 > p_1$. Then

$$\begin{aligned}\gamma^{j+1} &= \boldsymbol{Y}^{2,j+1} - \boldsymbol{Y}^{1,j+1} \\ &= \big[\boldsymbol{X}_{t_{j+1}} - \boldsymbol{Y}^{1,j+1}\big] - \big[\boldsymbol{X}_{t_{j+1}} - \boldsymbol{Y}^{2,j+1}\big] \\ &= \mathtt{err}^{j+1} + \mathcal{O}(h^{p_1+2}),\end{aligned}$$

where $\mathtt{err}^{j+1}$ is the local error at time t_{j+1}. Therefore, γ^{j+1} is an asymptotically correct estimate of the local consistency error of the lower order formula; see e.g. [Sha73]. A natural way to avoid the usage of different formulas is to look at two numerical approximations coming from the same method: one based on a single step with time step size τ^{j+1}, and the other based on two time steps with time step size $\tau^{j+1}/2$. Improved accuracy is then achieved via local extrapolation by computing a weighted average of both numerical approximations.

For both strategies EPS and EPUS, the local error estimator rests on the assumption that the higher order formula is more accurate. In general, the higher order method is used to advance the integration, and the lower order method is used for the purposes of error estimation and step size selection. Some of the effective codes that use EPS employ local extrapolation to achieve higher-order convergence by using lower-order methods instead of using two methods of the same order. This amounts to a generalized EPUS strategy. The most important difficulty with the local extrapolation is that the stability of the integration then depends on the stability of the higher order formula. This is due to the fact that performing local extrapolation on a low order method is equivalent to using a higher order formula and this high order formula may not be suitable for solving possibly stiff systems. In this context, deferred correction methods can be used to obtain solutions with improved orders of accuracy to ODEs; see e.g. [Ske82].

Let $\mathcal{I}_\tau := \{t_j\}_{j=0}^J$ be a partition of $[t_0, T]$, with adaptive time step size $\tau^{j+1} := t_{j+1} - t_j > 0$. By using Taylor's series, one can expect that the consistency error [DB02] of the implicit Euler method takes the form

$$\Theta^{j+1}\big(t_{j+1}, \boldsymbol{Y}^{j+1}; \tau^{j+1}\big) \approx \gamma^{j+1}\big(t_{j+1}, \boldsymbol{Y}^{j+1}\big)(\tau^{j+1})^2.$$

By performing two (trial) implicit Euler steps $\boldsymbol{Y}^{1,j+1}$ respectively $\boldsymbol{Y}^{2,j+1}$ with different time step sizes $\tau^{1,j+1} := \tau^{j+1}$ respectively $\tau^{2,j+1} := \tau^{j+1}/2$, where $\boldsymbol{Y}^{2,j+1}$ is considered as prior

approximation to $\boldsymbol{X}_{t_{j+1}}$, the function γ^{j+1} can be estimated via Richardson extrapolation

$$\gamma^{j+1}\big(t_{j+1}, \boldsymbol{Y}^{j+1}\big) \approx 2\boldsymbol{Y}^{2,j+1} - \boldsymbol{Y}^{1,j+1}.$$

This approximation of γ^{j+1} then provides a criterion to adaptively select the time step size via local extrapolation.

Algorithm 2.0.1 (Step size adaptivity via extrapolation)**.** *Fix $T > t_0$, $\tau_{\min} > 0$, $\texttt{Tol}_\tau > 0$, $\rho \in (0,1]$, and $\eta \in \mathbb{N}$. Choose $\tau^0 > 0$. Set $j := 0$, and $\boldsymbol{Y}^0 := \boldsymbol{x}_0 \in \mathbb{R}^L$.*
While $(t_j < T)$ do:

(I) Set $\tau^{1,j+1} := \min\{\tau^j, T - t_j\}$, $\tau^{2,j+1} := \tau^{1,j+1}/2$, and $t_{j+1} := t_j + \tau^{1,j+1}$.

(II) Compute (by solving the non-linear systems)

$$\boldsymbol{Y}^{1,j+1} := \boldsymbol{Y}^j + \tau^{1,j+1} f\big(t_{j+1}, \boldsymbol{Y}^{1,j+1}; \tau^{1,j+1}\big)$$

and

$$\begin{aligned}\boldsymbol{Z}^{2,j+1} &:= \boldsymbol{Y}^j + \tau^{2,j+1} f\big(t_{j+1}, \boldsymbol{Z}^{2,j+1}; \tau^{2,j+1}\big),\\ \boldsymbol{Y}^{2,j+1} &:= \boldsymbol{Z}^{2,j+1} + \tau^{2,j+1} f\big(t_{j+1} + \tau^{2,j+1}, \boldsymbol{Z}^{2,j+1}; \tau^{2,j+1}\big).\end{aligned}$$

(III) Compute $\gamma^{j+1} := |2\boldsymbol{Y}^{2,j+1} - \boldsymbol{Y}^{1,j+1}|$, and decide: If $\gamma^{j+1} \leqslant \tau^{1,j+1}\texttt{Tol}_\tau$ or $\tau^{1,j+1} \leqslant \tau_{min}$, set $\boldsymbol{Y}^{j+1} := \boldsymbol{Y}^{2,j+1}$, and

$$\tau^{j+1} := \max\left\{\tau_{min}, \min\left\{\eta\tau^{j+1}, \rho\sqrt{\frac{\texttt{Tol}_\tau}{\gamma^{j+1}}(\tau^{j+1})^2}\right\}\right\}, \tag{2.0.2}$$

and $j := j + 1$. Otherwise, set $\tau^{j+1} := \tau^{2,j+1}$.

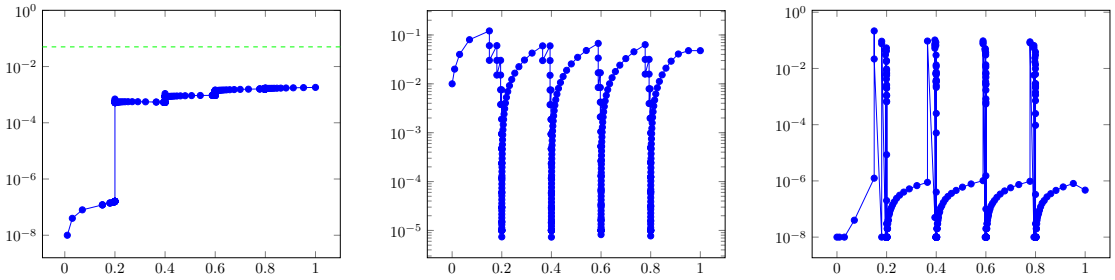
By shrinking τ^{j+1} to zero, it is possible, in principle, that γ^{j+1} in (2.0.2) becomes zero or, at least, so small that exponent overflow occurs (division by zero). Therefore, we force, due to stability reason, all time steps to remain above the minimal time step size $\tau_{\min} > 0$. Next to this, in order to keep the number of function evaluations small, the parameter ρ has also been introduced to increase the probability that the proposed time step size τ^{j+1} will be accepted; in fact, this means that some of the time steps $\{t_j\}_j$ are unnecessarily small. Since each failed step must be repeated, efficiency will be lost unless the number of such failures is kept small by a conservative choice of $\{\tau^j\}_j$.

We illustrate the behavior of Algorithm 2.0.1 for the following ODE:

$$\dot{\boldsymbol{X}}_t = \big[-\boldsymbol{J} + \boldsymbol{D}(t)\big]\boldsymbol{X}_t \quad \forall\, t \in (0,1], \quad \boldsymbol{X}_0 = (1,1,1,1)^T, \tag{2.0.3}$$

with $\boldsymbol{J} = 16 \cdot \text{tridiag}[-1, 2, -1] \in \mathbb{R}^{4\times 4}$, and $\boldsymbol{D}(t) = \text{diag}[\gamma_1(t), \ldots, \gamma_4(t)] \in \mathbb{R}^{4\times 4}$, where $\gamma_i(t) := {}^{\mathbb{1}_{(\beta_i,1]}(t)}/_{2\sqrt{t-\beta_i+10^{-8}}}$ and $\beta_i := {}^i/_5$. Here, $t \mapsto \boldsymbol{X}_t$ fastly changes in the vicinity of times $t \in \{\beta_i\}_{i=1}^4$; see Figure 2.1(a).

Example 2.0.2. *Figure 2.1(a) displays the behavior of $t_j \mapsto |\boldsymbol{X}_{t_j} - \boldsymbol{Y}^j|$ for the implicit Euler discretization of* (2.0.3) *on an adaptively refined/coarsened time mesh $\mathcal{I}_\tau$ with $J = 300$ time steps via Algorithm 2.0.1 to stay below the given error threshold $\texttt{Tol}_\tau$ (- - - -). In fact,* 70% *of these overall required time steps are concentrated in the vicinity of the times $t \in \{\beta_i\}_{i=1}^4$. The adapted local time step size $\{\tau^j\}_j$ rapidly decays near these times, and afterwards coarsens again; see Figure 2.1(b). Figure 2.1(c) shows the behavior of the consistency error to obtain an accumulated error $\mathcal{O}(\texttt{Tol}_\tau) = \sum_{j=-1}^{J-1} \tau^{j+1}\gamma^{j+1} \approx 0.03768$ at the terminal time T.*



(a) $t_j \mapsto |\boldsymbol{X}_{t_j} - \boldsymbol{Y}^j|$ **(b)** $t_j \mapsto \tau^j$ **(c)** $t_j \mapsto \gamma^j$

Figure 2.1: (Example 2.0.2 for $T = 1.0$, $\tau_{\min} = 10^{-5}$, $\texttt{Tol}_\tau = 0.05$, $\rho = 0.9$, $\eta = 2$, and $t_0 = 0.0$, $\tau_0 = 10^{-2}$) (a) Behavior of $t_j \mapsto |\boldsymbol{X}_{t_j} - \boldsymbol{Y}^j|$, and (b) corresponding adaptive time step size. (c) Evolution of the estimated consistency error $t_j \mapsto \gamma^j$.

Adapting the time step size or method to the solution is not just a matter of improving the efficiency of the integration; it makes practical the solution of stiff problems. If the solution change at different times (as in Figure 2.1(b)), it may be impractical to solve (2.0.1) with a constant mesh size.

Chapter 3

Adaptivity for stochastic differential equations

Stochastic modeling and simulations have become areas of intense research in recent years, as more sophisticated mathematical models of physical phenomenon became available. Stochastic differential equations arise in many applications; see e.g. [vK81]. Stochastic models are computationally much more challenging than deterministic models. One way to reduce the computational cost of an approximation algorithm is to use an adaptive time-stepping scheme to advance the numerical solution. In the framework of ODEs, mesh size adapting strategies have been proved to be essential in generating an optimal algorithm.

Let $T > t_0$, $n, L \in \mathbb{N}$, and $\mathfrak{P} := (\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a complete filtered probability space, on which a $\mathbb{R}^n$-valued Wiener process $\boldsymbol{W} \equiv \{\boldsymbol{W}_t;\, t \in [t_0, T]\}$ is defined. We shall consider an $\mathbb{R}^L$-valued Itô process $\boldsymbol{X} \equiv \{\boldsymbol{X}_t;\, t \in [t_0, T]\}$, satisfying the stochastic differential equation (SDE)

$$\mathrm{d}\boldsymbol{X}_t = f\big(t, \boldsymbol{X}_t\big)\mathrm{d}t + \sigma\big(t, \boldsymbol{X}_t\big)\mathrm{d}\boldsymbol{W}_t, \quad t \in (t_0, T], \quad \boldsymbol{X}_{t_0} = \boldsymbol{x}_0 \in \mathbb{R}^L, \tag{3.0.1}$$

where $f : \mathbb{R}_0^+ \times \mathbb{R}^L \to \mathbb{R}^L$ and $\sigma : \mathbb{R}_0^+ \times \mathbb{R}^L \to \mathbb{R}^{L\times n}$ are given drift and diffusion fluxes. This corresponds to a system of SDEs such that, for all $\mathtt{l} \in \{1, \ldots, L\}$,

$$\mathrm{d}X_t^{\mathtt{l}} = f^{\mathtt{l}}\big(t, \boldsymbol{X}_t\big)\mathrm{d}t + \sum_{i=1}^{n} \sigma_i^{\mathtt{l}}\big(t, \boldsymbol{X}_t\big)\mathrm{d}W_t^i, \quad t \in (t_0, T], \quad X_{t_0}^{\mathtt{l}} = x_0^{\mathtt{l}} \in \mathbb{R}, \tag{3.0.2}$$

with independent $\mathbb{R}$-valued Wiener processes $W^i \equiv \{W_t^i;\, t \in [t_0, T]\}$, $i \in \{1, \ldots, n\}$ on $\mathfrak{P}$, as well as $f^{\mathtt{l}} : \mathbb{R}_0^+ \times \mathbb{R}^L \to \mathbb{R}^L$, $\sigma_i^{\mathtt{l}} : \mathbb{R}_0^+ \times \mathbb{R}^L \to \mathbb{R}$, for $\mathtt{l} \in \{1, \ldots, L\}$. The $\mathbb{R}$-valued Wiener processes $\{W_t^i\}_i$ are also referred to as Brownian motions. If the coefficients $\{f^{\mathtt{l}}\}_{\mathtt{l}}$, and $\{\sigma_i^{\mathtt{l}}\}_{i,\mathtt{l}}$ satisfy the Lipschitz condition, i.e., there exist constants $K_1, K_2 > 0$ such that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^L$, and $t \in [t_0, T]$ holds

$$\begin{aligned} |f^{\mathtt{l}}(t, \boldsymbol{x}) - f^{\mathtt{l}}(t, \boldsymbol{y})| &\leqslant K_1|\boldsymbol{x} - \boldsymbol{y}|, \\ |\sigma_i^{\mathtt{l}}(t, \boldsymbol{x}) - \sigma_i^{\mathtt{l}}(t, \boldsymbol{y})| &\leqslant K_2|\boldsymbol{x} - \boldsymbol{y}|, \end{aligned}$$

then the solution $\boldsymbol{X}_t$ of problem (3.0.1) has a unique solution which is adapted to the filtration induced by the Wiener process $\boldsymbol{W}_t$. This solution satisfies $\sup_{t\in[t_0,T]} \mathbb{E}\big[|\boldsymbol{X}_t|^p\big] \leqslant C(T, |\boldsymbol{x}_0|^p)$ for each $p \in [1, \infty)$.

The Monte-Carlo (MC) method approximates the unknown process $\boldsymbol{X}$ by the explicit Euler discretization for (3.0.1), which is constructed as in [KP92, Mil95].

Scheme 3.0.1 (Explicit Euler method)**.** *Let $\mathcal{I}_\tau := \{t_j\}_{j=0}^J$ be a partition of $[t_0, T]$, with adaptive time step size $\tau^{j+1} := t_{j+1} - t_j > 0$, and $\boldsymbol{Y}^0 \in \mathbb{R}^L$ be given. For all $j \in \{0, 1, \dots, J\}$, and $\Delta_{j+1}\boldsymbol{W} := \boldsymbol{W}_{t_{j+1}} - \boldsymbol{W}_{t_j} \sim \mathcal{N}(\boldsymbol{0}, \tau^{j+1}\boldsymbol{Id})$, determine a $\mathcal{F}_{t_j}$-measurable $\mathbb{R}^L$-valued random variable $\boldsymbol{Y}^{j+1}$ such that $\mathbb{P}$-almost surely*

$$\boldsymbol{Y}^{j+1} - \boldsymbol{Y}^j = \tau^{j+1} f\big(t_j, \boldsymbol{Y}^j\big) + \sigma\big(t_j, \boldsymbol{Y}^j\big)\Delta_{j+1}\boldsymbol{W}.$$

To simulate $\Delta_{j+1}\boldsymbol{W}$, we approximate the $\mathbb{R}^n$-valued Wiener process $\boldsymbol{W}$ on $\mathfrak{P}$ by a $\mathbb{R}^n$-valued discretely distributed random walk. For this purpose, consider the (time-discrete) filtration $\mathfrak{F}_\tau := \{\mathcal{F}_{t_j};\, t_j \in \mathcal{I}_\tau\} \subset \mathfrak{F}$, and set $\mathfrak{P}_\tau := (\Omega, \mathcal{F}, \mathfrak{F}_\tau, \mathbb{P})$.

Definition 3.0.2. *Given $\mathcal{I}_\tau := \{t_j\}_{j=0}^J$. A random walk on $\mathfrak{P}_\tau$ is a sequence $\{\boldsymbol{\xi}_{j+1}\}_{j=0}^J$ of $\mathbb{R}^n$-valued independent identically distributed random variables such that for each $j \in \{0, 1, \dots, J\}$ the conditions below are satisfied:*

(1) $\boldsymbol{\xi}_{j+1}$ is $\mathcal{F}_{t_{j+1}}$-measurable and independent of $\big\{\mathcal{F}_{t_i};\, i \in \{1, \dots, j\}\big\}$.

(2) $\mathbb{E}\big[\boldsymbol{\xi}_{j+1}\big] = \boldsymbol{0}$, and $\mathbb{E}\big[\langle\boldsymbol{\xi}_{j+1}, \mathbf{x}\rangle\langle\boldsymbol{\xi}_{j+1}, \mathbf{y}\rangle\big] = \tau^{j+1}\langle\mathbf{x}, \mathbf{y}\rangle$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

(3) For every $p \in \mathbb{N}$, there exists $C_p > 0$ such that $\mathbb{E}\big[|\boldsymbol{\xi}_{j+1}|^{2p}\big] \leqslant C_p(\tau^{j+1})^p$.

Then, thanks to Donsker's invariance principle [KS91, p. 70], $\{\boldsymbol{\xi}_{j+1}\}_{j=0}^J$ converges weakly to the n-dimensional Wiener process $\boldsymbol{W}$. Note that Definition 3.0.2 generalizes the definition of an $\mathbb{R}$-valued random walk.

Adaptive time-stepping algorithms for strong or weak approximation of the SDE (3.0.1) depends crucially on the application. Weak approximations (cf. [KP92, Chapter 14]) are acceptable when only the moments of the exact solution $\boldsymbol{X}_t$ need to be estimated accurately, while strong approximations (cf. [KP92, Chapter 10]) are required when individual trajectories of the exact solution $\boldsymbol{X}_t$ need to be well approximated.

Section 3.1 is concerned with concepts of discrete Brownian motions and Brownian paths, which is relevant in the case of a strong approximation to (3.0.1). We use these strategies for temporal adaptivity in Chapter 4 to reduce the variance while performing local extrapolation along the same path. In Section 3.2, we recall in the literature existing adaptive time-stepping strategies for weak approximation of (3.0.1), where the time discretization error can be either approximated by deterministic or stochastic time steps. In this thesis we focus on deterministic time steps, however, for the sake of completeness, both strategies are stated.

3.1 Strong approximation

Adaptive time-stepping strategies for the strong (or 'pathwise') approximation of (3.0.1) driven by a one-dimensional ($n = 1$) Wiener process was considered in [Lam03, HMGR01]. Adaptivity for Stratonovich SDEs (cf. [KP92, Chapter 4.9]) with a multi-dimensional ($n > 1$) Wiener process was studied in [BB02]. In [HMGR01], adaptive time-stepping strategies in the mean-square sense were developed, which were optimal for asymptotically small mesh sizes. In [GL97], the authors showed that to guarantee the convergence of a variable time step size scheme applied to the SDE in (3.0.1), a strong order at least one discretization scheme is needed.

An important difficulty in integrating the pathwise solution of a system of SDEs is that, when rejection of time step sizes is allowed, the solution should remain on the same Brownian path. Otherwise a bias in the numerical solution is introduced. To overcome this biasedness, according to [GL97], the asymptotically Lévy-Ciesielski construction of a Brownian motion is used. Fix $\mathcal{I}_j := [t_j, t_{j+1}]$ for some $j \in \mathbb{N}_0$, and $i \in \{1, \dots, n\}$. For each level $l \in \mathbb{N}$, subdivide the interval $\mathcal{I}_j$ into 2^l equi-distant sub-intervals

$$\mathcal{I}_j^l := \left\{ s = t_j + i\frac{(t_{j+1} - t_j)}{2^l} \ : \ i \in \{0, 1, \dots, 2^l\} \right\}.$$

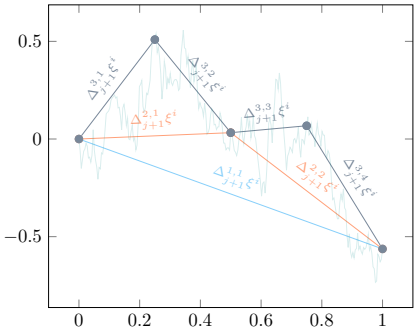
Note that $\mathcal{I}_j = \bigcup_{l \in \mathbb{N}} \mathcal{I}_j^l$. Fix $i \in \{1, \dots, n\}$. Then, for any $t \in \mathcal{I}_j$, the i-th random walk ξ_t^i is calculated iteratively as follows:

$$\xi_t^i := \frac{1}{2}\left[\xi^i\left(t - \frac{(t_{j+1} - t_j)}{2^l}\right) + \xi^i\left(t + \frac{(t_{j+1} - t_j)}{2^l}\right)\right] + Z_t^i \quad \forall\, t \in \mathcal{I}_j^l \backslash \mathcal{I}_j^{l-1},$$

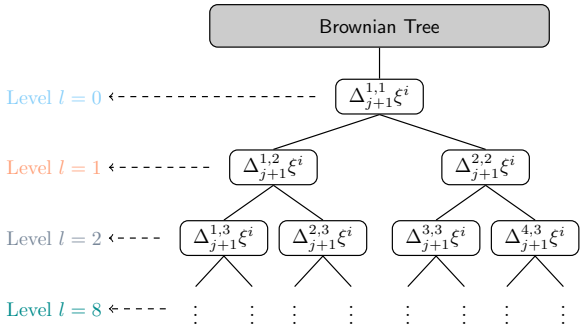
where $Z^i \equiv \{Z_t^i;\, t \in \mathcal{I}_j^l\}$, $i \in \{1, \dots, n\}$ is a stochastic process such that

$$Z_t^i \sim \mathcal{N}\left(0, \frac{t_j + t_{j+1}}{2^{l+1}}\right) \quad \forall\, t \in \mathcal{I}_j^l \backslash \mathcal{I}_j^{l-1}.$$

Due to the binary logic that the time integration step can be either halved or doubled, each Brownian path ξ_t^i is stored as the so-called Brownian tree; see Figure 3.1(b). Generalizations



(a) Brownian motion



(b) Brownian tree

Figure 3.1: (Fixed $i \in \{1, \dots, n\}$, $n \in \mathbb{N}$): (a) Discrete Brownian motion after $l = 0$ (——), $l = 1$ (——), $l = 2$ (——), and $l = 8$ (- - -) levels of sub-division of $[0, 1]$, and the corresponding Brownian tree to store the Brownian increments on the associated sub-interval.

of Brownian trees for flexible changes of time step sizes exist, but require: the calculation of covariances on each sub-interval as well as an extra effort to match the boundaries/alignment points. The Wiener increments corresponding to level $\mathtt{L} \in \mathbb{N}$ are constructed as follows.

Algorithm 3.1.1 (Discrete Brownian path)**.** *Let $\mathcal{I}_j := [t_j, t_{j+1}]$ for some $j \in \mathbb{N}_0$ be given. Choose $i \in \{1, \dots, n\}$, $\omega \in \Omega$, and $\mathtt{L} \in \mathbb{N}$. Set $\Delta_{j+1}^{1,1}\xi^i(\omega) := \xi_{t_{j+1}}^i(\omega) - \xi_{t_j}^i(\omega)$ on $\mathcal{I}_j$.*
For $l = 1, \dots, \mathtt{L}$ do:
For $m = 1, \dots, 2^{l-1}$ do:

(I) Compute $\Delta_{j+1}^{2m-1,l+1}\xi^i(\omega) := \frac{1}{2}\Delta_{j+1}^{m,l}\xi^i(\omega) + Z_{j+1}^{2m-1,l}(\omega)$, for $Z_{j+1}^{2m-1,l} \sim \mathcal{N}(0, \frac{t_j+t_{j+1}}{2^l})$.

(II) Compute $\Delta_{j+1}^{2m,l+1}\xi^i(\omega) := \frac{1}{2}\Delta_{j+1}^{m,l}\xi^i(\omega) - Z_{j+1}^{2m,l}(\omega)$, for $Z_{j+1}^{2m,l} \sim \mathcal{N}(0, \frac{t_j+t_{j+1}}{2^l})$.

The Brownian path has to pass through all the points generated for the smaller sub-intervals before progressing to an upper level in the Brownian tree; cf. 3.1(b). This condition imposes a serious restrictions on the selection of a time step size, and may slow down the integration process significantly.

The key to the construction of most higher-order numerical approximations is usually obtained from the truncated expansion of the variables of interest over the small increments. Taylor's formula provides the basis for the derivation of most deterministic numerical algorithms; cf. Chapter 2. In the stochastic case, a stochastic Taylor's expansion for Itô SDEs is described in [KP92], which is obtained by iterated applications of the Itô formula to the integrands in the integral version of the SDE (3.0.1). In general, one can say that for a high-order numerical scheme one requires adequate smoothness of the drift and diffusion coefficients, but also adequate information about the driving Wiener process. This information contained in the multiple stochastic integrals appearing in the stochastic Taylor's formula.

3.2 Weak approximation

In many applications it is not necessary to generate an almost exact replica of the sample path of the solution of the underlying SDE. The Monte-Carlo simulation of the option prices in mathematical finance (cf. [Gla04]) is a typical example, where simple random walks can be used to approximate option pricing functionals. Within this section, we discuss numerical methods that focus on approximating the probability distributions of solutions of SDEs. For the following, we define a piecewise affine, globally continuous process $\boldsymbol{Y}$ on $[t_0, T]$ by the iterates $\{\boldsymbol{Y}^j\}_j$ obtained via Scheme 3.0.1.

Definition 3.2.1. *Let $\mathcal{I}_\tau = \{t_j\}_{j=0}^J$ be a net of fineness $\{\tau^j\}_{j=1}^J$ covering $[t_0, T]$ with $T > t_0$, and $J \in \mathbb{N}$. For all $t \in [t_0, T]$ define*

$$\boldsymbol{Y}_t := \boldsymbol{Y}^j + \int_{t_j}^t \widetilde{f}\big(s, \boldsymbol{Y}^j\big)\mathrm{d}s + \int_{t_j}^t \widetilde{\sigma}\big(s, \boldsymbol{Y}^j\big)\mathrm{d}\boldsymbol{W}_s \tag{3.2.1}$$

with

$$\begin{aligned} \widetilde{f}\big(s, \boldsymbol{Y}^j\big) &:= \sum_{j=0}^{J-1} \mathbb{1}_{[t_j, t_{j+1})}(s) \cdot f\big(t_j, \boldsymbol{Y}^j\big) \quad \forall\, s \in [t_0, T], \\ \widetilde{\sigma}\big(s, \boldsymbol{Y}^j\big) &:= \sum_{j=0}^{J-1} \mathbb{1}_{[t_j, t_{j+1})}(s) \cdot \sigma\big(t_j, \boldsymbol{Y}^j\big) \quad \forall\, s \in [t_0, T]. \end{aligned} \tag{3.2.2}$$

Note that in view of (3.2.1), $\lim_{t\uparrow t_{j+1}} \boldsymbol{Y}_t = \boldsymbol{Y}^{j+1}$ *for all* $j \in \{0,\dots,J-1\}$. *Therefore,* $\boldsymbol{Y}_t$ *is a continuous process on* $[t_0,T]$, $\mathbb{P}$*-almost surely. We remark that since why* $\boldsymbol{Y}^{j+1}$ *is* $\mathcal{F}_{t_{j+1}}$*-adapted, and the filtration* $\mathbb{F}$ *is a increasing sequence of* σ*-algebras, for each* $t \in (t_j, t_{j+1})$ *the process* $\boldsymbol{Y}_t$ *is* $\mathcal{F}_{t_{j+1}}$*-measurable. Thus* $\boldsymbol{Y}_t$ *is not* $\mathbb{F}$*-adapted.*

As in the case of strong approximations, weak higher-order schemes can be constructed only if there is an adequate smoothness of the drift and diffusion coefficients, and a sufficiently rich set of random variables approximating the multiple stochastic integrals of the corresponding weak Taylor schemes, generated at each time step. Recall that the weak higher-order Taylor schemes involve higher-order derivatives of f and σ_i. Obviously, it would be desirable to have derivative-free or Runge-Kutta-type weak schemes.

In the deterministic numerical analysis, (local) extrapolation method represents an elegant way of achieving higher-order convergence by using lower-order methods, provided the numerical stability of these for a range of step sizes can be guaranteed. For the weak second-order approximation of the functional $\mathbb{E}\big[g(\boldsymbol{X}_T)\big]$, for smooth functional $g : \mathbb{R}^L \to \mathbb{R}$, the authors in [TT90] proposed a Richardson extrapolation of the form, cf. Algorithm 2.0.1 in the deterministic context,

$$v_T = 2\mathbb{E}\big[g(\boldsymbol{Y}_T^2)\big] - \mathbb{E}\big[g(\boldsymbol{Y}_T^1)\big],$$

where $v_T \in \mathbb{R}$ denotes the functional value at terminal time T of the explicit Euler iterates $\boldsymbol{Y}_T^1$ respectively $\boldsymbol{Y}_T^2$ obtained with steps sizes $\tau^{1,j+1} := \tau^{j+1}$ respectively $\tau^{2,j+1} := {}^{\tau^{j+1}}\!/_2$. Further, the weak higher-order extrapolation method has been developed in [TT90, Chapter 15.3], which require the existence of a leading error expansion for functionals of the underlying discrete time weak approximations $\{\boldsymbol{Y}^j\}_j$.

The basis for adaptive weak approximations [STZ01] is the expansion of the discretization error for the explicit Euler approximation in [TT90]. It states that for the Euler approximation $\boldsymbol{Y}$ of the SDE (3.0.1) with uniform time step size $\tau = \frac{T-t_0}{J}$, $J \in \mathbb{N}$, the error arising in the weak approximations can be expanded as,

$$\mathbb{E}\big[g(\boldsymbol{X}_T)\big] - \mathbb{E}\big[g(\boldsymbol{Y}_T)\big] = \tau \int_{t_0}^{T} \mathbb{E}\big[\Gamma(s,\boldsymbol{X}_s)\big]\mathrm{d}s + \mathcal{O}(J^{-2}), \tag{3.2.3}$$

where the function $\Gamma : \mathbb{R}_0^+ \times \mathbb{R}^L \to \mathbb{R}$ is given by

$$\begin{aligned}\Gamma(t,\boldsymbol{x}) = \frac{1}{2}\sum_{\mathtt{l},\mathtt{k}=1}^{L} f^{\mathtt{l}}(t,\boldsymbol{x}) f^{\mathtt{k}}(t,\boldsymbol{x}) \frac{\partial^2 u(t,\boldsymbol{x})}{\partial x_{\mathtt{l}} \partial x_{\mathtt{k}}} + \sum_{\mathtt{l},\mathtt{k},\mathtt{v}=1}^{L} f^{\mathtt{l}}(t,\boldsymbol{x}) d^{\mathtt{k}\mathtt{v}}(t,\boldsymbol{x}) \frac{\partial^3 u(t,\boldsymbol{x})}{\partial x_{\mathtt{l}} \partial x_{\mathtt{k}} \partial x_{\mathtt{v}}} \\ + \frac{1}{2}\frac{\partial u(t,\boldsymbol{x})}{\partial t} + \frac{1}{2}\sum_{\mathtt{l},\mathtt{k},\mathtt{v},\mathtt{w}=1}^{L} d^{\mathtt{l}\mathtt{k}}(t,\boldsymbol{x}) d^{\mathtt{v}\mathtt{w}}(t,\boldsymbol{x}) \frac{\partial^4 u(t,\boldsymbol{x})}{\partial x_{\mathtt{l}} \partial x_{\mathtt{k}} \partial x_{\mathtt{v}} \partial x_{\mathtt{w}}} \\ + \sum_{\mathtt{l}=1}^{L} f^{\mathtt{l}}(t,\boldsymbol{x}) \frac{\partial^2 u(t,\boldsymbol{x})}{\partial t \partial x_{\mathtt{l}}} + \sum_{\mathtt{l},\mathtt{k}=1}^{L} d^{\mathtt{l}\mathtt{k}}(t,\boldsymbol{x}) \frac{\partial^3 u(t,\boldsymbol{x})}{\partial t \partial x_{\mathtt{l}} \partial x_{\mathtt{k}}},\end{aligned}$$

with $d^{\mathtt{l}\mathtt{k}} \equiv \frac{1}{2}\sigma^{\mathtt{l}}\sigma^{\mathtt{k}}$ and $u(t,\boldsymbol{x}) = \mathbb{E}\big[g(\boldsymbol{X}_T) \,|\, \boldsymbol{X}_t = \boldsymbol{x}\big]$. Note that, here and for the rest of this chapter, the Einstein summation convection is used, i.e., if the same subscript appears twice in a term, the term denotes the sum over the range of this subscript. The proof of this error expansion extends to non-uniform time steps as well and can also be generalized to higher-order weak Itô-Taylor schemes (see [KP92, Theorem 14.6.1]).

The idea of adaptive weak approximations is to use the foregoing expansion to refine the time steps on those time intervals $[t_j, t_{j+1}]$, where

$$\int_{t_j}^{t_{j+1}} \mathbb{E}\big[\Gamma(s, \boldsymbol{X}_s)\big]\mathrm{d}s > \mathtt{Tol}_\tau \tag{3.2.4}$$

for some given tolerance $\mathtt{Tol}_\tau > 0$. Unfortunately, the function Γ depends on the unknown solution $\boldsymbol{X}$ and derivatives of the unknown solution $u : \mathbb{R}_0^+ \times \mathbb{R}^L \to \mathbb{R}$. This problem can, however, be solved with the aid of the solution to a dual problem. The use of dual functions is standard in optimal control theory and in particular for adaptive mesh control for ODEs and PDEs, see [EEHJ95], and was successfully applied in the probabilistic context [STZ01, MSTZ05].

Let $\mathcal{S}_{\boldsymbol{Y}_T} := \big\{g\big(\boldsymbol{Y}_T(\omega_k)\big)\big\}_{k=1}^M$ be an M-sample of independent identically distributed realizations of the related random variable $\boldsymbol{Y}_T$, for $M \in \mathbb{N}$. The aim is to choose both, the time step sizes and the sample size, such that

$$\left|\mathbb{E}\big[g(\boldsymbol{X}_T)\big] - \frac{1}{M}\sum_{k=1}^{M} g\big(\boldsymbol{Y}_T(\omega_k)\big)\right| \leqslant \mathtt{Tol}_\tau, \tag{3.2.5}$$

with high probability and for minimum number of time steps and realizations. The computational error naturally separates into two parts,

$$\begin{aligned}
&\mathbb{E}\big[g(\boldsymbol{X}_T)\big] - \frac{1}{M}\sum_{k=1}^{M} g\big(\boldsymbol{Y}_T(\omega_k)\big) \\
&\qquad = \Big(\mathbb{E}\big[g(\boldsymbol{X}_T)\big] - \mathbb{E}\big[g(\boldsymbol{Y}_T)\big]\Big) + \Big(\mathbb{E}\big[g(\boldsymbol{Y}_T)\big] - \frac{1}{M}\sum_{k=1}^{M} g\big(\boldsymbol{Y}_T(\omega_k)\big)\Big) \\
&\qquad \equiv \mathcal{E}_T + \mathcal{E}_S.
\end{aligned}$$

The time steps $\{t_j\}_j$ for each trajectory $\boldsymbol{Y}$ are determined from the statistical approximations of the time discretization error $\mathcal{E}_T$. The number of realizations M are determined from the statistical error $\mathcal{E}_S$, which can be asymptotically determined by the central limit theorem:

$$\forall a > 0 : \lim_{M\uparrow\infty} \mathbb{P}\left[\left|\mathbb{E}\big[g(\boldsymbol{Y}_T)\big] - \frac{1}{M}\sum_{k=1}^{M} g\big(\boldsymbol{Y}_T(\omega_k)\big)\right| \leqslant a\sqrt{\frac{\mathrm{Var}_M\big[g(\boldsymbol{Y}_T)\big]}{M}}\right] = 2\Phi(a) - 1,$$

where Φ is the standard normal distribution. To reduce $\mathcal{E}_S$ in the MC estimation of $\mathbb{E}[g(\boldsymbol{Y}_T)]$, choose $M \in \mathbb{N}$ according to $M = \lfloor a^2\,\mathrm{Var}_M[g(\boldsymbol{Y}_T)]/\mathtt{Tol}_\tau^2 \rfloor$, where $\lfloor x \rfloor := \max\{k \in \mathbb{Z} \,|\, k \leqslant x\}$. Here, $a > 0$ defines the confidence interval of the realizations of $\mathcal{S}_{\boldsymbol{Y}_T}$.

The main results of [STZ01] are the expansions of $\mathcal{E}_T$ with computational leading order term in *a posteriori* form. The main difference, compared to [TT90], is that the weight for the local error contribution (cf. (3.2.4)) to the global error can be computed by the stochastic flows and discrete dual backward problems, extending [MSTZ03] to SDEs. Thus, $\mathcal{E}_T$ is then approximated by either

$$\mathcal{E}_T \simeq \mathbb{E}\left[\sum_{j=-1}^{J-1} \rho(t_{j+1})\big(\tau^{j+1}\big)^2\right] \tag{3.2.6}$$

with an error density function $\rho : \mathbb{R} \times \Omega \to \mathbb{R}$ for stochastic time steps or

$$\mathcal{E}_T \simeq \sum_{j=-1}^{J-1} \mathbb{E}\big[\rho(t_{j+1})\big]\big(\tau^{j+1}\big)^2 \tag{3.2.7}$$

for deterministic time steps. We emphasize that, the computable local error density $\rho(t_{j+1}, \omega)$ measures the local contribution of the global error $\mathcal{E}_T$. They give information on where to refine $[t_j, t_{j+1}]$ (by successive halving) or not to reach an optimal mesh according to (3.2.5), based on the $\mathbb{P}$-almost sure convergence of the density $\rho(t_{j+1}, \omega)$ as $\mathcal{I}_\tau$ is refined. Both the deterministic and stochastic time-stepping algorithm are detailed in the following sub-sections. The adaptive algorithm with deterministic time steps is simpler and not as computationally demanding as the adaptive algorithm with stochastic time steps, where $J \equiv J(\omega)$ for $\omega \in \Omega$. The algorithm with stochastic time steps will, on the other hand, produce better results for problems where the initial data is perturbed from one trajectory to another according to some given distribution.

3.2.1 Adaptive approximations with deterministic time steps

The following Lemma 3.2.2 and Theorem 3.2.3, derived in [STZ01], describe the error expansion which is used in the adaptive Algorithms 3.2.4 and 3.2.6 below.

Lemma 3.2.2. *Suppose that, for some $m > \lfloor \frac{L}{2} \rfloor + 10$, there are constants $p \equiv p(\boldsymbol{\alpha}, T) > 0$, and $C \equiv C\big(\boldsymbol{\alpha}, T\big) > 0$, such that $g \in \mathcal{C}^m_{loc}\big(\mathbb{R}^L\big)$, $|\partial_{\boldsymbol{\alpha}} g(\boldsymbol{x})| \leqslant C\big(1 + |\boldsymbol{x}|^p\big)$, for all $|\boldsymbol{\alpha}| \leqslant m$, where $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_L)^T \in \mathbb{N}_0^L$ is a multi-index with $\mathbb{N}_0 \ni |\boldsymbol{\alpha}| := \sum_{i=1}^L \alpha_i$, and*

$$\partial_{\boldsymbol{\alpha}} := \frac{\partial^{|\boldsymbol{\alpha}|}}{\partial_1^{\alpha_1} \partial_2^{\alpha_2} \ldots \partial_L^{\alpha_L}}.$$

Furthermore, let $\mathbb{E}\big[|\boldsymbol{X}_{t_0}|^{2p+L+1} + |\boldsymbol{Y}|^{2p+L+1}\big] \leqslant C$, and bounded $f, \sigma \in \mathcal{C}^m\big([t_0, T] \times \mathbb{R}^L\big)$. Then the solution $\boldsymbol{X}$ of (3.0.1) and $\boldsymbol{Y}$ in (3.2.1), based on deterministic time steps, satisfy

$$\begin{aligned}
\mathbb{E}\big[g(\boldsymbol{X}_T) - g(\boldsymbol{Y}_T)\big] &= \mathbb{E}\big[u(t_0, \boldsymbol{X}_{t_0}) - u(t_0, \boldsymbol{Y}_{t_0})\big] \\
&\quad + \mathbb{E}\left[\int_{t_0}^T \sum_{\mathtt{l}=1}^L \left(\Big(f^{\mathtt{l}}(s, \boldsymbol{Y}_s) - \widetilde{f}^{\mathtt{l}}(s, \boldsymbol{Y}_s)\Big) \frac{\partial u(s, \boldsymbol{Y}_s)}{\partial x_{\mathtt{l}}}\right) \mathrm{d}s\right] \\
&\quad + \mathbb{E}\left[\int_{t_0}^T \sum_{\mathtt{l},\mathtt{k}=1}^L \left(\Big(d^{\mathtt{lk}}(s, \boldsymbol{Y}_s) - \widetilde{d}^{\mathtt{lk}}(s, \boldsymbol{Y}_s)\Big) \frac{\partial^2 u(s, \boldsymbol{Y}_s)}{\partial x_{\mathtt{l}} \partial x_{\mathtt{k}}}\right) \mathrm{d}s\right].
\end{aligned} \tag{3.2.8}$$

Here, $\widetilde{d}^{\mathtt{lk}} \equiv \frac{1}{2}\widetilde{\sigma}^{\mathtt{l}}\widetilde{\sigma}^{\mathtt{k}}$, $\mathtt{l}, \mathtt{k} \in \{1, \ldots, L\}$, for coefficients $\widetilde{f}^{\mathtt{l}}$, $\widetilde{\sigma}^{\mathtt{l}}$ as defined in (3.2.2).

Proof. According to [Fri64, Theorem 15], a standard energy estimate, using the regularity assumptions on f, σ, and g, can be combined with the Sobolev inequality to show that there exists a unique solution $u \in \mathcal{C}^{1,6}_{\mathrm{loc}}\big([t_0, T] \times \mathbb{R}^L\big)$ of the Kolmogorov backward equation

$$\begin{aligned}
-\frac{\partial}{\partial t} u(t, \boldsymbol{x}) - \sum_{\mathtt{l}=1}^L f^{\mathtt{l}}(t, \boldsymbol{x}) \frac{\partial}{\partial x_{\mathtt{l}}} u(t, \boldsymbol{x}) - \sum_{\mathtt{l},\mathtt{k}=1}^L d^{\mathtt{lk}}(t, \boldsymbol{x}) \frac{\partial^2}{\partial x_{\mathtt{l}} \partial x_{\mathtt{l}}} u(t, \boldsymbol{x}) &= 0, \\
u(T, \boldsymbol{x}) &= g(\boldsymbol{x}),
\end{aligned} \tag{3.2.9}$$

satisfying the polynomial growth condition

$$\max_{t\in[t_0,T]} |\partial_{\boldsymbol{\alpha}} u(t,\boldsymbol{x})| \leqslant C\Big(1+|\boldsymbol{x}|^{p+\frac{L+1}{2}}\Big), \quad |\boldsymbol{\alpha}|\leqslant 6,$$

for some $p, C > 0$. The Feynman-Kac formula without potential, implies that the solution u of (3.2.9) can be represented by the expected value $u(t,\boldsymbol{x}) = \mathbb{E}\big[g(\boldsymbol{Y}_T)\,|\,\boldsymbol{Y}_t = \boldsymbol{x}\big]$. The Itô formula applied to (3.2.1) (cf. [IW89, p. 66]) with respect to $g(t,\boldsymbol{x})$, taking expectation, and using the foregoing representation for u, gives

$$\begin{aligned}\mathrm{d}u(t,\boldsymbol{Y}_t) = \Bigg(&\frac{\partial}{\partial t}u(t,\boldsymbol{Y}_t) + \sum_{\mathtt{l}=1}^{L} \widetilde{f}^{\mathtt{l}}(t,\boldsymbol{Y}_t)\frac{\partial}{\partial x_{\mathtt{l}}}u(t,\boldsymbol{Y}_t)\\ &+ \sum_{\mathtt{l},\mathtt{k}=1}^{L} \widetilde{d}^{\mathtt{l}\mathtt{k}}(t,\boldsymbol{Y}_t)\frac{\partial^2}{\partial x_{\mathtt{l}}\partial x_{\mathtt{l}}}u(t,\boldsymbol{Y}_t)\Bigg)\mathrm{d}t + \sum_{\mathtt{l}=1}^{L}\widetilde{\sigma}^{\mathtt{l}}(t,\boldsymbol{Y}_t)\frac{\partial}{\partial t}u(t,\boldsymbol{Y}_t)\mathrm{d}\boldsymbol{W}_t.\end{aligned}$$

Substituting the Kolmogorov equation (3.2.9) to eliminate $\partial u/\partial t$, integrate over $[t_0,T]$, using the identity $u(T,\boldsymbol{x}) = \mathbb{E}\big[g(\boldsymbol{Y}_T)\,|\,\boldsymbol{Y}_T=\boldsymbol{x}\big] = \mathbb{E}\big[g(\boldsymbol{x})\big]$, and taking expectation yields

$$\begin{aligned}\mathbb{E}\big[u(t_0,\boldsymbol{Y}_{t_0})\big] - \mathbb{E}\big[g(\boldsymbol{Y}_T)\big] &= \mathbb{E}\left[\int_{t_0}^{T}\sum_{\mathtt{l}=1}^{L}\Big(\Big(f^{\mathtt{l}}(s,\boldsymbol{Y}_s) - \widetilde{f}^{\mathtt{l}}(s,\boldsymbol{Y}_s)\Big)\frac{\partial u(s,\boldsymbol{Y}_s)}{\partial x_{\mathtt{l}}}\Big)\mathrm{d}s\right]\\ &\quad + \mathbb{E}\left[\int_{t_0}^{T}\sum_{\mathtt{l},\mathtt{k}=1}^{L}\Big(\Big(d^{\mathtt{l}\mathtt{k}}(s,\boldsymbol{Y}_s) - \widetilde{d}^{\mathtt{l}\mathtt{k}}(s,\boldsymbol{Y}_s)\Big)\frac{\partial^2 u(s,\boldsymbol{Y}_s)}{\partial x_{\mathtt{l}}\partial x_{\mathtt{k}}}\Big)\mathrm{d}s\right]\\ &\quad + \mathbb{E}\left[\int_{t_0}^{T}\sum_{\mathtt{l}=1}^{L}\Big(\widetilde{\sigma}^{\mathtt{l}}(s,\boldsymbol{Y}_s)\frac{\partial}{\partial s}u(s,\boldsymbol{Y}_s)\mathrm{d}\boldsymbol{W}_s\Big)\right].\end{aligned}$$

According to the martingale property of Itô integrals, see [KP92, Lemma 3.2.2], the last term on the right hand side is zero. It holds $0 = \mathbb{E}\big[g(\boldsymbol{X}_T)\big] - \mathbb{E}\big[g(\boldsymbol{X}_T)\big] = \mathbb{E}\big[g(\boldsymbol{X}_T)\big] - \mathbb{E}\big[u(T,\boldsymbol{x})\big]$, since by the Feynman-Kac Formula $u(T,\boldsymbol{x}) = \mathbb{E}\big[g(\boldsymbol{X}_T)\,|\,\boldsymbol{X}_T=\boldsymbol{x}\big] = \mathbb{E}\big[g(\boldsymbol{x})\big]$. Insert this on the left hand side of the foregoing equation concludes the proof. □

Lemma 3.2.2 is combined with the stochastic flows to derive the *a posteriori* error expansion in Theorem 3.2.3 below. This error expansion is based on the variations of the processes $\boldsymbol{X}$ and $\boldsymbol{Y}$. For a process $\boldsymbol{Y}$, the first variation of a function $F(\boldsymbol{Y}_T)$ with respect to a perturbation in the initial location of the path $\boldsymbol{Y}$, at time $s\in[t_0,T]$, is denoted by

$$F'(T;s) := \frac{\partial F(\boldsymbol{Y}_T)}{\partial x(s)} \equiv \left(\frac{\partial}{\partial x_1}F\big(\boldsymbol{Y}_T;\ \boldsymbol{Y}_s=\boldsymbol{x}\big),\ldots,\frac{\partial}{\partial x_L}F\big(\boldsymbol{Y}_T;\ \boldsymbol{Y}_s=\boldsymbol{x}\big)\right).$$

The proof of Theorem 3.2.3 uses mainly that the error in replacing $g(\boldsymbol{X}_T)$ in Lemma 3.2.2 by $g(\boldsymbol{Y}_T)$, in the representation (3.2.8) of $\partial_{\boldsymbol{\alpha}}u$, yields the small deterministic remainder term $\int_{t_0}^{T}\mathcal{O}\big((\tau^{j+1})^2\big)\mathrm{d}s$ in (3.2.10) of Theorem 3.2.3, which is analogous to the $\mathcal{O}(J^{-2})$ term in (3.2.3), and needs some *a priori* estimate to be controlled. Lemma 3.2.2 can be applied to estimate this error. The second important ingredient in the proof is the Markov property of $\boldsymbol{Y}$ satisfied at the discrete times t_j. Based on the fact that $\boldsymbol{Y}_{t_j}$ is $\mathcal{F}_{t_j}$-measurable, the nested expected values, for $\mathtt{l}\in\{1,\ldots,L\}$,

$$\mathbb{E}\Big[f^{\mathtt{l}}(t_j,\boldsymbol{Y}_{t_j})\partial_{x_1(t_j)}\mathbb{E}\big[g(\boldsymbol{Y}_T)\big]\,\Big|\,\mathcal{F}_{t_j}\Big]$$

in (3.2.8) can, by the definition (3.2.11) of φ, be decoupled to

$$\mathbb{E}\big[f^{\mathtt{l}}(t_j, \boldsymbol{Y}_{t_j})\varphi_{\mathtt{l}}(t_j)\big],$$

which reduces the computational complexity substantially.

Theorem 3.2.3. *Suppose that f, σ, g, $\boldsymbol{X}$, and $\boldsymbol{Y}$ satisfy the assumptions in Lemma 3.2.2. Then, $\mathcal{E}_T$ has the following expansion: for each $\mathtt{l}, \mathtt{k}, \mathtt{v} \in \{1, \dots, L\}$,*

$$\begin{aligned}
\mathcal{E}_T &= \mathbb{E}\big[g(\boldsymbol{X}_T) - g(\boldsymbol{Y}_T)\big] \\
&= \sum_{j=0}^{J-1}\sum_{k=1}^{M}\bigg[\Big(f^{\mathtt{l}}\big(t_{j+1}, \boldsymbol{Y}_{t_{j+1}}(\omega_k)\big) - f^{\mathtt{l}}\big(t_j, \boldsymbol{Y}_{t_j}(\omega_k)\big)\Big)\varphi_{\mathtt{l}}\big(t_{j+1}, \omega_k\big)\bigg]\frac{\tau^{j+1}}{2M} \\
&\quad + \sum_{j=0}^{J-1}\sum_{k=1}^{M}\bigg[\Big(d^{\mathtt{lk}}\big(t_{j+1}, \boldsymbol{Y}_{t_{j+1}}(\omega_k)\big) - d^{\mathtt{lk}}\big(t_j, \boldsymbol{Y}_{t_j}(\omega_k)\big)\Big)\varphi'_{\mathtt{lv}}\big(t_{j+1}, \omega_k\big)\bigg]\frac{\tau^{j+1}}{2M} \\
&\quad + \sum_{j=-1}^{J-1}\big(\tau^{j+1}\big)^2\Big\{\mathcal{O}(\tau^{j+1}) + \sum_{m=j}^{J-1}\mathcal{O}\big((\tau^{m+1})^2\big)\Big\} + \int_{t_0}^{T}\big(\mathrm{I}_M + \mathrm{II}_M\big)\mathrm{d}s,
\end{aligned} \tag{3.2.10}$$

where I_M and II_M represent the statistical error of the first and second terms, respectively. The functions $\varphi_{\mathtt{l}} : \mathbb{R}_0^+ \times \Omega \to \mathbb{R}^L$ and $\varphi'_{\mathtt{lk}} : \mathbb{R}_0^+ \times \Omega \to \mathbb{R}^{L\times L}$ in (3.2.10) are determined as follows: For $\{t_j\}_{j=0}^{J-1}$, $\{\tau^{j+1}\}_{j=0}^{J-1}$ and $\boldsymbol{x} \in \mathbb{R}^L$, let

$$c_{\mathtt{l}}(t_j, \boldsymbol{x}) = \boldsymbol{x}^{\mathtt{l}} + f^{\mathtt{l}}(t_j, \boldsymbol{x})\tau^{j+1} + \sum_{i=1}^{n}\sigma_i^{\mathtt{l}}(t_j, \boldsymbol{x})\Delta_{j+1}\boldsymbol{W}^i,$$

and note that $c_{\mathtt{l}}(t_j, \boldsymbol{Y}_{t_j}) = \boldsymbol{Y}^{\mathtt{l}}_{t_{j+1}}$. The spatial derivatives of c quantify the sensitivity of $\boldsymbol{Y}_{t_{j+1}}$ with respect to perturbations in $\boldsymbol{Y}_{t_j}$ and can, hence, be used to measure the propagation of discretization error. The dual functions $\varphi_{\mathtt{l}}$, for $\mathtt{l}, \mathtt{v} \in \{1, \dots, L\}$, are recursively defined as means of the dual backward problem

$$\begin{aligned}
\varphi_{\mathtt{l}}(T, \omega_k) &= \frac{\partial}{\partial x_{\mathtt{l}}}g\big(\boldsymbol{Y}_T(\omega_k)\big), \\
\varphi_{\mathtt{l}}(t_j, \omega_k) &= \frac{\partial}{\partial x_{\mathtt{l}}}c_{\mathtt{v}}\big(t_j, \boldsymbol{Y}_T(\omega_k)\big)\varphi_{\mathtt{v}}\big(t_{j+1}, \omega_k\big), \quad t_j \in [t_0, T).
\end{aligned} \tag{3.2.11}$$

Similarly, the dual function $\varphi'_{\mathtt{lk}} : \mathbb{R}_0^+ \times \Omega \to \mathbb{R}^{L\times L}$, $\mathtt{l}, \mathtt{k} \in \{1, \dots, L\}$ are recursively defined as

$$\begin{aligned}
\varphi'_{\mathtt{lk}}(T, \omega_k) &= \frac{\partial^2}{\partial x_{\mathtt{l}}\partial x_{\mathtt{k}}}g\big(\boldsymbol{Y}_T(\omega_k)\big), \\
\varphi'_{\mathtt{lk}}(t_j, \omega_k) &= \frac{\partial}{\partial x_{\mathtt{l}}}c_{\mathtt{v}}\big(t_j, \boldsymbol{Y}_T(\omega_k)\big)\frac{\partial}{\partial x_{\mathtt{k}}}c_{\mathtt{w}}\big(t_j, \boldsymbol{Y}_T(\omega_k)\big)\varphi'_{\mathtt{vw}}\big(t_{j+1}, \omega_k\big) \\
&\quad + \frac{\partial^2}{\partial x_{\mathtt{l}}\partial x_{\mathtt{k}}}c_{\mathtt{v}}\big(t_j, \boldsymbol{Y}_T(\omega_k)\big)\varphi'_{\mathtt{v}}\big(t_{j+1}, \omega_k\big), \quad t_j \in [t_0, T).
\end{aligned}$$

The variances of the statistical errors I_M and II_M are of order $\mathcal{O}\big((\tau^{j+1})^2\big)/M$, so it is clear that the first two terms in (3.2.10) are solely responsible for the weak order of convergence of the Euler scheme being equal to one. Consequently these two terms can be used to construct a refinement criterion for the time discretization.

For a given M-sample $\mathcal{S}_Z := \{Z(\omega_k)\}_{k=1}^M$ of a related random variable Z, let

$$\mathrm{Var}_M[Z] := \sqrt{\mathbb{E}_M[Z^2] - \big(\mathbb{E}_M[Z]\big)^2}, \quad \mathbb{E}_M[Z] := \frac{1}{M}\sum_{k=1}^{M} Z(\omega_k),$$

as well as, for $j \in \{1,\dots,J\}$, and $k \in \{1,\dots,M\}$,

$$\begin{aligned}\rho(t_j,\omega_k) := \frac{1}{2\tau^{j+1}}\Big(f^{\mathtt{l}}\big(t_{j+1},\boldsymbol{Y}_{t_{j+1}}(\omega_k)\big) - f^{\mathtt{l}}\big(t_j,\boldsymbol{Y}_{t_j}(\omega_k)\big)\Big)\varphi_{\mathtt{l}}\big(t_{j+1},\omega_k\big)\\ + \frac{1}{2\tau^{j+1}}\Big(d^{\mathtt{lk}}\big(t_{j+1},\boldsymbol{Y}_{t_{j+1}}(\omega_k)\big) - d^{\mathtt{lk}}\big(t_j,\boldsymbol{Y}_{t_j}(\omega_k)\big)\Big)\varphi'_{\mathtt{lv}}\big(t_{j+1},\omega_k\big).\end{aligned}$$

The adaptive Algorithm 3.2.4 can be divided into two halves. In the first half the time discretization is refined until $\mathcal{E}_T$ and $\mathcal{E}_S$ is larger than $\mathtt{Tol}_\tau$. In the second half, the number of trajectories is increased, with fixed time discretization, until $\mathcal{E}_S < \mathtt{Tol}_\tau$.

Algorithm 3.2.4 (Adaptive control with deterministic time steps)**.** *Fix $T > t_0$, $\mathtt{Tol}_\tau > 0$, $M_{\max} \in \mathbb{N}$, $J_{\max} \in \mathbb{N}$, and $a > 0$. Choose $M^0 \in \mathbb{N}$, and $\mathcal{I}_\tau^0 = \{\tau^{j,0}\}_{j=1}^{J^0-1}$ with $J^0 \in \mathbb{N}$. Split $\mathtt{Tol}_\tau = \mathtt{Tol}_S + \mathtt{Tol}_T + \mathtt{Tol}_{ST}$, with statistical tolerance $\mathtt{Tol}_S$, time discretization tolerance $\mathtt{Tol}_T$, and statistical time discretization tolerance $\mathtt{Tol}_{ST}$. Set $\mathtt{err}_T^0 := 2\mathtt{Tol}_T$, $\mathtt{err}_{ST}^0 := 2\mathtt{Tol}_{ST}$, and $q := 0$.*
While $\big(\mathtt{err}_T^q + \mathtt{err}_{ST}^q\big) > \mathtt{Tol}_T + \mathtt{Tol}_{ST}$ do:

(I) Compute $\mathcal{S}^q := \{\mathcal{S}^{q,k}\}_{k=1}^{M^q}$, with $\mathcal{S}^{q,k} := \{\boldsymbol{Y}^j(\omega_k)\}_{j=0}^{J^q}$ on $\mathcal{I}_\tau^q$ via Scheme 3.0.1.

(II) Based on $\mathcal{S}^q$, compute

$$\begin{aligned}\mathtt{err}_T^{q+1} &:= \left|\mathbb{E}_{M^q}\left[\sum_{j=-1}^{J^q-1}\rho(t_{j+1})\big(\tau^{j+1,q}\big)^2\right]\right|,\\ \mathtt{err}_{ST}^{q+1} &:= \frac{a}{\sqrt{M^q}}\,\mathrm{Var}_{M^q}\left[\sum_{j=-1}^{J^q-1}\rho(t_{j+1})\big(\tau^{j+1,q}\big)^2\right],\end{aligned}$$

and decide:

(1) If $\mathtt{err}_{ST}^{q+1} > \mathtt{Tol}_{ST}$, set

$$M^{q+1} := \min\left\{\left\lfloor\left(a\cdot\mathrm{Var}_{M^q}\left[\sum_{j=-1}^{J^q-1}\rho(t_{j+1})\big(\tau^{j+1,q}\big)^2\right]\right)^2\mathtt{Tol}_{ST}^{-1}\right\rfloor, M^q M_{\max}\right\},$$

as well as $J^{q+1} := J^q$, and $q := q+1$.

(2) If $\mathtt{err}_T^{q+1} > \mathtt{Tol}_T$*, determine* $\mathcal{I}_\tau^{q+1}$ *based on* $\mathcal{I}_\tau^q$ *by dividing each interval* $[t_j, t_{j+1}]$*,* $j \in \{0, 1, \ldots, J^q - 1\}$ *into* $r_j \in \mathbb{N}$ *uniform sub-intervals, where*

$$r_j := \min\left\{\max\left\{\left\lfloor \frac{\tau^{j+1,q}}{\widetilde{\tau}^{j+1,q}} \right\rfloor, 1\right\}, J_{\max}\right\},$$

and

$$\widetilde{\tau}^{j+1,q} := \frac{\mathtt{Tol}_T}{\sqrt{\left|\mathbb{E}_{M^q}\left[\rho(t_j)\right]\right|} \int_{t_0}^T \sqrt{\left|\mathbb{E}_{M^q}\left[\rho(t_{j(s)})\right]\right|}\mathrm{d}s}.$$

Set $M^{q+1} := M^q$*, and* $q := q + 1$*.*

Set $M_S^0 := M^q$*,* $\mathtt{err}_S^0 := 2\mathtt{Tol}_{ST}$*, and* $p := 0$*. Below, keep* $\mathcal{I}_\tau^q$*, and* J^q *fixed.*
While $\mathtt{err}_S^p > \mathtt{Tol}_S$ *do:*

(III) Compute $\mathcal{S}^p := \{\mathcal{S}^{p,k}\}_{k=1}^{M_S^p}$*, with* $\mathcal{S}^{p,k} := \{\boldsymbol{Y}^j(t_j, \omega_k)\}_{j=0}^{J^q}$ *on* $\mathcal{I}_\tau^q$ *via Scheme 3.0.1.*

(IV) Based on $\mathcal{S}^p$*, compute* $\mathbb{E}_{M_S^p}\left[g(\boldsymbol{Y}_T)\right]$*,* $\mathtt{err}_S^{p+1} := a \cdot \left(\sqrt{M_S^p}\right)^{-1} \mathrm{Var}_{M_S^p}\left[g(\boldsymbol{Y}_T)\right]$*, and*

$$M_S^{p+1} := \min\left\{\left\lfloor \left(a \cdot \mathrm{Var}_{M_S^q}\left[g(\boldsymbol{Y}_T)\right]\right)^2 \mathtt{Tol}_S^{-1} \right\rfloor, M_S^p M_{\max}\right\}.$$

Set $p := p + 1$*.*

Accept $\mathbb{E}_{M_S^p}\left[g(\boldsymbol{Y}_T)\right]$ *as an approximation to* $\mathbb{E}\left[g(\boldsymbol{X}_T)\right]$ *since the computational error is bounded by* $\mathtt{Tol}_\tau$*, at least with probability* $2\Phi(a) - 1$*.*

Compared to the deterministic time steps obtained by Algorithm 3.2.4, the stochastic time steps are advantageous for problems with possible singularities at random times. The criterion (3.2.6) is achieved by through a test performed at each interval at each realization, to decide whether to refine or not the given interval. In this case, when a node is added, the interpolation is carried through the consideration of a Brownian bridge; cf. Algorithm 3.1.1. Besides, since their use entails more work per realization than the deterministic time steps they should be judiciously used.

3.2.2 Adaptive approximations with stochastic time steps

The expansion (3.2.10) represents the error $\mathcal{E}_T$ as a sample mean over $M \in \mathbb{N}$ simulated trajectories and cannot be used to determine the error arising from a certain trajectory. For this, we need the error expansion derived in [STZ01, Theorem 3.3].

Theorem 3.2.5. *Suppose that* f*,* σ*,* g*,* $\boldsymbol{X}$ *satisfy the assumptions in Lemma 3.2.2, and that* $\boldsymbol{Y}$ *is constructed via Scheme 3.0.1 with time step sizes* $\{\tau^j\}_{j=1}^J$ *satisfying* $\mathbb{P}$*-almost surely for all* $s \in [t_0, T]$*,* $c(\mathtt{Tol}_\tau) \leqslant |\rho(s)| \leqslant C(\mathtt{Tol}_\tau)$*, for some positive functions* c *and* C*, with* $\frac{\mathtt{Tol}_\tau}{c(\mathtt{Tol}_\tau)} \to 0$ *as* $\mathtt{Tol}_\tau \to 0$*, and the corresponding* $\Delta_{j+1}\boldsymbol{W}$ *are generated by Brownian bridges. Then,* $\mathcal{E}_T$ *has the expansion*

$$\begin{aligned}\mathbb{E}\left[g(\boldsymbol{X}_T) - g(\boldsymbol{Y}_T)\right] = \mathbb{E}&\left[\sum_{j=-1}^{J-1} \widetilde{\rho}\left(t_{j+1}, \boldsymbol{Y}\right)\left(\tau^{j+1}\right)^2\right] \\ &+ \mathcal{O}\left(\sqrt{\frac{\mathtt{Tol}_\tau}{c(\mathtt{Tol}_\tau)}}\left(\frac{C(\mathtt{Tol}_\tau)}{c(\mathtt{Tol}_\tau)}\right)^{\frac{8}{c_0}}\right)\sum_{j=-1}^{J-1}\left(\tau^{j+1}\right)^2,\end{aligned} \tag{3.2.12}$$

for some $c_0 > 16$, *and*

$$\begin{aligned}\widetilde{\rho}(t_j,\omega_k) = \frac{1}{2}\sum_{\mathtt{l}=1}^{L}\left(\frac{\partial}{\partial t}f^{\mathtt{l}} + \sum_{\mathtt{k}=1}^{L} f^{\mathtt{k}}\frac{\partial}{\partial x_{\mathtt{l}}}f^{\mathtt{l}} + \sum_{\mathtt{v}=1}^{L} d^{\mathtt{kf}}\frac{\partial^2}{\partial x_{\mathtt{k}}\partial x_{\mathtt{v}}}f^{\mathtt{l}}\right)\varphi_{\mathtt{l}}(t_{j+1},\omega_k)\\ + \frac{1}{2}\sum_{\mathtt{l},\mathtt{k}=1}^{L}\left(\frac{\partial}{\partial t}d^{\mathtt{lk}} + \sum_{\mathtt{v}=1}^{L} f^{\mathtt{v}}\frac{\partial}{\partial x_{\mathtt{v}}}d^{\mathtt{lk}} + \sum_{\mathtt{vw}=1}^{L} d^{\mathtt{vw}}\frac{\partial^2}{\partial x_{\mathtt{v}}\partial x_{\mathtt{w}}}d^{\mathtt{lk}}\right.\\ \left. +2\sum_{\mathtt{v}=1}^{L} d^{\mathtt{vk}}\frac{\partial}{\partial x_{\mathtt{v}}}f^{\mathtt{l}}\right)\varphi'_{\mathtt{lk}}(t_{j+1},\omega_k) + \sum_{\mathtt{l},\mathtt{k},\mathtt{v},\mathtt{w}=1}^{L} d^{\mathtt{vw}}\frac{\partial}{\partial x_{\mathtt{v}}}d^{\mathtt{lk}}\varphi''_{\mathtt{lkw}}(t_{j+1},\omega_k),\end{aligned} \tag{3.2.13}$$

where the terms in (3.2.13) *are evaluated at the a posteriori known points* $\big(t_j, \boldsymbol{Y}_{t_j}(\omega_k)\big)$. *Here, the functions* φ *and* φ' *are the dual functions in Theorem 3.2.3, and* φ'' *are recursively defined as, for* $\mathtt{l},\mathtt{k},\mathtt{u},\mathtt{w},\mathtt{z} \in \{1,\dots,L\}$,

$$\begin{aligned}\varphi''_{\mathtt{lkw}}(T,\omega_k) &= \frac{\partial^3}{\partial x_{\mathtt{l}}\partial x_{\mathtt{k}}\partial x_{\mathtt{w}}}g\big(\boldsymbol{Y}_T(\omega_k)\big),\\ \varphi'_{\mathtt{lkw}}(t_j,\omega_k) &= \sum_{\mathtt{v},\mathtt{u}=1}^{L}\Bigg[\frac{\partial}{\partial x_{\mathtt{l}}}c_{\mathtt{v}}\frac{\partial}{\partial x_{\mathtt{k}}}c_{\mathtt{u}}\sum_{\mathtt{z}=1}^{L}\frac{\partial}{\partial x_{\mathtt{w}}}c_{\mathtt{z}}\varphi'_{\mathtt{vwz}}\big(t_{j+1},\omega_k\big)\\ &\qquad + \frac{\partial^2}{\partial x_{\mathtt{l}}\partial x_{\mathtt{k}}}c_{\mathtt{v}}\frac{\partial}{\partial x_{\mathtt{w}}}c_{\mathtt{u}}\varphi'_{\mathtt{vu}}\big(t_{j+1},\omega_k\big)\\ &\qquad + \frac{\partial}{\partial x_{\mathtt{l}}}c_{\mathtt{v}}\frac{\partial^2}{\partial x_{\mathtt{k}}\partial x_{\mathtt{w}}}c_{\mathtt{u}}\varphi'_{\mathtt{vu}}\big(t_{j+1},\omega_k\big)\\ &\qquad + \frac{\partial^2}{\partial x_{\mathtt{l}}\partial x_{\mathtt{w}}}c_{\mathtt{v}}\frac{\partial}{\partial x_{\mathtt{k}}}c_{\mathtt{u}}\varphi'_{\mathtt{vu}}\big(t_{j+1},\omega_k\big)\\ &\qquad + \frac{\partial^3}{\partial x_{\mathtt{l}}\partial x_{\mathtt{k}}\partial x_{\mathtt{w}}}\varphi_{\mathtt{v}}(t_j,\omega_k)\Bigg] \quad t_j \in [t_0,T).\end{aligned}$$

There are some practical differences to mention. On the one hand, the variance $\mathrm{Var}_M\big[\widetilde{\rho}\big]$ is of order $\mathcal{O}\big((\tau^{j+1}M)^{-1}\big)$. This feature has been observed in [STZ01], where a local filtering procedure was proposed to reduce the variance of the error density estimator. A positive feature of this error density is that it does not require the computation of the second variation, φ'', which may be computationally expensive for large L. On the other hand, the error density $\widetilde{\rho}$ in (3.2.13) has a much smaller variance $\mathcal{O}\big(M^{-1}\big)$ which does not need filtering but it requires the computation of φ''.

It is sometimes possible to use sparse matrix structure so that the additional work to compute the stochastic flows φ, φ', and φ'' is low compared to $\boldsymbol{Y}$; see [BSTZ13]. The additional storage of one realization is clearly also a drawback. At the expense of losing optimal mesh adaptivity, the additional work and storage (of $\boldsymbol{Y}$) can be removed by setting $\varphi \equiv \varphi' \equiv \varphi'' \equiv 1$, which corresponds to use only the local error as indicator for the adaptive refinements.

Algorithm 3.2.6 (Adaptive control with stochastic time steps)**.** *Fix $T > t_0$, $\mathtt{Tol}_\tau > 0$, $M_{\max} \in \mathbb{N}$, $J_{\max} \in \mathbb{N}$, and $a > 0$. Choose $M^0 \in \mathbb{N}$, and $\mathcal{I}_\tau^{m,0} = \{\tau_{m,0}^j\}_{j=1}^{J^{m,0}-1}$ with $J^{m,0} \in \mathbb{N}$. Split $\mathtt{Tol}_\tau = \mathtt{Tol}_S + \mathtt{Tol}_T$, with statistical tolerance $\mathtt{Tol}_S$, and time discretization tolerance $\mathtt{Tol}_T$. Set $\delta := \mathtt{Tol}_T/J^{m,0}$, $\mathtt{err}_S^0 := 2\mathtt{Tol}_S$, and $q := 0$.*
While $\mathtt{err}_S^q > \mathtt{Tol}_S$ do:

(I) For $m = 1, \dots, M^q$ do:

(1) Generate the increments $\{\boldsymbol{\xi}_{j+1}^{m,0}\}_{j=0}^{J^{m,0}-1}$ on $\mathcal{I}_\tau^{m,0}$. Set $p := 0$, and $\mathtt{err}_{T,j+1}^{m,p} := 2\delta$ for all $j \in \{0, 1, \dots, J^{m,p} - 1\}$.

(2) While $\max_j \mathtt{err}_{T,j+1}^{m,p} \geqslant \delta$ do:

(a) Compute $\mathcal{S}_{\boldsymbol{Y}}^{m,p} := \{\mathcal{S}_{\boldsymbol{Y}}^{m,p,k}\}_{k=1}^{M^q}$, $\mathcal{S}_{\boldsymbol{Y}}^{m,p,k} := \{\boldsymbol{Y}^j(\omega_k)\}_{j=0}^{J^{m,p}}$ via Scheme 3.0.1, and $\mathcal{S}_\rho^{m,p} := \{\widetilde{\rho}(t_j, \omega_k)\}_{k=1}^{M^q}$ based on $\mathcal{S}_{\boldsymbol{Y}}^{m,p}$ via (3.2.13) using $\{\boldsymbol{\xi}_{j+1}^{m,p}\}_{j=0}^{J^{m,p}-1}$ on $\mathcal{I}_\tau^{m,p}$. Set $J^{m,p+1} := J^{m,p}$, and $p := p + 1$.

(b) For $j = 0, \dots, J^{m,p} - 1$ do:

(i) Set $\mathtt{err}_{T,j+1}^{m,p+1} := |\widetilde{\rho}(t_{j+1})|(\tau_{m,p}^{j+1})^2$. If $\mathtt{err}_{T,j+1}^{m,p+1} \geqslant \delta$, set $J^{m,p+1} := J^{m,p} + 1$, $\mathcal{I}_\tau^{m,p+1} := \mathcal{I}_\tau^{m,p} \cup \{\tau_{m,p}^{j+1}/2\}$, and construct $\{\boldsymbol{\xi}_{j+1}^{m,p+1}\}_{j=0}^{J^{m,p}-1}$ by Brownian bridges based on $\{\boldsymbol{\xi}_{j+1}^{m,p}\}_{j=0}^{J^{m,p}-1}$ via Algorithm 3.1.1.

(c) If $J^{m,p+1} > J^{m,p}$, set $p := p + 1$.

(II) Compute $\mathbb{E}_{M^q}\big[g(\boldsymbol{Y}_T)\big]$, $\mathtt{err}_S^{q+1} := a\big(\sqrt{M^q}\big)^{-1} \mathrm{Var}_{M^q}\big[g(\boldsymbol{Y}_T)\big]$,

$$M^{q+1} := \min\left\{\left\lfloor (a \cdot \mathrm{Var}_{M^q}\left[g(\boldsymbol{Y}_T)\right])^2 \mathtt{Tol}_S^{-1} \right\rfloor, M^q M_{\max}\right\}.$$

Set $\delta := \mathtt{Tol}_T \cdot \big(\mathbb{E}[J^{M^q}]\big)^{-1}$, and $q := q + 1$.

Accept $\mathbb{E}_{M^q}\big[g(\boldsymbol{Y}_T)\big]$ as an approximation to $\mathbb{E}\big[g(\boldsymbol{X}_T)\big]$ since the computational error is bounded by $\mathtt{Tol}_\tau$, at least with probability $2\Phi(a) - 1$.

The computation of the stochastic flows up to the third order requires the solution of linear dual SDEs in dimension L^3, and to store a realization of $\boldsymbol{Y}$ and $\boldsymbol{W}$ for all time levels. This additional work is clearly a drawback, especially for high-dimensional problems ($L \gg 1$). Next to this, due to the use of Kolmogorov's backward equation, which is a partial differential equation (PDE), the practical applicability of the foregoing algorithms is restricted to SDE system of dimension $L \leqslant 3$.

The strategies in [STZ01] have the aim to obtain optimal meshes in the sense of (3.2.5) at terminal time T with a high probability. However, these strategies suffer from the fact that inequality (3.2.5) is not satisfied with a high probability for times $t \in [t_0, T)$; in particular, if the solution possesses possible discrete singularities or fast changes at several times (as in Figure 2.1), all indicators $\widetilde{\rho}$ must be re-calculated by simulating the trajectories $\boldsymbol{Y}$ again. Next to this, the use of a local high-dimensional density estimator is based solely on the realizations of $\boldsymbol{Y}^{1,j+1}$ and $\boldsymbol{Y}^{2,j+1}$, independent of the functional g, such that the high-dimensional approach can be regarded as a generalization of the strategies in [STZ01] which, in particular, lead to non adapted meshes with respect to the behavior of the functional g.

Chapter 4

Adaptivity for stochastic partial differential equations

We develop new space-time adaptive concepts based on local changes of empirical laws of approximate solutions of the general SPDE

$$\begin{aligned} \mathrm{d}X_t &= \mathscr{A}(X_t)\mathrm{d}t + \iota\sigma(X_t)\mathrm{d}W_t \qquad && \text{in } D_T := (0,T)\times D, \\ X_0 &= x_0 && \text{in } D, \end{aligned} \tag{4.0.1}$$

with proper boundary conditions. Here, $D \subset \mathbb{R}^d$, $d \in \{1,2,3\}$ is a bounded polyhedral domain, $0 < T < \infty$ denotes the terminal time, and $\mathscr{A}$ a non-linear second order differential operator; see (4.4.1) and (4.4.5). We denote by $W \equiv \{W_t; t \in [0,T]\}$ a Hilbert space-valued trace-class Wiener process; $\iota \in \mathbb{R}$ is later referred to as noise intensity, and σ is a diffusion operator.

For deterministic PDEs, there is a rich literature on adaptive methods to automatically generate space-time meshes which capture the structure of the solution and thus significantly reduce the computational effort. For low-dimensional SDEs, different concepts of time adaptivity have been proposed e.g. in [LMS07] to accurately approximate pathwise solutions, or e.g. in [STZ01] to approximate corresponding laws of iterates. The construction [MSTZ05, STZ01] of an adaptive mesh is based on an (asymptotic) *a posteriori* estimate, whose computation involves global, coupled random dual backward equations; conceptually, the construction rests on using Kolmogorov's backward equation, which restricts the practical applicability to low dimensional SDEs. The advantage, however, is a theoretical backup of this adaptive algorithm.

In this work, we sample local distributions to steer a space-time adaptive method for a finite element discretization of (4.0.1); see Section 4.5 for the used notation. For $j \in \mathbb{N}_0$, and a given random variable $Y^j \in L^2(\Omega; \mathbb{V}_h^j)$, we determine the new time step size $\tau^{j+1} := t_{j+1} - t_j > 0$ with the help of the following adaptive concepts:

- (Distance of subsequent empirical laws) We compute the distance $\mathtt{d}$ (either Hellinger $\mathtt{d}_{\mathtt{H}}$, Kullback-Leibler $\mathtt{d}_{\mathtt{KL}}$, total variation $\mathtt{d}_{\mathtt{TV}}$, or chi-square $\mathtt{d}_{\chi^2}$) of empirical approximations $\widehat{\mu}_\tau^{\mathfrak{s},j+1} := \widehat{\mathcal{L}}(Y^{\mathfrak{s},j+1})$ of laws $\mathcal{L}(Y^{\mathfrak{s},j+1})$ $\big(\mathfrak{s} \in \{\mathtt{1},\mathtt{2}\}\big)$, where the iterate $Y^{\mathtt{1},j+1} \in L^2(\Omega; \mathbb{V}_h^j)$ solves the j-th step of a discretization of problem (4.0.1) with local time step size $\tau^{\mathtt{1},j+1} := \tau^j$, while $Y^{\mathtt{2},j+1} \in L^2(\Omega; \mathbb{V}_h^j)$ results from using two time steps of this scheme with step size $\tau^{\mathtt{2},j+1} := \tau^j/2$, and extrapolation; cf. Figure 4.1. The empirical laws

$\{\widehat{\mu}_\tau^{\mathfrak{s},j+1};\mathfrak{s}\in\{1,2\}\}$ are assembled via the data-dependent partition $\widehat{\mathcal{P}}_{\tau;R_\tau}^{j+1}:=\bigcup_{r=1}^{R_\tau}\widehat{C}_{\tau;r}^{j+1}$ of the state space $\mathbb{R}^{L^j}$, where $L^j:=\dim\mathbb{V}_h^j$. The distance $\mathtt{d}(\widehat{\mu}_\tau^{1,j+1},\widehat{\mu}_\tau^{2,j+1})$ is then computed to steer refinement (or coarsening) of $\tau^{1,j+1}$, and thus eventually attain the new time step size $\tau^{j+1}>0$; see Sections 4.6.2 and 4.6.3 for further details.

- (Gradient based ZZ estimator in space) Once $\tau^{j+1}:=t_{j+1}-t_j>0$ has been found, elements K of the regular mesh $\mathcal{T}_h^j$ covering D are refined or coarsened depending on the distance $\mathtt{d}(\widehat{\mu}_{h;K}^{1,j+1},\widehat{\mu}_{h;K}^{2,j+1})$ of empirical laws $\widehat{\mu}_{h;K}^{1,j+1}\approx\mathcal{L}\big(|\nabla Y^{j+1}|_K\big)$ and $\widehat{\mu}_{h;K}^{2,j+1}\approx\mathcal{L}\big(|G_h(\nabla Y^{j+1})|_K\big)$ on each $K\in\mathcal{T}_h^j$; see Section 4.6.4.

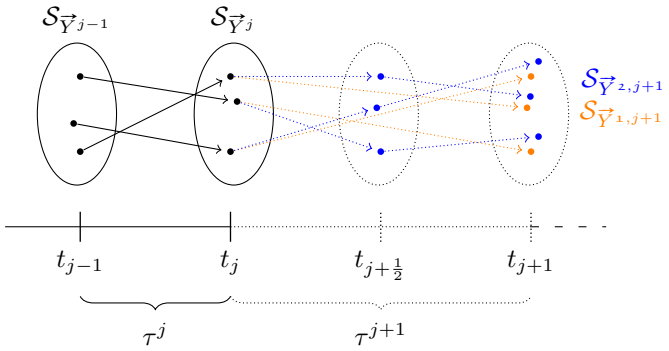


Figure 4.1: Computation of the new local time step size $\tau^{j+1}:=t_{j+1}-t_j>0$ based on the old one τ^j: Samples $\mathcal{S}_{\vec{Y}^{1,j+1}}$ and $\mathcal{S}_{\vec{Y}^{2,j+1}}$ computed by the same scheme, but with different time step sizes $\tau^{1,j+1}$ and $\tau^{2,j+1}$; see Algorithm 4.6.3 for details.

These empirical laws are of the form

$$\widehat{\mu}_\tau^{\mathfrak{s},j+1}:=\frac{1}{M_\tau}\sum_{r=1}^{R_\tau}\widehat{\nu}_{\tau;r}^{\mathfrak{s},j+1}\cdot\widehat{\xi}_{\tau;r}^{\mathfrak{s},j+1}\quad\mathfrak{s}\in\{1,2\},$$

where

$$\widehat{\xi}_{\tau;r}^{\mathfrak{s},j+1}:=\frac{1}{\widehat{\nu}_{\tau;r}^{\mathfrak{s},j+1}}\sum_{k\in\widehat{A}_r^{\mathfrak{s}}}\delta_{\vec{Y}^{\mathfrak{s},j+1}(\omega_k)}\quad\mathfrak{s}\in\{1,2\},$$

with $\vec{Y}^{\mathfrak{s},j+1}:=\vec{\Upsilon}_{\mathbb{V}_h^j}(Y^{\mathfrak{s},j+1})$ for $Y^{\mathfrak{s},j+1}\in L^2(\Omega;\mathbb{V}_h^j)$; see Section 4.5 for notation. Here,

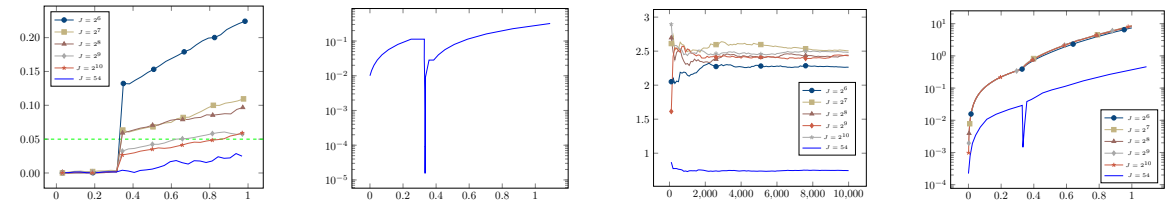
$$\widehat{A}_r^{\mathfrak{s}}:=\big\{k;\,\vec{Y}^{\mathfrak{s},j+1}(\omega_k)\in\widehat{C}_{\tau;r}^{j+1}\big\}\quad\mathfrak{s}\in\{1,2\}$$

denotes the associated index set of realizations in sample $\mathcal{S}_{\vec{Y}^{\mathfrak{s},j+1}}$ in the r-th cell $\widehat{C}_{\tau;r}^{j+1}\subset\mathbb{R}^{L^j}$, and $\widehat{\nu}_{\tau;r}^{\mathfrak{s},j+1}:=\#\widehat{A}_r^{\mathfrak{s}}$ the corresponding frequency; cf. Section 4.6.2. The histogram-based distance estimator for $\{\widehat{\mu}_\tau^{\mathfrak{s},j+1};\mathfrak{s}\in\{1,2\}\}$ then rests on a data-dependent partition $\widehat{\mathcal{P}}_{\tau;R_\tau}^{j+1}$ of $\mathbb{R}^{L^j}$ into $R_\tau\in\mathbb{N}$ many statistically equivalent disjoint cells $\{\widehat{C}_{\tau;r}^{j+1}\}_{r=1}^{R_\tau}$, which is generated via the 'nodal based' Binary Tree Cuboids (**BTC**) method [DGL96, Chapter 20] from [DP16]: the strategy here is to initiate a recursive subdivision of the initial sample set

$\mathcal{S}_{\vec{Y}^{1,j+1}} \cup \mathcal{S}_{\vec{Y}^{2,j+1}}$ by repeatedly identifying those spatial nodal points $\mathbf{x}_\ell \in D$, $\ell \in \{1, \dots, L^j\}$ for sub-samples where the empirical standard deviation $\widehat{\sigma}_\ell$ is largest; see Section 4.6.1 and Figure 4.8. Then, the proposal $\tau^{1,j+1}$ is refined when the probability of the (bootstrap) estimator $\mathtt{d}^\star(\widehat{\mu}_\tau^{1,j+1,\star}, \widehat{\mu}_\tau^{2,j+1,\star}) \geqslant 2\mathtt{Tol}_\tau$ is large, is accepted when the distance is approximately equal $\mathtt{Tol}_\tau$, or is coarsened in case $\mathtt{d}^\star(\widehat{\mu}_\tau^{1,j+1,\star}, \widehat{\mu}_\tau^{2,j+1,\star}) \leqslant \frac{1}{2}\mathtt{Tol}_\tau$ with high probability. For the computation of the frequencies $\{\widehat{\nu}_{\tau;r}^{\mathfrak{s},j+1}\}_{r=1}^{R_\tau}$ $\big(\mathfrak{s} \in \{\mathtt{1}, \mathtt{2}\}\big)$ we use $M_\tau \gg R_\tau$ realizations, which is a slight sharpening of the condition [Coc52] that at least ten realizations per cell is required to set M_τ. We illustrate these concepts for the following SDE from [MSTZ05, Example 5.1],

$$\mathrm{d}X_t = \frac{\mathbb{1}_{(1/3,1]}(t)}{2\sqrt{t - 1/3}} X_t \mathrm{d}t + X_t \mathrm{d}W_t \quad \forall\, t \in (0,1], \quad X_0 = 1; \tag{4.0.2}$$

note that $t \mapsto \mathbb{E}\big[|X_t|\big]$ fastly changes in the vicinity of $t = 1/3$, and thus favors a higher resolution by a refined mesh in its neighborhood; see Figure 4.2(b).



(a) $t_j \mapsto \big|\mathbb{E}[|X_{t_j}|] - \mathbb{E}_{M_\tau}\big[|Y^j|\big]\big|$ (b) $t_j \mapsto \tau^j$ (c) $M_\tau \mapsto \mathbb{E}_{M_\tau}\Big[\max_{t_j} \big||Y^j| - \mathbb{E}_{M_\tau}\big[|Y^j|\big]\big|^2\Big]$ (d) $t_j \mapsto \mathrm{Var}_{M_\tau}\big[|Y^j|\big]$

Figure 4.2: (Example 4.0.1 for $\mathtt{d} = \mathtt{d}_{\mathtt{H}}$, $\mathtt{Tol}_\tau = 0.05$ (- - - -), $\mathtt{N}_\tau = 10$, and $R_\tau = 2^6$, $M_\tau = 10^5$) (a) Error for uniform (, , , ,) vs. adaptive (——) time meshes via Algorithm 4.6.3, and (b) corresponding adaptive time step size. (c) Evolution of $M_\tau \mapsto \mathbb{E}_{M_\tau}\Big[\max_{t_j} \big||Y^j| - \mathbb{E}_{M_\tau}\big[|Y^j|\big]\big|^2\Big]$, and (d) empirical variance $t_j \mapsto \mathrm{Var}_{M_\tau}\big[|Y^j|\big]$ for uniform vs. adaptive meshes.

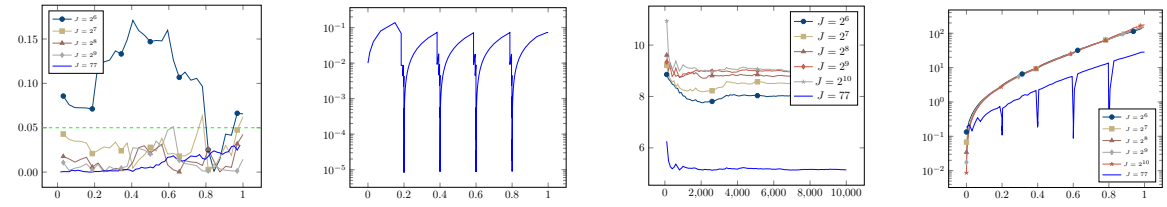
Example 4.0.1. *Figure 4.2(a) displays errors for a drift-implicit Euler discretization of* (4.0.2) *on uniform vs. adaptive time meshes ($M_\tau = 10^5$ MC simulations). A uniform mesh with $J = 2^{10}$ time steps is needed vs. an adaptively refined mesh with $J = 54$ time steps via Algorithm 4.6.3 (——) to stay below the given error threshold (- - - -). The adapted local time step size $\{\tau^j\}_j$ rapidly decays near $t = 1/3$, and afterwards coarsens again (see Figure 4.2(b)); compare also with [MSTZ05, Figure 2]. Figure 4.2(c) shows a significantly reduced empirical variance on adaptive meshes $\big(\mathbb{E}_{M_\tau}\big[|Y^j|\big] := \frac{1}{M_\tau}\sum_{k=1}^{M_\tau} |Y^j(\omega_k)|\big)$; Figure 4.2(d) displays the behavior of the empirical variance $t_j \mapsto \mathrm{Var}_{M_\tau}\big[|Y^j|\big] := \mathbb{E}_{M_\tau}\Big[\big||Y^j| - \mathbb{E}_{M_\tau}\big[|Y^j|\big]\big|^2\Big]$ to attribute this observation to a smaller empirical variance near $t = 1/3$ via Algorithm 4.6.3, as opposed to uniform grids.*

These adaptive concepts can be easily generalized to systems of SDEs.

Example 4.0.2. *Consider*

$$\mathrm{d}\boldsymbol{X}_t = \big[-\boldsymbol{J} + \boldsymbol{D}(t)\big]\boldsymbol{X}_t \mathrm{d}t + \big[1 + |\boldsymbol{X}_t|\big]\mathrm{d}\boldsymbol{W}_t \quad \forall\, t \in (0,1], \quad \boldsymbol{X}_0 = (1,1,1,1)^T, \tag{4.0.3}$$

with $\boldsymbol{J} = 16 \cdot \mathrm{tridiag}[-1, 2, -1] \in \mathbb{R}^{4\times 4}$, *and* $\boldsymbol{D}(t) = \mathrm{diag}[\gamma_1(t), \dots, \gamma_4(t)] \in \mathbb{R}^{4\times 4}$, *where* $\gamma_i(t) := \mathbb{1}_{(\beta_i,1]}(t)/2\sqrt{t-\beta_i+10^{-8}}$ *and* $\beta_i := i/5$. *Here,* $t \mapsto \mathbb{E}\big[|\boldsymbol{X}_t|\big]$ *rapidly changes in the vicinity of times* $t \in \{\beta_i\}_{i=1}^4$*; see Figure 4.3(b). Computational studies for other choices* $\mathtt{d} \in \{\widetilde{\mathtt{d}}_{\mathtt{KL}}, \mathtt{d}_{\mathtt{TV}}\}$ *(see Section 4.6.2) show comparable results.*



(a) $t_j \mapsto \big|\mathbb{E}[|\boldsymbol{X}_{t_j}|] - \mathbb{E}_{M_\tau}[|\boldsymbol{Y}^j|]\big|$ (b) $t_j \mapsto \tau^j$ (c) $M_\tau \mapsto \mathbb{E}_{M_\tau}\big[\max_{t_j} \big||\boldsymbol{Y}^j| - \mathbb{E}_{M_\tau}[|\boldsymbol{Y}^j|]\big|^2\big]$ (d) $t_j \mapsto \mathrm{Var}_{M_\tau}[|\boldsymbol{Y}^j|]$

Figure 4.3: (Example 4.0.2 for $\mathtt{d} = \mathtt{d}_{\mathtt{H}}$, $\mathtt{Tol}_\tau = 0.05$ (----), $\mathtt{N}_\tau = 10$, and $R_\tau = 2^6$, $M_\tau = 10^5$) (a) Error for uniform (, , , ,) vs. adaptive (——) time meshes via Algorithm 4.6.3, and (b) corresponding adaptive time step size. (c) Evolution of $M_\tau \mapsto \mathbb{E}_{M_\tau}\Big[\max_{t_j} \big||\boldsymbol{Y}^j| - \mathbb{E}_{M_\tau}\big[|\boldsymbol{Y}^j|\big]\big|^2\Big]$, and (d) empirical variance $t_j \mapsto \mathrm{Var}_{M_\tau}\big[|\boldsymbol{Y}^j|\big]$ for uniform vs. adaptive meshes.

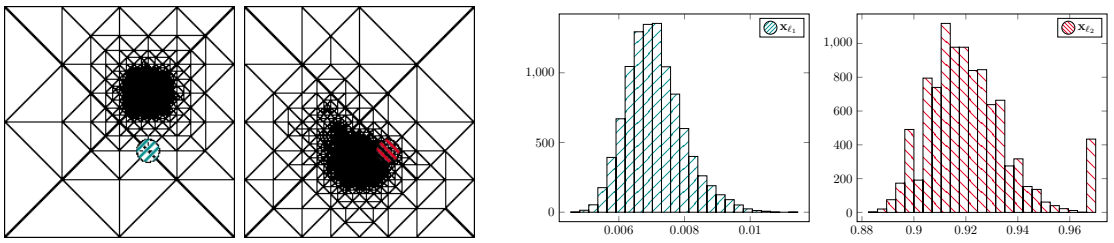
Larger systems of SDEs occur after a spatial discretization of SPDEs, where finer partitions $\widehat{\mathcal{P}}_{\tau;R_\tau}^{j+1}$ are needed to properly resolve the high-dimensional state space $\mathbb{R}^{L^j}$. The simulations of SPDE (4.0.1) in Example 4.7.1 below e.g. suggest $R_\tau \approx 2^{12}$ for a stable selection of variable step sizes; see Figure 4.13. Next to the noise intensity ι in (4.0.1), a proper choice R_τ also depends on the drift operator: for the SPDE in Example 4.7.1 with convection dominated drift, steep spatial gradients inside diffuse layers need to be properly resolved, which requires a fine partition of (a large subset of) $\mathbb{R}^{L^j}$. In contrast, comparative studies for the drift operator $\mathscr{A}(X) = \Delta X$ in (4.0.1) evidence a much smaller number $R_\tau \approx 2^7$ to reliably detect changes in subsequent empirical distributions.

Next to temporal adaptivity, we use space adaptivity to discretize (4.0.1). For this purpose, we adapt the idea of the ZZ-estimator [ZZ92] to possibly refine/coarsen an element $K \in \mathcal{T}_h^j$, by measuring the distance $\mathtt{d}^\star(\widehat{\mu}_{h;K}^{1,j+1}, \widehat{\mu}_{h;K}^{2,j+1})$ of the empirical measures

$$\widehat{\mu}_{h;K}^{1,j+1} := \widehat{\mu}_{|\nabla Y^{j+1}|_K},$$
$$\widehat{\mu}_{h;K}^{2,j+1} := \widehat{\mu}_{|G_h(\nabla Y^{j+1})|_K},$$

see Figure 4.4. Therefore, a data-dependent partition $\widehat{\mathcal{P}}_{h;R_h;K}^{j+1} := \bigcup_{r=1}^{R_h} \widehat{C}_{h;r;K}^{j+1}$ of $\mathbb{R}_0^+$ in $R_h \in \mathbb{N}$ many cells is generated via (**BTC**) for each $K \in \mathcal{T}_h^j$, on which the above two empirical measures are composed. The distance of these measures then essentially steers the re-distribution of spatial nodal points, leading e.g. to a better resolution of diffuse layers in Figure 4.4(a) in the case of the convection-dominated SPDE given in Example 4.7.1. The comparative studies in Figure 4.4(a) and 4.16(a) evidence that the choice of the distance $\mathtt{d} \in \{\mathtt{d}_{\mathtt{H}}, \mathtt{d}_{\mathtt{KL}}, \mathtt{d}_{\mathtt{TV}}\}$ matters, and that $\mathtt{d}_{\mathtt{TV}}$ should be given preference; see also Table 4.5. However, spatial adaptivity not only enhances the resolution of the computation (see Figure 4.15), but also further reduces the empirical variance of estimators, which was already an outcome of adaptivity

in time; see Figure 4.14. This effect thus lowers the number of required MC simulations M_τ (resp. M_h), as well as of R_τ (resp. R_h), if compared to uniform space-time meshes. So



(a) $\mathcal{T}_h^{j+1}$ at $t_{j+1} \in \{0.25, 3.75\}$ **(b)** Histogram of $\mathtt{d}_{\mathtt{H}}^\star(\widehat{\mu}_{h;K}^{\mathfrak{1},j+1,\star}, \widehat{\mu}_{h;K}^{\mathfrak{2},j+1,\star})$ for $K \in \mathcal{T}_h^{j+1}$

Figure 4.4: (Example 4.7.1 for $\iota = 0.3$, as well as $\mathtt{d} = \mathtt{d}_{\mathtt{H}}$, $\mathtt{Tol}_h = 0.05$, $\mathtt{N}_h = 1$, and $R_h = 2^5$, $M_h = 10^3$ for each $K \in \mathcal{T}_h^{j+1}$): (a) Snapshots of spatial meshes $\mathcal{T}_h^{j+1}$ at evaluated times $t_{j+1} \in [0, 2\pi]$ with (b) corresponding histograms of $\mathtt{d}^\star(\widehat{\mu}_{h;K}^{\mathfrak{1},j+1,\star}, \widehat{\mu}_{h;K}^{\mathfrak{2},j+1,\star})$ with respect to $K \in \mathcal{T}_h^{j+1}$ containing two fixed points $\mathbf{x}_{\ell_1}$ (◎) and $\mathbf{x}_{\ell_2}$ (◎) as indicated in (a).

far, the number of cells R_τ (resp. R_h) was fixed. Evidently, the required resolution of an image $\vec{Y}^{j+1}[\Omega]$ (resp. $|\nabla Y^{j+1}|_K[\Omega]$) may also change in time (respectively in space), which, in turn, affects the required numbers M_τ (resp. M_h). The computational studies in Section 4.7 (see Figure 4.17) favor a heuristic strategy (4.7.5) to generate a non-constant sequence $\{\widehat{R}_\tau^{j+1}\}$ (resp. $\{\widehat{R}_h^{j+1}\}$); corresponding partitions $\{\widehat{\mathcal{P}}_{\tau;\widehat{R}_\tau^{j+1}}^{j+1}\}$ (resp. $\{\widehat{\mathcal{P}}_{h;\widehat{R}_h^{j+1}}^{j+1}\}$) are then determined by parallel computation using OpenMP [DM98]; see Table 4.4.

The overall computational effort per time step is distributed across different tasks: it starts with the computation of initial M_τ-samples $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$ $\big(\mathfrak{s} \in \{\mathfrak{1}, \mathfrak{2}\}\big)$ of complexity [GN88] $\mathcal{O}\big(M_\tau (L^j)^{\frac{3}{2}}\big)$. To build the data-dependent partition $\widehat{\mathcal{P}}_{\tau;R_\tau}^{j+1}$ to initiate time adaptivity causes low storage requirements $\mathcal{O}\big(R_\tau(L^j + \log(R_\tau))\big)$, since $R_\tau \ll M_\tau$. We then sample again — and refer below again to the new samples as $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$ $\big(\mathfrak{s} \in \{\mathfrak{1}, \mathfrak{2}\}\big)$ — , to afterwards generate an amount $B_\tau \in \mathbb{N}$ of bootstrap M_τ-samples $\{\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}^{b,\star}\}_{b=1}^{B_\tau}$ $\big(\mathfrak{s} \in \{\mathfrak{1}, \mathfrak{2}\}\big)$ for the histogram based bootstrap distance estimator $\mathtt{d}^\star$ (see Figure 4.4(b)) of the related empirical probability measures $\{\widehat{\mu}_\tau^{\mathfrak{s},j+1,b,\star}\}_{b=1}^{B_\tau}$ $\big(\mathfrak{s} \in \{\mathfrak{1}, \mathfrak{2}\}\big)$ on $\widehat{\mathcal{P}}_{\tau;R_\tau}^{j+1}$. We adjust the local time step size τ^{j+1} then based upon these estimators; see Section 4.6. This part requires no additional storage effort, since bootstrap samples $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}^{b,\star}$ are internally identified via associated index sets $\mathfrak{I}^{\mathfrak{s},b,\star}$ (see Section 4.3); for example, to provide a single sample of size $M_\tau = 10^5$ with $\mathcal{T}_h^j = \mathcal{T}_h^0$ for $h^0 = 2^{-6}$ (Example 4.7.1) requires approximatively 10 minutes of computational time, whereas its generation via bootstrap only takes approximatively 4 seconds. Corresponding considerations hold for space adaptivity and each element $K \in \mathcal{T}_h^j$. Next to bootstrap, we adaptively choose the statistical parameters (M_τ, B_τ, R_τ) (analogously in space) to save computer resources (Section 4.7.1.3): this strategy particularly benefits from differing solution structures resp. scales in space and time (e.g. uniformly $M_\tau = 10^5$ for Example 4.7.1, as opposed to adaptive $M_\tau \approx 10^2$ in Figure 4.17(b)), which e.g. in this example reduces the overall computation by 20%, preserving accuracy. Our simulations show that the combination of space-time and statistical adaptivity with the bootstrap method leads to a practical, accurate (see Figure 4.15), and efficient overall method.

These concepts are tested for two prototype problems: we start with a convection dominated SPDE with linear drift in Section 4.4.1. In Section 4.5.1 we propose a stable discretization by a stochastic version of the SUPG method (see Section 4.5.1) to avoid spurious solutions which appear for the classical Galerkin scheme; see Figure 4.10. It turns out that the SUPG method not only yields a more accurate (see Figure 4.11), stable approximation, see Lemma 4.5.2, but also reduces the empirical variance of related estimators. An SPDE with non-linear drift is the stochastic harmonic map flow to the sphere $\mathbb{S}^2$, for which blow-up is known in the presence of super-critical initial data. For $D \subset \mathbb{R}^d$, $d \in \{2, 3\}$ this SPDE has a weak martingale solution [Hoc16] rather than a probabilistically strong solution as for the former problem; however, our concept of space-time adaptivity is based on distributions rather than single trajectories, and therefore is applicable also here. The simulations in Section 4.7 illustrate refinement of space-time meshes close to singular behaviors of the solution, and a fast coarsening again beyond; concomitantly, the adaptively chosen number of cells R_τ^j (resp. R_h^j) grows close to the (discrete) blow-up time, and rapidly decays beyond it again. A corresponding dependence is observed for M_τ^j (resp. M_h^j).

4.1 Adaptive finite element methods

In the context of the adaptive finite element method (AFEM) there are three main techniques for the error reduction. The 'h-version' of AFEM modifies the mesh size (h-refinement) of the elements while keeping the polynomial degrees fixed. The 'p-version' of AFEM adjusts the polynomial degrees in the element (p-refinement) while keeping the mesh size of the elements fixed. The 'r-version' of AFEM performs a relocation (r-refinement) of the spatial nodes of the original mesh by distributing them in a more appropriate fashion. This procedure is quite time consuming and can produce highly distorted element geometries. According to approximation theory, p-refinement should be performed on elements in which the solution to the PDEs is smooth, and h-refinement should be performed on elements in which the solution is non-smooth [MW01], as for solutions that may exhibit localized singular behavior. Therefore, in this thesis, we are concerned with evaluating the error and its reduction by h-refinement.

Adaptive methods, and more precisely, adaptive methods for finite element approximations have been extensively used and studied in the last decades with emphasis on both the theoretical and computational aspects of the methods. Concerning the theoretical aspects, the research in the area starts with the pioneering work by [BR78] and continues with studies devoted to the Residual Based method (see [Ver94]). In this approach certain local residuals are evaluated and then the *a posteriori* error indicator is obtained by solving local Dirichlet or Neumann problems, taking the residuals as data [BR78, BS93]. Another approach of the method uses the Galerkin orthogonality, *a priori* interpolation estimates, and global stability in order to get error estimators in global $\mathbb{L}^2$- and $\mathbb{H}^1$-norms; see e.g. [EEHJ96]. Furthermore, solving appropriate dual backward problems leads to error estimators controlling various kinds of error functionals [BR96]. Solving finite element problems in an enriched function space (by hierarchical bases) gives rise to the so-called 'Hierarchical Based error estimators' [BS93]. There are also error estimators that control the error or its gradient in the $\mathbb{L}^\infty$-norm, which are based on optimal *a priori* estimates for the error in maximum norm. However, the above strategies have in common that the solution must be sufficiently smooth.

Another error indicator, which is widely (and in most cases heuristically) used in many adaptive finite element codes, is the Zienkiewicz-Zhu (ZZ) estimator [ZZ92]. This estimator is based on post-processing of the computed solution gradient in order to get a better approximation, which is later used instead of the exact solution to bound the error, which is part of the next Section 4.1.1. Some analysis of the method can be found in [Rod94] and references therein.

4.1.1 Gradient recovery-based estimator

The Zienkiewicz-Zhu (ZZ) estimator [ZZ92] is based on smoothing in some way the discontinuous finite element gradient ∇Y^j to obtain some continuous 'recovered gradient' $G_h(\nabla Y^j) \in \mathbb{V}_h^j$ for some $j \in \mathbb{N}_0$. The hope behind this idea is that the smooth recovered gradient could be a much better approximation to ∇X_{t_j} than ∇Y^j itself, i.e., suppose, for a fixed element $K \in \mathcal{T}_h^j$,

$$\|\nabla X_{t_j} - G_h(\nabla Y^j)\|_{\mathbb{L}^2(K)} \leqslant \gamma \|\nabla X_{t_j} - \nabla Y^j\|_{\mathbb{L}^2(K)} \quad \gamma \in (0,1].$$

It holds

$$\Big| \|\nabla X_{t_j} - \nabla Y^j\|_{\mathbb{L}^2(K)} - \|\nabla X_{t_j} - G_h(\nabla Y^j)\|_{\mathbb{L}^2(K)} \Big| \leqslant \|\nabla Y^j - G_h(\nabla Y^j)\|_{\mathbb{L}^2(K)}.$$

Thus the two-sided estimate of the true error

$$\frac{1}{1+\gamma} \|\nabla Y^j - G_h(\nabla Y^j)\|_{\mathbb{L}^2(K)} \leqslant \|\nabla X_{t_j} - \nabla Y^j\|_{\mathbb{L}^2(K)} \leqslant \frac{1}{1-\gamma} \|\nabla Y^j - G_h(\nabla Y^j)\|_{\mathbb{L}^2(K)}$$

is valid and the local ZZ estimator is

$$\|\nabla Y^j - G_h(\nabla Y^j)\|_{\mathbb{L}^2(K)} \quad \forall\, K \in \mathcal{T}_h^j. \tag{4.1.1}$$

Originally, the ZZ-estimator was proposed to approximate solution of the Poisson problem. In this case, the computation of $G_h(\nabla Y^j)$ with respect to the $\mathbb{L}^2$-scalar product would be as expensive as the approximate solution of solving the problem itself. Thus, the authors in [ZZ92] proposed to use reduced integration via the h-scalar product (see [BBNP14b, Section 2.1]), such that $G_h(\nabla Y^j)$ can be explicitly written in the form

$$G_h(\nabla Y^j)(\mathbf{x}) := \sum_{K \in w_{\mathbf{x}}} \frac{|K|}{|w_{\mathbf{x}}|} \nabla Y^j\big|_K \quad \forall\, \mathbf{x} \in D,$$

where $w_{\mathbf{x}} := \bigcup \{K \in \mathcal{T}_h^j : \mathbf{x} \in \overline{K}\}$ denotes the patch associated to $\mathbf{x} \in D$. Furthermore, $|w_{\mathbf{x}}| := \sum_{K \in w_{\mathbf{x}}} \mathrm{vol}(K)$. That is, $G_h(\nabla Y^j)(\mathbf{x})$ is simply a weighted average of ∇Y^j on the elements $K \in w_{\mathbf{x}}$.

The advantages of the ZZ estimator are that it is possible to obtain the error estimation for the whole domain, for element sub-domains, and even a point-wise definition for the estimation. The ZZ estimator can be readily implemented in a finite element code and it is computationally effective with easy extension to vector-valued functions. Since no assumptions are needed concerning the regularity of the solution, the local ZZ estimator in (4.1.1) can be used for various problems, which makes it computationally attractive.

4.2 Partitioning

This section is concerned about binary tree based data-dependent classification strategies of the state space $\mathbb{R}^L$ ($L \in \mathbb{N}$). The consistency of tree-structured classification and regression was investigated in [GO80] in a general framework, and was extended by [BFOS84].

Let $\vec{Y}$ be a $\mathbb{R}^L$-valued random variable, and $\mu := \mathcal{L}(\vec{Y}) \in \mathscr{P}(\mathbb{R}^L)$ be the related probability measure of $\vec{Y}$ on $\big(\mathbb{R}^L, \mathcal{B}(\mathbb{R}^L)\big)$, where $\mathcal{B}(\mathbb{R}^L)$ is the Borel σ-algebra on $\mathbb{R}^L$, as well as $\widehat{\mu}$ its empirical counterpart on the sample $\mathcal{S}_{\vec{Y}} := \{\vec{Y}(\omega_k)\}_{k=1}^M$, for $M \in \mathbb{N}$. Below, we assume that a smooth density $\rho : \mathbb{R}^L \to \mathbb{R}$ with respect to Lebesgue measure λ of the law of $\vec{Y}$ exists.

4.2.1 Consistent partitioning

In machine learning, see e.g. [Das11], classification is the problem of identifying the region of the state space $\mathbb{R}^L$ to which a new realization of $\vec{Y}$ belongs to. One way to define such a classification rule $\widehat{\mathcal{P}}_{M;R} : \mathbb{R}^L \to \{1, \ldots, R\}$, later referred to as partition, is to define a finite collection $\{\widehat{C}_r\}_{r=1}^R$ of Borel-measurable subsets of $\mathbb{R}^L$, referred to as cells, each of them defined as level set $\widehat{C}_r := \big\{\vec{y} \in \mathbb{R}^L;\, \widehat{\mathcal{P}}_{M;R}(\vec{y}) = r\big\}$, such that $\mathbb{R}^L = \bigcup_{r=1}^R \widehat{C}_r$.

Definition 4.2.1. *Given an M-sample $\mathcal{S}_{\vec{Y}} := \{\vec{Y}(\omega_k)\}_{k=1}^M$, with $M \in \mathbb{N}$. A classification rule $\widehat{\mathcal{P}}_{M;R}$ is a partition of the state space $\mathbb{R}^L$ of $\vec{Y}$ into $R \in \mathbb{N}$ cells, i.e.*

$$\widehat{\mathcal{P}}_{M;R} \equiv \widehat{\mathcal{P}}_{M;R}\big(\mathcal{S}_{\vec{Y}}\big) = \big\{\widehat{C}_r\big\}_{r=1}^R, \quad \mathring{\widehat{C}}_r \cap \mathring{\widehat{C}}_s = \emptyset, \quad r \neq s, \quad \forall\, r, s \in \{1, \ldots, R\},$$

such that $\mathbb{R}^L = \bigcup_{r=1}^R \widehat{C}_r$, and for every $\vec{x} \in \widehat{C}_r$, $r \in \{1, \ldots, R\}$ the predicted class is r.

Finding a classifier $\widehat{\mathcal{P}}_{M;R}$ that performs optimally is a fundamental problem in non-parametric statistics. For M sufficiently large, by using Bayes' formula (law of total probability),

$$\mathbb{P}\big[\widehat{\mathcal{P}}_{M;R}(\vec{Y})\big] = \sum_{r=1}^R \mathbb{P}\big[\widehat{\mathcal{P}}_{M;R}(\vec{Y}) \in \widehat{C}_r\big] \cdot \mathbb{P}\big[\widehat{C}_r\big],$$

with $\mathbb{P}\big[\widehat{C}_r\big] := \int_{\widehat{C}_r} \rho(\mathbf{x}) \mathrm{d}\lambda(\mathbf{x})$, $r \in \{1, \ldots, R\}$. Thus, the probability of misclassification is minimized by selecting the cell $\widehat{C}_r \in \widehat{\mathcal{P}}_{M;R}$, $r \in \{1, \ldots, R\}$ satisfying the inequality

$$\mathbb{P}\big[\widehat{\mathcal{P}}_{M;R}(\vec{Y}) \in \widehat{C}_r\big] \cdot \mathbb{P}\big[\widehat{C}_r\big] \geqslant \mathbb{P}\big[\widehat{\mathcal{P}}_{M;R}(\vec{Y}) \in \widehat{C}_s\big] \cdot \mathbb{P}\big[\widehat{C}_s\big], \quad s \neq r, \quad \forall\, s \in \{1, \ldots, R\}.$$

This is the fundamental decision formula to obtain an optimally performing classification rule, which minimizes the probability of misclassification. Let

$$\widehat{\mathcal{L}}_{M;R} \equiv \widehat{\mathcal{L}}\big(\widehat{\mathcal{P}}_{M;R}(\vec{Y})\big) := \mathbb{P}\big[\widehat{\mathcal{P}}_{M;R}(\vec{Y}) \neq \widehat{C}_r\big]$$

be the risk (probability of misclassification) of the corresponding classifier $\widehat{\mathcal{P}}_{M;R}$. There is a best possible classification rule, $\mathcal{P}_R^\star$, which is defined by

$$\mathcal{P}_R^\star \equiv \mathcal{P}_R^\star\big(\vec{Y}\big) := \underset{\mathcal{P}_R\,:\,\mathbb{R}^L \to \{\widehat{C}_r\}_{r=1}^R}{\arg\min} \mathbb{P}\big[\mathcal{P}_R\big(\vec{Y}\big) \neq \widehat{C}_r\big].$$

Here, $\mathcal{P}_R^\star$ depends upon the distribution of $\vec{Y}$. The problem of finding $\mathcal{P}_R^\star$ is Bayes' problem, and the classifier $\mathcal{P}_R^\star$ is called the Bayes classifier (or the Bayes rule) which minimizes this risk. The minimal probability of error is called the Bayes error and is denoted by $\mathcal{L}_R^\star := \mathcal{L}\big(\mathcal{P}_R^\star\big)$. The natural question is then: What is the best we can expect from the classifier $\widehat{\mathcal{P}}_{M;R}$? Since it is not possible to build a classifier that makes no errors, the accuracy of the classifier depends crucially on how separated the cells are with respect to $\mathcal{S}_{\vec{Y}}$. In accordance with [DGL96], a classification rule is consistent if it minimizes the expected cost of misclassification, that is,

$$\lim_{M\to\infty} \mathbb{E}\big[\widehat{\mathcal{L}}_{M;R}\big] = \mathcal{L}_R^\star.$$

Since $\widehat{\mathcal{L}}_{M;R}$ is a random variable which takes values in $[\mathcal{L}_R^\star, 1]$ due to its definition, Markov's inequality implies convergence in probability

$$\lim_{M\to\infty} \sup_{t>0} \mathbb{P}\Big[\big|\widehat{\mathcal{L}}_{M;R} - \mathcal{L}_R^\star\big| > t\Big] = 0. \tag{4.2.1}$$

Following [DGL96, Chapter 6], the weak and strong consistency of a classification rule are defined as follows.

Definition 4.2.2 (Weak and strong consistency)**.** *A classification rule $\widehat{\mathcal{P}}_{M;R}$ is weak consistent (or asymptotically Bayes risk efficient) for a certain law $\mathcal{L}(\vec{Y})$ if*

$$\lim_{M\to\infty} \mathbb{E}\big[\widehat{\mathcal{L}}_{M;R}\big] = \mathcal{L}_R^\star,$$

and strongly consistent if

$$\mathbb{P}\left[\lim_{M\to\infty} \widehat{\mathcal{L}}_{M;R} = \mathcal{L}_R^\star\right] = 1.$$

To deduce consistency of data-dependent partitions requires two crucial conditions of the given partition $\widehat{\mathcal{P}}_{M;R}$: first, cells should be small enough to resolve spatial changes of the distribution accurately. On the other hand, cells should be large enough to contain a large number of realizations such that Monte-Carlo estimation is effective; cf. [DGL96, Theorem 6.1].

A general $\mathbb{P}$-almost sure L^1-consistency for histogram rules with data-dependent partitions is obtained by extending the Vapnik-Chervonenkis inequality [VC15] to families of partitions from families to sets (cf. [DGL96, Theorem 21.1]).

Let Π be a (possibly infinite) family of partitions of $\mathbb{R}^L$, and $\mathfrak{r} \in \mathbb{R}_0^+$. Furthermore, let $\widehat{\mathcal{P}}^\mathfrak{r}$ be a partition of the ball $B_\mathfrak{r} := \big\{\vec{x} \in \mathbb{R}^L \,\big|\, |\vec{x}| \leqslant \mathfrak{r}\big\}$, whose cells are obtained by intersecting the cells of $\widehat{\mathcal{P}}_{M;R}$ with $B_\mathfrak{r}$, i.e., $\widehat{\mathcal{P}}^\mathfrak{r} := \bigcup_{r=1}^R \big\{\widehat{C}_r \cap B_\mathfrak{r}\big\}$. We denote by $\Pi^\mathfrak{r} := \big\{\widehat{\mathcal{P}}^\mathfrak{r} \,:\, \widehat{\mathcal{P}} \in \Pi\big\}$ the family of partitions of $B_\mathfrak{r}$ obtained by restricting members of Π to $B_\mathfrak{r}$. For each $\mathfrak{r} \in \mathbb{R}_0^+$, define the class $\mathcal{A}^\mathfrak{r}$ of sub-sets of $\mathbb{R}^L$ by

$$\mathcal{A}^\mathfrak{r} := \big\{A \in \mathcal{B}(\widehat{\mathcal{P}}^\mathfrak{r});\, \widehat{\mathcal{P}}^\mathfrak{r} \in \Pi^\mathfrak{r}\big\},$$

and define

$$\Delta_M\big(\Pi^\mathfrak{r}\big) := \mathscr{S}(\mathcal{A}^\mathfrak{r}, M),$$

where $\mathscr{S}$ is the shatter coefficient [DGL96, Chapter 13] of $\mathcal{A}^\mathfrak{r}$, i.e., the number of different classifiers in the class $\mathcal{A}^\mathfrak{r}$. Here, $\mathcal{A}^\mathfrak{r}$ is thus the set of all sets which can be obtained as union of cells of some partition of $B^\mathfrak{r}$ in the collection $\Pi^\mathfrak{r}$. Then the following theorem is a consequence of the Vapnik-Chervonenkis inequality, which is a powerful generalization of the Glivenko-Cantelli theorem.

Theorem 4.2.3. *Given an M-sample $\mathcal{S}_{\vec{Y}} = \{\vec{Y}(\omega_k)\}_{k=1}^{M}$ with $M \in \mathbb{N}$, and measures μ, $\widehat{\mu}$. Let Π be a family of partitions of $\mathbb{R}^L$. Then for every $\mathfrak{r} \in \mathbb{R}^+$ and $t > 0$,*

$$\mathbb{P}\left[\sup_{\widehat{\mathcal{P}}^{\mathfrak{r}} \in \Pi^{\mathfrak{r}}} \sum_{\widehat{C}_r \in \widehat{\mathcal{P}}^{\mathfrak{r}}} \left|\mu\big[\widehat{C}_r\big] - \widehat{\mu}\big[\widehat{C}_r\big]\right| > t\right] \leqslant 8\Delta_M\big(\Pi^{\mathfrak{r}}\big) \exp\left\{-\frac{Mt^2}{512}\right\} + \exp\left\{-\frac{Mt^2}{2}\right\}.$$

We will use a special application of Theorem 4.2.3 (cf. [DGL96, Corollary 21.1]).

Corollary 4.2.4. *Given an M-sample $\mathcal{S}_{\vec{Y}} = \{\vec{Y}(\omega_k)\}_{k=1}^{M}$ with $M \in \mathbb{N}$, and measures μ, $\widehat{\mu}$. Let $\widehat{\mathcal{P}}_{M;R} = \{\widehat{C}_r\}_{r=1}^{R}$ be a partition of $\mathbb{R}^L$, and $\mathfrak{r} \in \mathbb{R}_0^+$. If*

$$\lim_{M\to\infty} \frac{\log\big(\Delta_M(\Pi^{\mathfrak{r}})\big)}{M} = 0,$$

then

$$\lim_{M\to\infty} \sup_{t>0} \mathbb{P}\left[\sum_{r=1}^{R} \left|\mu\big[\widehat{C}_r\big] - \widehat{\mu}\big[\widehat{C}_r\big]\right| > t\right] = 0.$$

Several properties of shatter coefficients $\mathscr{S}$ are collected in [DGL96, Chapter 13]. We refer to [Nob96, LN96] for inequalities of shatter coefficients $\mathscr{S}$ in the case of nearest-neighbor and tree-structured partitions as discussed in Section 4.2.2.

A natural method of estimating local properties of data in non-parametric statistics is to partition the space of observations into cells and then compute statistics locally within each cell. This leads to histogram estimates of an unknown density (and consequently distribution) and to partition-based classification rules. The simplest histogram methods partition the state space into congruent intervals or cubes whose size and position depends on the number of available data points, but not on the data itself. These methods provide estimates that are consistent, regardless of the underlying distribution of the given data. In [AJ76], the author gave the necessary and sufficient conditions under which a sequence of regular partitions gives rise to L^1-consistent estimates for every density; see also [DG85] and references therein. Statistical practice suggests that the histogram estimators based on data-dependent partitions will provide a better performance than those based on a fixed sequence of partitions. Theoretical evidence for this superiority was given by [Sto85]. The simplest data-dependent partitioning methods are based on statistically equivalent blocks [And66], in which each cell contains the same number of points. Many other data-dependent partitioning schemes have been introduced in the literature [Dev88]. For an overview of the theoretical aspects of the estimation of distribution functions and densities, we refer to [DG02].

Existing conditions for the consistency of histogram classification and density estimation using data-dependent partitions require significant restrictions. The conditions derived in [BFOS84] for consistent classification require that each cell of every partition belongs to a fixed Vapnik-Chervonenkis class of sets and that every cell must contain at least $R \equiv R(M)$ points, where $R/\log(M) \to \infty$ as $M \to \infty$. The authors in [CZ87, ZKC90] restrict their attention to density estimates based on rectangular partitions. General sufficient conditions for the almost sure L^1-consistency of histogram classification and density estimates that are based on data-dependent partitions were given in [LN96]. Analogous conditions for the consistency of histogram regression estimates are addressed in [Nob96].

4.2.2 Multivariate partitioning

Let $\mathcal{S}_{\vec{Y}} := \{\vec{Y}(\omega_k)\}_{k=1}^M$ be a M-sample of the related $\mathbb{R}^L$-valued random variable $\vec{Y}$, as well as $M, R \in \mathbb{N}$. The choice of how to construct a partition $\widehat{\mathcal{P}}_{M;R}$ as in Section 4.2.1 is crucial: for small dimension L, a uniform mesh (so-called Hypercubes (**HC**)) are usually used; see e.g. [DP16] and references therein. Let $\lceil x \rceil := \min\{k \in \mathbb{Z} \,|\, k \geqslant x\}$.

Algorithm 4.2.5 (Hypercubes (**HC**))**.**

(I) Define $\widehat{\mathsf{D}} := \prod_{\ell=1}^L [\widehat{a}_\ell, \widehat{b}_\ell] \subset \mathbb{R}^L$*, where*

$$\widehat{a}_\ell := \min_{k\in\{1,\dots,M\}} \left\lfloor \vec{Y}_\ell(\omega_k) \right\rfloor, \quad \widehat{b}_\ell := \max_{k\in\{1,\dots,M\}} \left\lceil \vec{Y}_\ell(\omega_k) \right\rceil.$$

In addition, let $\widehat{\mathsf{P}} := (\widehat{\mathsf{p}}_1, \dots, \widehat{\mathsf{p}}_L)^T \in \mathbb{R}^L$ *with* $\widehat{\mathsf{p}}_\ell = {}^{(\widehat{a}_\ell + \widehat{b}_\ell)}/_2$*, and* $\widehat{\mathsf{Q}} := (\widehat{\mathsf{q}}_1, \dots, \widehat{\mathsf{q}}_L)^T \in \mathbb{R}^L$ *with* $\widehat{\mathsf{q}}_\ell = {}^{(\widehat{b}_\ell - \widehat{a}_\ell)}/_2$ *for every* $\ell \in \{1, \dots, L\}$*, as well as*

$$\gamma := \frac{1}{R}\left[\left(\max_{\ell\in\{1,\dots,L\}} \widehat{b}_\ell\right) - \left(\min_{\ell\in\{1,\dots,L\}} \widehat{a}_\ell\right)\right] \geqslant 0.$$

(II) Partition $\widehat{\mathsf{D}}$ *into* R *hypercubes of edge length* $\gamma > 0$ *according to*

$$\widehat{\mathcal{P}}_R := \bigcup_{\mathbf{i}\in\Theta} \widehat{C}_\mathbf{i}, \quad \widehat{C}_\mathbf{i} := \bigtimes_{\ell\in\{1,\dots,L\}} \big[\widehat{\mathsf{p}}_\ell - \widehat{\mathsf{q}}_\ell + i_\ell\gamma, \widehat{\mathsf{p}}_\ell - \widehat{\mathsf{q}}_\ell + (i_\ell+1)\gamma\big],$$

where $\Theta := \big\{\mathbf{i} := (i_1, \dots, i_L) \,\big|\, i_\ell \in \{0, 1, \dots, R-1\},\ \ell \in \{1, \dots, L\}\big\}$ *is the set of all possible* (R^L *many) permutations with repetition of the multi-index* $\mathbf{i}$*.*

This strategy is illustrated in Figure 4.5(a). As shown in [DGL96, Chapter 21.2], the shatter coefficient satisfies

$$\Delta_M\big(\widehat{\mathcal{P}}_{M;R}\big) \leqslant 2^{\gamma^L} \binom{M+\gamma}{M}^L,$$

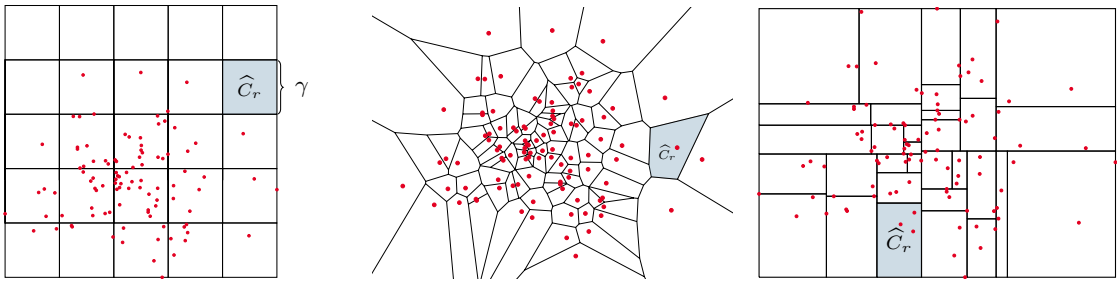
such that Corollary 4.2.4 is fulfilled when (see [DGL96, Chapter 21.4])

$$\lim_{M\to\infty} \frac{\gamma^L}{M} = 0.$$

However, this uniform strategy is not able to adapt spatially varying smoothness adequately. This motivates to adaptively choose cells to be thinner in the direction where the distribution varies more; see Figure 4.5(b)–(c). In addition, choosing the partition $\widehat{\mathcal{P}}_{M;R}$ in a adaptive way reduces curse of dimensionality; see [GKKW02, Chapter 2.2]. A lower bound of the expected closeness of the nearest neighbor with respect to hypercubes is now more detailed.

Suppose that $\vec{y}, \{\vec{Y}(\omega_k)\}_{k=1}^M$ be independent and identically distributed $\mathbb{R}^L$-valued random variables, with $\vec{y} = (\vec{y}_1, \dots, \vec{y}_L)^T \in \mathbb{R}^L$ uniformly distributed in the hypercube $[0,1]^L$, and $\vec{Y}(\omega_k) = \big(\vec{Y}_1(\omega_k), \dots, \vec{Y}_L(\omega_k)\big)^T \in \mathbb{R}^L$. Then

$$\mathbb{E}\left[\min_{k\in\{1,\dots,M\}} \big|\vec{y} - \vec{Y}(\omega_k)\big|_\infty\right] = \int_0^\infty \mathbb{P}\left[\min_{k\in\{1,\dots,M\}} \big|\vec{y} - \vec{Y}(\omega_k)\big|_\infty > t\right] \mathrm{d}t.$$



(a) HC **(b) V** **(c) BTC**

Figure 4.5: (M-sample $\mathcal{S}_{\vec{Y}}$(•) for $M = 100$) (a) Uniform Hypercubes (**HC**) partition with edge length $\gamma > 0$. Adaptive (b) Voronoi (**V**) partition via the nearest neighbor clustering, and (c) Binary Tree Cuboid (**BTC**) partition to obtain statistically equivalent cells.

Splitting the event $\min_{k\in\{1,\dots,M\}} |\vec{y} - \vec{Y}(\omega_k)|_\infty$ into M disjoint events, and using independence of the random variables, it holds

$$\begin{aligned}\mathbb{P}\Big[\min_{k\in\{1,\dots,M\}} |\vec{y} - \vec{Y}(\omega_k)|_\infty > t\Big] &= \mathbb{P}\Big[\prod_{k=1}^{M}\Big\{|\vec{y} - \vec{Y}(\omega_k)|_\infty > t\Big\}\Big] \\ &= \prod_{k=1}^{M}\mathbb{P}\Big[|\vec{y} - \vec{Y}(\omega_k)|_\infty > t\Big] \\ &= \prod_{k=1}^{M}\Big(1 - \mathbb{P}\Big[|\vec{y} - \vec{Y}(\omega_k)|_\infty \leqslant t\Big]\Big).\end{aligned}$$

Fix $k \in \{1,\dots,M\}$. Again, split the event $|\vec{y} - \vec{Y}(\omega_k)|_\infty = \max_{\ell\in\{1,\dots,L\}} |\vec{y}_\ell - \vec{Y}_\ell(\omega_k)| \leqslant t$ into L disjoint events, such that

$$\mathbb{P}\Big[\max_{\ell\in\{1,\dots,L\}} |\vec{y}_\ell - \vec{Y}_\ell(\omega_k)| \leqslant t\Big] = \mathbb{P}\Big[\prod_{\ell=1}^{L}\Big\{|\vec{y}_\ell - \vec{Y}_\ell(\omega_k)| \leqslant t\Big\}\Big] = \prod_{\ell=1}^{L}\mathbb{P}\Big[|\vec{y}_\ell - \vec{Y}_\ell(\omega_k)| \leqslant t\Big].$$

Since both $\vec{y}_\ell \sim \mathcal{U}\big([0,1]\big)$, and $\vec{Y}_\ell(\omega_k) \sim \mathcal{U}\big([0,1]\big)$, it holds $\mathbb{P}\big[|\vec{y}_\ell - \vec{Y}_\ell(\omega_k)| \leqslant t\big] = 1-(1-t)^2$ for all $\ell \in \{1,\dots,L\}$. Hence $(1-(1-t)^2)^L = (2t-t^2)^L \leqslant (2t)^L$ for all $t \in [0,1]$. Collecting the foregoing results yields

$$\mathbb{P}\Big[\min_{k\in\{1,\dots,M\}} |\vec{y} - \vec{Y}(\omega_k)|_\infty > t\Big] \leqslant M\big(1 - (2t)^L\big).$$

The definition of the probability implies $M \cdot (2t)^L \leqslant 1 \Leftrightarrow t \leqslant 1/2\,\sqrt[L]{M}$, such that

$$\mathbb{E}\Big[\min_{k\in\{1,\dots,M\}} |\vec{y} - \vec{Y}(\omega_k)|_\infty\Big] \geqslant \int_0^{1/2\,\sqrt[L]{M}} M\big(1 - (2t)^L\big)\mathrm{d}t = \frac{L}{2(L+1)} \cdot \frac{1}{\sqrt[L]{M}}.$$

As one can see in Table 4.1 this lower bound is not close to zero if the sample size M is large; especially, if $L \gg 1$. The only way to overcome this curse of dimensionality is to incorporate additional assumptions about the measure besides the sample, which is implicitly

Table 4.1: Lower bound for different constellations (M, L) of the expected closeness of nearest neighbor $\eta(L, M) := \mathbb{E}\big[\min_{k\in\{1,\dots,M\}} |\vec{y} - \vec{Y}(\omega_k)|_\infty\big]$ with respect to Hypercubes (**HC**).

	M	10^2	10^3	10^4	10^5	10^6
$L = 10^1$	$\eta(L, M)$	0.287	0.228	0.181	0.144	0.114
$L = 10^2$	$\eta(L, M)$	0.473	0.462	0.451	0.441	0.431
$L = 10^3$	$\eta(L, M)$	0.497	0.496	0.495	0.494	0.493

done by nearly all multivariate estimation procedures.

To allow a more accurately estimation of the empirical probability measure $\widehat{\mu}$ on $\widehat{\mathcal{P}}_{M;R}$, two further commonly used methods are proposed [DGL96] to generate a data-dependent partition of the state space $\mathbb{R}^L$:

- the Voronoi (**V**) method constructs one cell $\widehat{C}_r \subset \mathbb{R}^L$, $r \in \{1, \dots, R\}$ based on a single realization of $\mathcal{S}_{\vec{Y}}$ in combination with nearest neighbor clustering; see Figure 4.5(b).
- the Binary Tree Cuboid (**BTC**) method constructs one cell $\widehat{C}_r \subset \mathbb{R}^L$, $r \in \{1, \dots, R\}$ for an ensemble of realizations of $\mathcal{S}_{\vec{Y}}$: this makes the partition $\widehat{\mathcal{P}}_{M;R}$ more robust; see Figure 4.5(c).

In contrast to partitioning via Hypercubes (**HC**), both the Voronoi (**V**) and Binary Tree Cuboid (**BTC**) method construct the cells $\{\widehat{C}_r\}_{r=1}^R$ according to the distribution of realizations of $\mathcal{S}_{\vec{Y}}$, which only requires to simulate $\vec{Y}$, while the knowledge of the law $\mathcal{L}(\vec{Y})$ is not needed.

Algorithm 4.2.6 (Voronoi (**V**))**.** *Set* $R := M$*. Define*

$$\widehat{C}_r := \Big\{ \vec{y} \in \mathbb{R}^L;\, |\vec{y} - \vec{Y}(\omega_r)| < \inf_{k\neq r} |\vec{y} - \vec{Y}(\omega_k)| \Big\} \quad \forall\, r \in \{1, \dots, R\}.$$

The centers are of three types, depending on how many nearest neighbors in $\mathcal{S}_{\vec{Y}}$ they have. A point with exactly one nearest neighbor is in the interior of a cell, a point with two nearest neighbors is on the boundary between cells, and a point with three nearest neighbors is a vertex where three cells meet. For a survey and comprehensive treatment of Voronoi diagrams; see e.g. [OBSC00, Aur91].

According to [DGLW17], independently of the density ρ, for μ-almost all $\vec{y} \in \mathbb{R}^L$,

$$\lim_{t\to\infty} \limsup_{M\to\infty} \mathbb{P}\Big[\operatorname{diam}\big[\widehat{C}_r\big] \geqslant t M^{-1/L} \,\Big|\, \vec{Y}(\omega_r) = \vec{y}\Big] = 0 \quad \forall\, r \in \{1, \dots, R\},$$

where the diameter of the cell $\widehat{C}_r$, $r \in \{1, \dots, R\}$ is defined through

$$\operatorname{diam}\big[\widehat{C}_r\big] := \sup_{\vec{x},\vec{y}\in\widehat{C}_r} |\vec{x} - \vec{y}|_\infty.$$

In particular, for μ-almost all $\vec{y} \in \mathbb{R}^L$,

$$\lim_{M\to\infty} \sup_{t>0} \frac{1}{\sqrt[L]{M}} \mathbb{P}\Big[\operatorname{diam}\big[\widehat{C}_r\big] > t \,\Big|\, \vec{Y}(\omega_r) = \vec{y}\Big] = 0 \quad \forall\, r \in \{1, \dots, R\},$$

Table 4.2: Lower bound for different constellations (M, L) of the expected closeness of nearest neighbor $\eta(L, M) := \mathbb{E}\big[\min_{k\in\{1,\dots,M\}} |\,\vec{y} - \vec{Y}(\omega_k)|_\infty\big]$ with respect to Voronoi (**V**) partition.

	M	10^2	10^3	10^4	10^5	10^6
$L = 10^1$	$\eta(L, M)$	0.631	0.501	0.398	0.316	0.251
$L = 10^2$	$\eta(L, M)$	0.955	0.933	0.912	0.891	0.871
$L = 10^3$	$\eta(L, M)$	0.995	0.993	0.991	0.989	0.986

such that the diameter of the r-th cell $\widehat{C}_r \in \widehat{\mathcal{P}}_{M;R}$ centered at $\vec{Y}(\omega_r) \in \mathbb{R}^L$ converges to zero, in probability, at a rate of $M^{-1/L}$; see Table 4.2. However, according to [DGLW17, Theorem 6.1], it holds $M \cdot \mathbb{E}\big[\widehat{\mu}[\widehat{C}_r] \,|\, \vec{Y}(\omega_r) = \vec{y}\big] \to 1$ if $(M \uparrow \infty)$ for μ-almost all $\vec{y} \in \mathbb{R}^L$, and $r \in \{1, \dots, R\}$. Thus, according to Markov's inequality, $\lim_{M\to\infty} \sup_{t>0} \mathbb{P}\big[M\widehat{\mu}[\widehat{C}_r] > t \,|\, \vec{Y}(\omega_r) = \vec{y}\big] = 1$, such that the second assumption of [DGL96, Theorem 6.1] is violated, i.e., the random cells $\{\widehat{C}_r\}_{r=1}^R$ has to contain many points with large probability. Thus, in the case of $M = R$, consistency is not ensured for Voronoi (**V**) partitions.

As a second method we partition the sample $\mathcal{S}_{\vec{Y}}$ into cells which are statistically equally likely visited; more precisely, $\mathbb{P}\big[\vec{Y} \in \widehat{C}_r\big] = {}^M\!/_R$, for ${}^M\!/_R \in \mathbb{N}$. In this case, the partition $\widehat{\mathcal{P}}_{M;R}$ will be stored as a binary tree, see e.g. [DGL96, Chapter 20], also known in the literature as so-called k-d-tree [Ben75]. Suppose we want to partition $\mathbb{R}^L$ into $R = 2^\kappa$, $\kappa \in \mathbb{N}$ many disjoint cells $\{\widehat{C}_r\}_{r=1}^R$.

Algorithm 4.2.7 (Binary Tree Cuboids (**BTC**))**.** *Choose $\kappa \in \mathbb{N}$. Set $\mathcal{S}_{0,1} := \mathcal{S}_{\vec{Y}}$.*
For $p = 1, \dots, \kappa + 1$ do:
For $q = 1, \dots, 2^{p-1}$ do:

(I) Define $\mathcal{S} := \mathcal{S}_{p-1,q}$, consisting of $\big\{\vec{Y}(\omega_k)\big\}_{k=1}^{2^{1-p}\cdot M}$.

(II) Find the component $\ell \in \{1, \dots, L\}$ in the set of vectors $\mathcal{S}$ which possesses the largest empirical standard deviation $\widehat{\sigma}_\ell \in \mathbb{R}$; denote this component by $\ell_{p,q} \in \mathbb{N}$.

(III) Compute the median $\mathtt{med}_{p,q} \in \mathbb{R}$ of $\big\{\vec{Y}_{\ell_{p,q}}(\omega_k)\big\}_{k=1}^{2^{1-p}\cdot M}$.

(IV) Divide the (sub-)sample $\mathcal{S}$ into two equal parts $\mathcal{S} = \mathcal{S}_{p,q} \cup \mathcal{S}_{p,2^{p-1}+q}$ according to the criterion $\vec{Y}_{\ell_{p,q}}(\omega_k) \leqslant \mathtt{med}_{p,q}$ $\big(k \in \{1, \dots, 2^{1-p}M\}\big)$.

The (**BTC**) mesh is stored a balanced multi-dimensional search tree which is computationally attractive due to low storage efforts $\mathcal{O}(R)$, as well as fast range queries $\mathcal{O}\big(\log(R)\big)$ to check in which a new realization of $\vec{Y}$ lies. In this case, the r-th leaf node of the binary tree represents the r-th cell $\widehat{C}_r$; see e.g. Figure 4.6(a).

As shown in [DGL96, Chapter 21.7], in the case of statistically equivalent cells, the shatter coefficient can be bounded by

$$\Delta_M\big(\widehat{\mathcal{P}}_{M;R}\big) \leqslant 2^\kappa 2^{L(\kappa+1)\log(M)},$$

such that Corollary 4.2.4 is fulfilled when (see [DGL96, Chapter 21.7])

$$\lim_{M\to\infty} \frac{2^\kappa \log(M)}{M} = 0.$$

Since both the (**HC**) and (**BTC**) strategies obtain the partition by dichotomizing the given sample $\mathcal{S}_{\vec{Y}}$ via hyper-planes along the dimension axes, related volumes of the cells $\{\widehat{C}_r\}_{r=1}^R$ for the histogram based density estimates can be easily determined.

Suppose, $\mathcal{S}_{\vec{Y}}$ is given from the *a priori* unknown multivariate density $\rho : \mathbb{R}^L \to \mathbb{R}$. According to the construction of the (**BTC**) (analogously for (**HC**)) mesh, each cell $\widehat{C}_r := \prod_{\ell=1}^L [a_{r\ell}, b_{r\ell}]$, $r \in \{1, \ldots, R\}$, satisfying $b_{r\ell} > b_{r\ell}$ for all constellations of (r, ℓ), is a hyper-rectangle in $\mathbb{R}^L$, where each tuple $(a_{r\ell}, b_{r\ell}) \in \mathbb{R}^2$ specifies the lower and upper bound of the hyper-rectangle along dimension $\ell \in \{1, \ldots, L\}$. Thus, the piecewise constant Monte-Carlo based density estimator $\widehat{\rho}$ of ρ is

$$\widehat{\rho}\big[\widehat{C}_r\big] \equiv \widehat{\rho}_{\mathcal{S}_{\vec{Y}}}\big[\widehat{C}_r\big] := \frac{\#\{k;\, \vec{Y}(\omega_k) \in \widehat{C}_r\}}{M \cdot \mathrm{vol}\big(\widehat{C}_r\big)} = \frac{\widehat{\nu}_r}{M \cdot \mathrm{vol}\big(\widehat{C}_r\big)} \quad \forall\, r \in \{1, \ldots, R\}, \tag{4.2.2}$$

where $\mathrm{vol}\big(\widehat{C}_r\big) := \prod_{\ell=1}^L \lambda\big([a_{r\ell}, b_{r\ell}]\big)$ with Lebesgue measure λ, for $r \in \{1, \ldots, R\}$. In the case of adaptive, data-dependent partitions, an error bound for Monte-Carlo based integration is proposed in [LYW16] by the use of star discrepancy, which is a concept originates from the analysis of Quasi-Monte-Carlo methods, to formally measure the degree of uniformity of points in a hyper-rectangle.

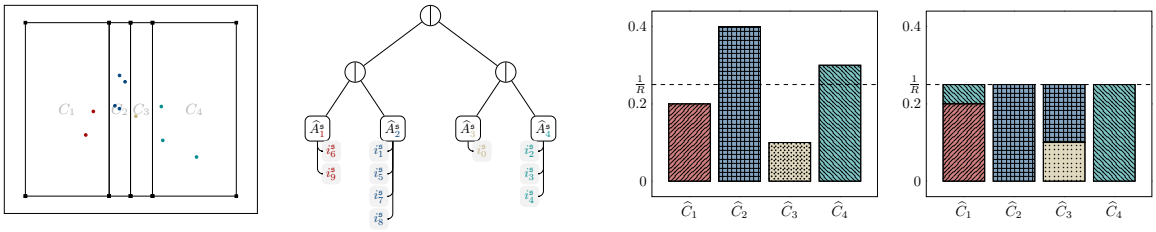
In this thesis, we focus on partitions constructed via the (**BTC**) strategy, where we benefit from the crucial property of statistically equivalent cells, which is, in particular, essentially needed for the adaptive choice of the number of cells $R \equiv R^j$ at time t_j. Moreover, due to its low need of computer resources (low storage effort, fast range queries), it is computationally very attractive to increase the overall efficiency.

4.3 Bootstrapping

To approximate standard estimators (e.g. the probability in (4.6.2)) requires several independent M-samples $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$ $\big(\mathfrak{s} \in \{\mathtt{1}, \mathtt{2}\}\big)$, which is costly. An alternative way to obtain new independent M-samples is by a simple random draw with replacement from existent samples $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$ via the bootstrap method [Das08]. We use the 'Alias method' [Wal77] as an improved alternative to inversion sampling [Gla04] to draw according to a given empirical measure, which does not require time-consuming search techniques.

The general bootstrap procedure involves drawing samples with replacement, using weights that ascribe to the k-th realization $\vec{Y}(\omega_k)$ a probability equal $1/M$. In this section, we suggest a weighted bootstrap approach, in which the weights are chosen according to their relative frequency to reduce the computational effort.

Suppose that $\widehat{\mathcal{P}}_{M;R}$, and $\mathbb{R}^L$-valued M-samples $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}} := \{\vec{Y}^{\mathfrak{s}}(\omega_k)\}_{k=1}^M$ $\big(\mathfrak{s} \in \{\mathtt{1}, \mathtt{2}\}\big)$ are given as in Section 4.2. We then compute the empirical probability measures $\big\{\widehat{\mu}^{\mathfrak{s}};\, \mathfrak{s} \in \{\mathtt{1}, \mathtt{2}\}\big\}$ on $\widehat{\mathcal{P}}_{M;R}$ via (4.6.1). As part of the grouping of $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$, to compute the values $\{\widehat{q}_r^{\mathfrak{s}}\}_{r=1}^R$, realizations of the sample $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$ which are localized in the cell $\widehat{C}_r \in \widehat{\mathcal{P}}_{M;R}$, $r \in \{1, \ldots, R\}$ are represented by the corresponding index set $\widehat{A}_r^{\mathfrak{s}} := \big\{k;\, \vec{Y}^{\mathfrak{s}}(\omega_k) \in \widehat{C}_r\big\}$; see Figure 4.6(a).



(a) $\widehat{\mathcal{P}}_{M;R}$ and associated index sets $\{\widehat{A}^{\mathfrak{s}}_r\}_{r=1}^R$ **(b)** Re-distribution of $\widehat{\mu}^{\mathfrak{s}}$

Figure 4.6: (a) Index sets $\{\widehat{A}^{\mathfrak{s}}_r\}_{r=1}^R$ associated to the cells $\{\widehat{C}_r\}_{r=1}^R$ of the grouped M-sample $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$ (•); partitioned along the dimensions (here ⨁; $M = 10$, $R = 4$). (b) Statistically equivalent re-distribution of $\widehat{\mu}^{\mathfrak{s}}$ over R cells such that $\mathbb{P}[\vec{Y}^{\mathfrak{s}} \in \widehat{C}_r] = 1/R$ for all $r \in \{1, \dots, R\}$.

The improved 'Alias method' to sample according to $\widehat{\mu}^{\mathfrak{s}}$ on $\widehat{\mathcal{P}}_{M;R}$, is based on the fact that an arbitrary R-point distribution may be represented as an equally weighted mixture of R two-point distributions; see e.g. [PK80], where the idea is to re-distribute the probability mass into R cells of equal weight $1/R$, and Figure 4.6(b). To see that such a re-distribution of $\widehat{\mu}^{\mathfrak{s}}$ can be done generally, consider the probability distribution $\mathbb{P}\big[\vec{Y}^{\mathfrak{s}} \in \widehat{C}_r\big] = \widehat{q}^{\mathfrak{s}}_r \in [0,1]$ assigned to the r-th cell $\widehat{C}_r \in \widehat{\mathcal{P}}_{M;R}$, $r \in \{1, \dots, R\}$. If $\widehat{q}^{\mathfrak{s}}_1 = \ldots = \widehat{q}^{\mathfrak{s}}_R = \frac{1}{R}$, then trivially, the original distribution is an equal mixture of one-point (and hence two-point) distributions; to be more precisely, for every $r \in \{1, \dots, R\}$,

$$\widehat{\mu}^{\mathfrak{s}}\big[\widehat{C}_r\big] = \mathbb{P}\big[\vec{Y}^{\mathfrak{s}} \in \widehat{C}_r\big] = \widehat{q}^{\mathfrak{s}}_r = \frac{1}{R}\sum_{l=1}^{R}\Big[\widehat{q}^{\mathfrak{s}}_l \mathbb{1}_{\widehat{C}_{\widehat{a}^{\mathfrak{s}}_l}}\big(\vec{Y}^{\mathfrak{s}}\big) + \big(1 - \widehat{q}^{\mathfrak{s}}_l\big)\mathbb{1}_{\widehat{C}_{\widehat{c}^{\mathfrak{s}}_l}}\big(\vec{Y}^{\mathfrak{s}}\big)\Big]$$

with empirical random integer variables $\widehat{a}^{\mathfrak{s}}_l, \widehat{c}^{\mathfrak{s}}_l : \Omega \to \{1, \dots, R\}$. If not all $\{\widehat{q}^{\mathfrak{s}}_r\}_r$ are equal, then there must exist integers $i^{\mathfrak{s}}, j^{\mathfrak{s}} \in \{1, \dots, R\}$ such that $\widehat{q}^{\mathfrak{s}}_{i^{\mathfrak{s}}} < 1/R$ and $\widehat{q}^{\mathfrak{s}}_{j^{\mathfrak{s}}} \geqslant 1/R$. Thus, fill the cell $\widehat{C}_{i^{\mathfrak{s}}}$ by first adding $\widehat{q}^{\mathfrak{s}}_{i^{\mathfrak{s}}}$ and then transferring an amount $(1/R) - \widehat{q}^{\mathfrak{s}}_{i^{\mathfrak{s}}}$ from $\widehat{q}^{\mathfrak{s}}_{j^{\mathfrak{s}}}$. This leaves $R-1$ cells to be filled with $R-1$ probabilities that sum up to $(R-1)/R$, which can be done recursively in exactly the same way by choosing recursively $i^{\mathfrak{s}}, j^{\mathfrak{s}} \in \{1, \dots, R\}$ from the remaining indices; see Figure 4.6(b). Below, we refer to $\{\widehat{a}^{\mathfrak{s}}_r\}_{r=1}^R$ as the 'alias values', and the $\{\widehat{c}^{\mathfrak{s}}_r\}_{r=1}^R$ as the 'cut-off values'. These can be determined by the following algorithm [Vos91], which formalizes the cell-filling procedure described in Figure 4.6(b).

Algorithm 4.3.1 (Set-up for the Alias Method)**.** *Given $M, R \in \mathbb{N}$, and $\widehat{\mu}^{\mathfrak{s}}$ as in* (4.6.3)*.*

(I) Set $\{\widehat{a}^{\mathfrak{s}}_r\}_{r=1}^R$ with $\widehat{a}^{\mathfrak{s}}_r := 0$, and $\{\widehat{c}^{\mathfrak{s}}_r\}_{r=1}^R$ with $\widehat{c}^{\mathfrak{s}}_r := M\widehat{q}^{\mathfrak{s}}_r$ for all $r \in \{1, \dots, R\}$.

(II) Determine $\mathfrak{S}^{\mathfrak{s}} := \big\{r \in \{1, \dots, R\};\, \widehat{c}^{\mathfrak{s}}_r < 1\big\}$, and $\mathfrak{L}^{\mathfrak{s}} := \big\{r \in \{1, \dots, R\};\, \widehat{c}^{\mathfrak{s}}_r \geqslant 1\big\}$.

(III) While $\mathfrak{S}^{\mathfrak{s}}$ and $\mathfrak{L}^{\mathfrak{s}}$ are not empty do:

(1) Choose some $i^{\mathfrak{s}} \in \mathfrak{S}^{\mathfrak{s}}$ and $j^{\mathfrak{s}} \in \mathfrak{L}^{\mathfrak{s}}$.

(2) Set $\widehat{a}^{\mathfrak{s}}_{i^{\mathfrak{s}}} = j^{\mathfrak{s}}$ and $\widehat{c}^{\mathfrak{s}}_{j^{\mathfrak{s}}} = \widehat{c}^{\mathfrak{s}}_{j^{\mathfrak{s}}} - (1 - \widehat{c}^{\mathfrak{s}}_{i^{\mathfrak{s}}})$.

(3) If $\widehat{c}^{\mathfrak{s}}_{j^{\mathfrak{s}}} < 1$, remove $j^{\mathfrak{s}}$ from $\mathfrak{L}^{\mathfrak{s}}$ and add to $\mathfrak{S}^{\mathfrak{s}}$.

(4) Remove $i^{\mathfrak{s}}$ from $\mathfrak{S}^{\mathfrak{s}}$.

The Set-up Algorithm 4.3.1 can be implemented to run in $\mathcal{O}(R)$ time; see e.g. [PK80]. Given such a equi-probable re-distribution of $\widehat{\mu}^{\mathfrak{s}}$ via Algorithm 4.3.1, the generation of $\mathcal{S}^{\star}_{\vec{Y}^{\mathfrak{s}}}$ based on $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$, which is internally stored on the computer as index set $\widehat{A}^{\mathfrak{s}} := \bigcup_{r=1}^{R} \widehat{A}^{\mathfrak{s}}_r$ (see Figure 4.6(a)), requires to draw according to the random integer variable $\mathfrak{i}^{\mathfrak{s}} : \Omega \to \widehat{A}^{\mathfrak{s}}$ with probability distribution [PK80]

$$\mathbb{P}\big[\mathfrak{i}^{\mathfrak{s}} \in \widehat{A}^{\mathfrak{s}}_r\big] := \frac{1}{R}\sum_{l=1}^{R}\Big[\widehat{q}^{\mathfrak{s}}_l \mathbb{1}_{\widehat{a}^{\mathfrak{s}}_l}(r) + \big(1-\widehat{q}^{\mathfrak{s}}_l\big)\mathbb{1}_{\widehat{c}^{\mathfrak{s}}_l}(r)\Big] \quad \forall\, r \in \{1,\ldots,R\},$$

where $\{\widehat{a}^{\mathfrak{s}}_l\}_{l=1}^{R}$ and $\{\widehat{c}^{\mathfrak{s}}_l\}_{l=1}^{R}$ are determined in Algorithm 4.3.1. These steps are made precise in Algorithm 4.3.2. Note that the alias and cut-off values need only be calculated once and are identical for all bootstrap replications of $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$. Below, let notations indexed by $\star$ indicate probability calculations related to the bootstrap distribution of the given sample $\mathcal{S}^{\star}_{\vec{Y}^{\mathfrak{s}}}$.

Algorithm 4.3.2 (Alias Method). *Given $B, M, R \in \mathbb{N}$, $\widehat{\mathcal{P}}_{M;R}$, and $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$. Choose $M_b \geqslant M$.*

(I) Compute $\widehat{\mu}^{\mathfrak{s}}$ with respect to $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$ on $\widehat{\mathcal{P}}_{M;R}$ via (4.6.3).

(II) Compute the alias values $\{\widehat{a}^{\mathfrak{s}}_r\}_{r=1}^{R}$, and cut-off values $\{\widehat{c}^{\mathfrak{s}}_r\}_{r=1}^{R}$ according to Algorithm 4.3.1 to achieve a equi-probable re-distribution of $\widehat{\mu}^{\mathfrak{s}}$ on $\widehat{\mathcal{P}}_{M;R}$.

(III) For $b = 1,\ldots,B$ do:

(1) Set $\mathfrak{I}^{b,\mathfrak{s},\star} := \emptyset$.

(2) For $k = 1,\ldots,M_b$ do:

(a) Draw $v^{b,\mathfrak{s}}_k \equiv v^{b,\mathfrak{s}}(\omega_k)$ for $v^{b,\mathfrak{s}} \sim \mathcal{U}\big([0,1]\big)$.

(b) Compute $r^{b,\mathfrak{s}}_k \equiv r^{b,\mathfrak{s}}(\omega_k)$ for $r^{b,\mathfrak{s}} := \big\lfloor (R-1)v^{b,\mathfrak{s}} + 1 \big\rfloor \sim \mathcal{U}\big(\{1,\ldots,R\}\big)$.

(c) Draw $u^{b,\mathfrak{s}}_k \equiv u^{b,\mathfrak{s}}(\omega_k)$ for $u^{b,\mathfrak{s}} \sim \mathcal{U}\big([0,1]\big)$ independent of $v^{b,\mathfrak{s}}_k$ and decide: If $u^{b,\mathfrak{s}}_k \leqslant \widehat{c}_{r^{b,\mathfrak{s}}_k}$, set $\mathfrak{I}^{b,\mathfrak{s},\star} := \mathfrak{I}^{b,\mathfrak{s},\star} \cup \big\{\mathfrak{i}^{b,\mathfrak{s}}_k\big\}$ with $\mathfrak{i}^{b,\mathfrak{s}}_k \equiv \mathfrak{i}^{b,\mathfrak{s}}(\omega_k)$ for $\mathfrak{i}^{b,\mathfrak{s}} \sim \mathcal{U}\big(\widehat{A}^{\mathfrak{s}}_{r^{b,\mathfrak{s}}_k}\big)$; otherwise set $\mathfrak{i}^{b,\mathfrak{s}}_k = \widehat{a}^{\mathfrak{s}}_{r^{b,\mathfrak{s}}_k}$.

(3) Obtain $\mathcal{S}^{b,\star}_{\vec{Y}^{\mathfrak{s}}}$ via the index set $\mathfrak{I}^{b,\mathfrak{s},\star}$, i.e., $\mathcal{S}^{b,\star}_{\vec{Y}^{\mathfrak{s}}} := \{\vec{Y}(\omega_{\mathfrak{i}^{b,\mathfrak{s}}_k})\}_{k=1}^{M_b}$.

The effort to compute a sample via bootstrap is by far less expensive, and the computational studies in Section 4.7 confirm these savings. According to [Das08, Theorem 29.1], $\widehat{\mu}_{\mathcal{S}^{\star}_{\vec{Y}^{\mathfrak{s}}}}$ is a consistent perturbation of $\widehat{\mu}_{\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}}$. Statistical inferences about the distribution of the used distance $\mathtt{d}$ are now obtained via bootstrap estimators $\mathtt{d}^{\star}$: the idea is to generate $B \in \mathbb{N}$ replications $\{\mathcal{S}^{b,\star}_{\vec{Y}^{\mathfrak{s}}}\}_{b=1}^{B}$ $\big(\mathfrak{s} \in \{\mathtt{1},\mathtt{2}\}\big)$ of the existing sample $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$ via Algorithm 4.3.2, each of sample size $M_b \geqslant M$; see Figure 4.7. For each tuple $\big(\mathcal{S}^{b,\star}_{\vec{Y}^{\mathtt{1}}}, \mathcal{S}^{b,\star}_{\vec{Y}^{\mathtt{2}}}\big)$ of these $B \in \mathbb{N}$ replications, we then compute the bootstrap estimator $\mathtt{d}^{\star}$ to approximate the probability in (4.6.2) through

$$\mathbb{P}^{\star}\big[\mathtt{d}^{\star}(\widehat{\mu}^{\mathtt{1},\star}, \widehat{\mu}^{\mathtt{2},\star}) \leqslant \mathtt{Tol}\big] \approx \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}_{(-\infty,\mathtt{Tol}]}\big(\mathtt{d}^{\star}(\widehat{\mu}^{\mathtt{1},b,\star}, \widehat{\mu}^{\mathtt{2},b,\star})\big), \tag{4.3.1}$$

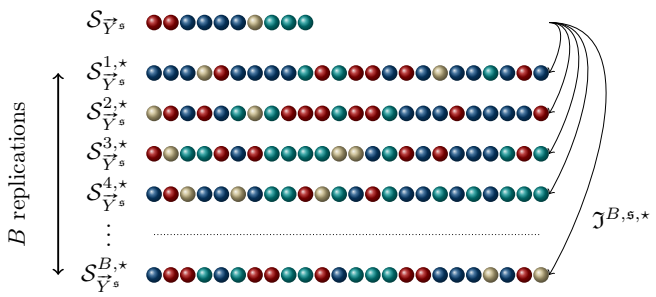


Figure 4.7: (Same setup as in Figure 4.6 for fixed $\mathfrak{s} \in \{1, 2\}$) Bootstrap replications $\{\mathcal{S}^{b,\star}_{\vec{Y}^{\mathfrak{s}}}\}_{b=1}^{B}$ drawn with replacement according to $\widehat{\mu}^{\mathfrak{s}}$, conditionally on the given M-sample $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$, via the index sets $\mathfrak{I}^{b,\mathfrak{s},\star} = \{\mathfrak{i}_k^{b,\mathfrak{s}}\}_k$ via Algorithm 4.3.2 (here $M = 10$, and $M_b = 24$).

as well as expectation values

$$\mathbb{E}^{\star}\big[\mathtt{d}^{\star}(\widehat{\mu}^{1,\star}, \widehat{\mu}^{2,\star})\big] \approx \mathbb{E}^{\star}_{B}\big[\mathtt{d}^{\star}(\widehat{\mu}^{1,\star}, \widehat{\mu}^{2,\star})\big] := \frac{1}{B}\sum_{b=1}^{B} \mathtt{d}^{\star}(\widehat{\mu}^{1,b,\star}, \widehat{\mu}^{2,b,\star}), \tag{4.3.2}$$

where $\mathtt{d}^{\star}\big(\widehat{\mu}^{1,b,\star}, \widehat{\mu}^{2,b,\star}\big) := \mathtt{d}\big(\widehat{\mu}^{1}_{\mathcal{S}^{b,\star}_{\vec{Y}^{1}}}, \widehat{\mu}^{2}_{\mathcal{S}^{b,\star}_{\vec{Y}^{2}}}\big)$.

As in Section 3.2, the number of required bootstrap replications $B \in \mathbb{N}$ can be asymptotically determined by the central limit theorem:

$$\forall a > 0 : \lim_{B\uparrow\infty} \mathbb{P}\left[\left|\mathbb{E}^{\star}\big[\mathtt{d}^{\star}\big] - \mathbb{E}^{\star}_{B}\big[\mathtt{d}^{\star}\big]\right| \leqslant a\sqrt{\frac{\mathrm{Var}^{\star}_{B}\big[\mathtt{d}^{\star}\big]}{B}}\right] = 2\Phi(a) - 1,$$

where Φ is the standard normal distribution, $\mathbb{E}^{\star}_{B}\big[\mathtt{d}^{\star}\big] \equiv \mathbb{E}^{\star}_{B}\big[\mathtt{d}^{\star}(\widehat{\mu}^{1,b,\star}, \widehat{\mu}^{2,b,\star})\big]$ as defined in (4.3.2), and empirical variance

$$\mathrm{Var}^{\star}_{B}\big[\mathtt{d}^{\star}\big] \equiv \mathrm{Var}^{\star}_{B}\big[\mathtt{d}^{\star}(\widehat{\mu}^{1,b,\star}, \widehat{\mu}^{2,b,\star})\big] := \mathbb{E}^{\star}_{B}\Big[\big|\mathtt{d}^{\star} - \mathbb{E}^{\star}_{B}\big[\mathtt{d}^{\star}\big]\big|^2\Big].$$

Let $\mathtt{Tol} > 0$, and $a := \Phi^{-1}(1 - \frac{\alpha}{2}) > 0$ for some level of significance $\alpha \in (0, 1)$. The idea is then to accept $\mathbb{E}^{\star}_{B}\big[\mathtt{d}^{\star}\big]$ as an approximation of $\mathbb{E}^{\star}\big[\mathtt{d}^{\star}\big] \equiv \mathbb{E}^{\star}\big[\mathtt{d}^{\star}(\widehat{\mu}^{1,b,\star}, \widehat{\mu}^{2,b,\star})\big]$, which is *a priori* unknown, up to $\mathtt{Tol}$, at least with probability $2\Phi(a) - 1$. Thus, to reduce the statistical error (cf. $\mathcal{E}_S$ in Section 4.5) in the Monte-Carlo estimation of $\mathbb{E}^{\star}\big[\mathtt{d}^{\star}\big]$, choose $B \in \mathbb{N}$ according to $B := \lfloor a^2\,\mathrm{Var}^{\star}_{B}[\mathtt{d}^{\star}]/\mathtt{Tol}^2 \rfloor$.

For some given level of significance $\alpha \in (0, 1)$, according to [DH97, Chapter 4.2.5] and references therein, the performed Monte-Carlo error is bounded from below by

$$\mathtt{err}(\alpha, B) := 1 - \sqrt{\frac{1-\alpha}{2\pi(B+1)\alpha}}.$$

The values in Table 4.3 suggest that the choice $B = 99$ is not sufficient for $\alpha \geqslant 0.05$, such that in this case $B \geqslant 999$ should generally be chosen. Reducing the error probability of first kind to $\alpha \geqslant 0.01$, at least $B = 9999$ bootstrap replications should be used to for the bootstrap estimators $\mathtt{d}^{\star}$. This observation motivates to choose in the simulations

Table 4.3: Lower bound for the Monte-Carlo error $\mathtt{err}(\alpha, B)$ based on B bootstrap replications, and some given level of significance $\alpha \in (0,1)$.

	B	19	39	99	199	499	999	9999
$\alpha = 0.05$	$\mathtt{err}(\alpha, B)$	0.61	0.73	0.83	0.88	0.92	0.95	0.98
$\alpha = 0.01$	$\mathtt{err}(\alpha, B)$	–	–	0.60	0.72	0.82	0.87	0.96

$B \geqslant \min\{10^3, \lfloor a^2 \operatorname{Var}_B^\star[\mathtt{d}^\star]/\mathtt{Tol}^2 \rfloor\}$, cf. with the discussion in [Ber94], since $\alpha = 0.05$ is convenient in practice as a standard level of significance. However, there is no scientific basis for this choice of α; see e.g. [LR05].
In general, the foregoing point estimate $\mathtt{d}^\star$ of $\mathtt{d}$ is expected to vary for each sample. This can be contrasted with interval estimates by specifying a range within which $\mathtt{d}$ is estimated to lie. To show the reliability of the estimates $\{\mathtt{d}^\star\}$, confidence intervals are commonly used; more precisely, based on $\{\mathcal{S}_{\vec{Y}_\mathfrak{s}}^{b,\star}\}_{b=1}^B$ $\big(\mathfrak{s} \in \{1, 2\}\big)$ estimate the so-called coverage probability $\mathbb{P}\big[\mathtt{d} \in \mathsf{C}^\star\big]$ via the confidence interval $\mathsf{C}^\star := [\widehat{\mathsf{L}}^\star, \widehat{\mathsf{U}}^\star]$ with endpoints

$$\widehat{\mathsf{L}}^\star \equiv \widehat{\mathsf{L}}^\star\big[\mathtt{d}^\star(\widehat{\mu}^{1,b,\star}, \widehat{\mu}^{2,b,\star})\big] := \mathbb{E}_B^\star\big[\mathtt{d}^\star(\widehat{\mu}^{1,b,\star}, \widehat{\mu}^{2,b,\star})\big] - a\sqrt{\operatorname{Var}_B^\star\big[\mathtt{d}^\star(\widehat{\mu}^{1,b,\star}, \widehat{\mu}^{2,b,\star})\big]},$$
$$\widehat{\mathsf{U}}^\star \equiv \widehat{\mathsf{U}}^\star\big[\mathtt{d}^\star(\widehat{\mu}^{1,b,\star}, \widehat{\mu}^{2,b,\star})\big] := \mathbb{E}_B^\star\big[\mathtt{d}^\star(\widehat{\mu}^{1,b,\star}, \widehat{\mu}^{2,b,\star})\big] + a\sqrt{\operatorname{Var}_B^\star\big[\mathtt{d}^\star(\widehat{\mu}^{1,b,\star}, \widehat{\mu}^{2,b,\star})\big]}.$$

How frequently the confidence interval $\mathsf{C}^\star$ contains $\mathtt{d}$ is determined by the confidence level $1 - \alpha = 95\%$ (complement of the level of significance α), i.e., in repeated sampling 95% of the confidence intervals include $\mathtt{d}$. Thus, a confidence interval indicates the uncertainty of the estimate $\mathtt{d}^\star$ of $\mathtt{d}$ which allows to consider the practical importance, rather than just statistical significance.

4.4 Two prototype SPDEs

We emphasize that the results from Section 4.2 are satisfied if the solution of (4.0.1) admits a smooth density. Spatial discretization of (4.0.1), together with the truncation of the driving Wiener process as e.g. in (4.7.1) for some $n \in \mathbb{N}$, leads to an system of SDEs of the form (3.0.1). In this case, if the related vector fields $\{f, \sigma_1, \ldots, \sigma_n\}$ satisfy Hörmander's condition [Hör67] for $n = L$, then the solution $\boldsymbol{X}_t$ of the SDE (3.0.1) admits a smooth density with respect to Lebesgue measure. Unfortunately, since $L \equiv L^j$ and n is fixed in our simulations, thus, in general $n \ll L$, these arguments certainly do not apply. Nevertheless, we use the proposed partitioning strategies from Section 4.2.2 for the simulations.

4.4.1 An SPDE with linear drift

Let $T > 0$, and $\mathscr{A} : \mathbb{W}_0^{1,2} \to \mathbb{W}^{-1,2}$ in (4.0.1) of type

$$\mathscr{A}(u) = \operatorname{div}(\boldsymbol{A}(t,\mathbf{x})\nabla u) - \boldsymbol{\beta}(t,\mathbf{x}) \cdot \nabla u, \tag{4.4.1}$$

with measurable $\boldsymbol{A} : D_T \to \mathbb{R}^{d\times d}_{\text{spd}}$, and $\boldsymbol{\beta} \in L^\infty\big(D_T; [\mathbb{W}^{1,\infty}]^d\big)$. Let $\mathbb{K}$ be a separable Hilbert space, and $\sigma : \mathbb{L}^2 \to \mathscr{L}(\mathbb{L}^2, \mathbb{K})$ be Lipschitz continuous, i.e., there exist constants $K_1, K_2 > 0$ such that for all $u, v \in \mathbb{L}^2$ holds

$$\|\sigma(u)\|_{\mathscr{L}(\mathbb{L}^2,\mathbb{K})} \leqslant K_1\big(1 + \|u\|_{\mathbb{L}^2}\big), \qquad \|\sigma(u) - \sigma(v)\|^2_{\mathscr{L}(\mathbb{L}^2,\mathbb{K})} \leqslant K_2\|u - v\|^2_{\mathbb{L}^2}. \tag{4.4.2}$$

By $\mathscr{L}_1$ we denote the space of trace class operators on $\mathbb{K}$. Let $\mathfrak{P} := (\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a complete filtered probability space, on which a $\mathbb{K}$-valued $\boldsymbol{Q}$-Wiener process $W \equiv \big\{W_t; t \in [0,T]\big\}$ with covariance operator $\boldsymbol{Q} \in \mathscr{L}_1(\mathbb{K})$ is defined through

$$W(t,\mathbf{x}) = \sum_{|\mathbf{k}|\geqslant 0} \sqrt{\lambda_{\mathbf{k}}} e_{\mathbf{k}}(\mathbf{x})\beta^{\mathbf{k}}(t), \quad \forall\, t \in [0,T] \quad \forall\, \mathbf{x} \in D, \tag{4.4.3}$$

where $\{e_{\mathbf{k}}\}_{\mathbf{k}}$ is an orthonormal basis of $\mathbb{K}$ consisting of eigen-functions of $\boldsymbol{Q}$, $\{\lambda_{\mathbf{k}}\}_{\mathbf{k}}$ are related eigen-values, and $\big\{\beta^{\mathbf{k}}(t); t \in [0,T]\big\}_{\mathbf{k}}$ are independent $\mathbb{R}$-valued Brownian motions, where $\mathbf{k} := (k_1, \ldots, k_d)^T \in \mathbb{N}_0^d$ is a multi-index with $\mathbb{N}_0 \ni |\mathbf{k}| := \sum_{i=1}^d k_i$; see e.g. [DPZ14, Chapter 4]. Problem 4.0.1 with $x_0 \in \mathbb{L}^2(D)$ may then be recast into the following form: Find the process $X \equiv \big\{X_t; t \in [0,T]\big\}$ such that $X \in L^2_{\mathbb{F}}\big(\Omega; \mathcal{C}([0,T]; \mathbb{L}^2) \bigcap L^2(0,T; \mathbb{W}_0^{1,2})\big)$ satisfies $\mathbb{P}$-almost surely for every $t \in [0,T]$, and all $\Psi \in \mathbb{W}_0^{1,2}$,

$$\begin{aligned}\big(X_t, \Psi\big) - \big(x_0, \Psi\big) = &- \int_0^t \Big[\big(\boldsymbol{A}_s \nabla X_s, \nabla\Psi\big) + \big(\boldsymbol{\beta}_s \cdot \nabla X_s, \Psi\big)\Big]\mathrm{d}s \\ &+ \iota \int_0^t \big(\sigma(X_s)\mathrm{d}W_s, \Psi\big).\end{aligned} \tag{4.4.4}$$

Example 4.7.1 studies the numerical discretization via Scheme 4.5.1 and adaptivity in the 'convection-dominated case', i.e. $\boldsymbol{A} \equiv \epsilon \boldsymbol{I}$, and $\boldsymbol{\beta} \neq \mathbf{0}$.

We approximate the $\mathbb{K}$-valued $\boldsymbol{Q}$-Wiener process $\boldsymbol{W}$ on $\mathfrak{P}$ by a $\mathbb{K}$-valued $\boldsymbol{Q}$-random walk. For this purpose, recall $\mathcal{I}_\tau := \{t_j\}_{j=0}^J$ with $\tau^{j+1} := t_{j+1} - t_j > 0$, consider the (time-discrete) filtration $\mathfrak{F}_\tau := \big\{\mathcal{F}_{t_j};\, t_j \in \mathcal{I}_\tau\big\} \subset \mathfrak{F}$, and set $\mathfrak{P}_\tau := (\Omega, \mathcal{F}, \mathfrak{F}_\tau, \mathbb{P})$.

Definition 4.4.1. *Let $\boldsymbol{Q} \in \mathscr{L}_1(\mathbb{K})$ be symmetric and non-negative definite, and $\mathcal{I}_\tau := \{t_j\}_{j=0}^J$. A $\boldsymbol{Q}$-random walk on $\mathfrak{P}_\tau$ is a sequence $\{\xi_{j+1}\}_{j=0}^{J-1}$ of $\mathbb{K}$-valued independent identically distributed random variables such that for each $j \in \{0, \ldots, J-1\}$ the following conditions are satisfied:*

(1) ξ_{j+1} is $\mathcal{F}_{t_{j+1}}$-measurable and independent of $\big\{\mathcal{F}_{t_i};\, i \in \{1, \ldots, j\}\big\}$.

(2) $\mathbb{E}\big[\xi_{j+1}\big] = 0$, and $\mathbb{E}\big[(\xi_{j+1}, x)_{\mathbb{K}}(\xi_{j+1}, y)_{\mathbb{K}}\big] = \tau^{j+1}(\boldsymbol{Q}x, y)_{\mathbb{K}}$ for all $x, y \in \mathbb{K}$.

(3) For every $p \in \mathbb{N}$, there exists $C_p > 0$ such that $\mathbb{E}\big[\|\xi_{j+1}\|^{2p}_{\mathbb{K}}\big] \leqslant C_p(\tau^{j+1})^p$.

Definition 4.4.1 generalizes the definition of an $\mathbb{R}$-valued random walk. In the finite-dimensional case, the weak limit of a random walk is a Wiener-process by the Donsker's invariance principle; see [KS91, p. 70].

4.4.2 An SPDE with non-linear drift — the stochastic Landau-Lifshitz-Gilbert equation

Let $T > 0$. We look for an $\mathbb{S}^2$-valued process $\boldsymbol{X} \equiv \{\boldsymbol{X}_t;\, t \in [0,T]\}$ that satisfies the non-linear SPDE ($\zeta_1, \zeta_2 \geqslant 0$)

$$\begin{aligned} \mathrm{d}\boldsymbol{X}_t - \boldsymbol{X}_t \times \big[\zeta_1 \Delta \boldsymbol{X}_t - \zeta_2 \boldsymbol{X}_t \times \Delta \boldsymbol{X}_t\big]\mathrm{d}t &= \iota \boldsymbol{X}_t \times \circ \mathrm{d}\boldsymbol{W}_t \qquad \text{in } D_T := (0,T) \times D, \\ \partial_{\mathbf{n}} \boldsymbol{X}_t &= 0 \qquad \text{on } \partial D_T := (0,T) \times \partial D, \\ \boldsymbol{X}_0 &= \boldsymbol{x}_0 \qquad \text{in } D, \end{aligned} \tag{4.4.5}$$

for $\mathbf{x}_0 \in \mathbb{W}^{1,2}(D;\mathbb{S}^2)$, and three independent $\mathbb{K}$-valued $\boldsymbol{Q}$-Wiener processes $\{W_i\}_{i=1}^3$, such that $\boldsymbol{W} = (W_1, W_2, W_3)^T$, where $\boldsymbol{Q} \in \mathscr{L}_1(\mathbb{K})$. In order to accommodate for the pathwise sphere property $|\boldsymbol{X}_t| = 1$, the stochastic term is understood in Stratonovich sense (indicated by '$\circ$'). Problem (4.4.5) is only known to have a weak martingale solution [BBNP14b], which relaxes the pathwise solution concept that applies in (4.4.4).

Definition 4.4.2. *Let $T > 0$. A weak martingale solution $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \boldsymbol{X}, \boldsymbol{W})$ of problem (4.4.5) consists of*

(1) a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, where the filtration $\mathbb{F} := \{\mathcal{F}_t;\, t \in [0,T]\}$ satisfies the usual conditions,

(2) a $\mathbb{F}$-adapted $\mathbb{K}$-valued $\boldsymbol{Q}$-Wiener process $\boldsymbol{W} = \{\boldsymbol{W}_t;\, t \in [0,T]\}$ on it, and

(3) a progressively measurable $\mathbb{F}$-adapted process $\boldsymbol{X} : [0,T] \times \Omega \to \mathbf{L}^2(D;\mathbb{R}^3)$ such that

(a) for $\mathbb{P}$-almost surely every $\omega \in \Omega$, and all $t \in [0,T]$,

$$\boldsymbol{X}(\cdot,\omega) \in \mathcal{C}\big([0,T];\, \mathbf{L}^2(D;\mathbb{R}^3)\big).$$

(b) $\mathbb{E}\Big[\sup_{t\in[0,T]} \|\nabla \boldsymbol{X}_t(\cdot)\|_{\mathbf{L}^2}^2\Big] < \infty$, and for every $t \in [0,T]$, the equality $|\boldsymbol{X}_t(\cdot,\cdot)| = 1$ is satisfied Lebesgue almost everywhere in D_T, $\mathbb{P}$-almost surely.

(c) for every $t \in [0,T]$, for all $\boldsymbol{\Psi} \in \mathcal{C}_0^\infty(D;\mathbb{R}^3)$, there holds $\mathbb{P}$-almost surely

$$\begin{aligned} \big(\boldsymbol{X}_t, \boldsymbol{\Psi}\big) - \big(\boldsymbol{x}_0, \boldsymbol{\Psi}\big) &= \zeta_1 \int_0^t \big(\boldsymbol{X}_s \times \Delta \boldsymbol{X}_s, \boldsymbol{\Psi}\big)\mathrm{d}s \\ &\quad - \zeta_2 \int_0^t \big(\boldsymbol{X}_s \times \big[\boldsymbol{X}_s \times \Delta \boldsymbol{X}_s\big], \boldsymbol{\Psi}\big)\mathrm{d}s \\ &\quad + \iota \int_0^t \big(\boldsymbol{X}_s \times \circ \mathrm{d}\boldsymbol{W}_s, \boldsymbol{\Psi}\big). \end{aligned} \tag{4.4.6}$$

Related computational experiments in [BBNP14b] motivate possible blow-up dynamics of a solution of (4.4.5). A recent analytical study of this phenomenon for $\zeta_1 = 0$, and $\zeta_2 = 1$ is [Hoc16] characterizes a proper (see [Hoc16, Theorem 3]) weak martingale solution ($d = 2$) as a classical, smooth solution, apart from no more but finitely many space-time points. The computational studies in Section 4.7.3 and 4.7.4 illustrate a proper space-time resolution of singular behaviors of the solution of (4.4.5) by the adaptive method discussed in Section 4.6.

4.5 Space-time discretization of the SPDE (4.0.1)

We propose stable space-time discretizations for problems (4.4.4) and (4.4.5).

Let $\{(\tau^{j+1}, \mathcal{T}_h^{j+1})\}_{j=-1}^J$ be a space-time mesh of $[0,T] \times \overline{D}$. For $j \in \{-1, 0, \dots, J\}$, we define sequences of $\mathbb{W}_0^{1,2}$-conforming lowest order finite element spaces,

$$\mathbb{V}_h^{j+1} := \Big\{ \Psi \in \mathcal{C}(\overline{D}) : \Psi\big|_K \in \mathbb{P}_1(K) \; \forall \, K \in \mathcal{T}_h^{j+1} \Big\},$$

with $L^{j+1} := \dim \mathbb{V}_h^{j+1}$ and d-simplexes $K \in \mathcal{T}_h^{j+1}$. Moreover, let $\{\Psi_\ell\}_{\ell=1}^{L^{j+1}}$ denote the nodal basis of the finite element space $\mathbb{V}_h^{j+1}$, and $\vec{\Upsilon}_{\mathbb{V}_h^{j+1}} : \mathbb{V}_h^{j+1} \to \mathbb{R}^{L^{j+1}}$ the corresponding coordinate map. A finite element discretization of problem (4.4.5) uses $\mathbb{R}^3$-valued functions $\boldsymbol{\Psi} \in \big[\mathbb{V}_h^{j+1}\big]^3$. The bilinear form $(\cdot,\cdot)_h : \mathcal{C}(\overline{D};\mathbb{R}^3) \times \mathcal{C}(\overline{D};\mathbb{R}^3) \to \mathbb{R}$ is defined via [BBNP14b]

$$(\cdot,\cdot)_h := \int_D \mathcal{I}_h^{j+1}\Big[\big\langle \boldsymbol{\Psi}(\mathbf{x}), \boldsymbol{\Xi}(\mathbf{x})\big\rangle_{\mathbb{R}^3}\Big] \mathrm{d}\mathbf{x} = \sum_{\ell=1}^{L^{j+1}} w_\ell \big\langle \boldsymbol{\Psi}(\mathbf{x}_\ell), \boldsymbol{\Xi}(\mathbf{x}_\ell)\big\rangle_{\mathbb{R}^3} \qquad \forall\, \boldsymbol{\Psi}, \boldsymbol{\Xi} \in \mathcal{C}(\overline{D};\mathbb{R}^3)$$

for certain weights $w_\ell = \int_D \Psi_\ell \, \mathrm{d}\mathbf{x} > 0$, $\ell \in \{1, \dots, L^j\}$. The (affine) nodal interpolation operator $\mathcal{I}_h^{j+1}$ is a bounded map from $\mathcal{C}(\overline{D};\mathbb{R})$ to $\mathbb{V}_h^{j+1}$. The discrete Laplace operator $\Delta_h : \big[\mathbb{V}_h^{j+1}\big]^3 \to \big[\mathbb{V}_h^{j+1}\big]^3$ is defined by $-(\Delta_h \boldsymbol{\Xi}, \boldsymbol{\Psi})_h = (\nabla \boldsymbol{\Xi}, \nabla \boldsymbol{\Psi})_{\mathbf{L}^2}$ for all $\boldsymbol{\Psi}, \boldsymbol{\Xi} \in \big[\mathbb{V}_h^{j+1}\big]^3$.

4.5.1 Discretization of the SPDE (4.0.1) with (4.4.1)

The standard Galerkin FEM solution is usually globally polluted by spurious oscillations, since it is equivalent to a central finite difference discretization which is known to give rise to spurious oscillations in convection dominated regimes, i.e., $\|\boldsymbol{\beta}\|_{\mathbb{L}^\infty(D)} \gg \epsilon$.; see e.g. [RST08]. The basic stabilized method in the framework of finite element discretizations is the SUPG (Streamline Upwind Petrov-Galerkin) method; see [BH82].

The discretization in [JNP84] uses the continuous Galerkin method of order $\geqslant 1$, with (local) space-time test functions $\Psi + \delta_K(\partial_t \Psi + \boldsymbol{\beta} \cdot \nabla \Psi)$ on strips $S_j := \big\{(t,\mathbf{x});\, t \in (t_j, t_{j+1}),\, \mathbf{x} \in D\big\}$. For the convection-dominated SPDE (4.4.4), we rather prefer the SUPG method with spatial test functions $\Psi + \delta_K \boldsymbol{\beta} \cdot \nabla \Psi$, in combination with the implicit Euler-Maruyama scheme to control amplification of oscillatory numerical artefacts by the noise term.

Scheme 4.5.1. *Let $Y^0 \in \mathbb{V}_h^0$, and $\{(\tau^{j+1}, \mathcal{T}_h^{j+1}, \{\delta_K^{j+1}\}_K)\}_{j=-1}^J$ be given. For all $j \in \{0, \dots, J\}$, and ξ_{j+1}, determine a $\mathbb{V}_h^{j+1}$-valued $\mathcal{F}_{t_{j+1}}$-measurable random variable Y^{j+1} such that $\mathbb{P}$-almost surely for all $\Psi \in \mathbb{V}_h^{j+1}$*

$$\begin{aligned} \big(Y^{j+1} - Y^j, \Psi + \delta_K^{j+1} \boldsymbol{\beta}_{t_{j+1}} \cdot \nabla \Psi\big)_{\mathbb{L}^2} + \tau^{j+1} \epsilon \big(\nabla Y^{j+1}, \nabla \Psi\big)_{\mathbb{L}^2} \\ + \tau^{j+1} \big(\boldsymbol{\beta}_{t_{j+1}} \cdot \nabla Y^{j+1}, \Psi + \delta_K^{j+1} \boldsymbol{\beta}_{t_{j+1}} \cdot \nabla \Psi\big)_{\mathbb{L}^2} \\ = \big(\sigma(Y^j)\xi_{j+1}, \Psi + \delta_K^{j+1} \boldsymbol{\beta}_{t_{j+1}} \cdot \nabla \Psi\big)_{\mathbb{L}^2}. \end{aligned} \tag{4.5.1}$$

The term scaled by $\epsilon > 0$ forgoes a further term '$\tau^{j+1}\epsilon\big(\Delta Y^{j+1}, \delta_K^{j+1} \boldsymbol{\beta}_{t_{j+1}} \cdot \nabla \Psi\big)_{\mathbb{L}^2}$' due to the use of piecewise affine finite element functions. Computations for the deterministic counterpart in [JN11, Bur10] favor the uniform choice $\delta_K^{j+1} = \max_K h_K^{j+1}/2|\boldsymbol{\beta}_{t_{j+1}}|$ to achieve stable results on coarse meshes $\mathcal{T}_h^{j+1} \equiv \mathcal{T}_h^0$, since the small scales which require a stabilization are the

spatial ones. Stability for this deterministic case and possible driving right-hand sides $f \in \mathrm{BV}\big(0,T;\mathbb{L}^2(D)\big)$ of the PDE has been shown in [Bur10]; according to [Bur10, Remark 5], a different choice for δ_K^{j+1} of order $\mathcal{O}(\tau^{j+1})$ should however be chosen for a less regular f which, in particular, applies in the stochastic setting in Scheme 4.5.1; see also the simulations in Section 4.7 for computational evidence.

Lemma 4.5.2. *Let $\{Y^{j+1}\}_{j=-1}^J$ be the solution of Scheme 4.5.1, and $\max_K \delta_K^{j+1} \leqslant \tau^{j+1}/2$ for each $j \in \{-1,0,\dots,J\}$. There exists $C \equiv C\big(T,\mathrm{Tr}(\boldsymbol{Q}),\boldsymbol{\beta}\big) > 0$ such that*

$$\max_{0\leqslant j\leqslant J}\mathbb{E}\Big[\|Y^{j+1}\|_{\mathbb{L}^2}^2\Big] + \sum_{j=0}^J \mathbb{E}\Big[\|Y^{j+1}-Y^j\|_{\mathbb{L}^2}^2\Big] + \epsilon\sum_{j=0}^J \tau^{j+1}\mathbb{E}\Big[\|\nabla Y^{j+1}\|_{\mathbb{L}^2}^2\Big] + \sum_{j=0}^J \tau^{j+1}\mathbb{E}\left[\left\|\sqrt{\delta_K^{j+1}}\boldsymbol{\beta}_{t_{j+1}}\cdot\nabla Y^{j+1}\right\|_{\mathbb{L}^2}^2\right] \leqslant C\mathbb{E}\Big[\|Y^0\|_{\mathbb{L}^2}^2\Big].$$

The Lax-Milgram lemma implies the existence and uniqueness of $Y^{j+1} \in L^2_{\mathcal{F}_{t_{j+1}}}(\Omega;\mathbb{V}_h^{j+1})$ for each $j \in \mathbb{N}_0$. We remark that the argument in [Bur10] to prove stability in the deterministic case may not be applied in the present stochastic case, since (discrete) temporal derivatives of the iterates $\{Y^{j+1}\}_j$ in Scheme 4.5.1 grow unboundedly for $\tau^{j+1} \downarrow 0$. For our numerical experiments in Section 4.7, we later choose $\delta_K^{j+1} := \min\{h_K^{j+1}, \tau^{j+1}/2\}$.

Proof. Consider (4.5.1) for a fixed $\omega \in \Omega$, and choose $\Psi = Y^{j+1}(\omega) \in \mathbb{V}_h^{j+1}$ as test function in equation (4.5.1). By (4.4.2), we obtain

$$\begin{aligned}
&\frac{1}{2}\Big[\big\|Y^{j+1}\big\|_{\mathbb{L}^2}^2 - \big\|Y^j\big\|_{\mathbb{L}^2}^2 + \big\|Y^{j+1}-Y^j\big\|_{\mathbb{L}^2}^2\Big] \\
&\qquad + \big(Y^{j+1}-Y^j, \delta_K^{j+1}\boldsymbol{\beta}_{t_{j+1}}\cdot\nabla Y^{j+1}\big)_{\mathbb{L}^2} + \tau^{j+1}\epsilon\big\|\nabla Y^{j+1}\big\|_{\mathbb{L}^2}^2 \\
&\qquad - \tau^{j+1}\big(\boldsymbol{\beta}_{t_{j+1}}\cdot\nabla Y^{j+1}, Y^{j+1}\big)_{\mathbb{L}^2} + \tau^{j+1}\left\|\sqrt{\delta_K^{j+1}}\boldsymbol{\beta}_{t_{j+1}}\cdot\nabla Y^{j+1}\right\|_{\mathbb{L}^2}^2 \\
&\qquad \leqslant \big(\sigma(Y^j)\xi_{j+1}, \{Y^{j+1}-Y^j\} + \delta_K^{j+1}\boldsymbol{\beta}_{t_{j+1}}\cdot\nabla Y^{j+1}\big)_{\mathbb{L}^2} + \mathrm{I}_j \\
&\qquad \leqslant 2K_1\Big(1+\big\|Y^j\big\|_{\mathbb{L}^2}\Big)^2\big\|\xi_{j+1}\big\|_{\mathbb{K}}^2 + \frac{1}{8}\big\|Y^{j+1}-Y^j\big\|_{\mathbb{L}^2}^2 \\
&\qquad\quad + \frac{\tau^{j+1}}{8}\left\|\sqrt{\delta_K^{j+1}}\boldsymbol{\beta}_{t_{j+1}}\cdot\nabla Y^{j+1}\right\|_{\mathbb{L}^2}^2 + \mathrm{I}_j,
\end{aligned}$$

where $\mathbb{E}\big[\mathrm{I}_j\big] = \mathbb{E}\big[\big(\sigma(Y^j)\xi_{j+1}, Y^j\big)_{\mathbb{L}^2}\big] = 0$. We bound

$$\tau^{j+1}\big|\big(\boldsymbol{\beta}_{t_{j+1}}\cdot\nabla Y^{j+1}, Y^{j+1}\big)_{\mathbb{L}^2}\big| = \frac{\tau^{j+1}}{2}\big|\big(\mathrm{div}(\boldsymbol{\beta}_{t_{j+1}}), |Y^{j+1}|^2\big)_{\mathbb{L}^2}\big| \leqslant C\frac{\tau^{j+1}}{2}\big\|Y^{j+1}\big\|_{\mathbb{L}^2}^2,$$

and, by the assumption $\max_K \delta_K^{j+1} \leqslant \tau^{j+1}/2$,

$$\big|\big(Y^{j+1}-Y^j, \delta_K^{j+1}\boldsymbol{\beta}_{t_{j+1}}\cdot\nabla Y^{j+1}\big)_{\mathbb{L}^2}\big| \leqslant \frac{1}{4}\|Y^{j+1}-Y^j\|_{\mathbb{L}^2}^2 + \frac{\tau^{j+1}}{2}\left\|\sqrt{\delta_K^{j+1}}\boldsymbol{\beta}_{t_{j+1}}\cdot\nabla Y^{j+1}\right\|_{\mathbb{L}^2}^2.$$

After absorbing terms, we sum over all iteration steps, take expectations, and use independence and the discrete Gronwall inequality. □

4.5.2 Discretization of the non-linear SPDE (4.4.5)

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a complete filtered probability space. A stable discretization for (4.4.5) was constructed in [BBNP14b]. We restrict ourselves to the case $\zeta_1 = 1$.

Scheme 4.5.3. *Let $\boldsymbol{Y}^0 \in \big[\mathbb{V}_h^0\big]^3$ be given such that $|\boldsymbol{Y}^0(\mathbf{x}_\ell)| = 1$ for all $\ell \in \{1, \dots, L^0\}$, and meshes $\{(\tau^{j+1}, \mathcal{T}_h^{j+1})\}_{j=-1}^J$. For all $j \in \{0, 1, \dots, J\}$, and $\boldsymbol{\xi}_{j+1}$, determine a $\big[\mathbb{V}_h^{j+1}\big]^3$-valued $\mathcal{F}_{t_{j+1}}$-measurable random variable $\boldsymbol{Y}^{j+1}$ such that $\mathbb{P}$-almost surely*

$$\begin{aligned}\big(\boldsymbol{Y}^{j+1} - \boldsymbol{Y}^j, \boldsymbol{\Psi}\big)_h - \tau^{j+1}\big(\boldsymbol{Y}^{j+1/2} \times \Delta_h \boldsymbol{Y}^{j+1}, \boldsymbol{\Psi}\big)_h \\ + \zeta_2 \tau^{j+1}\big(\boldsymbol{Y}^{j+1/2} \times \big[\boldsymbol{Y}^{j+1/2} \times \Delta_h \boldsymbol{Y}^{j+1}\big], \boldsymbol{\Psi}\big)_h \\ = \iota\big(\boldsymbol{Y}^{j+1/2} \times \boldsymbol{\xi}_{j+1}, \boldsymbol{\Psi}\big)_h \quad \forall\, \boldsymbol{\Psi} \in \big[\mathbb{V}_h^{j+1}\big]^3.\end{aligned} \tag{4.5.2}$$

Here, $\boldsymbol{Y}^{j+1/2} := \frac{1}{2}\big(\boldsymbol{Y}^j + \boldsymbol{Y}^{j+1}\big)$.

The following stability properties for Scheme 4.5.3 were proved in [BBNP14b].

Theorem 4.5.4. *For every $j \in \{0, 1, \dots, J\}$, there exists a $\big[\mathbb{V}_h^{j+1}\big]^3$-valued $\mathcal{F}_{t_{j+1}}$-measurable random variable $\boldsymbol{Y}^{j+1} \subset L^2\big(\Omega; \mathbb{W}^{1,2}(D; \mathbb{R}^3)\big)$ of Scheme 4.5.3. Moreover, there exists a constant $C \equiv C(T, \mathrm{Tr}(\boldsymbol{Q})) > 0$ such that*

(1) $|\boldsymbol{Y}^j(\mathbf{x}_\ell)| = 1$ for all $\ell \in \{1, \dots, L^{j+1}\}$, and all $j \in \{1, \dots, J\}$, $\mathbb{P}$-almost surely,

(2) $$\mathbb{E}\bigg[\sup_{1 \leqslant j \leqslant J} \|\nabla \boldsymbol{Y}^j\|_{\mathbf{L}^2}^2 + \sum_{j=1}^J \tau^j \|\boldsymbol{Y}^{j+1/2} \times \Delta_h \boldsymbol{Y}^{j+1}\|_h^2\bigg] \leqslant C.$$

Note that a relevant property of iterates of (4.5.2) is that the length of initial profiles is preserved; cf. (1) in Theorem 4.5.4.

A second stable discretization for (4.4.5) is obtained by formally applying $\boldsymbol{Id} - \zeta_2 \boldsymbol{X}_t \times$ and some elementary calculations; see [Hoc15, Chapter 6]. Following the deterministic procedure [AJ06] to overcome the difficulty of solving a non-linear system at each time step, and motivated by the orthogonal property $\langle \partial \boldsymbol{X}_t / \partial t, \boldsymbol{X}_t \rangle = 0$ of the analytical equation (1.0.4), we seek the unknown solution $\boldsymbol{V}^{j+1} \approx \partial \boldsymbol{X}_{t_{j+1}} / \partial t$ in the 'tangential space'

$$\mathbb{T}_h^{j+1} := \Big\{ \boldsymbol{\Phi} \in \big[\mathbb{V}_h^{j+1}\big]^3 \,\big|\, \langle \boldsymbol{\Phi}(\mathbf{x}_\ell), \boldsymbol{Y}^j(\mathbf{x}_\ell, \omega) \rangle,\ \ell \in \{1, \dots, L^{j+1}\},\ \omega \in \Omega \Big\}.$$

The algorithm reads then as follows.

Scheme 4.5.5. *Let $\boldsymbol{Y}^0 \in \big[\mathbb{V}_h^0\big]^3$ be given such that $|\boldsymbol{Y}^0(\mathbf{x}_\ell)| = 1$ for all $\ell \in \{1, \dots, L^0\}$, $\zeta_2 > 0$, $\theta \in (1/2, 1]$, and meshes $\big\{(\tau^{j+1}, \mathcal{T}_h^{j+1})\big\}_{j=-1}^J$. For all $j \in \{0, 1, \dots, J\}$, and $\boldsymbol{\xi}_{j+1}$:*

(1) Determine $\boldsymbol{U}^{j+1} \in \big[\mathbb{V}_h^{j+1}\big]^3$ such that

$$\big(\nabla \boldsymbol{U}^{j+1}, \nabla \boldsymbol{\Psi}\big)_h = \big(\boldsymbol{\xi}_{j+1}, \boldsymbol{\Psi}\big)_h \quad \forall\, \boldsymbol{\Psi} \in \big[\mathbb{V}_h^{j+1}\big]^3,$$

and $\boldsymbol{V}^{j+1} \in \mathbb{T}_h^{j+1}$ such that for all $\boldsymbol{\Phi} \in \mathbb{T}_h^{j+1}$

$$\begin{aligned}\big(\zeta_2 \boldsymbol{V}^{j+1} - \boldsymbol{Y}^j \times \boldsymbol{V}^{j+1}, \boldsymbol{\Phi}\big)_h + \tau^{j+1}\theta(1 + \zeta_2^2)\big(\nabla \boldsymbol{V}^{j+1}, \nabla \boldsymbol{\Phi}\big)_h \\ = -\tau^{j+1}(1 + \zeta_2^2)\big(\nabla \boldsymbol{Y}^j, \nabla \boldsymbol{\Phi}\big)_h \\ + \big(\boldsymbol{Y}^j \times \boldsymbol{U}^{j+1}, \boldsymbol{\Phi}\big)_h.\end{aligned} \tag{4.5.3}$$

(2) Set $\mathbb{P}$-almost surely in $\big[\mathbb{V}_h^{j+1}\big]^3$:

$$\boldsymbol{Y}^{j+1}(\mathbf{x}) := \sum_{\ell=1}^{L^{j+1}} \frac{\boldsymbol{Y}^j(\mathbf{x}_\ell) + \boldsymbol{V}^{j+1}(\mathbf{x}_\ell)}{|\boldsymbol{Y}^j(\mathbf{x}_\ell) + \boldsymbol{V}^{j+1}(\mathbf{x}_\ell)|} \Psi_\ell(\mathbf{x}) \quad \forall\, \mathbf{x} \in D.$$

The Lax-Milgram lemma implies the existence and uniqueness of $\boldsymbol{V}^{j+1} \in \mathbb{T}_h^{j+1}$ for each $j \in \mathbb{N}_0$. As a consequence of the orthogonality relation $\langle \boldsymbol{V}^{j+1}(\mathbf{x}_\ell), \boldsymbol{Y}^j(\mathbf{x}_\ell)\rangle = 0$ for all $\ell \in \{1, \dots, L^j\}$, it holds $\mathbb{P}$-almost surely $|\boldsymbol{Y}^j(\mathbf{x}_\ell) + \boldsymbol{V}^{j+1}(\mathbf{x}_\ell)| = 1 + |\boldsymbol{V}^{j+1}(\mathbf{x}_\ell)| \geqslant 1$, such that $\boldsymbol{Y}^{j+1} \in \big[\mathbb{V}_h^{j+1}\big]^3$ in Scheme 4.5.5 is well-defined. In addition, let $\{\Psi_\ell\}_{\ell=1}^{L^{j+1}}$ fulfill

$$\int_D \big\langle \nabla \Psi_\ell, \nabla \Psi_{\ell'} \big\rangle \,\mathrm{d}\mathbf{x} \leqslant 0 \quad \forall\, \ell, \ell' \in \{1, \dots, L^{j+1}\}, \quad \ell \neq \ell',$$

to ensure compatibility of the re-normalization step in Scheme 4.5.5 with the energy estimates of [Hoc15, Chapter 5]. For $\theta \in (1/2, 1]$, the analysis (see [Hoc15, Lemma 5.1]) reveals

$$\mathbb{E}\Big[\big\|\nabla \boldsymbol{Y}^{j+1}\big\|_{\mathbf{L}^2}^2\Big] \leqslant \mathbb{E}\Big[\big\|\nabla \boldsymbol{Y}^{j}\big\|_{\mathbf{L}^2}^2\Big] - \big(2\theta - 1 - \varepsilon\big)\mathbb{E}\Big[\big\|\nabla \boldsymbol{V}^{j}\big\|_{\mathbf{L}^2}^2\Big] \quad \forall\, \varepsilon \in (0, 2\theta - 1), \tag{4.5.4}$$

where the additional term on the right hand side with negative sign simply rises from the construction of the scheme and by no means follows from the equation.

4.6 Space-time adaptivity based on the distance of empirical laws

We construct adaptive meshes in time ($\{\tau^{j+1}\}_j$) and space ($\{\mathcal{T}_h^{j+1}\}_j$) to compute iterates $\{Y^{j+1}\}_j$ from Scheme 4.5.1 and 4.5.3. At time $t_{j+1} = \sum_{i=0}^{j} \tau^{i+1}$, we determine sequences of time step sizes $\{\tau_n^{j+1}\}_n$, of regular spatial meshes $\{\mathcal{T}_{h;n}^{j+1}\}_n$, and finite element spaces $\{\mathbb{V}_{h;n}^{j+1}\}_n$, indexed by $n \in \mathbb{N}_0$ until a threshold criterion is met. For this purpose, we propose

- the (**BTC**) strategy to build a data-dependent partition $\widehat{\mathcal{P}}_{\tau;R_\tau;n}^{j+1}$ (resp. $\widehat{\mathcal{P}}_{h;R_h;K;n}^{j+1}$) of the state space $\mathbb{R}^{L^j}$ (resp. $\mathbb{R}_0^+$) into $R_\tau \equiv R_\tau(M_\tau)$ (resp. $R_h \equiv R_h(M_h)$) many cells, depending on the sample size M_τ (resp. M_h) which is adjusted via the empirical variance of the iterates $\{Y^j\}_j$. This partition at time $t_j + \tau_n^{j+1}$ will allow a comparison of empirical probability measures $\{\widehat{\mu}_{\tau;n}^{\mathfrak{s},j+1}\}_{\mathfrak{s}}$ (resp. $\{\widehat{\mu}_{h;K;n}^{\mathfrak{s},j+1}; K \in \mathcal{T}_{h;n}^{j+1}\}_{\mathfrak{s}}$) to detect temporal (resp. spatial) changes.

- This temporal strategy uses different distances (Hellinger $\mathtt{d_H}$, Kullback-Leibler $\mathtt{d_{KL}}$, or total variation $\mathtt{d_{TV}}$) of subsequent empirical measures that are defined on the partition $\widehat{\mathcal{P}}_{\tau;R_\tau;n}^{j+1}$ (resp. $\widehat{\mathcal{P}}_{h;R_h;K;n}^{j+1}$).

- the re-sampling strategy bootstrap (see Section 4.3) is used for the estimator $\mathtt{d}^\star$ of this distance to prevent the ample computation of independent new samples.

- The basic sample size M_τ (resp. M_h), as well as the number of bootstrap replications B_τ (resp. B_h) are chosen depending on the empirical variance of computed realizations.

- a space adaptive strategy is based on the ZZ-estimator (see Section 4.1.1), where the new regular spatial mesh $\mathcal{T}_{h;n+1}^{j+1}$ is obtained by refining/coarsening elements $K \in \mathcal{T}_{h;n}^{j+1}$ according to the distance of empirical probability measures of $\mathbb{R}_0^+$-valued random variables $|\nabla Y_n^{j+1}|_K$ and $|G_h(\nabla Y_n^{j+1})|_K$ for a random variable $Y_n^{j+1} \in L^2(\Omega; \mathbb{V}_{h;n}^{j+1})$.

- an empirical partition $\widehat{\mathcal{P}}_{\tau;\widehat{R}_{\tau;n}^{j+1};n}^{j+1}$ (resp. $\widehat{\mathcal{P}}_{h;\widehat{R}_{h;K;n}^{j+1};K;n}^{j+1}$), where $\widehat{R}_{\tau;n}^{j+1}$ (resp. $\widehat{R}_{h;K;n}^{j+1}$) is chosen depending on the empirical variance of computed realizations.

4.6.1 Partitioning

Let $Y \in L^2(\Omega; \mathbb{V}_h)$ be a random variable, with corresponding $\mathbb{R}^L$-valued $\vec{Y} = \vec{\Upsilon}_{\mathbb{V}_h}(Y)$ coordinate map, where the finite element space $\mathbb{V}_h$ is from Section 4.5, and $L := \dim \mathbb{V}_h$. We fix $R \equiv R(M)$, and choose $M \gg R$ for the sample $\mathcal{S}_{\vec{Y}} := \{\vec{Y}(\omega_k)\}_{k=1}^M$. We use the Binary Tree Cuboid (**BTC**) method as described in Section 4.2.2 to generate a data-dependent partition of the state space $\mathbb{R}^L$:

$$\widehat{\mathcal{P}}_R \equiv \widehat{\mathcal{P}}_{M;R}(\mathcal{S}_{\vec{Y}}) = \{\widehat{C}_r\}_{r=1}^R, \quad \mathring{\widehat{C}}_r \cap \mathring{\widehat{C}}_s = \emptyset, \quad r \neq s, \quad \forall\, r, s \in \{1, \ldots, R\}, \tag{4.6.1}$$

such that all (closed) cells $\widehat{C}_r \subset \mathbb{R}^L$ are equally likely visited, i.e.,

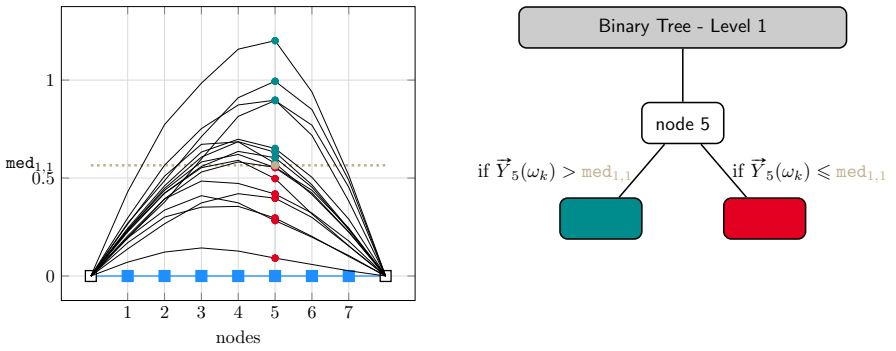
$$\widehat{\mu}[\widehat{C}_r] \equiv \widehat{\mu}_{\mathcal{S}_{\vec{Y}}}[\widehat{C}_r] := \frac{\#\{k;\, \vec{Y}(\omega_k) \in \widehat{C}_r\}}{M} = \frac{\widehat{\nu}_r}{M} \quad \forall\, r \in \{1, \ldots, R\},$$

where the frequency $\widehat{\nu}_r \in \mathbb{N}_0$ counts the number of realizations in $\mathcal{S}_{\vec{Y}}$ that lie in $\widehat{C}_r \in \widehat{\mathcal{P}}_R$. Grouping the events uses geometry-based splittings of the index set of the spatial nodal points $\{\mathbf{x}_\ell\}_{\ell=1}^L$; see Figure 4.8. Finding the entry $\ell_{p,q} \in \mathbb{N}$ is the most expensive part to create the binary tree; the computation of all empirical standard deviations at all spatial nodal points $\mathbf{x}_\ell \in D$, $\ell \in \{1, \ldots, L\}$ may be done through $\mathcal{O}(LM)$ many operations in parallel to increase the efficiency; see Table 4.4. To determine the median $\mathtt{med}_{p,q} \in \mathbb{R}$ at entry $\ell_{p,q}$ is accomplished

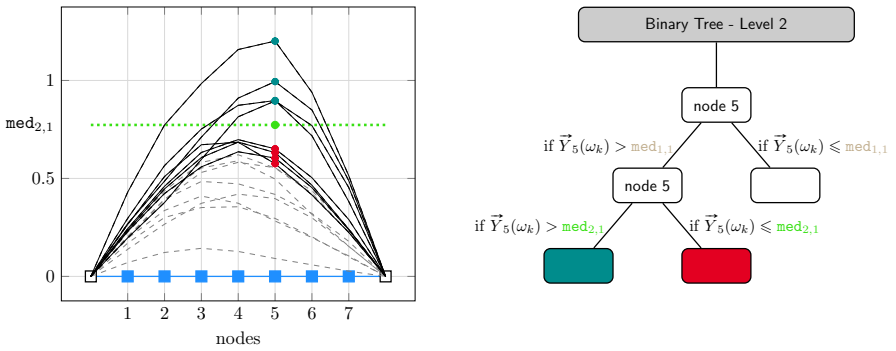
Table 4.4: Different number of cells R to build $\widehat{\mathcal{P}}_R$: the absolute simulation time (in seconds) in double precision arithmetic. Parallelized calculations are performed using OpenMP, cf. [DM98].

R		1024	2048	4096	8192
	BTC	$246s$	$468s$	$812s$	$1510s$
(OpenMP)	**BTC**	$51s$	$61s$	$123s$	$201s$

by sorting all function values $\{\vec{Y}_{\ell_{p,q}}(\omega_k)\}_{k=1}^{2^{1-p}\cdot M}$, which may be done through $\mathcal{O}(M \log(M))$ many comparisons. Computational evidence supports that this division of (sub-)samples is stable against statistical outliers; cf. also [DP16]. Localizing the cell $\widehat{C}_r \in \widehat{\mathcal{P}}_R$, $r \in \{1, \ldots, R\}$ where a new realization of $\vec{Y}$ lies in requires $\mathcal{O}(R)$ many checks. We remark that to uniformly partition a large subset of $\mathbb{R}^L$ via hypercubes [GLW05], or Voronoi meshes [DP16] is not favorable here due to the high dimensionality ($L \gg 1$), and the equi-probability of the cells is a relevant property of the (**BTC**) based partition.



(a) Level 1 of (**BTC**)



(b) Level 2 of (**BTC**)

Figure 4.8: (a) First partition step to construct the (**BTC**) mesh: M many realizations of $\vec{Y}$ (——) are divided according to their value at node $\mathbf{x}_5$ (■) and the median $\texttt{med}_{1,1}$ (·····) into two subsets $\mathcal{S}_{1,1}$(●) and $\mathcal{S}_{1,2}$(●); see step (IV) in Algorithm 4.2.7. (b) Second partition step to construct the (**BTC**) mesh: the realizations in $\mathcal{S}_{1,1}$ (——) are divided according to their value at node $\mathbf{x}_5$ (■) and the median $\texttt{med}_{2,1}$ (·····) into two subsets $\mathcal{S}_{2,1}$(●) and $\mathcal{S}_{2,2}$(●).

4.6.2 The distance between two $\mathbb{R}^L$-valued samples

Let $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$ be a M-sample of realizations of $\vec{Y}^{\mathfrak{s}} \in L^2(\Omega;\mathbb{R}^L)$, where $\mathfrak{s} \in \{\mathtt{1},\mathtt{2}\}$. We want to approximately compute the distance $\mathtt{d}\big(\mathcal{L}(\vec{Y}^{\mathtt{1}}),\mathcal{L}(\vec{Y}^{\mathtt{2}})\big)$ of the related probability measures

$$\mu^{\mathfrak{s}} := \mathcal{L}\big(\vec{Y}^{\mathfrak{s}}\big) \in \mathscr{P}\big(\mathbb{R}^L\big) \quad \big(\mathfrak{s} \in \{\mathtt{1},\mathtt{2}\}\big)$$

on $\big(\mathbb{R}^L,\mathcal{B}(\mathbb{R}^L)\big)$, where $\mathcal{B}(\mathbb{R}^L)$ is the Borel σ-algebra on $\mathbb{R}^L$. For this purpose, we replace them by their empirical counterparts $\big\{\widehat{\mu}^{\mathfrak{s}};\ \mathfrak{s} \in \{\mathtt{1},\mathtt{2}\}\big\}$, which are obtained from sampling on the underlying partition $\widehat{\mathcal{P}}_R$, and check whether

$$\mathbb{P}\big[\mathtt{d}\big(\widehat{\mu}^{\mathtt{1}},\widehat{\mu}^{\mathtt{2}}\big) \leqslant \mathtt{Tol}\big] > p, \tag{4.6.2}$$

for some given $\mathtt{Tol} > 0$, and $p \in [0,1]$. The data-dependent partition $\widehat{\mathcal{P}}_R$ of $\mathbb{R}^L$ uses the (**BTC**) method from Section 4.6.1,

$$\widehat{\mathcal{P}}_R \equiv \widehat{\mathcal{P}}_R\big(\mathcal{S}_{\vec{Y}^{\mathtt{1},\mathtt{2}}}\big) = \big\{\widehat{C}_r\big\}_{r=1}^R \qquad \text{where} \qquad \mathcal{S}_{\vec{Y}^{\mathtt{1},\mathtt{2}}} := \mathcal{S}_{\vec{Y}^{\mathtt{1}}} \cup \mathcal{S}_{\vec{Y}^{\mathtt{2}}}.$$

We then sample again — and refer below again to the new samples as $\mathcal{S}_{\vec{Y}^{\mathfrak{s}}}$ $\big(\mathfrak{s} \in \{\mathtt{1},\mathtt{2}\}\big)$ — , to now compute frequency vectors $\widehat{\boldsymbol{\nu}}^{\mathfrak{s}} := \big(\widehat{\nu}_1^{\mathfrak{s}},\ldots,\widehat{\nu}_R^{\mathfrak{s}}\big)^T$, where $\widehat{\nu}_r^{\mathfrak{s}} := \#\widehat{A}_r^{\mathfrak{s}}$, $r \in \{1,\ldots,R\}$, with associated index set $\widehat{A}_r^{\mathfrak{s}} := \big\{k;\ \vec{Y}^{\mathfrak{s}}(\omega_k) \in \widehat{C}_r\big\}$, and thus obtain empirical probability measures $\widehat{\mu}^{\mathfrak{s}} \in \mathscr{P}(\mathbb{R}^L)$ via

$$\widehat{\mu}^{\mathfrak{s}} := \sum_{r=1}^R \widehat{q}_r^{\mathfrak{s}} \cdot \widehat{\xi}_r^{\mathfrak{s}}, \quad \text{where} \quad \widehat{q}_r^{\mathfrak{s}} := \frac{\widehat{\nu}_r^{\mathfrak{s}}}{M}, \quad \text{and} \quad \widehat{\xi}_r^{\mathfrak{s}} := \frac{1}{\widehat{\nu}_r^{\mathfrak{s}}} \sum_{k \in \widehat{A}_r^{\mathfrak{s}}} \delta_{\vec{Y}^{\mathfrak{s}}(\omega_k)} \quad \big(\mathfrak{s} \in \{\mathtt{1},\mathtt{2}\}\big). \tag{4.6.3}$$

The Hellinger distance of the discrete measures $\big\{\widehat{\mu}^{\mathfrak{s}};\ \mathfrak{s} \in \{\mathtt{1},\mathtt{2}\}\big\}$ is

$$\mathtt{d}_{\mathtt{H}}\big(\widehat{\mu}^{\mathtt{1}},\widehat{\mu}^{\mathtt{2}}\big) = \left(\frac{1}{2}\sum_{r=1}^R \Big(\sqrt{\widehat{q}_r^{\mathtt{1}}} - \sqrt{\widehat{q}_r^{\mathtt{2}}}\Big)^2\right)^{1/2}.$$

Other choices are the Kullback-Leibler distance resp. the total variation metric,

$$\mathtt{d}_{\mathtt{KL}}\big(\widehat{\mu}^{\mathtt{1}},\widehat{\mu}^{\mathtt{2}}\big) = \sum_{r=1}^R \widehat{q}_r^{\mathtt{1}} \log\left(\frac{\widehat{q}_r^{\mathtt{1}}}{\widehat{q}_r^{\mathtt{2}}}\right) \qquad \text{resp.} \qquad \mathtt{d}_{\mathtt{TV}}\big(\widehat{\mu}^{\mathtt{1}},\widehat{\mu}^{\mathtt{2}}\big) = \frac{1}{2}\sum_{r=1}^R \big|\widehat{q}_r^{\mathtt{1}} - \widehat{q}_r^{\mathtt{2}}\big|.$$

We refer to [GS02] for further details, and a comparison of these different distance functions. For comparison of the different distances, since the Kullback-Leibler distance is only a relative measure and its value ranging from 0 to ∞, a normalization of $\mathtt{d}_{\mathtt{KL}}$ is achieved below via the non-linear transformation $1-\exp(-\mathtt{d}_{\mathtt{KL}})$ to take values in $[0,1]$, and is denoted by $\widetilde{\mathtt{d}}_{\mathtt{KL}}$.

Remark 4.6.1. Likelihood-based inference for $\big\{\widehat{\mu}^{\mathfrak{s}};\ \mathfrak{s} \in \{\mathtt{1},\mathtt{2}\}\big\}$ via Fisher's non-parametric χ^2-test of homogeneity uses the χ^2-distance, see e.g. [Pit79, Chapter 2],

$$\mathtt{d}_{\chi^2}^2\big(\widehat{\mu}^{\mathtt{1}},\widehat{\mu}^{\mathtt{2}}\big) = \sum_{r=1}^R \frac{\big(\widehat{\nu}_r^{\mathtt{1}} - \widehat{\nu}_r^{\mathtt{2}}\big)^2}{\widehat{\nu}_r^{\mathtt{1}} + \widehat{\nu}_r^{\mathtt{2}}} = M\sum_{r=1}^R \frac{\big(\widehat{q}_r^{\mathtt{1}} - \widehat{q}_r^{\mathtt{2}}\big)^2}{\widehat{q}_r^{\mathtt{1}} + \widehat{q}_r^{\mathtt{2}}}. \tag{4.6.4}$$

Recall that if $H_0 : \mu^{\mathtt{1}} = \mu^{\mathtt{2}}$ is not rejected, there is no statistical evidence for $\mu^{\mathtt{1}} = \mu^{\mathtt{2}}$. As will be discussed in Remark 4.6.4, this fact limits the use of the χ^2-test for time adaptivity.

2) Contrary to testing the H_0-hypothesis of homogeneity via $\mathtt{d}_{\chi^2}^2$ in (4.6.4) with well-known asymptotic distribution χ_{R-1}^2, probabilities $\mathbb{P}[\mathtt{d}(\widehat{\mu}^{\mathtt{1}},\widehat{\mu}^{\mathtt{2}}) \geqslant \mathtt{Tol} \,|\, H_0]$ for other choices of $\mathtt{d}$ must be approximated via MC simulations; here, the bootstrapping method (see Remark 4.6.4) is the relevant tool to drastically reduce the computational effort.

4.6.3 Time adaptivity

Let t_j be fixed, $M_\tau \in \mathbb{N}$, $\mathtt{N}_\tau \in \mathbb{N}_0$, $\mathtt{d} \in \{\mathtt{d_H}, \widetilde{\mathtt{d}}_{\mathtt{KL}}, \mathtt{d_{TV}}\}$, and $\mathtt{Tol}_\tau > 0$. Below, we generate a finite sequence $\{\tau_n^{j+1}\}_{n=0}^{\mathtt{N}_\tau}$ with $\tau_0^{j+1} := \tau^j$ to (possibly) adaptively refine/coarsen τ_n^{j+1}. Therefore, for each $n \in \{0, \dots, \mathtt{N}_\tau\}$, we obtain empirical probability measures $\widehat{\mu}_{\tau;n}^{\mathfrak{s},j+1} \in \mathscr{P}_{\tau;n}^{j+1}(\mathbb{R}^{L^j})$, where $L^j := \dim \mathbb{V}_h^j$.

Let $\mathcal{S}_{\vec{Y}_n^{\mathfrak{s},j+1}}$ be M_τ-samples from random variables $Y^{\mathfrak{s},j+1} \in L^2(\Omega; \mathbb{V}_h^j)$ with unknown laws $\mu^{\mathfrak{s},j+1}$ $\big(\mathfrak{s} \in \{\mathtt{1}, \mathtt{2}\}\big)$, where the first is the solution of Scheme 4.5.1 with coarser time step size $\tau_n^{\mathtt{1},j+1} := \tau_n^{j+1}$, while the latter is obtained by extrapolation using the additional scale $\tau_n^{\mathtt{2},j+1} := \tau_n^{\mathtt{1},j+1}/2$; see e.g. [Abo13], and Algorithm 4.6.2 below. We approximate the distance $\mathtt{d}(\widehat{\mu}_{\tau;n}^{\mathtt{1},j+1}, \widehat{\mu}_{\tau;n}^{\mathtt{2},j+1})$ with the help of the related empirical measures $\big\{\widehat{\mu}_{\tau;n}^{\mathfrak{s},j+1}; \mathfrak{s} \in \{\mathtt{1}, \mathtt{2}\}\big\}$ to then steer refinement or coarsening of $\tau_n^{j+1} > 0$. For this purpose, we sample on the underlying partition $\widehat{\mathcal{P}}_{\tau;R_\tau;n}^{j+1} := \bigcup_{r=1}^{R_\tau} \widehat{C}_{\tau;r;n}^{j+1}$ of $\mathbb{R}^{L^j}$ which is obtained via Algorithm 4.2.7, i.e.,

$$\widehat{\mathcal{P}}_{\tau;R_\tau;n}^{j+1} \equiv \widehat{\mathcal{P}}_{\tau;R_\tau;n}^{j+1}\big(\mathcal{S}_{\vec{Y}_n^{\mathtt{1},\mathtt{2};j+1}}\big) = \big\{\widehat{C}_{\tau;r;n}^{j+1}\big\}_{r=1}^{R_\tau},$$

where $\mathcal{S}_{\vec{Y}_n^{\mathtt{1},\mathtt{2};j+1}} := \mathcal{S}_{\vec{Y}_n^{\mathtt{1},j+1}} \cup \mathcal{S}_{\vec{Y}_n^{\mathtt{2},j+1}}$ and $\mathcal{S}_{\vec{Y}_n^{\mathfrak{s},j+1}} := \{\vec{Y}_n^{\mathfrak{s},j+1}(\omega_k)\}_{k=1}^{M_\tau}$. We then sample again — and refer below again to the new samples as $\mathcal{S}_{\vec{Y}_n^{\mathfrak{s},j+1}}$ $\big(\mathfrak{s} \in \{\mathtt{1}, \mathtt{2}\}\big)$ — , to compute frequency vectors $\widehat{\boldsymbol{\nu}}_{\tau;n}^{\mathfrak{s},j+1} := (\widehat{\nu}_{\tau;1;n}^{\mathfrak{s},j+1}, \dots, \widehat{\nu}_{\tau;R_\tau;n}^{\mathfrak{s},j+1})^T$, where $\widehat{\nu}_{\tau;r;n}^{\mathfrak{s},j+1} := \#\widehat{A}_{\tau;r;n}^{\mathfrak{s},j+1}$, $r \in \{1, \dots, R_\tau\}$, with associated index set $\widehat{A}_{\tau;r;n}^{\mathfrak{s},j+1} := \{k;\, \vec{Y}_n^{\mathfrak{s},j+1}(\omega_k) \in \widehat{C}_{\tau;r;n}^{j+1}\}$, and thus obtain empirical probability measures $\widehat{\mu}_{\tau;n}^{\mathfrak{s},j+1} \in \mathscr{P}_{\tau;n}^{j+1}(\mathbb{R}^{L^j})$ via

$$\widehat{\mu}_{\tau;n}^{\mathfrak{s},j+1} = \sum_{r=1}^{R_\tau} \widehat{q}_{\tau;r;n}^{\mathfrak{s},j+1} \cdot \widehat{\xi}_{\tau;r;n}^{\mathfrak{s},j+1} \quad \big(\mathfrak{s} \in \{\mathtt{1}, \mathtt{2}\}\big),$$

where

$$\widehat{q}_{\tau;r;n}^{\mathfrak{s},j+1} := \frac{\widehat{\nu}_{\tau;r;n}^{\mathfrak{s},j+1}}{M_\tau}, \quad \text{and} \quad \widehat{\xi}_{\tau;r;n}^{\mathfrak{s},j+1} := \frac{1}{\widehat{\nu}_{\tau;r;n}^{\mathfrak{s},j+1}} \sum_{k \in \widehat{A}_{\tau;r;n}^{\mathfrak{s},j+1}} \delta_{\vec{Y}_n^{\mathfrak{s},j+1}(\omega_k)} \quad \big(\mathfrak{s} \in \{\mathtt{1}, \mathtt{2}\}\big).$$

We then generate $B_\tau \in \mathbb{N}$ many new bootstrap samples $\mathcal{S}_{\vec{Y}_n^{\mathfrak{s},j+1,\star}}$ $\big(\mathfrak{s} \in \{\mathtt{1}, \mathtt{2}\}\big)$ according to the measure $\widehat{\mu}_{\tau;n}^{\mathfrak{s},j+1}$ on $\widehat{\mathcal{P}}_{\tau;R_\tau;n}^{j+1}$ from $\mathcal{S}_{\vec{Y}_n^{\mathfrak{s},j+1}}$ to eventually determine the distance of both via (4.3.1), and then choose the local time step size τ_n^{j+1}.

In order to reduce the empirical variance of these samples $\mathcal{S}_{\vec{Y}_n^{\mathfrak{s},j+1}}$ $\big(\mathfrak{s} \in \{\mathtt{1}, \mathtt{2}\}\big)$, we interpolate the same Wiener process on $[t_j, t_j + \tau_n^{\mathtt{1},j+1}]$; cf. Algorithm 3.1.1. The following algorithm precises the generation of these two samples at time t_j.

Algorithm 4.6.2 (Richardson extrapolation)**.** *Let $\tau_n^{j+1} > 0$, and $M_\tau \in \mathbb{N}$ be given. Initialize the samples $\mathcal{S}_{\vec{Y}_n^{\mathtt{1},j+1}} := \emptyset$, $\mathcal{S}_{\vec{Y}_n^{\mathtt{2},j+1}} := \emptyset$, as well as $\tau_n^{\mathtt{1},j+1} := \tau_n^{j+1}$ and $\tau_n^{\mathtt{2},j+1} := \tau_n^{\mathtt{1},j+1}/2$.*
For $k = 1, \dots, M_\tau$ do:

(I) Compute $\xi_{j+1}^{1,1;k} := \xi_{t_{j+1}}(\omega_k) - \xi_{t_j}(\omega_k)$ on $[t_j, t_j + \tau_n^{\mathtt{1},j+1}]$.

(II) Compute the realization $Y_{n;k}^{1,j+1} \equiv Y_n^{\mathtt{1},j+1}(\omega_k)$ via one step of Scheme 4.5.1 with time step size $\tau_n^{\mathtt{1},j+1}$ and $\xi_{j+1}^{1,1;k}$. Set $\mathcal{S}_{\vec{Y}_n^{\mathtt{1},j+1}} := \mathcal{S}_{\vec{Y}_n^{\mathtt{1},j+1}} \cup \{\vec{Y}_{n;k}^{\mathtt{1},j+1}\}$.

(III) *Compute* $\xi_{j+1}^{1,2;k} := \frac{1}{2}\xi_{j+1}^{1,1;k} + Z_{j,n}^{1,2;k}$, *with* $Z_{j,n}^{1,2;k} \equiv Z_{j,n}^{1,2}(\omega_k)$, *for* $Z_{j,n}^{1,2} \sim \mathcal{N}\big(0, t_j + \tau_n^{2,j+1}\big)$.

(IV) *Compute* $\xi_{j+1}^{2,2;k} := \frac{1}{2}\xi_{j+1}^{1,1;k} - Z_{j,n}^{2,2;k}$, *with* $Z_{j,n}^{2,2;k} \equiv Z_{j,n}^{2,2}(\omega_k)$, *for* $Z_{j,n}^{2,2} \sim \mathcal{N}\big(0, t_j + \tau_n^{2,j+1}\big)$.

(V) *Compute the iterate* $\mathsf{Y}_{n;k}^{2,j+1} \equiv \mathsf{Y}_n^{2,j+1}(\omega_k)$ *via two steps of Scheme 4.5.1 with time step size* $\tau_n^{2,j+1}$, *and increments* $\xi_{j+1}^{1,2;k}$, $\xi_{j+1}^{2,2;k}$. *Define* $Y_{n;k}^{2,j+1} := 2\mathsf{Y}_{n;k}^{2,j+1} - Y_{n;k}^{1,j+1}$, *and set* $\mathcal{S}_{\vec{Y}_n^{2,j+1}} := \mathcal{S}_{\vec{Y}_n^{2,j+1}} \cup \big\{\vec{Y}_{n;k}^{2,j+1}\big\}$.

We may now apply the tools from Section 4.6.2 to compute $\mathtt{d}\big(\widehat{\mu}_{\tau;n}^{1,j+1}, \widehat{\mu}_{\tau;n}^{2,j+1}\big)$ from (4.6.3). These steps are made precise in the following algorithm.

Algorithm 4.6.3 (Adaptivity in time). *Fix* $j \geqslant 0$, $\mathtt{Tol}_\tau > 0$, $M_\tau \in \mathbb{N}$, *and* $\mathtt{N}_\tau \in \mathbb{N}_0$. *Choose a distance* $\mathtt{d} \in \{\mathtt{d}_\mathtt{H}, \widetilde{\mathtt{d}}_\mathtt{KL}, \mathtt{d}_\mathtt{TV}\}$, $B_\tau \in \mathbb{N}$, *and* $p_\tau^r, p_\tau^c \in [0,1]$. *Set* $\tau_0^{j+1} := \tau^j$.
For $n = 0, \dots, \mathtt{N}_\tau$ *do:*

(I) *Compute* M_τ*-samples* $\mathcal{S}_{\vec{Y}_n^{\mathfrak{s},j+1}}$ $\big(\mathfrak{s} \in \{\mathtt{1}, \mathtt{2}\}\big)$ *for* τ_n^{j+1} *via Algorithm 4.6.2.*

(II) *Create* $\widehat{\mathcal{P}}_{\tau;R_\tau;n}^{j+1} \equiv \widehat{\mathcal{P}}_{\tau;R_\tau;n}^{j+1}\big(\mathcal{S}_{\vec{Y}_n^{\mathtt{1},\mathtt{2};j+1}}\big)$ *via Algorithm 4.2.7.*

(III) *Compute new* M_τ*-samples* $\mathcal{S}_{\vec{Y}_n^{\mathfrak{s},j+1}}$ $\big(\mathfrak{s} \in \{\mathtt{1}, \mathtt{2}\}\big)$ *for* τ_n^{j+1} *via Algorithm 4.6.2.*

(IV) *Generate* B_τ *independent bootstrap* M_τ*-samples* $\{\mathcal{S}_{\vec{Y}_n^{\mathfrak{s},j+1}}^{b,\star}\}_{b=1}^{B_\tau}$ $\big(\mathfrak{s} \in \{\mathtt{1}, \mathtt{2}\}\big)$ *from* $\mathcal{S}_{\vec{Y}_n^{\mathfrak{s},j+1}}$ *via Algorithm 4.3.2 to obtain* B_τ *measures* $\{\widehat{\mu}_{\tau;n}^{\mathfrak{s},j+1,b,\star}\}_{b=1}^{B_\tau}$. *Then, approximate involved probabilities via* (4.3.1), *and decide:*

(1) *If* $\mathbb{P}^\star\big[\mathtt{d}^\star(\widehat{\mu}_{\tau;n}^{\mathtt{1},j+1,\star}, \widehat{\mu}_{\tau;n}^{\mathtt{2},j+1,\star}) \geqslant 2\mathtt{Tol}_\tau\big] > p_\tau^r$, *set* $\tau_n^{j+1} := {}^{\tau_n^{j+1}}\!/_2$.

(2) *If* $\mathbb{P}^\star\big[\mathtt{d}^\star(\widehat{\mu}_{\tau;n}^{\mathtt{1},j+1,\star}, \widehat{\mu}_{\tau;n}^{\mathtt{2},j+1,\star}) < \frac{1}{2}\mathtt{Tol}_\tau\big] > p_\tau^c$, *set* $\tau^{j+1} := 2\tau_n^{j+1}$ *and stop.*

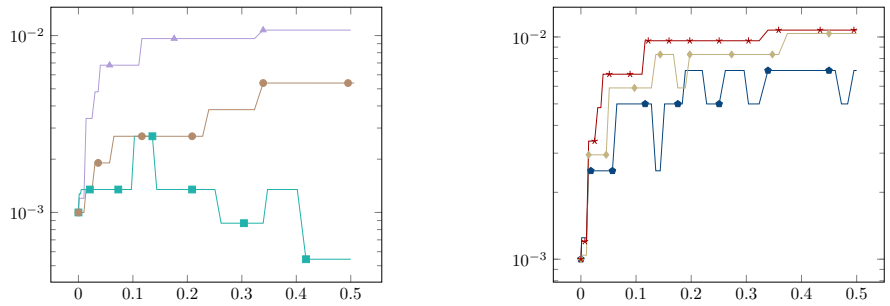
The probabilities $p_\tau^r, p_\tau^c \in [0,1]$ determine the probability of τ_n^{j+1} being refined/coarsened. If not specified otherwise, we choose $p_\tau^r = p_\tau^c = 95\%$.

Remark 4.6.4. The decision criteria in Step (IV) of Algorithm 4.6.3 involve the probability $\mathbb{P}^\star$ for the estimators $\mathtt{d}^\star$, for which the bootstrap method is used. Alternatively, a frequent sampling may be avoided by Fisher's χ^2-test of homogeneity for which the asymptotic distribution χ_{R-1}^2 is well-known (see Remark 4.6.1, item 1): for a fixed level of significance $\alpha \in (0,1)$, a local refinement of the time step size τ_n^{j+1} could be based on the rejection of $H_0 : \mathcal{L}(Y_{\tau;n}^{\mathtt{1},j+1}) = \mathcal{L}(Y_{\tau;n}^{\mathtt{2},j+1})$ by a *single* realization of $\mathtt{d}_{\chi^2}^2$ in (4.6.4), which is achieved by a modification of Step (IV) in Algorithm 4.6.3 to

- If $\mathtt{d}_{\chi^2}^{2,\star}\big(\widehat{\mu}_{\tau;n}^{\mathtt{1},j+1,\star}, \widehat{\mu}_{\tau;n}^{\mathtt{2},j+1,\star}\big) \geqslant c_\alpha$, set $\tau_n^{j+1} := {}^{\tau_n^{j+1}}\!/_2$.

Coarsening of the previously found τ_n^{j+1} (obtained by successive refinement) may be carried out by performing an additional *single* (i.e., $\mathtt{N}_\tau = 0$) corresponding Fisher χ^2-test of homogeneity with $\tau_0^{j+1} := 2\tau^{j+1}$. The simulations in Figure 4.9(b) evidence an oscillatory behavior of $t_j \mapsto \tau^j$ for larger values of α in the case of Fisher's χ^2-test. In fact, we see the following principle drawbacks that spontaneous refining/coarsening of τ_n^{j+1} via Fisher's χ^2-test suffers from:

1) The computational studies show a dependence of the adaptive time step size $\{\tau^j\}_j$ on the sample size M_τ; see Figure 4.9(a). In particular, Figure 4.9(a) illustrates smaller time step sizes for growing values of M_τ, which restricts the flexibility of the adaptive sampling algorithm, in particular for singular dynamics (see Examples 4.7.1, 4.7.3); see also Figure 4.15(a) as reference.



(a) $t_j \mapsto \tau^j$ **(b)** $t_j \mapsto \tau^j$

Figure 4.9: (Example 4.7.1 for the Algorithm in Remark 4.6.4, item 1) for fixed $\mathcal{T}_h^j \equiv \mathcal{T}_h^0$ with $h^0 = 2^{-6}$, $\tau^0 = 10^{-3}$, $\mathtt{N}_\tau = 0$, and $R_\tau = 2^{12}$) (a) Behavior of $t_j \mapsto \tau^j$ for varying $M_\tau \in \{10^3$ (—▲—$), 10^4$ (—●—$), 10^5$ (—■—$)\}$ and fixed $\alpha = 0.01$. (b) Adaptive time meshes for varying level of significance $\alpha \in \{0.10$ (—⬟—$), 0.05$ (—◆—$), 0.01$ (—★—$)\}$ and fixed $M_\tau = 10^3$.

2) The error of second kind in this adaptive strategy is not controllable, which affects the stable selection of local time step sizes $\{\tau^j\}_j$. For example, in the case of refinement, this leads to unnecessarily small time step sizes.

4.6.4 Space adaptivity

Let $j \in \mathbb{N}_0$ be fixed, $M_h \in \mathbb{N}$, $\mathtt{N}_h \in \mathbb{N}_0$, and $\tau^{j+1} > 0$ be chosen as detailed in Section 4.6.3. We generate a finite sequence $\{\mathcal{T}_{h;n}^{j+1}\}_{n=0}^{\mathtt{N}_h}$ resp. $\{\mathbb{V}_{h;n}^{j+1}\}_{n=0}^{\mathtt{N}_h}$, with $\mathcal{T}_{h;0}^{j+1} := \mathcal{T}_h^j$ resp. $\mathbb{V}_{h;0}^{j+1} := \mathbb{V}_h^j$ via the ZZ estimator by Zienkiewicz and Zhu [ZZ92] to (possibly) adaptively refine/coarsen each element $K \in \mathcal{T}_{h;n}^{j+1}$. Therefore, for each $n \in \{0, \dots, \mathtt{N}_h\}$, we obtain a family of parametrized empirical probability measures

$$\widehat{\mu}_{h;n}^{\mathfrak{s},j+1} = \Big\{\widehat{\mu}_{h;K;n}^{\mathfrak{s},j+1};\, K \in \mathcal{T}_{h;n}^{j+1}\Big\} \quad \text{with} \quad \widehat{\mu}_{h;K;n}^{\mathfrak{s},j+1} \in \mathscr{P}_{h;K;n}^{j+1}\big(\mathbb{R}_0^+\big) \quad \big(\mathfrak{s} \in \{\mathtt{1}, \mathtt{2}\}\big)$$

whose construction is now detailed. For each element $K \in \mathcal{T}_{h;n}^{j+1}$, we compare $|\nabla Y_n^{j+1}|_K$ with $|G_h(\nabla Y_n^{j+1})|_K$ to detect spatial changes in the solution, where the recovered gradient denoted by $G_h(\nabla Y_n^{j+1}) \in \mathbb{V}_{h;n}^{j+1}$ is computed via (cf. Section 4.1.1)

$$G_h\big(\nabla Y_n^{j+1}\big)(\mathbf{x}_{\ell;n}) := \frac{1}{|w_{\mathbf{x}_{\ell;n}}|} \int_{w_{\mathbf{x}_{\ell;n}}} \nabla Y_n^{j+1}(\mathbf{x})\,\mathrm{d}\mathbf{x}\,,$$

and $w_{\mathbf{x}_{\ell;n}} := \bigcup\big\{K \in \mathcal{T}_{h;n}^{j+1} : \mathbf{x}_{\ell;n} \in \overline{K}\big\}$ denotes the patch associated to $\mathbf{x}_{\ell;n} \in D$.

Let $\mathcal{S}_{Y_n^{j+1}}$ be an M_h-sample of realizations of $Y_n^{j+1} \in L^2\big(\Omega; \mathbb{V}_{h;n}^{j+1}\big)$ computed with τ^{j+1} from Scheme 4.5.1. Our goal is now to measure the distance $\mathtt{d}$ between the (unknown) probability

measures

$$\mu^{\mathfrak{1},j+1}_{h;K;n} := \mathcal{L}\big(|\nabla Y^{j+1}_n|_K\big),$$
$$\mu^{\mathfrak{2},j+1}_{h;K;n} := \mathcal{L}\big(|G_h(\nabla Y^{j+1}_n)|_K\big)$$

for each element $K \in \mathcal{T}^{j+1}_{h;n}$. We want to approximately compute the distance $\mathtt{d}\big(\widehat{\mu}^{\mathfrak{1},j+1}_{h;K;n}, \widehat{\mu}^{\mathfrak{2},j+1}_{h;K;n}\big)$ of related empirical laws $\big\{\widehat{\mu}^{\mathfrak{s},j+1}_{h;K;n};\, \mathfrak{s} \in \{\mathfrak{1},\mathfrak{2}\}\big\}$ to steer refinement/coarsening of $K \in \mathcal{T}^{j+1}_{h;n}$. For this purpose, the M_h-sample $\mathcal{S}_{Y^{j+1}_n}$ is restricted to a single $K \in \mathcal{T}^{j+1}_{h;n}$, to sample on the underlying partition $\widehat{\mathcal{P}}^{j+1}_{h;R_h;K;n} := \bigcup_{r=1}^{R_h} \widehat{C}_{h;r;K;n}$ of $\mathbb{R}^+_0$ which is obtained via Algorithm 4.2.7,

$$\widehat{\mathcal{P}}^{j+1}_{h;R_h;K;n} \equiv \widehat{\mathcal{P}}^{j+1}_{h;R_h;n}\big(\mathcal{S}^K_{Y^{\mathfrak{1},\mathfrak{2};j+1}_n}\big) = \big\{\widehat{C}^{j+1}_{h;r;K;n}\big\}_{r=1}^{R_h},$$

where $\mathcal{S}^K_{Y^{\mathfrak{1},\mathfrak{2};j+1}_n} := \mathcal{S}^K_{Y^{\mathfrak{1},j+1}_n} \cup \mathcal{S}^K_{Y^{\mathfrak{2},j+1}_n}$ and

$$\begin{aligned}
\mathcal{S}^K_{Y^{\mathfrak{1},j+1}_n} &:= \big\{|\nabla Y^{j+1}_n(\omega_k)|_K\big\}_{k=1}^{M_h},\\
\mathcal{S}^K_{Y^{\mathfrak{2},j+1}_n} &:= \big\{|G_h(\nabla Y^{j+1}_n(\omega_k))|_K\big\}_{k=1}^{M_h}
\end{aligned} \tag{4.6.5}$$

for each element $K \in \mathcal{T}^{j+1}_{h;n}$. We then sample again — and refer below again to the new samples as $\mathcal{S}^K_{Y^{\mathfrak{s};j+1}_n}$ $\big(\mathfrak{s} \in \{\mathfrak{1},\mathfrak{2}\}\big)$ — , to compute frequency vectors $\widehat{\boldsymbol{\nu}}^{\mathfrak{s},j+1}_{h;K;n} := (\widehat{\nu}^{\mathfrak{s},j+1}_{h;1;K;n}, \ldots, \widehat{\nu}^{\mathfrak{s},j+1}_{h;R_h;K;n})^T$ and associated index sets $\widehat{A}^{\mathfrak{1},j+1}_{h;r;K;n}$ for $r \in \{1,\ldots,R_h\}$, where

$$\begin{aligned}
\widehat{\nu}^{\mathfrak{1},j+1}_{h;r;K;n} &:= \#\widehat{A}^{\mathfrak{1},j+1}_{h;r;K;n}, \quad \widehat{A}^{\mathfrak{1},j+1}_{h;r;K;n} := \big\{k;\, |\nabla Y^{j+1}_n(\omega_k)|_K \in \widehat{C}^{j+1}_{h;r;K;n}\big\},\\
\widehat{\nu}^{\mathfrak{2},j+1}_{h;r;K;n} &:= \#\widehat{A}^{\mathfrak{2},j+1}_{h;r;K;n}, \quad \widehat{A}^{\mathfrak{2},j+1}_{h;r;K;n} := \big\{k;\, |G_h(\nabla Y^{j+1}_n(\omega_k))|_K \in \widehat{C}^{j+1}_{h;r;K;n}\big\},
\end{aligned}$$

and thus obtain empirical measures $\widehat{\mu}^{\mathfrak{s},j+1}_{h;K;n} \in \mathscr{P}^{j+1}_{h;K;n}(\mathbb{R}^+_0)$ via

$$\widehat{\mu}^{\mathfrak{s},j+1}_{h;K;n} = \sum_{r=1}^{R_h} \widehat{q}^{\mathfrak{s},j+1}_{h;r;K;n} \cdot \widehat{\xi}^{\mathfrak{s},j+1}_{h;r;K;n} \quad \big(\mathfrak{s} \in \{\mathfrak{1},\mathfrak{2}\}\big), \tag{4.6.6}$$

see e.g. Figure 4.4(b) for illustration, where $\widehat{q}^{\mathfrak{s},j+1}_{h;r;K;n} := \widehat{\nu}^{\mathfrak{s},j+1}_{h;r;K;n}/M_h$, and

$$\begin{aligned}
\widehat{\xi}^{\mathfrak{1},j+1}_{h;r;K;n} &:= \frac{1}{\widehat{\nu}^{\mathfrak{1},j+1}_{h;r;K;n}} \sum_{k \in \widehat{A}^{\mathfrak{1},j+1}_{h;r;K;n}} \delta_{|\nabla Y^{j+1}_n(\omega_k)|_K},\\
\widehat{\xi}^{\mathfrak{2},j+1}_{h;r;K;n} &:= \frac{1}{\widehat{\nu}^{\mathfrak{2},j+1}_{h;r;K;n}} \sum_{k \in \widehat{A}^{\mathfrak{2},j+1}_{h;r;K;n}} \delta_{|G_h(\nabla Y^{j+1}_n(\omega_k))|_K}.
\end{aligned}$$

Thus, we obtain two families of empirical measures $\big\{\widehat{\mu}^{\mathfrak{s},j+1}_{h;K;n};\, K \in \mathcal{T}^j_{h;n}\big\}$ $\big(\mathfrak{s} \in \{\mathfrak{1},\mathfrak{2}\}\big)$ for every $n \in \{0,\ldots,\mathtt{N}_h\}$. To handle degenerate noise, we weight the probabilities in (4.6.2) with the difference of the first moments of measures $\big\{\widehat{\mu}^{\mathfrak{s},j+1}_{h;K;n};\, K \in \mathcal{T}^j_{h;n}\big\}$ $\big(\mathfrak{s} \in \{\mathfrak{1},\mathfrak{2}\}\big)$. Let

$$\begin{aligned}
\widehat{\eta}^{\mathfrak{1},j+1}_{h;K;n} &:= \mathbb{E}\big[|\nabla Y^{j+1}_n|_K\big],\\
\widehat{\eta}^{\mathfrak{2},j+1}_{h;K;n} &:= \mathbb{E}\big[|G_h(\nabla Y^{j+1}_n)|_K\big].
\end{aligned} \tag{4.6.7}$$

Algorithm 4.6.5 (Adaptivity in space). *Fix $j \geqslant 0$, $\mathtt{Tol}_h > 0$, $M_h \in \mathbb{N}$, and $\mathtt{N}_h \in \mathbb{N}_0$. Choose a distance $\mathtt{d} \in \{\mathtt{d}_{\mathtt{H}}, \widetilde{\mathtt{d}}_{\mathtt{KL}}, \mathtt{d}_{\mathtt{TV}}\}$, $B_h \in \mathbb{N}$, and $p_h^r, p_h^c, p_h \in [0,1]$. Set $\mathcal{T}_{h;0}^{j+1} := \mathcal{T}_h^j$, as well as $\mathbb{V}_{h;0}^{j+1} := \mathbb{V}_h^j$.*
For $n = 0, \dots, \mathtt{N}_h$ do:

(I) Compute a M_h-sample $\mathcal{S}_{Y_n^{j+1}} := \{Y_n^{j+1}(\omega_k)\}_{k=1}^{M_h}$ for τ^{j+1} with new $\{\xi_{j+1}(\omega_k)\}_k$.

(II) For each $K \in \mathcal{T}_{h;n}^{j+1}$ do:

(1) Create $\widehat{\mathcal{P}}_{h;R_h;K;n}^{j+1}$ based on $\mathcal{S}_{Y_n^{1,z;j+1}}^K$ in (4.6.5) via Algorithm 4.2.7.

(2) Generate B_h new bootstrap samples $\{\mathcal{S}_{Y_n^{\mathfrak{s},j+1}}^{K,b,\star}\}_{b=1}^{B_h}$ $\big(\mathfrak{s} \in \{\mathtt{1},\mathtt{2}\}\big)$ from $\mathcal{S}_{Y_n^{\mathfrak{s},j+1}}^{K,b,\star}$ via Algorithm 4.3.2 to approximate the first moments $\{\widehat{\eta}_{h;K;n}^{\mathfrak{s},j+1,\star}\}_{b=1}^{B_h}$ in (4.6.7).

(3) Generate B_h new bootstrap samples $\{\mathcal{S}_{Y_n^{\mathfrak{s},j+1}}^{K,b,\star}\}_{b=1}^{B_h}$ $(\mathfrak{s} \in \{\mathtt{1},\mathtt{2}\})$ from $\mathcal{S}_{Y_n^{\mathfrak{s},j+1}}^{K,b,\star}$ via Algorithm 4.3.2 to obtain B_h measures $\{\widehat{\mu}_{h;K;n}^{\mathfrak{s},j+1}\}_{b=1}^{B_h}$. Then, approximate involved probabilities via (4.3.1), and decide:

(a) If $\big|\widehat{\eta}_{h;K;n}^{\mathtt{1},j+1,\star} - \widehat{\eta}_{h;K;n}^{\mathtt{2},j+1,\star}\big| \mathbb{P}^\star\big[\mathtt{d}^\star(\widehat{\mu}_{h;K;n}^{\mathtt{1},j+1,\star}, \widehat{\mu}_{h;K;n}^{\mathtt{2},j+1,\star}) > 2\mathtt{Tol}_h\big] > p_h^r$, mark K for refinement.

(b) If $\big|\widehat{\eta}_{h;K;n}^{\mathtt{1},j+1,\star} - \widehat{\eta}_{h;K;n}^{\mathtt{2},j+1,\star}\big| \mathbb{P}^\star\big[\mathtt{d}^\star(\widehat{\mu}_{h;K;n}^{\mathtt{1},j+1,\star}, \widehat{\mu}_{h;K;n}^{\mathtt{2},j+1,\star}) < \frac{1}{2}\mathtt{Tol}_h\big] > p_h^c$, mark K for coarsening.

(III) If $\max_{K \in \mathcal{T}_{h;n}^{j+1}} \big|\widehat{\eta}_{h;K;n}^{\mathtt{1},j+1,\star} - \widehat{\eta}_{h;K;n}^{\mathtt{2},j+1,\star}\big| \mathbb{P}^\star\big[\mathtt{d}^\star(\widehat{\mu}_{h;K;n}^{\mathtt{1},j+1,\star}, \widehat{\mu}_{h;K;n}^{\mathtt{2},j+1,\star}) \leqslant \mathtt{Tol}_h\big] > p_h$, set $\mathcal{T}_h^{j+1} := \mathcal{T}_{h;n}^{j+1}$, $\mathbb{V}_h^{j+1} := \mathbb{V}_{h;n}^{j+1}$ and stop; otherwise continue.

(IV) Obtain the new mesh $\mathcal{T}_{h;n+1}^{j+1}$ from $\mathcal{T}_{h;n}^{j+1}$ by local refinement resp. coarsening of the elements $K \in \mathcal{T}_{h;n}^{j+1}$ marked in the previous steps.

In accordance with time adaptivity, the values $p_h^r, p_h^c, p_h \in [0,1]$ determine the probability of $K \in \mathcal{T}_{h;n}^{j+1}$ being refined/coarsened. For each $K \in \mathcal{T}_h^j$, the first moments in Algorithm 4.6.5 are approximated via

$$\begin{aligned} \widehat{\eta}_{h;K;n}^{\mathtt{1},j+1,\star} &:= \mathbb{E}^\star\big[|\nabla Y_n^{j+1}|_K\big] \approx \frac{1}{B_h}\sum_{b=1}^{B_h} \big|\nabla Y_n^{j+1}(\omega_b)\big|_K, \\ \widehat{\eta}_{h;K;n}^{\mathtt{2},j+1,\star} &:= \mathbb{E}^\star\big[\big|G(\nabla Y_n^{j+1})\big|_K\big] \approx \frac{1}{B_h}\sum_{b=1}^{B_h} \big|G\big(\nabla Y_n^{j+1}(\omega_b)\big)\big|_K. \end{aligned} \tag{4.6.8}$$

4.7 Computational experiments

We computationally study stability/accuracy of the adaptive algorithms from Sections 4.6.3–4.6.4 with respect to the parameters (M_τ, B_τ, R_τ) (analogously for M_h, B_h, R_h in space), as well as the distances $\mathtt{d} \in \{\mathtt{d}_{\mathtt{H}}, \widetilde{\mathtt{d}}_{\mathtt{KL}}, \mathtt{d}_{\mathtt{TV}}\}$. For this purpose, we employ random number generators from the GNU Scientific Library [Gal09]. Local mesh refinement and coarsening of $\mathcal{T}_{h;n}^{j+1}$ in Algorithm 4.6.5 is performed using a bisection algorithm, and are based on the finite element code ALBERTA, cf. [SS05]. All computations are performed on an Intel Core i5-4670 3.40GHz

processor with 16GB RAM in double precision arithmetic. Parallelized calculations are performed using OpenMP, cf. [DM98]. Arising linear algebraic systems are solved by Gaussian elimination, and the Software package [Dav04].

The computational studies for Example 4.7.1 resp. Example 4.7.3 evidence that

- a robust adaptive time mesh may be obtained for all distances $\mathtt{d} \in \{\mathtt{d_H}, \widetilde{\mathtt{d}}_{\mathtt{KL}}, \mathtt{d_{TV}}\}$ in Algorithm 4.6.3 via partitions $\{\widehat{\mathcal{P}}^j_{\tau;R^j_\tau}\}_j$ (e.g. uniformly $M^j_\tau \approx 10^4$, and $R^j_\tau \approx 10^2$ for Example 4.7.1, as opposed to $M^j_\tau \approx 10^6$, and $R^j_\tau \approx 10^3$ near the discrete blow-up time for Example 4.7.3) on which the empirical probability measures $\{\widehat{\mu}^{\mathfrak{s}}_\tau;\, \mathfrak{s} \in \{\mathtt{1}, \mathtt{2}\}\}$ may be compared with the help of the bootstrap method.
- an efficient adaptive space mesh may be obtained for $\mathtt{d} = \mathtt{d_{TV}}$ in Algorithm 4.6.5, while $\mathtt{d} \in \{\mathtt{d_H}, \widetilde{\mathtt{d}}_{\mathtt{KL}}\}$ leads to unnecessarily fine meshes; cf. Table 4.5.
- fast computations of bootstrap estimators as $\mathbb{P}^\star$, $\mathbb{E}^\star$ and $\mathtt{d}^\star$ require bootstrap.
- next to much smaller errors on adaptive space-time meshes, the empirical variance of the iterates $\{Y^j\}_j$ is significantly reduced, leading to reduced samples sizes M^j_τ (resp. M^j_h) and less cells R^j_τ (resp. R^j_h) for a coarser partition $\widehat{\mathcal{P}}^j_{\tau;R^j_\tau}$ (resp. $\widehat{\mathcal{P}}^j_{h;R^j_h;K}$).
- adaptivity of all involved discretization and statistical parameters is necessary to accurately resolve singular behaviors of the solution.

4.7.1 Computational experiments for an SPDE (4.4.4)

We use Scheme 4.5.1 in combination with space-time adaptivity to approximate the solution of the convection-dominated linear SPDE (4.4.4).

Example 4.7.1. *Let $D = (0,1)^2$, $\boldsymbol{\beta} \equiv \boldsymbol{\beta}(x_1, x_2) = (\frac{1}{2} - x_2, x_1 - \frac{1}{2})^T$, and $\iota \in \{0.1, 0.3, 0.5\}$. Consider ($t \in (0,T]$)*

$$\mathrm{d}X_t - \epsilon\Delta X_t \mathrm{d}t + \theta\boldsymbol{\beta}\cdot\nabla X_t \mathrm{d}t = \iota\big[1 + |X_t|\big] \sum_{0\leqslant|\mathbf{k}|\leqslant n} \sigma_{\mathbf{k}}(t)\mathrm{d}W^{\mathbf{k}}(t), \qquad X_0 = x_0, \tag{4.7.1}$$

with $X_t = 0$ on ∂D_T, $\varepsilon = 10^{-8}$, $\theta = 1.0$ (if not specified otherwise), and $n = 3$,

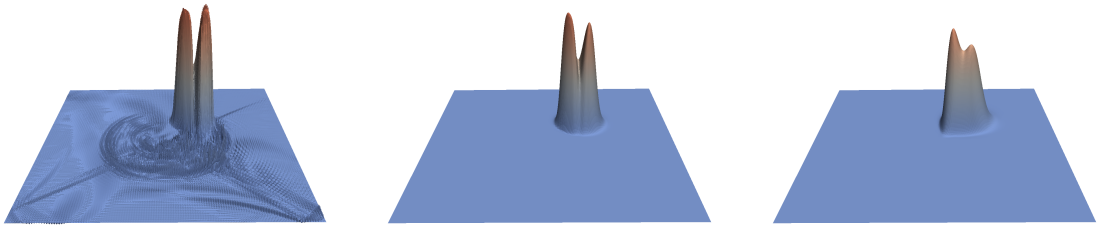
$$\sigma_{\mathbf{k}}(t,\mathbf{x}) = \sqrt{2}\prod_{i=1}^{2}\left(\frac{2}{(2k_i+1)\pi}\right)^2 \sin(k_i\pi x_i)$$

for all $\mathbf{x} = (x_1, x_2)^T \in D$, and multi-indices $\mathbf{k} = (k_1, k_2)^T \in \mathbb{N}_0^2$ with $|\mathbf{k}| = k_1 + k_2$. The initial datum x_0 is given by the slotted cylinder; see [JS08] for an explicit formula of it.

4.7.1.1 Time adaptivity

Figure 4.10 illustrates the stability properties of Scheme 4.5.1 for a uniform space-time mesh: the standard Galerkin method (i.e., $\delta^j_K \equiv 0$) yields highly oscillatory expectations of iterates to approximate (4.7.1); this (global Gibbs) phenomenon is well-known for convection dominated PDEs, and here is amplified by the noise; see Figure 4.10(a). Corresponding simulations of higher moments show a significant reduction of spurious oscillations outside the diffuse layers.

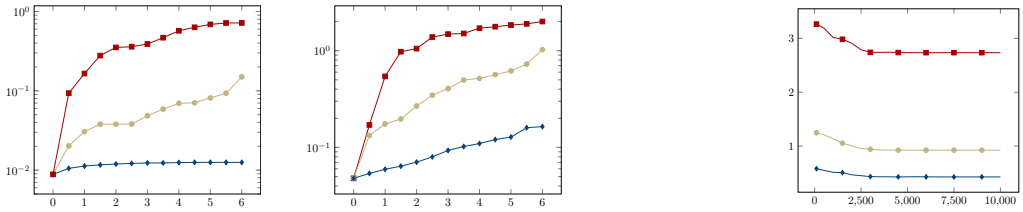
The simulations in Figure 4.10(b) via the SUPG based Scheme 4.5.1 show improved stability properties without oscillatory patterns attached to the cylindric profile. The stabilization parameter $\delta_K^j = \min\{h_K^j, \tau^j/2\}$ suitably balances stabilization effects both, with respect to temporal and spatial discretization scales; see Lemma 4.5.2, and Figure 4.11(a)–4.11(b). In



(a) $\delta_K^j = 0$ **(b)** $\delta_K^j = \min\{h_K^j, \tau^j/2\}$ **(c)** $\delta_K^j = \mathcal{O}(h_K^j)$

Figure 4.10: (Example 4.7.1 for $T = \pi/2$, and $\texttt{Tol}_\tau = 0.05$, $M_\tau = 10^5$) $\mathbb{E}_{M_\tau}\left[|Y^j|\right]$ obtained with the Scheme 4.5.1 for uniform space-time meshes $\{(\tau^j, \mathcal{T}_h^j)\}_j$ with $h^j \equiv h^0 = 2^{-8}$, $\tau^j \equiv \tau^0 = 10^{-4}$, and $\iota = 0.1$: (a) Standard Galerkin FEM ($\delta_K^j = 0$), (b) $\delta_K^j = \min\{h_K^j, \tau^j/2\}$, (c) $\delta_K^j = \mathcal{O}(h_K^j)$.

addition, it leads to smaller empirical variances of the computed iterates $\{Y^j\}_j$; see Figure 4.11(c). Small-scale effects through W are not accurately recovered for $\delta_K^j = \mathcal{O}(h_K^j)$ which heavily diffuses the solution structure; see Figure 4.10(c). A comparison with the reference solution (uniform meshes with $h^j \equiv h^0 = 2^{-10}$, $\tau^j \equiv \tau^0 = 10^{-5}$ and $M_\tau = 10^5$) in Figure 4.11(a)–(b) shows that the solution obtained via $\delta_K^j = \min\{h_K^j, \tau^j/2\}$ is most accurate. Thus, the SUPG scheme with $\delta_K^j = \min\{h_K^j, \tau^j/2\}$ nicely combines accuracy with stability properties and is chosen for the simulations to follow.



(a) $t_j \mapsto \texttt{err}_0$ **(b)** $t_j \mapsto \texttt{err}_1$ **(c)**

$$M_\tau \mapsto \Big(\sum_{j=1}^{J} \tau^j \mathbb{E}_{M_\tau}\left[\left|\|Y^j\|_{\mathbb{H}^1} - \mathbb{E}_{M_\tau}\left[\|Y^j\|_{\mathbb{H}^1}\right]\right|^2\right]\Big)^{\frac{1}{2}}$$

Figure 4.11: (Example 4.7.1 with reference space-time meshes $\left\{(\tau^j, \mathcal{T}_h^j)\right\}_j$ for $h^j \equiv h^0 = 2^{-10}$, $\tau^j \equiv \tau^0 = 10^{-5}$. Time adaptivity for $T = 2\pi$, $\iota = 0.3$, as well as $\texttt{d} = \texttt{d}_\mathtt{H}$, $\texttt{Tol}_\tau = 0.05$, and $R_\tau = 2^{12}$, $M_\tau = 10^5$) Behavior of $t_j \mapsto \texttt{err}_k := \left|\mathbb{E}_{M_\tau}\left[\|X_{t_j}\|_{\mathbb{W}^{k,\infty}}\right] - \mathbb{E}_{M_\tau}\left[\|Y^j\|_{\mathbb{W}^{k,\infty}}\right]\right|$ for $(k = 0)$ (a), $(k = 1)$ (b), and empirical variance (c) for the standard Galerkin FEM ($\delta_K^j = 0$) (—■—), $\delta_K^j = \min\{h_K^j, \tau^j/2\}$ (—◆—), and $\delta_K^j = \mathcal{O}(h_K^j)$ (—●—).

Next, we study the dependence on the parameters (M_τ,R_τ) to obtain a stable time-adaptive

mesh on D_T for different $\mathtt{d} \in \{\mathtt{d_H}, \widetilde{\mathtt{d}}_{\mathtt{KL}}, \mathtt{d_{TV}}\}$. Consider

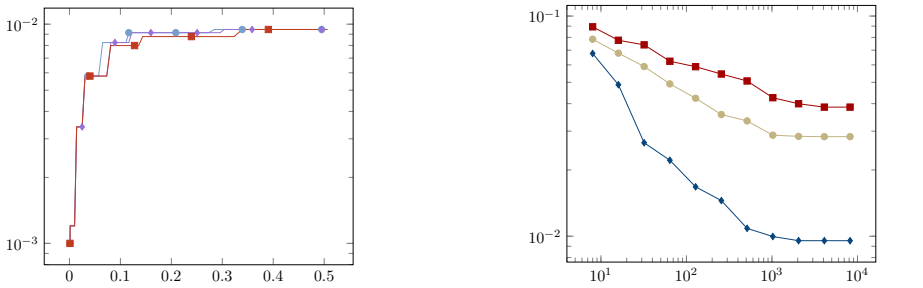
$$\mathcal{R}_\tau := \arg\min_{R_\tau} \left[\max_{t\in[0,T]} \left| \tau^{R_\tau}(t) - \tau^{2R_\tau}(t) \right| \leqslant \mathtt{Tol} \right] \tag{4.7.2}$$

with $\mathtt{Tol} := 10^{-2} \min_{t\in[0,T]} |\tau(t)|$, where $\tau(t) \equiv \tau^{R_\tau}(t)$ is the piecewise affine interpolation of the sequence $\{(t_j, \tau^j)\}_j$. Criterion (4.7.2) identifies the minimum value $\mathcal{R}_\tau \geqslant 1$, where the adaptive time mesh is not sensitive to refinement any more. The results in Figure 4.12(b) show that at least $\mathcal{R}_\tau \geqslant 2 \cdot 10^3$ cells are necessary to partition $\mathbb{R}^{L^j}$.

The empirical variances

$$\mathrm{Var}_{M_\tau}\left[\|Y^j\|_{\mathbb{W}^{k,2}} \right] := \mathbb{E}_{M_\tau}\left[\left| \|Y^j\|_{\mathbb{W}^{k,2}} - \mathbb{E}_{M_\tau}\left[\|Y^j\|_{\mathbb{W}^{k,2}} \right] \right|^2 \right] \quad k \in \{0,1\} \tag{4.7.3}$$

of iterates $\{Y^j\}_j$ increase in time, see Figure 4.14(a)–(c). Without convection ($\theta = 0.0$), a large number of cells R_τ is required only at small times, and the empirical variance decays at later times, which motivates coarser partitions and hence non-constant cell numbers $R_\tau \equiv R_\tau^j$ in time; see Section 4.7.1.3. This observation is common for all choices of $\mathtt{d}$; see Figure 4.12(a).

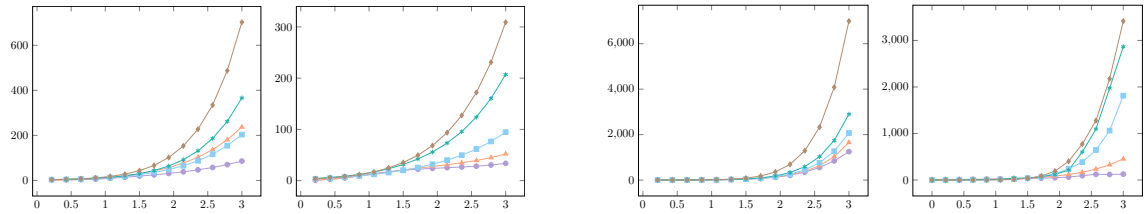


(a) $t_j \mapsto \tau^j$ **(b)** $R_\tau \mapsto \max_{t\in[0,T]} \left| \tau^{R_\tau}(t) - \tau^{2R_\tau}(t) \right|$

Figure 4.12: (Example 4.7.1 with uniform $\mathcal{T}_h^j \equiv \mathcal{T}_h^0$ for $h^j \equiv h^0 = 2^{-6}$. Time adaptivity for $T = 1/2$, $\iota = 0.3$, as well as $\tau^0 = 10^{-3}$, $\mathtt{Tol}_\tau = 0.05$, $\mathtt{N}_\tau = 8$, and $R_\tau = 2^{12}$, $M_\tau = 10^5$) (a) Adaptive time meshes for $\mathtt{d} = \mathtt{d_H}$ (—●—), $\mathtt{d} = \mathtt{d_{TV}}$ (—♦—), $\mathtt{d} = \widetilde{\mathtt{d}}_{\mathtt{KL}}$ (—■—), (b) and behavior of $R_\tau \mapsto \max_{t\in[0,T]} \left| \tau^{R_\tau}(t) - \tau^{2R_\tau}(t) \right|$ for $\iota = 0.1$ (—♦—), $\iota = 0.3$ (—●—), and $\iota = 0.5$ (—■—) with respect to $\mathtt{d} = \mathtt{d_H}$.

4.7.1.2 Space adaptivity

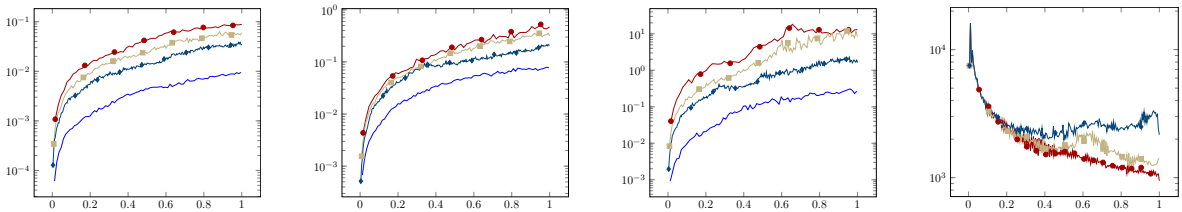
So far, time adaptivity was performed for a fixed uniform spatial mesh. The results in Figure 4.13 display the required number of cells $\mathcal{R}_\tau$ to meet (4.7.2) for varying convection resp. noise intensity θ resp. ι in Example 4.7.1, both, for uniform and adaptive space-time meshes, and dimensions $d \in \{1, 2\}$. The studies indicate that the number of required cells $\mathcal{R}_\tau$ to partition each $\mathbb{R}^{L^j}$ is significantly reduced in the case of space-time adaptivity: here, the meshes $\mathcal{T}_h^j$ resolve regions in space and time where large gradients are likely to occur; see also Figure 4.4(a). Besides an increased spatial resolution, we observe smaller empirical variances of computed realizations $\{Y^j\}_j$ (see Figure 4.14(a)–(c)) such that coupled space-time adaptivity can also be regarded as an importance sampling strategy. As a consequence, smaller



(a) $d = 1$: (left) uniform and (right) adaptive **(b)** $d = 2$: (left) uniform and (right) adaptive

Figure 4.13: (Example 4.7.1 for $T = 1.0$, as well as $\mathtt{d} = \mathtt{d_H}$, $\mathtt{Tol}_\tau = \mathtt{Tol}_h = 0.05$, $\mathtt{N}_\tau = \mathtt{N}_h = 8$, $h^0 = 2^{-5}$, $\tau^0 = 10^{-3}$, and $M_\tau = 10^5$) Number of required cells $\mathcal{R}_\tau$ to meet (4.7.2) for varying convection resp. noise intensity (θ, ι) with $\iota \in [0, 3]$ and $\theta = 1.0$ ( ), $\theta = 2.0$ ( ), $\theta = 3.0$ ( ), $\theta = 4.0$ ( ) and $\theta = 5.0$ ( ) on uniform and space-time adaptive meshes for $d = 1$ (a) and $d = 2$ (b).

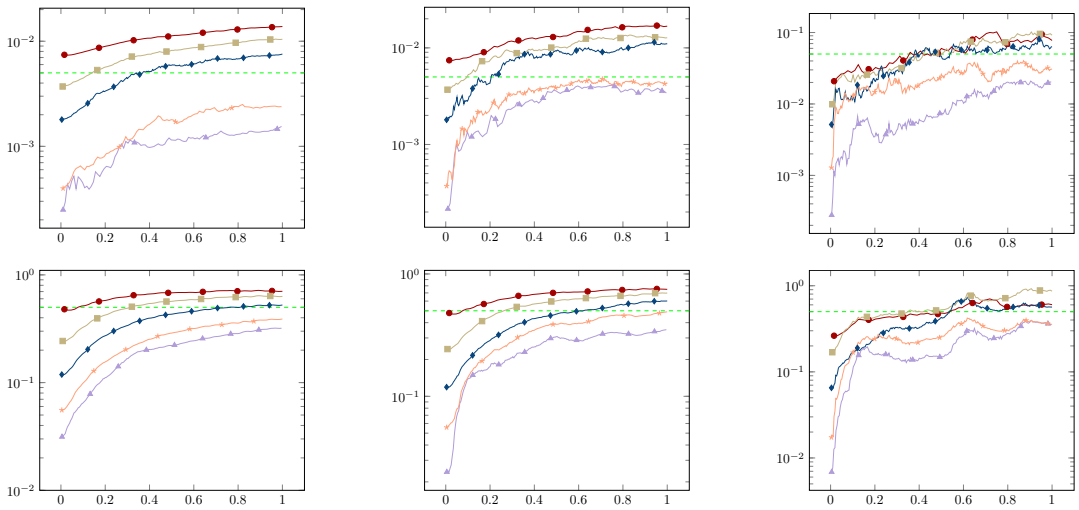
empirical variances of Y^j imply smaller confidence intervals and smaller image ranges $\vec{Y}^j[\Omega]$, such that smaller sample sizes M_τ (resp. M_h) and less cells R_τ (resp. R_h) are necessary to build an appropriate partition $\widehat{\mathcal{P}}^j_{\tau;R_\tau}$ (resp. $\widehat{\mathcal{P}}^j_{h;R_h;K}$).



(a) $\iota = 0.1$ **(b)** $\iota = 0.3$ **(c)** $\iota = 0.5$ **(d)** $t_j \mapsto L^j$

Figure 4.14: (Example 4.7.1 for $T = 1.0$, as well as $\mathtt{d} = \mathtt{d_H}$, $\mathtt{Tol}_\tau = \mathtt{Tol}_h = 0.05$, $\mathtt{N}_\tau = \mathtt{N}_h = 8$, $h^0 = 2^{-5}$, $\tau^0 = 10^{-3}$, and $M_\tau = 10^5$, $R_\tau = 10^3$) (a)–(c) Behavior of the empirical variance in (4.7.3) with $k = 0$ on uniform meshes of sizes $(2^{-i}h^0, 2^{-i}\tau^0)$ with $i = 0$ ( ), $i = 1$ ( ) and $i = 2$ ( ), and for adaptive ( ) space-time meshes. (d) Behavior of $t_j \mapsto L^j$ for noise intensity $\iota = 0.1$ ( ), $\iota = 0.3$ ( ), and $\iota = 0.5$ ( ).

The evolution of $t_j \mapsto L^j$ for different noise intensities ι is plotted in Figure 4.14(d). Related weak errors displayed in Figure 4.15 are much smaller if compared to a uniform discretization, or only time adaptivity: a significantly smaller number of performed time steps ($J = 83$) vs. the uniform grid is required to underrun a given threshold criterion for the error $\max_j \left| \mathbb{E}\left[\|X_{t_j}\|_{\mathbb{W}^{k,2}} \right] - \mathbb{E}\left[\|Y^j\|_{\mathbb{W}^{k,2}} \right] \right|$ ($k \in \{0, 1\}$).
So far, the distance $\mathtt{d} = \mathtt{d_H}$ was used for both, time and space adaptivity. Computational experiments indicate that the behavior of the adaptive time meshes is similar for all distances $\mathtt{d} \in \{\mathtt{d_H}, \widetilde{\mathtt{d}}_{\mathtt{KL}}, \mathtt{d_{TV}}\}$ in Algorithm 4.6.3. However, the choice is more critical for space adaptivity in Algorithm 4.6.5; cf. Figure 4.4(a) and Figure 4.16(a). We observe that the resolution of the diffuse layer in Example 4.7.1, and its width depend on $\mathtt{d}$. According to [GS02], $\mathtt{d_{TV}} \leqslant \mathtt{d_H} \leqslant \sqrt{\mathtt{d_{KL}}}$, and we indeed observe more elements $K \in \mathcal{T}^j_h$ for $\widetilde{\mathtt{d}}_{\mathtt{KL}}$ to resolve diffusive layers, if e.g. compared to meshes obtained via $\mathtt{d_{TV}}$; see Figure 4.16(a) and Table 4.5. This ordering for used distances is also reflected in the related histograms, see Figure 4.16(b), since e.g. small values of the realization of $\mathtt{d}$ favor coarsening of the corresponding $K \in \mathcal{T}^j_h$. These



(a) $\iota = 0.1$ **(b)** $\iota = 0.3$ **(c)** $\iota = 0.5$

Figure 4.15: (Example 4.7.1 for $\mathtt{d} = \mathtt{d_H}$, $\mathtt{Tol}_\tau = \mathtt{Tol}_h = 0.05$, $\mathtt{N}_\tau = \mathtt{N}_h = 8$, $h^0 = 2^{-5}$, $\tau^0 = 10^{-3}$, and $M_\tau = 10^5$, $M_h = 10^3$, $R_\tau = 10^3$, $R_h = 2^6$) Behavior of $t_j \mapsto \left|\mathbb{E}\left[\|X_{t_j}\|_{\mathbb{W}^{k,2}}\right] - \mathbb{E}\left[\|Y^j\|_{\mathbb{W}^{k,2}}\right]\right|$ for $t_j \in [0,1]$ on uniform time meshes with $J = 2^8$ ( ), $J = 2^9$ ( ) and $J = 2^{10}$ ( ) required time steps with respect to $k = 0$ (upper row) and $k = 1$ (lower row) for a given threshold ( ). Corresponding space-time adaptive meshes for uniform ( ) vs. adaptive ( ) statistical parameters (M_τ, B_τ, R_τ); see Section 4.7.1.3 for the latter.

results motivate to use $\mathtt{d_{TV}}$ which properly balances costs and accuracy; cf. also Figure 4.4 and 4.16.

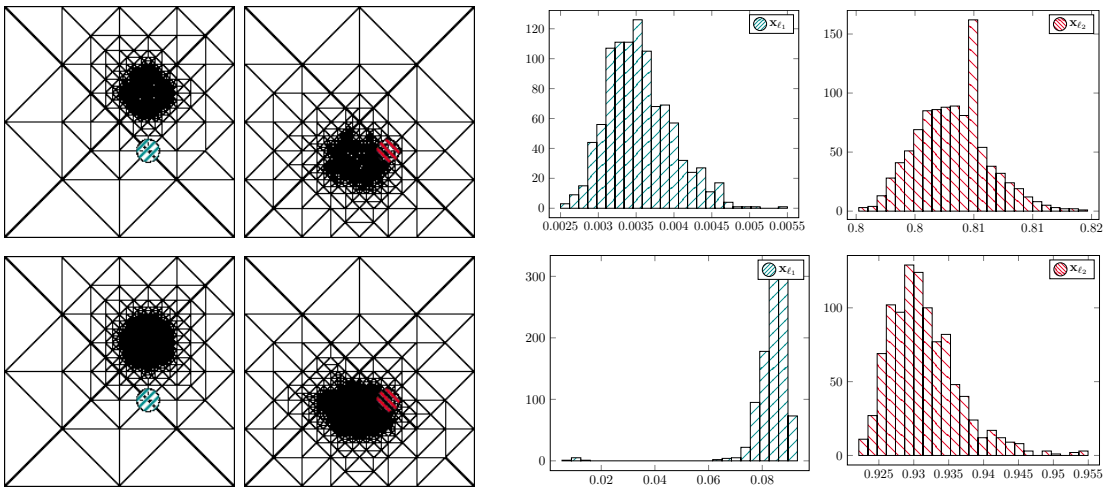
Table 4.5: (Example 4.7.1 for the same setup as in Figure 4.11) Error indicators $\mathtt{err}_k := \max_j \left|\mathbb{E}_{M_\tau}\left[\|X_{t_j}\|_{\mathbb{W}^{k,2}}\right] - \mathbb{E}_{M_\tau}\left[\|Y^j\|_{\mathbb{W}^{k,2}}\right]\right|$ for $(k \in \{0,1\})$, and maximum number of degrees of freedom $L_{\max} := \max_j L^j$.

$\mathtt{d}$	$\mathtt{d_H}$	$\mathtt{d_{TV}}$	$\widetilde{\mathtt{d}}_{\mathtt{KL}}$
$L_{\max}$	29700	20600	41200
$\mathtt{err}_0$	0.0223	0.0231	0.0244
$\mathtt{err}_1$	0.0941	0.0987	0.1022

The performed simulations discussed so far used a fixed number of cells $R_\tau \equiv R_\tau^j$, as well as M_τ, B_τ. In the following, we discuss the relevancy to select adaptively the statistical parameters M_τ, B_τ, and R_τ (analogously in space) also to further increase the efficiency of the numerical scheme.

4.7.1.3 Adaptive choice of statistical parameters M_τ, B_τ, R_τ

The following investigations refer to M_τ, R_τ, and B_τ, with analogous results for M_h, R_h, and B_h for each $K \in \mathcal{T}_h^j$. We recall the basic estimate for M_τ^j in terms of $\mathrm{Var}_{M_\tau^j}\left[\|Y^j\|_{\mathbb{L}^2}\right]$ to



(a) $\mathcal{T}_h^{j+1}$ at $t_{j+1} \in \{0.25, 3.75\}$ **(b)** Histogram of $\mathtt{d}^\star(\widehat{\mu}_{h;K}^{1,j+1,\star}, \widehat{\mu}_{h;K}^{2,j+1,\star})$ on $K \in \mathcal{T}_h^{j+1}$

Figure 4.16: (Example 4.7.1 for $\iota = 0.3$, as well as $\mathtt{Tol}_h = 0.05$, $\mathtt{N}_h = 1$, and $R_h = 2^5$, $M_h = 10^3$ for each $K \in \mathcal{T}_h^{j+1}$): (a) Snapshots of meshes $\mathcal{T}_h^{j+1}$ at evaluated times $t_{j+1} \in [0, 2\pi]$ with (b) corresponding histograms of $\mathtt{d}^\star(\widehat{\mu}_{h;K}^{1,j+1,\star}, \widehat{\mu}_{h;K}^{2,j+1,\star})$ with respect to $K \in \mathcal{T}_h^{j+1}$ containing two fixed points $\mathbf{x}_{\ell_1}$ (▨) and $\mathbf{x}_{\ell_2}$ (▨) for $\mathtt{d}_{\mathtt{TV}}$ (upper row) and $\widetilde{\mathtt{d}}_{\mathtt{KL}}$ (lower row).

achieve a certain accuracy:

$$\forall a > 0 : \lim_{M_\tau^j \uparrow \infty} \mathbb{P}\left[\left|\mathbb{E}\big[\|Y^j\|_{\mathbb{L}^2}\big] - \mathbb{E}_{M_\tau^j}\big[\|Y^j\|_{\mathbb{L}^2}\big]\right| \leqslant a\sqrt{\frac{\mathrm{Var}_{M_\tau^j}\left[\|Y^j\|_{\mathbb{L}^2}\right]}{M_\tau^j}}\right] = 2\Phi(a) - 1, \quad (4.7.4)$$

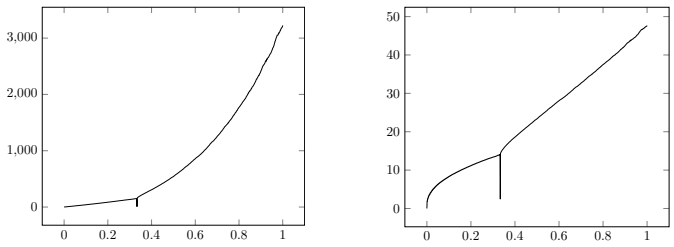
where Φ is the standard normal distribution. Let $\mathtt{Tol} > 0$, and $a := \Phi^{-1}(1 - \frac{\alpha}{2}) > 0$ for some level of significance $\alpha \in (0, 1)$. To reduce the statistical error in the MC estimation of $\mathbb{E}\big[\|Y^j\|_{\mathbb{L}^2}\big]$, we choose $M_\tau^j \in \mathbb{N}$ according to $M_\tau^j = \big\lfloor a^2 \,\mathrm{Var}_{M_\tau^j}\big[\|Y^j\|_{\mathbb{L}^2}\big] / \mathtt{Tol}^2 \big\rfloor$. Here, $a > 0$ defines the confidence interval with probability $1 - \alpha = 95\%$. If compared to uniformly chosen M_τ to e.g. achieve the results in Figure 4.15, the adaptive choice of the number of MC realizations $t_j \mapsto M_\tau^j$ is mainly influenced by the empirical variance reduction obtained via space-time adaptivity (see Figure 4.14(a)–(c)), which leads to overall smaller sample sizes in time; see Figure 4.17(b) below. Corresponding estimates for M_h^j are obtained in case we use $\|\nabla Y^j\|_{\mathbb{L}^2}$ (rather than $\|Y^j\|_{\mathbb{L}^2}$) in (4.7.4).

Analogously, we base the selection of the number of required bootstrap replications B_τ in (4.3.1) resp. (4.6.8) on the empirical variance of the estimators $\mathtt{d}^\star$, by replacing $\big(M_\tau^j, \|Y^j\|_{\mathbb{L}^2}\big)$ with $\big(B_\tau^j, \mathtt{d}^\star(\widehat{\mu}_{\tau;n}^{1,j,\star}, \widehat{\mu}_{\tau;n}^{2,j,\star})\big)$ in (4.7.4). For the simulations, if not specified otherwise, we choose at least $B_\tau^j \geqslant 10^3$ at each time t_j; cf. with the discussion in Section 4.3.

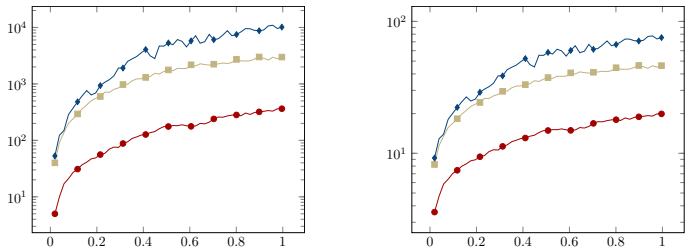
In order to adjust the number of cells $R_\tau = R_\tau^j$, we use a heuristic strategy suggested in [DS86, Chapter 3.2.4], which in particular benefits from the statistically equivalent cell property of the partition via (**BTC**): assuming that the adaptive sample size M_τ^j is large enough, we set

$$R_\tau^j = 4\sqrt[5]{2}\big(M_\tau^j / a\big)^{2/5}. \quad (4.7.5)$$

Hence, R_τ^j again depends on $\mathrm{Var}_{M_\tau^j}\big[\|Y^j\|_{\mathbb{L}^2}\big]$ through the choice of M_τ^j via (4.7.4). This behavior is also reflected for M_τ^j and R_τ^j in Example 4.0.1. The direct comparison of results



(a) $t_j \mapsto M_\tau^j$ (left) and $t_j \mapsto R_\tau^j$ (right).



(b) $t_j \mapsto M_\tau^j$ (left) and $t_j \mapsto R_\tau^j$ (right).

Figure 4.17: (Example 4.0.1 (a) and Example 4.7.1 (b) for $\|\cdot\|_{\mathbb{W}^{0,2}}$): Behavior of the required number of MC iterations M_τ^j resp. number of cells R_τ^j according to (4.7.4) resp. (4.7.5) for $\iota = 0.1$ (—●—), $\iota = 0.3$ (—■—), $\iota = 0.5$ (—♦—).

for uniformly vs. adaptively chosen statistical parameters in Figure 4.15 (resp. Figure 4.2) evidences an almost equal accuracy, while the overall computing time of 121380s in the first setting reduces to 98220s in the fully adaptive setting for the setup in Figure 4.17(b) with $\iota = 0.3$; moreover, by far less MC realizations are needed if compared to uniform statistical parameters (see Figure 4.17(b)) to meet the given error threshold.

4.7.2 Solving (non)-linear algebraic systems

The non-linear Newton scheme to approximate iterates $\{\boldsymbol{Y}^j\}_j$ from Scheme 4.5.3 can be formulated as follows.

Algorithm 4.7.2 (Non-linear Newton scheme)**.** *Fix* $j \in \mathbb{N}_0$. *Choose* $\mathtt{Tol} > 0$, $\mathtt{N} \in \mathbb{N}_0$, *and set* $\boldsymbol{Z}^{j+1,0} := \boldsymbol{Y}^j$.
For $n = 0, \dots, \mathtt{N}$ *do:*

(I) Compute $\boldsymbol{Z}^{j+1,n+1}$ *according to,* $\mathbb{P}$*-almost surely,*

$$D\boldsymbol{\mathcal{F}}^{j+1}\big(\boldsymbol{Z}^{j+1,n}\big)\big(\boldsymbol{Z}^{j+1,n+1} - \boldsymbol{Z}^{j+1,n}\big) = -\boldsymbol{\mathcal{F}}^{j+1}\big(\boldsymbol{Z}^{j+1,n}\big),$$

where $D\boldsymbol{\mathcal{F}}^{j+1}$ *denotes the Jacobian of the mapping* $\boldsymbol{\mathcal{F}}^{j+1} : \mathbb{V}_h^{j+1} \to \mathbb{V}_h^{j+1}$.

(II) If $\|\boldsymbol{Z}^{j+1,n+1} - \boldsymbol{Z}^{j+1,n}\| < \mathtt{Tol}$, *set* $\boldsymbol{Y}^{j+1} := 2\boldsymbol{Z}^{j+1,n+1} - \boldsymbol{Y}^j$ *and stop.*

As in Section 4.7, arising linear algebraic systems are solved by Gaussian elimination, and the Software package [Dav04]. According to Theorem 4.5.4, the convergence of the Newton method is ensured by the Kantorovich theorem [Deu11]. Computational studies show that the iterates obtained by Algorithm 4.7.2 are generally too long, i.e., $|\boldsymbol{Y}^{j+1}(\mathbf{x}_\ell)| \geqslant 1$ $\mathbb{P}$-almost surely for

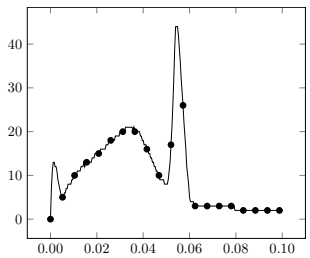
all $\ell \in \{1, \dots, L^j\}$; see Figure 4.18(b) in which the behavior of $t_j \mapsto \max_\ell \mathbb{E}\Big[\big|1 - |\boldsymbol{Y}^j(\mathbf{x}_\ell)|\big|\Big]$ is plotted. To overcome this, a point-wise re-normalization of Newton iterates onto $\mathbb{S}^2$ is proposed in [GR15]; more precisely, in Step (I) of Algorithm 4.7.2, set $\mathbb{P}$-almost surely

$$\boldsymbol{Z}^{j+1,n+1}(\mathbf{x}) := \sum_{\ell=1}^{L^{j+1}} \frac{\boldsymbol{Z}^{j+1,n+1}(\mathbf{x}_\ell)}{|\boldsymbol{Z}^{j+1,n+1}(\mathbf{x}_\ell)|} \Psi_\ell(\mathbf{x}) \quad \forall\, \mathbf{x} \in D.$$

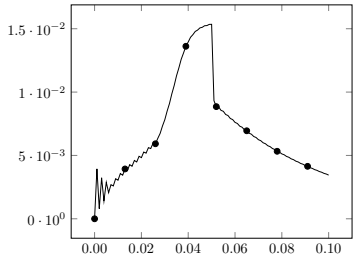
Furthermore, we observe in the simulations carried out in Section 4.7.3 and 4.7.4 the convergence condition $\tau^j = \mathcal{O}(h_{\min}^2)$, where $h_{\min} := \min_j \min_{K \in \mathcal{T}_h^j} h_K^j$, for the Newton method. However, this restrictive condition applies only to a small neighborhood around $j^* \in \mathbb{N}$, i.e.

$$\tau^j \approx \begin{cases} h_{\min} & \text{for } j \gg j^* \text{ or } j \ll j^*, \\ h_{\min}^2 & \text{for } j \approx j^*, \end{cases} \tag{4.7.6}$$

see Figure 4.18(a) where the evolution $t_j \mapsto \mathbb{E}\big[\#\{n \in \mathbb{N}_0;\, \|\boldsymbol{Z}^{j+1,n+1} - \boldsymbol{Z}^{j+1,n}\| < \mathtt{Tol}\}\big]$ is plotted to compute $\mathbb{E}\big[\boldsymbol{Y}^{j+1}\big]$. It can be seen that less iterations are needed away from $j^* \in \mathbb{N}$. This observation leads to the assumption that τ^j is not restricted by h_K^j, but by the rate of change in the discrete solution.



(a) $t_j \mapsto \mathbb{E}\big[\#\{n \in \mathbb{N}_0;\, \big\|\boldsymbol{Z}^{j+1,n+1} - \boldsymbol{Z}^{j+1,n}\big\| < 10^{-8}\}\big]$



(b) $t_j \mapsto \max_\ell \mathbb{E}\big[|1 - |\boldsymbol{Y}^j(\mathbf{x}_\ell)||\big]$

Figure 4.18: (Example 4.7.3 for $T = 0.1$, $\iota = 1.0$, as well as $\mathtt{d} = \mathtt{d}_{\mathtt{TV}}$, $\mathtt{Tol}_\tau = \mathtt{Tol}_h = 0.05$, $\mathtt{N}_\tau = \mathtt{N}_h = 10$, $M_\tau^0 = M_h^0 = 10^3$, $R_\tau^0 = 2^8$, $R_h = 2^5$, and $h^0 = 2^{-4}$, $\tau^0 = 10^{-4}$, $h_{\min} = 2^{-6}$) (a) Behavior of $t_j \mapsto \mathbb{E}\big[\#\{n \in \mathbb{N}_0;\, \|\boldsymbol{Z}^{j+1,n+1} - \boldsymbol{Z}^{j+1,n}\| < 10^{-8}\}\big]$ to compute $\mathbb{E}\big[\boldsymbol{Y}^{j+1}\big]$ via Algorithm 4.7.2, and (b) corresponding distance of iterates to $\mathbb{S}^2$.

4.7.3 Computational experiments for the harmonic map heat flow to $\mathbb{S}^2$

We use the ($\mathbb{P}$-almost sure) length-preserving discretization Scheme 4.5.3 for $\zeta_2 = 1$ without precession term in combination with space-time and statistical adaptivity. The non-linear algebraic systems at each time step of Scheme 4.5.3 are solved by Newton's method from Section 4.7.2.

In [vdBW13], the authors studied (in-)stability of blow-up dynamics for (4.4.5) in the case $\iota = 0$. Below, we computationally study the role that noise exerts on the formation of this singular behavior, and 'discrete blow-up dynamics'. We start the evolution from a super-critical initial datum.

Example 4.7.3. *Let* $D = B_1(\mathbf{0}) := \{\mathbf{x} \in \mathbb{R}^2 \,|\, |\mathbf{x}| \leqslant 1\}$, *and* $\boldsymbol{x}_0 \in \mathbf{H}^1(D;\mathbb{S}^2)$ *be given by*

$$\boldsymbol{x}_0(\mathbf{x}) = \begin{cases} (0,0,-1)^T, & \text{for } |\mathbf{x}| > 1/2, \\ \left(\frac{2x_1 A}{A^2+|\mathbf{x}|^2}, \frac{2x_2 A}{A^2+|\mathbf{x}|^2}, \frac{A^2-|\mathbf{x}|^2}{A^2+|\mathbf{x}|^2}\right)^T & \text{for } |\mathbf{x}| \leqslant 1/2, \end{cases} \quad \forall\, \mathbf{x} \in D, \tag{4.7.7}$$

and $A \equiv A(\mathbf{x}) := (1-2|\mathbf{x}|)^4$. *The noise term in* (4.5.2) *is simulated by* ($n = 3$)

$$\boldsymbol{\xi}_{j+1}(\mathbf{x}) = \sum_{0\leqslant|\mathbf{k}|\leqslant n} \sum_{l=1}^{3} \sqrt{2} \prod_{i=1}^{2} \left(\frac{2}{(2k_i+1)\pi}\right)^2 \sin(k_i \pi x_i) \mathbf{e}_l \xi_{j+1}^{\mathbf{k},l} \quad \forall\, \mathbf{x} \in D,$$

where $\xi_{j+1}^{\mathbf{k},l}$ *are i.i.d.* $\mathbb{R}$*-valued Brownian increments, and* $\mathbf{e}_l$ *for* $l \in \{1,2,3\}$ *are the canonical basis vectors of* $\mathbb{R}^3$.

For $\iota = 0$, a corresponding computational study on uniform space-time meshes for (4.7.7) in [BBNP14b] supports a discrete blow-up at the origin $\mathbf{x}_{\ell^*} := (0,0)^T$ of the domain D at time $\widetilde{t}_{j^*} := \min\{t_j \in [0,T];\, \|\nabla \boldsymbol{Y}^j\|_{\mathbf{L}^\infty} \geqslant 1/h\}$. Once the maximum possible 'discrete gradient' is reached, it is followed by a rapid switching of $\boldsymbol{Y}^{j^*}(\mathbf{x}_{\ell^*})$, and the solution becomes almost homogeneous. This dynamics favors Example 4.7.3 to test the adaptive concepts described in Section 4.6. According to Figure 4.22 below, the computational studies for Example 4.7.3 via Scheme 4.5.3 and adaptivity indicate the occurrence of a 'discrete blow-up time' $t_{j^*} \approx 0.05$, where

$$t_{j^*} := \min\Big\{t_j \in [0,T];\, \exists K \in \mathcal{T}_h^j :\, \big|\widehat{\eta}_{h;K}^{1,j,\star} - \widehat{\eta}_{h;K}^{2,j,\star}\big| \mathbb{P}^\star\big[\mathtt{d}^\star(\widehat{\mu}_{h;K}^{1,j,\star}, \widehat{\mu}_{h;K}^{2,j,\star}) \geqslant 95\%\big] \geqslant 95\%\Big\} \tag{4.7.8}$$

and $\{\widehat{\eta}_{h;K}^{\mathfrak{s},j,\star};\, \mathfrak{s} \in \{1,2\}\}$ are the first moments of the empirical measures $\{\widehat{\mu}_{h;K}^{\mathfrak{s},j,\star};\, \mathfrak{s} \in \{1,2\}\}$ defined in (4.6.6) resp. (4.6.7). Below, let

$$h_{\min} := \min_j \min_{K \in \mathcal{T}_h^j} h_K^j \quad \text{and} \quad \mathcal{K}^j := \bigcup_{K \in \mathcal{T}_h^j} \{\mathbf{x}_\ell \in \overline{K} :\, \mathbf{x}_{\ell^*} \in \overline{K}\}. \tag{4.7.9}$$

Close to t_{j^*}, $\mathbb{E}_{M_\tau^j}\big[\boldsymbol{Y}^j(\mathbf{x}_\ell)\big]$ enters a state as depicted in Figure 4.19. In this constellation, $\mathbb{E}_{M_\tau^j}\big[|\nabla \boldsymbol{Y}^j(\mathbf{x}_{\ell^*})|\big] = \mathcal{O}(h_{\min}^{-1})$, as well as $\mathbb{E}_{M_\tau^j}\big[|G_h(\nabla \boldsymbol{Y}^j(\mathbf{x}_{\ell^*}))|\big] = \mathcal{O}(1)$ by local averaging. Thus, according to (4.6.8), the prefactor $\big|\widehat{\eta}_{h;K}^{1,j,\star} - \widehat{\eta}_{h;K}^{2,j,\star}\big| = \mathcal{O}(h_{\min}^{-1})$ in (4.7.8) is in accordance with the definition of $\widetilde{t}_{j^*}$ for uniform $h > 0$. Instantaneous shrinking of $\mathbb{E}_{M_\tau^j}\big[\boldsymbol{Y}^j(\mathbf{x}_{\ell^*})\big]$ exhibits values of realizations of $\mathtt{d}^\star(\widehat{\mu}_{h;K}^{1,j,\star}, \widehat{\mu}_{h;K}^{2,j,\star})$ close to one, which justifies the value of 95%.

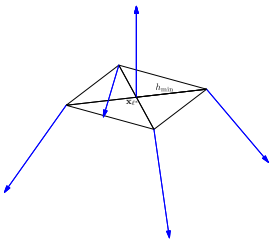
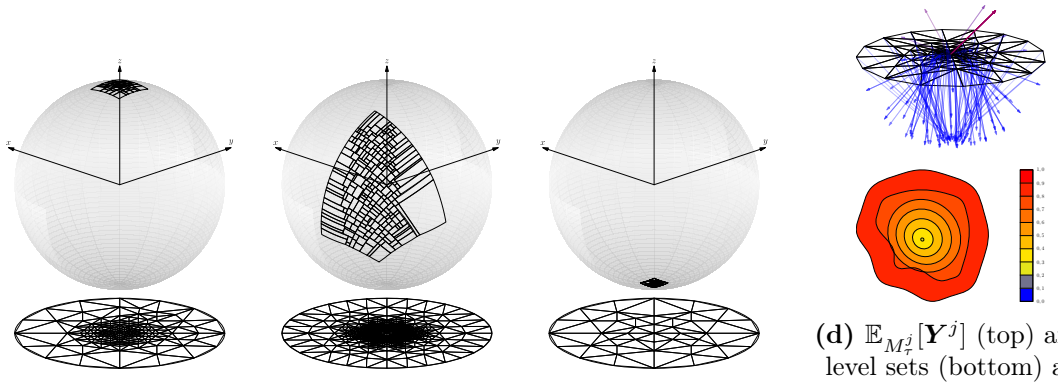


Figure 4.19: $\mathbb{E}_{M_\tau^j}\big[\boldsymbol{Y}^j(\mathbf{x}_\ell)\big]$ near t_{j^*} for the set of spatial nodal points $\mathcal{K}^j$ in (4.7.9).

For time adaptivity, we use a modification of the (**BTC**) strategy in Section 4.6.1 to partition the state space $(\mathbb{S}^2)^{L^j} = \bigcup_{r=1}^{R_\tau} \widehat{C}_{\tau;r}^j$: grouping the events is based on finding the tuple (ℓ, i), $\ell \in \{1,\ldots,L^j\}$, $i \in \{1,2,3\}$ in the set of L^j many $\mathbb{S}^2$-vectors $\mathcal{S} := \{\vec{\boldsymbol{Y}}(\omega_k)\}_k$ which possesses the largest empirical standard deviation. The snapshots in Figure 4.20(a)–(c) display partitions to resolve the marginal distributions $\big\{\mathcal{L}\big(\vec{\boldsymbol{Y}}^j(\mathbf{x}_{\ell^*})\big)\big\}_j$ on $\mathbb{S}^2$ at different times

near t_{j^*}: the local discrete blow-up of $\boldsymbol{Y}^j(\mathbf{x}_{\ell^*})$ is well-detected and resolved by the adaptive algorithm; more cells are created in areas where the random variable $\boldsymbol{Y}^j(\mathbf{x}_{\ell^*})$ is more likely to take values. The snapshot in Figure 4.20(d) displays both $\mathbb{E}_{M_\tau^j}[\boldsymbol{Y}^j]$ (top) and level sets $\big\{\mathbf{x}_\ell : |\mathbf{x}_{\ell^*} - \mathbf{x}_\ell| \leqslant 0.1 \,\big|\, \big|\mathbb{E}_{M_\tau^j}[\boldsymbol{Y}^j(\mathbf{x}_\ell)]\big| = c\big\}$ (bottom) for values $c \in \{0.0, 0.1, \ldots, 1.0\}$ at time $t_j \approx t_{j^*}$, with $\big|\mathbb{E}_{M_\tau^j}[\boldsymbol{Y}^j(\mathbf{x}_{\ell^*})]\big| \approx 0.282$. For nodal points $\mathbf{x}_\ell$ satisfying $|\mathbf{x}_{\ell^*} - \mathbf{x}_\ell| > 0.1$, $\big|\mathbb{E}_{M_\tau^j}[\boldsymbol{Y}^j]\big|$ is almost 1.0. This shrinking of statistical averages of vectors in the neighborhood of $(t_{j^*},\mathbf{x}_{\ell^*})$ is another indication of the discrete blow-up phenomenon; we also refer to [BBNP14b] for corresponding studies of single trajectories in the case of uniform space-time meshes.



(a) $t_j < t_{j^*}$ **(b)** $t_j \approx t_{j^*}$ **(c)** $t_j > t_{j^*}$ **(d)** $\mathbb{E}_{M_\tau^j}[\boldsymbol{Y}^j]$ (top) and level sets (bottom) at $t_j \approx t_{j^*}$

Figure 4.20: (Example 4.7.3 for $T = 0.1$, $\iota = 0.1$, as well as $\mathtt{d} = \mathtt{d_{TV}}$, $\mathtt{Tol}_\tau = \mathtt{Tol}_h = 0.05$, $\mathtt{N}_\tau = \mathtt{N}_h = 10$, $M_\tau^0 = M_h^0 = 10^3$, $R_\tau^0 = 2^8$, $R_h = 2^5$, and $h^0 = 2^{-4}$, $\tau^0 = 10^{-4}$, $h_{\min} = 2^{-6}$) Snapshots of the partitioning of $\mathbb{S}^2$ supporting $\mathcal{L}\big(\vec{\boldsymbol{Y}}^j(\mathbf{x}_{\ell^*})\big)$ and corresponding spatial meshes $\mathcal{T}_h^j$ for $t_j < t_{j^*}$ (a), $t_j \approx t_{j^*}$ (b), and $t_j > t_{j^*}$ (c). (d) Snapshots of $\mathbb{E}_{M_\tau^j}[\boldsymbol{Y}^j]$ (top) and corresponding level sets $\big\{\mathbf{x}_\ell : |\mathbf{x}_{\ell^*} - \mathbf{x}_\ell| \leqslant 0.1 \,\big|\, \big|\mathbb{E}_{M_\tau^j}[\boldsymbol{Y}^j(\mathbf{x}_\ell)]\big| = c\big\}$ (bottom) for values $c \in \{0.0, 0.1, \ldots, 1.0\}$ at time $t_j \approx t_{j^*}$.

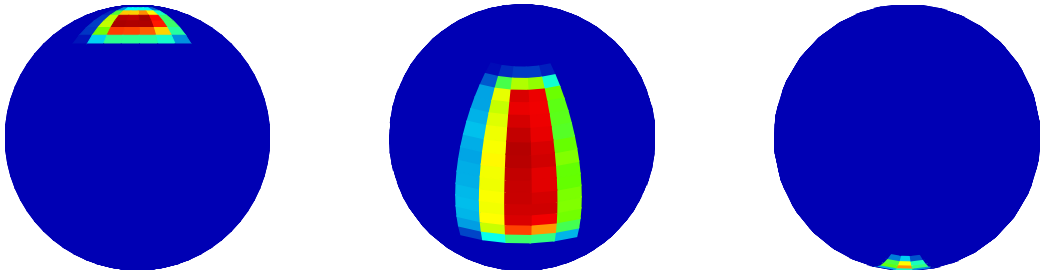
The snapshots in Figure 4.21 display the histogram based density of $\boldsymbol{Y}^j(\mathbf{x}_{\ell^*})$ at different times near t_{j^*}. There, the sphere $\mathbb{S}^2$ is divided into segments $w_{\mathsf{ij}} \subset \mathbb{S}^2$ associated with points

$$\mathbf{x}_{\mathsf{ij}} := \begin{pmatrix} \sin\big(\mathsf{i}\pi/16\big)\cos\big(\mathsf{j}\pi/16\big) \\ \sin\big(\mathsf{i}\pi/16\big)\sin\big(\mathsf{j}\pi/16\big) \\ \cos\big(\mathsf{i}\pi/16\big) \end{pmatrix},$$

for $\mathsf{i} \in \{1, \ldots, 16\}$, $\mathsf{j} \in \{1, \ldots, 32\}$ such that $w_{\mathsf{ij}} := \big\{\mathbf{x} \in \mathbb{S}^2 \,\big|\, \mathbf{x}_{\mathsf{ij}} = \arg\min_{\mathbf{x}_{\mathsf{lm}}} |\mathbf{x} - \mathbf{x}_{\mathsf{lm}}|\big\}$. Thus, at time t_j, we construct a empirical probability density function $\widehat{\rho} : \mathbb{S}^2 \to \mathbb{R}$ via

$$\widehat{\rho}|_{w_{\mathsf{ij}}}(\mathbf{x}) = \widehat{\rho}(\mathbf{x}_{\mathsf{ij}}) := \frac{\#\big\{k;\, \boldsymbol{Y}^j(\mathbf{x}_{\ell^*}, \omega_k) \in w_{\mathsf{ij}}\big\}}{|w_{\mathsf{ij}}| \cdot M_\tau^j} \quad \forall\, \mathbf{x} \in \mathbb{S}^2.$$

We observe a closer concentration for times $t_j < t_{j^*}$ and $t_j > t_{j^*}$ due to the small variance of the iterates, see Figure 4.22(d), whereas the density function covers a larger area of the spherical surface for $t_j \approx t_{j^*}$.
The discrete blow-up phenomenon is also computationally detected by shrinking local step sizes $\big(\tau^j, \{h_K^j\}_K\big)$ to $\tau^j \approx h_{\min}^2$ near $j \approx j^*$, where $h_{\min} = 2^{-6}$ was chosen to terminate



(a) $t_j < t_{j^*}$ **(b)** $t_j \approx t_{j^*}$ **(c)** $t_j > t_{j^*}$

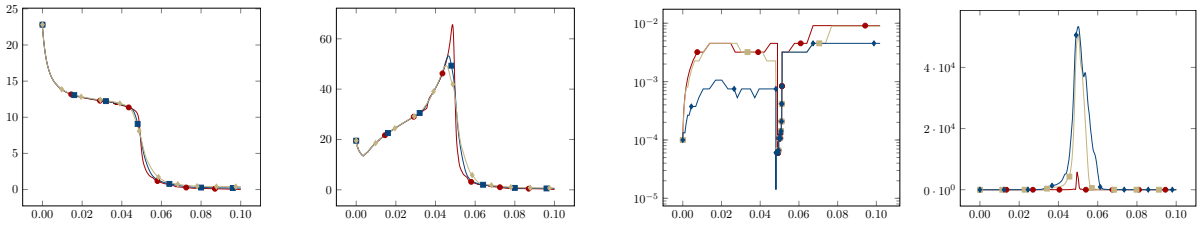
Figure 4.21: (Example 4.7.3 for the same setup as in Figure 4.20) Snapshots of the probability density function of the random variable $\boldsymbol{Y}^j(\mathbf{x}_{\ell^*})$ at certain times $t_j < t_{j^*}$ (a), $t_j \approx t_{j^*}$ (b), and $t_j > t_{j^*}$ (c) for the same setup as in Figure 4.20.

repeated refinement; the Newton method requires the most iterations (up to 50 to meet a $\mathbb{P}$-almost sure threshold criterion) in the neighborhood of t_{j^*}, and its number decreases rapidly afterwards again, where the iterates approximate the asymptotic solution $\boldsymbol{X} \equiv (0, 0, -1)^T$; see Figure 4.18(a). As for Example 4.7.1, and motivated by the results in Table 4.6, we again prefer $\mathtt{d}_{\mathtt{TV}}$ to perform adaptivity in time and space.

Table 4.6: (Example 4.7.3 for the same setup as in Figure 4.20) Error indicators $\mathbf{err}_k := \max_j \left| \mathbb{E}_{M_\tau^j}\left[\|\boldsymbol{X}_{t_j}\|_{\mathbf{W}^{k,2}}\right] - \mathbb{E}_{M_\tau^j}\left[\|\boldsymbol{Y}^j\|_{\mathbf{W}^{k,2}}\right] \right|$ for $\big(k \in \{0,1\}\big)$, and maximum number of degrees of freedom $L_{\max} := \max_j L^j$.

$\mathtt{d}$	$\mathtt{d}_{\mathtt{H}}$	$\mathtt{d}_{\mathtt{TV}}$	$\widetilde{\mathtt{d}}_{\mathtt{KL}}$
$L_{\max}$	81300	74600	85500
$\mathbf{err}_0$	0.0344	0.0337	0.0421
$\mathbf{err}_1$	0.1456	0.1417	0.1502

In a vicinity of t_{j^*}, the expected energy loss of iterates is approximatively 4π; see Figure 4.22(a). For times $t_j \uparrow t_{j^*}$, the simulations show a growing concentration of spatial nodal points at $\mathbf{x}_{\ell^*}$ of D (see Figure 4.20), and smaller time steps sizes τ^j; larger M_τ^j are needed here due to an increased empirical variance $\mathrm{Var}_{M_\tau^j}\left[\|\nabla \boldsymbol{Y}^j\|_{\mathbf{L}^\infty}\right]$ of the iterates (see Figure 4.22(d)). This behavior goes along with smaller values $\mathbb{E}_{M_\tau^j}\left[\|\nabla \boldsymbol{Y}^j(\mathbf{x}_{\ell^*})\|_{\mathbf{L}^\infty}\right]$ if compared to single realizations with values $\|\nabla \boldsymbol{Y}^j(\mathbf{x}_{\ell^*})\|_{\mathbf{L}^\infty}$ near t_{j^*}. Beyond the time t_{j^*}, the time step size τ^j rapidly increases, and also $\mathcal{T}_h^j$ coarsens again; accordingly, adaptive selection of large statistical parameters M_h^j, B_h^j, and R_h^j recovers again for $j \gg j^*$. For example, using the setup in Figure 4.22, the number of required MC simulations rapidly changes from $M_\tau^j \approx 3000$ to $M_\tau^j \approx 180000$ close to t_{j^*}, where $L^j \approx 75000$; cf. Table 4.6. Hence, the adaptive concept from Section 4.7.1.3 concentrates computer resources to the discrete blow-up phenomenon: in fact, 76% of the overall required time steps ($J = 534$) are concentrated here. However, as for Example 4.7.1, a direct comparison with uniform discretization and statistical parameters again leads to a drastically reduced computational effort when the proposed adaptive concepts are applied.



(a) $t_j \mapsto \frac{1}{2}\mathbb{E}_{M_\tau^j}\big[\|\nabla \boldsymbol{Y}^j\|_{\mathbf{L}^2}^2\big]$ **(b)** $t_j \mapsto \mathbb{E}_{M_\tau^j}\big[\|\nabla \boldsymbol{Y}^j\|_{\mathbf{L}^\infty}\big]$ **(c)** $t_j \mapsto \tau^j$ **(d)** $t_j \mapsto \mathrm{Var}_{M_\tau^j}\big[\|\nabla \boldsymbol{Y}^j\|_{\mathbf{L}^\infty}\big]$

Figure 4.22: (Example 4.7.3 for the same setup as in Figure 4.20) (a) Behavior of $t_j \mapsto \frac{1}{2}\mathbb{E}_{M_\tau^j}\big[\|\nabla \boldsymbol{Y}^j\|_{\mathbf{L}^2}^2\big]$, (b) $t_j \mapsto \mathbb{E}_{M_\tau^j}\big[\|\nabla \boldsymbol{Y}^j\|_{\mathbf{L}^\infty}\big]$, (c) $t_j \mapsto \tau^j$, and (d) $t_j \mapsto \mathrm{Var}_{M_\tau^j}\big[\|\nabla \boldsymbol{Y}^j\|_{\mathbf{L}^\infty}\big]$ with $\iota = 1.0$ (—●—), $\iota = 2.0$ (—■—), $\iota = 3.0$ (—♦—).

4.7.4 Computational experiments for the stochastic Landau-Lifshitz-Gilbert equation

As in Section 4.7.3, we use the ($\mathbb{P}$-almost sure) length-preserving discretization Scheme 4.5.3 and 4.5.5 in combination with space-time and statistical adaptivity, as well as the Newton method from Section 4.7.2. Analogously to Section 4.7.3, we choose $h_{\min} = 2^{-6}$. Furthermore, we choose $\theta = 1/2$ such that the artificial damping in (4.5.4) vanish.

Concerning the implementation with respect to Scheme 4.5.5, the nodal-wise tangent plane constraint $\langle \boldsymbol{Y}^j(\mathbf{x}_\ell), \boldsymbol{V}^{j+1}(\mathbf{x}_\ell)\rangle = 0$ for all $\ell \in \{1, \dots, L^{j+1}\}$ is treated by means of a Lagrangian multiplier; more precisely, for each $j \in \{0, 1, \dots, J\}$ we solve a linear algebraic system of the form (cf. [GHPS12, Chapter 3] for the deterministic case $\iota = 0$)

$$\begin{pmatrix} \boldsymbol{M}^j & (\boldsymbol{\Lambda}^j)^T \\ \boldsymbol{\Lambda}^j & \mathbf{0} \end{pmatrix} \begin{pmatrix} \vec{v}^{j+1} \\ \vec{\lambda}^{j+1} \end{pmatrix} = \begin{pmatrix} \vec{b}^j \\ \mathbf{0} \end{pmatrix}. \tag{4.7.10}$$

Here, $\vec{\lambda}^{j+1} := \vec{\Upsilon}_{\mathbb{V}_h^{j+1}}(\lambda^{j+1}) \in \mathbb{R}^{L^{j+1}}$, for $\lambda^{j+1} \in \mathbb{V}_h^{j+1}$, is the Lagrangian multiplier associated to the sought coefficient vector $\vec{v}^{j+1} := \vec{\Upsilon}_{\mathbb{V}_h^{j+1}}(\boldsymbol{V}^{j+1}) \in \mathbb{R}^{3L^{j+1}}$, and

$$\boldsymbol{\Lambda}^j := \big(\Lambda^{1,j}, \Lambda^{2,j}, \Lambda^{3,j}\big) \in \mathbb{R}^{L^{j+1} \times 3L^{j+1}}$$

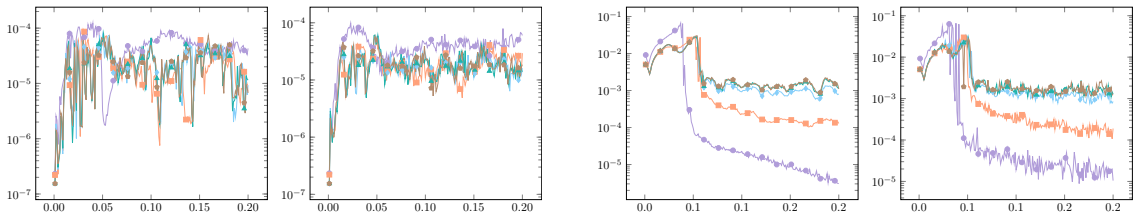
the Lagrangian matrix, with sub-matrices

$$\Lambda^{i,j} \equiv \big(\Lambda_{\ell,\ell'}^{i,j}\big)_{\ell,\ell'=1}^{L^{j+1}}, \quad \Lambda_{\ell,\ell'}^{i,j} := \big\langle \boldsymbol{Y}^j(\mathbf{x}_\ell), \mathbf{e}_i \big\rangle_{\mathbb{R}^3} \Psi_\ell \Psi_{\ell'},$$

where $\mathbf{e}_i$ for $i \in \{1, 2, 3\}$ are the canonical basis vectors of $\mathbb{R}^3$. The line $\boldsymbol{\Lambda}^j \vec{v}^{j+1} = \mathbf{0}$ thus realizes the nodal-wise tangential plane constraint, and the sub-matrix $\boldsymbol{M}^j \in \mathbb{R}^{3L^{j+1} \times 3L^{j+1}}$ encodes the left hand side in (4.5.3), which is well-conditioned for $\zeta_2 > 0$.

The linear algebraic system in (4.7.10) is solved by Gaussian elimination, where the full system matrix on the left hand side in (4.7.10), referred to as $\boldsymbol{S}^j \in \mathbb{R}^{4L^j \times 4L^j}$, is well conditioned, see Figure 4.23(a), in which the behavior of $t_j \mapsto \mathbb{E}_{M_\tau^j}\big[\mathrm{cond}(\boldsymbol{S}^j)\big] := \mathbb{E}_{M_\tau^j}\big[|||\boldsymbol{S}^j||| \cdot |||(\boldsymbol{S}^j)^{-1}|||\big]$ is plotted for different (ζ_2, ι) with $|||\boldsymbol{S}^j||| := \sup_{\mathbf{0} \neq \vec{z} \in \mathbb{R}^{4L^j}} |\boldsymbol{S}^j \vec{z}|/|\vec{z}|$. We observe unique solvability in the tangent space $\mathbb{T}_h^j$ with respect to space-time and statistical parameters due to the use of

the h-scalar product. Since $\boldsymbol{W}$ enters only in the right hand side $\vec{b}^{\,j}$, realizations of $\mathrm{cond}(\boldsymbol{S}^j)$ are only affected by the perturbation of $\boldsymbol{Y}^j$; in fact, we observe $\mathbb{E}_{M_\tau^j}\big[\mathrm{cond}(\boldsymbol{S}^j)\big] \leqslant 10^{-4}$ uniformly in time, such that system in (4.7.10) thus admits inversion. Especially in the discrete blow-up case, this observation propagates well-conditioned behavior of the solution; however for times $t_j > t_{j^*}$, when the solution $\boldsymbol{Y}^j$ becomes homogeneous, Scheme 4.5.5 suffers from instability indicated by oscillatory behavior if $\zeta_2 \downarrow 0$. In particular, for $\zeta_2 \leqslant 10^{-2}$, the quantity $\mathbb{E}_{M_\tau^j}\big[\|\nabla \boldsymbol{Y}^j\|_{\mathbf{L}^\infty}\big]$ oscillates and $\frac{1}{2}\mathbb{E}_{M_\tau^j}\big[\|\nabla \boldsymbol{Y}^j\|_{\mathbf{L}^2}^2\big]$ decreases very slowly after the discrete blow-up time $t_{j^*} \approx 0.05$, meaning possible instability leaving unclear long time dynamics. It seems that this oscillatory behavior is caused by the evolution of $t_j \mapsto \mathbb{E}_{M_\tau^j}\big[\|\nabla \lambda^j\|_{\mathbb{L}^\infty}\big]$ for the Lagrangian multiplier, see Figure 4.23(b), where the amplitudes are affected by ι. Compared to analogous simulations on uniform meshes, computational studies show that these oscillations occur regardless of whether time and space adaptivity takes place or not. As suggested in [GHPS12], in the case of uniform meshes and $\iota = 0$, possible stabilizing is achieved by multiplying the matrix $\boldsymbol{\Lambda}^j$ with τh; however, stabilizing is not clear in the stochastic context.



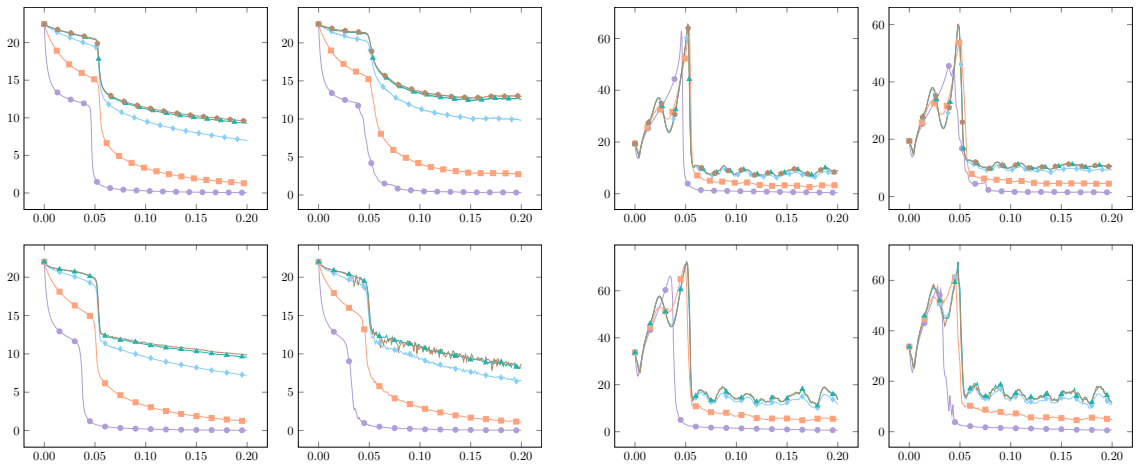
(a) $\iota = 1.0$ (left) and $\iota = 3.0$ (right)

(b) $\iota = 1.0$ (left) and $\iota = 3.0$ (right)

Figure 4.23: (Example 4.7.3 for $T = 0.2$, as well as $\mathtt{d} = \mathtt{d_{TV}}$, $\mathtt{Tol}_\tau = \mathtt{Tol}_h = 0.05$, $\mathtt{N}_\tau = \mathtt{N}_h = 10$, $M_\tau^0 = M_h^0 = 10^3$, $R_\tau^0 = 2^8$, $R_h = 2^5$, and $h^0 = 2^{-4}$, $\tau^0 = 10^{-4}$, $h_{\min} = 2^{-6}$) Behavior of $t_j \mapsto \mathbb{E}_{M_\tau^j}\big[\mathrm{cond}(\boldsymbol{S}^j)\big]$ (a), and $t_j \mapsto \mathbb{E}_{M_\tau^j}\big[\|\nabla \lambda^j\|_{\mathbb{L}^\infty}\big]$ (b) for noise intensities $\iota \in \{1.0, 3.0\}$ with $\zeta_2 = 10^{-0}$(—●—), $\zeta_2 = 10^{-1}$(—■—), $\zeta_2 = 10^{-2}$(—◆—), $\zeta_2 = 10^{-3}$(—▲—), $\zeta_2 = 10^{-4}$(—⬟—).

We observe for both Scheme 4.5.3 and 4.5.5 that, in a small neighborhood of t_{j^*}, the expected energy loss of iterates is approximatively 4π, and $t_{j^*} \downarrow 0$ if $\zeta_2 \uparrow \infty$ respectively $t_{j^*} \uparrow 0.0514$ if $\zeta_2 \downarrow 0$. Next to this, for Scheme 4.5.5, shrinking as well as the helical dynamic of $\mathbb{E}_{M_\tau^j}\big[\boldsymbol{Y}^j(\mathbf{x}_{\ell^*})\big]$ take place much faster for different (ζ_2, ι), especially if ι is large; see also the snapshots in Figure 4.25 for the behavior of $t_j \mapsto \mathbb{E}_{M_\tau^j}\big[\boldsymbol{Y}^j(\mathbf{x}_{\ell^*})\big]$ at different times near t_{j^*} for the set of spatial nodal points $\mathcal{K}^j$ in (4.7.9). This is contrary to simulations performed via Scheme 4.5.3, where the expected time period of homogeneous alignment of $\mathbb{E}_{M_\tau^j}\big[\boldsymbol{Y}^j(\mathbf{x}_{\ell^*})\big]$ with surrounding vectors proportionally increases with the noise intensity ι. For example, shrinking of $\mathbb{E}_{M_\tau^j}\big[\boldsymbol{Y}^j(\mathbf{x}_{\ell^*})\big]$ via Scheme 4.5.5 requires 14 time steps, whereas 32 time steps are needed for Scheme 4.5.3. This observation underlines again the faster shrinking behavior of $\mathbb{E}_{M_\tau^j}\big[\boldsymbol{Y}^j(\mathbf{x}_{\ell^*})\big]$ in a vicinity to the time t_{j^*}.

The snapshots in Figure 4.26 display partitions to resolve the marginal distribution $\mathcal{L}\big(\vec{\boldsymbol{Y}}^j(\mathbf{x}_{\ell^*})\big)$ on $\mathbb{S}^2$ at different times near t_{j^*}: the local discrete blow-up of $\vec{\boldsymbol{Y}}^j(\mathbf{x}_{\ell^*})$ is well-detected and resolved by the adaptive algorithm. We observe a similar behavior of the empirical variance for both schemes, see Figure 4.27(b), whereas, in the case of Scheme 4.5.5, cells are concentrated more close to some specific longitude between the two states $(0, 0, 1)^T$ and $(0, 0, -1)^T$ caused by faster shrinking as well as smaller variance of $\vec{\boldsymbol{Y}}^j(\mathbf{x}_{\ell^*})$, see Figure 4.27(b) below.



(a) $\iota = 1.0$ (left) and $\iota = 3.0$ (right) **(b)** $\iota = 1.0$ (left) and $\iota = 3.0$ (right)

Figure 4.24: (Example 4.7.3 for the same setup as in Figure 4.23: Scheme 4.5.3 (upper row), and Scheme 4.5.5 (lower row)) Behavior of $t_j \mapsto \frac{1}{2}\mathbb{E}_{M_\tau^j}\big[\|\nabla \boldsymbol{Y}^j\|_{\mathbf{L}^2}^2\big]$ (a), and $t_j \mapsto \mathbb{E}_{M_\tau^j}\big[\|\nabla \boldsymbol{Y}^j\|_{\mathbf{L}^\infty}\big]$ (b) for noise intensities $\iota \in \{1.0, 3.0\}$ with $\zeta_2 = 10^{-0}$(—●—), $\zeta_2 = 10^{-1}$(—■—), $\zeta_2 = 10^{-2}$(—◆—), $\zeta_2 = 10^{-3}$(—▲—), $\zeta_2 = 10^{-4}$(—⬟—).

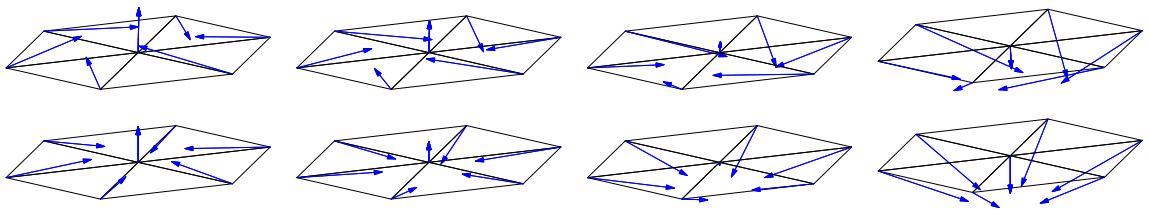
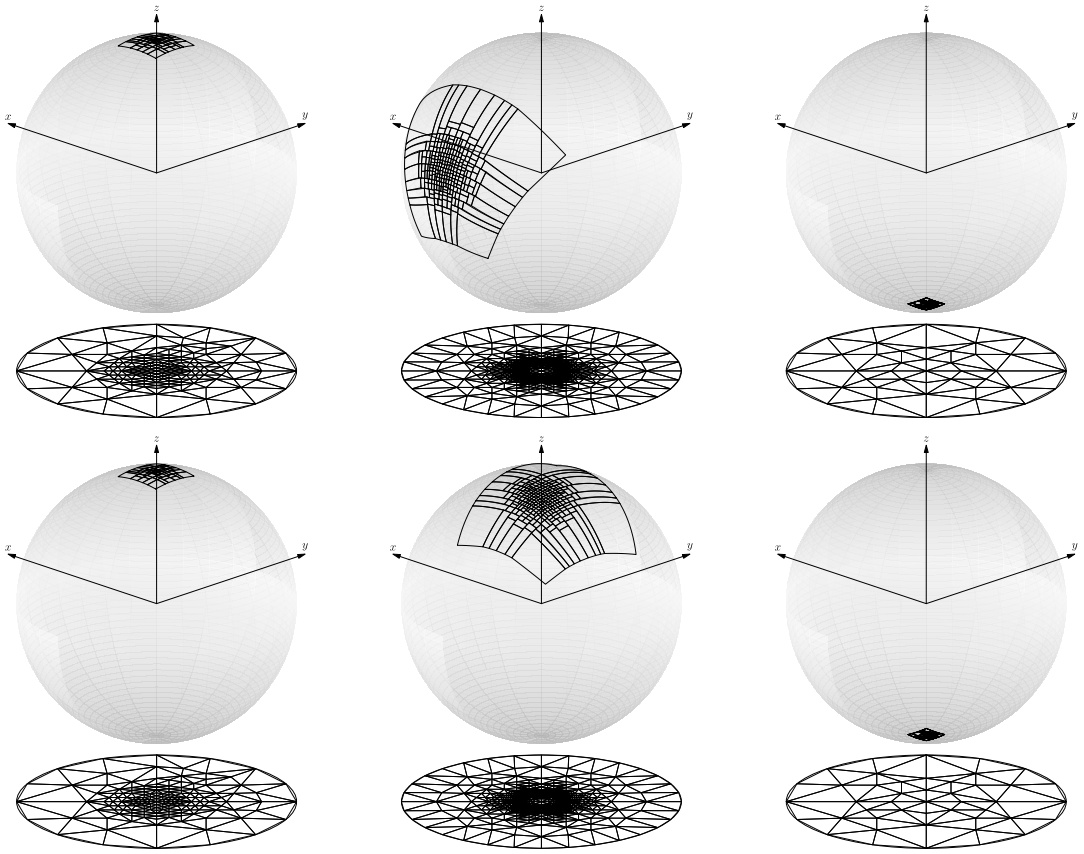


Figure 4.25: (Example 4.7.3 for the same setup as in Figure 4.26: Scheme 4.5.3 (upper row), and Scheme 4.5.5 (lower row)) Snapshots of $\mathbb{E}_{M_\tau^j}\big[\boldsymbol{Y}^j(\mathbf{x}_\ell)\big]$ at different times near t_{j^*} for the set of nodal spatial points $\mathcal{K}^j$ in (4.7.9).

For times $t_j > t_{j^*}$, the oscillating behavior of the solution causes smaller time steps as well as finer partition of D. According to the Figure 4.20, for times $t_j \uparrow t_{j^*}$, the simulations show



(a) $t_j < t_{j^*}$ **(b)** $t_j \approx t_{j^*}$ **(c)** $t_j > t_{j^*}$

Figure 4.26: (Example 4.7.3 for $T = 0.2$, $\zeta_2 = 1.0$, $\iota = 1.0$, as well as $\mathtt{d} = \mathtt{d}_{\mathtt{TV}}$, $\mathtt{Tol}_\tau = \mathtt{Tol}_h = 0.05$, $\mathtt{N}_\tau = \mathtt{N}_h = 10$, $M_\tau^0 = M_h^0 = 10^3$, $R_\tau^0 = 2^8$, $R_h = 2^5$, and $h^0 = 2^{-4}$, $\tau^0 = 10^{-4}$, $h_{\min} = 2^{-6}$: Scheme 4.5.5 (upper row), and Scheme 4.5.3 (lower row)) Snapshots of the partitioning of $\mathbb{S}^2$ supporting $\mathcal{L}\big(\vec{\boldsymbol{Y}}^j(\mathbf{x}_{\ell^*})\big)$ and corresponding spatial meshes $\mathcal{T}_h^j$ for $t_j < t_{j^*}$ (a), $t_j \approx t_{j^*}$ (b), and $t_j > t_{j^*}$ (c).

a growing concentration of spatial nodal points at $\mathbf{x}_{\ell^*} \in D$. However, for example, close to t_{j^*}, Scheme 4.5.5 requires only $L^j \approx 74000$ spatial nodal points compared to $L^j \approx 83000$ for Scheme 4.5.3 to retain the same accuracy; see Table 4.7. In fact, the choice of $\widetilde{\mathtt{d}}_{\mathtt{KL}}$ has less impact on L^j for Scheme 4.5.5 compared to Scheme 4.5.3.

As in Section 4.7.3, the discrete blow-up phenomenon is also computationally detected by shrinking local step sizes $\big(\tau^j, \{h_K^j\}_K\big)$ to zero. While the foregoing results motivate that $t_{j^*} \downarrow 0$ respectively $t_{j^*} \uparrow 0.0514$ for $\zeta_2 \uparrow \infty$ respectively $\zeta_2 \downarrow 0$, it is quite surprising that Scheme 4.5.5 seems to react fairly sensible to the mesh sizes $\big(\tau^j, \{h_K^j\}_K\big)$. In particular, we cannot confirm the observed conjecture stating from Scheme 4.5.5 that the discrete blow-up time t_{j^*} converges towards a specific point in time by shrinking $\big(\tau^j, \{h_K^j\}_K\big)$ for Scheme 4.5.3.

Solving a single algebraic system in each time step has a drastic practical impact regarding reduced computational complexity to obtain related M_τ-samples; cf. Table 4.8. Comparative

Table 4.7: (Example 4.7.3 for the same setup as in Figure 4.26): Error indicators $\mathbf{err}_k := \max_j \left|\mathbb{E}_{M_\tau^j}\left[\|\boldsymbol{X}_{t_j}\|_{\mathbf{W}^{k,2}}\right] - \mathbb{E}_{M_\tau^j}\left[\|\boldsymbol{Y}^j\|_{\mathbf{W}^{k,2}}\right]\right|$ for $\big(k \in \{0,1\}\big)$, and maximum number of degrees of freedom $L_{\max} := \max_j L^j$; cf. Table 4.6.

(a) Scheme 4.5.3

$\mathtt{d}$	$\mathtt{d}_{\mathtt{H}}$	$\mathtt{d}_{\mathtt{TV}}$	$\widetilde{\mathtt{d}}_{\mathtt{KL}}$
$L_{\max}$	86040	82800	91100
$\mathbf{err}_0$	0.0367	0.0358	0.0401
$\mathbf{err}_1$	0.1503	0.1512	0.1577

(b) Scheme 4.5.5

$\mathtt{d}$	$\mathtt{d}_{\mathtt{H}}$	$\mathtt{d}_{\mathtt{TV}}$	$\widetilde{\mathtt{d}}_{\mathtt{KL}}$
$L_{\max}$	75600	74300	77200
$\mathtt{err}_0$	0.0456	0.0438	0.0487
$\mathtt{err}_1$	0.1578	0.1581	0.1602

Table 4.8: (Example 4.7.3): Different number of realizations M_τ to generate M_τ-samples $\mathcal{S}_{\overrightarrow{Y}}$ for uniform $\tau = 10^{-5}$: the absolute simulation time (in seconds) with respect to the uniform mesh size h in double precision arithmetic.

(a) Scheme 4.5.3

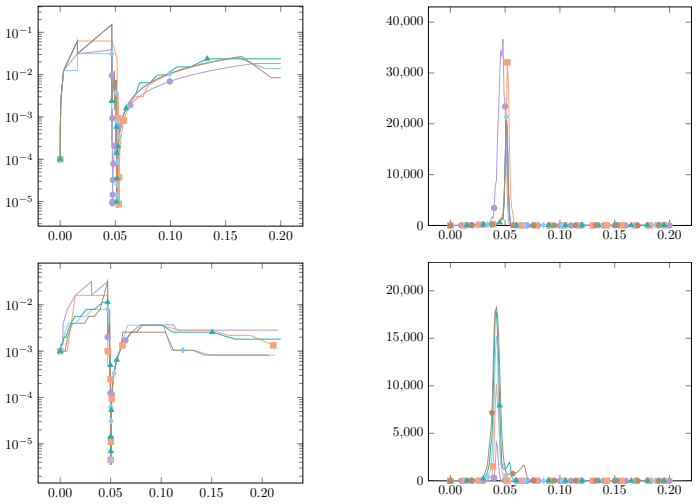
M_τ	10^2	10^3	10^4	10^5
$h = 2^{-4}$	$2s$	$28s$	$240s$	$1954s$
$h = 2^{-5}$	$11s$	$109s$	$1047s$	$9738s$
$h = 2^{-6}$	$46s$	$380s$	$2415s$	$13361s$

(b) Scheme 4.5.5

M_τ	10^2	10^3	10^4	10^5
$h = 2^{-4}$	$1s$	$21s$	$174s$	$1437s$
$h = 2^{-5}$	$6s$	$59s$	$572s$	$5245s$
$h = 2^{-6}$	$35s$	$289s$	$1837s$	$10165s$

computational studies between Scheme 4.5.3 and 4.5.5 indicate a time saving of approximatively 23%, which is mainly caused by assembling less contributions to the linear algebraic system: the Jacobian $D\boldsymbol{\mathcal{F}}^{j+1}$ in Algorithm 4.7.2 requires 7 contributions (the Jacobian is evaluated directly) compared to 2 contributions in order to assemble $\boldsymbol{S}^j$, while assembling the corresponding right side, the update as well as the projection requires nearly the same computational effort for both schemes. These savings of computer resources through Scheme 4.5.5 therefore allow larger statistical parameters (in time and space), and thus a better resolution of $\mathbb{E}\big[\boldsymbol{Y}^j(\mathbf{x}_{\ell^*})\big]$ near t_{j^*}.

Unfortunately, Scheme 4.5.5 requires smaller time step sizes beyond t_{j^*}, see Figure 4.27(a). Caused by the oscillatory behavior of $\mathbb{E}_{M_\tau^j}\left[\|\nabla \boldsymbol{Y}^j\|_{\mathbf{L}^\infty}\right]$ (see Figure 4.24(b)) for $t_j > t_{j^*}$, especially in the case $\zeta_2 \leqslant 10^{-2}$, the empirical variance is larger compared to Scheme 4.5.3. This implies unnecessary large statistical parameters $(M_\tau^j, B_\tau^j, R_\tau^j)$ respectively (M_h^j, B_h^j, R_h^j) for times where the solution is homogeneously aligned. The results show that to determine a realization of $\boldsymbol{Y}^j$ via Scheme 4.5.5 at least 3 Newton iterations are needed to meet the $\mathbb{P}$-almost surely threshold 10^{-8}. This multiple solving of algebraic systems, together with the enhanced number of contributions to assemble the Jacobian, leads to an computational effort which is comparable to that of Scheme 4.5.5. In summary, the proposed adaptive space-time and statistical concepts together with Scheme 4.5.5 would allow a more thorough investigation of discrete blow-up dynamics.



(a) $t_j \mapsto \tau^j$ **(b)** $t_j \mapsto \mathrm{Var}_{M_\tau^j}\left[\|\nabla \boldsymbol{Y}^j\|_{\mathbf{L}^\infty}\right]$

Figure 4.27: (Example 4.7.3 for the same setup as in Figure 4.26: Scheme 4.5.3 (upper row), and Scheme 4.5.5 (lower row)) (a) Behavior of $t_j \mapsto \tau^j$, and (b) $t_j \mapsto \mathrm{Var}_{M_\tau^j}\left[\|\nabla \boldsymbol{Y}^j\|_{\mathbf{L}^\infty}\right]$ with damping parameters $\zeta_2 = 10^{-0}$( ), $\zeta_2 = 10^{-1}$( ), $\zeta_2 = 10^{-2}$( ), $\zeta_2 = 10^{-3}$( ), $\zeta_2 = 10^{-4}$( ).

Bibliography

[Abo13] Omar Aboura. Weak error expansion of the implicit euler scheme. 2013.

[AdBH14] François Alouges, Anne de Bouard, and Antoine Hocquet. A semi-discrete scheme for the stochastic Landau-Lifshitz equation. *Stoch. Partial Differ. Equ. Anal. Comput.*, 2(3):281–315, 2014.

[AG07] Søren Asmussen and Peter W. Glynn. *Stochastic simulation: algorithms and analysis*, volume 57 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2007.

[AJ76] Saab Abou-Jaoude. Sur la convergence L^1 et L^∞ de l'estimateur de la partition aléatoire pour une densité. *Ann. Inst. H. Poincaré Sect. B (N.S.)*, 12(4):299–317, 1976.

[AJ06] François Alouges and Pascal Jaisson. Convergence of a finite element discretization for the Landau-Lifshitz equations in micromagnetism. *Math. Models Methods Appl. Sci.*, 16(2):299–316, 2006.

[And66] T. W. Anderson. Some nonparametric multivariate procedures based on statistically equivalent blocks. In *Multivariate Analysis (Proc. Internat. Sympos., Dayton, Ohio, 1965)*, pages 5–27. Academic Press, New York, 1966.

[AS92] François Alouges and Alain Soyeur. On global weak solutions for Landau-Lifshitz equations: existence and nonuniqueness. *Nonlinear Anal.*, 18(11):1071–1084, 1992.

[Aur91] Franz Aurenhammer. Voronoi diagrams: a survey of a fundamental geometric data structure. *ACM Comput. Surv.*, 23(3):345–405, 1991.

[BB02] P. M. Burrage and K. Burrage. A variable stepsize implementation for stochastic differential equations. *SIAM J. Sci. Comput.*, 24(3):848–864, 2002.

[BBNP14a] Lúbomír Baňas, Zdzisław Brzeźniak, Mikhail Neklyudov, and Andreas Prohl. A convergent finite-element-based discretization of the stochastic Landau-Lifshitz-Gilbert equation. *IMA J. Numer. Anal.*, 34(2):502–549, 2014.

[BBNP14b] Lúbomír Baňas, Zdzisław Brzeźniak, Mikhail Neklyudov, and Andreas Prohl. *Stochastic ferromagnetism*, volume 58 of *De Gruyter Studies in Mathematics*. De Gruyter, Berlin, 2014. Analysis and numerics.

[BBP08] Lúbomír Baňas, Sören Bartels, and Andreas Prohl. A convergent implicit finite element discretization of the Maxwell-Landau-Lifshitz-Gilbert equation. *SIAM J. Numer. Anal.*, 46(3):1399–1422, 2008.

[Ben75] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Commun. ACM*, 18:509–517, 1975.

[Ber94] Rudolf Beran. Seven stages of bootstrap. In *Computational statistics (Schloß Reisensburg, 1993)*, Contrib. Statist., pages 143–157. Physica, Heidelberg, 1994.

[Ber02] D. V. Berkov. Fast switching of magnetic nanoparticles: simulation of thermal noise effects using the langevin dynamics. *IEEE Transactions on Magnetics*, 38(5):2489–2495, 2002.

[BFOS84] Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. *Classification and regression trees.* Wadsworth Statistics/Probability Series. Wadsworth Advanced Books and Software, Belmont, CA, 1984.

[BGJ13] Zdzisław Brzeźniak, Beniamin Goldys, and Terence Jegaraj. Weak solutions of a stochastic Landau-Lifshitz-Gilbert equation. *Appl. Math. Res. Express. AMRX*, 2013(1):1–33, 2013.

[BGJ17] Zdzisław Brzeźniak, Ben Goldys, and Terence Jegaraj. Large deviations and transitions between equilibria for stochastic Landau-Lifshitz-Gilbert equation. *Arch. Ration. Mech. Anal.*, 226(2):497–558, 2017.

[BH82] Alexander N. Brooks and Thomas J. R. Hughes. Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations. *Comput. Methods Appl. Mech. Engrg.*, 32(1-3):199–259, 1982. FENOMECH '81, Part I (Stuttgart, 1981).

[BKP08] Sören Bartels, Joy Ko, and Andreas Prohl. Numerical analysis of an explicit approximation scheme for the Landau-Lifshitz-Gilbert equation. *Math. Comp.*, 77(262):773–788, 2008.

[BMS09] Giorgio Bertotti, Isaak D. Mayergoyz, and Claudio Serpico. *Nonlinear magnetization dynamics in nanosystems.* Elsevier Series in Electromagnetism. Elsevier B. V., Amsterdam, 2009.

[BP06] Sören Bartels and Andreas Prohl. Convergence of an implicit finite element method for the Landau-Lifshitz-Gilbert equation. *SIAM J. Numer. Anal.*, 44(4):1405–1419, 2006.

[BR78] I. Babuška and W. C. Rheinboldt. Error estimates for adaptive finite element computations. *SIAM J. Numer. Anal.*, 15(4):736–754, 1978.

[BR96] R. Becker and R. Rannacher. A feed-back approach to error control in finite element methods: basic analysis and examples. *East-West J. Numer. Math.*, 4(4):237–264, 1996.

[Bro63] William Fuller Brown. Thermal fluctuations of a single-domain particle. *Phys. Rev.*, 130:1677–1686, 1963.

[BS93] Randolph E. Bank and R. Kent Smith. A posteriori error estimates based on hierarchical bases. *SIAM J. Numer. Anal.*, 30(4):921–935, 1993.

[BSTZ13] T. Björk, A. Szepessy, R. Tempone, and G. E. Zouraris. Monte Carlo Euler approximations of HJM term structure financial models. *BIT*, 53(2):341–383, 2013.

[Bur10] Erik Burman. Consistent SUPG-method for transient transport problems: stability and convergence. *Comput. Methods Appl. Mech. Engrg.*, 199(17-20):1114–1123, 2010.

[BvdHH11] Michiel Bertsch, Rein van der Hout, and Josephus Hulshof. Energy concentration for 2-dimensional radially symmetric equivariant harmonic map heat flows. *Commun. Contemp. Math.*, 13(4):675–695, 2011.

[Coc52] William G. Cochran. The χ^2 test of goodness of fit. *Ann. Math. Statistics*, 23:315–345, 1952.

[CZ87] X. R. Chen and L. C. Zhao. Almost sure L_1-norm convergence for data-based histogram density estimates. *J. Multivariate Anal.*, 21(1):179–188, 1987.

[Das08] Anirban DasGupta. *Asymptotic theory of statistics and probability*. Springer Texts in Statistics. Springer, New York, 2008.

[Das11] Anirban DasGupta. *Probability for statistics and machine learning*. Springer Texts in Statistics. Springer, New York, 2011. Fundamentals and advanced topics.

[Dav04] Timothy A. Davis. Algorithm 832: UMFPACK V4.3—an unsymmetric-pattern multifrontal method. *ACM Trans. Math. Software*, 30(2):196–199, 2004.

[DB02] Peter Deuflhard and Folkmar Bornemann. *Scientific computing with ordinary differential equations*, volume 42 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 2002. Translated from the 1994 German original by Werner C. Rheinboldt.

[Deu11] Peter Deuflhard. *Newton methods for nonlinear problems*, volume 35 of *Springer Series in Computational Mathematics*. Springer, Heidelberg, 2011. Affine invariance and adaptive algorithms, First softcover printing of the 2006 corrected printing.

[Dev88] L. Devroye. Automatic pattern recognition: a study of the probability of error. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):530–543, 1988.

[DG85] Luc Devroye and László Györfi. *Nonparametric density estimation*. Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics. John Wiley & Sons, Inc., New York, 1985. The L_1 view.

[DG02] L. Devroye and L. Györfi. Distribution and density estimation. In *Principles of nonparametric learning (Udine, 2001)*, volume 434 of *CISM Courses and Lect.*, pages 211–270. Springer, Vienna, 2002.

[DGL96] Luc Devroye, László Györfi, and Gábor Lugosi. *A probabilistic theory of pattern recognition*, volume 31 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1996.

[DGLW17] Luc Devroye, László Györfi, Gábor Lugosi, and Harro Walk. On the measure of Voronoi cells. *J. Appl. Probab.*, 54(2):394–408, 2017.

[DH97] A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*, volume 1 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1997. With 1 IBM-PC floppy disk (3.5 inch; HD).

[DM98] Leonardo Dagum and Ramesh Menon. Openmp: An industry-standard api for shared-memory programming. *IEEE Comput. Sci. Eng.*, 5(1):46–55, 1998.

[DP16] Thomas Dunst and Andreas Prohl. The Forward-Backward Stochastic Heat Equation: Numerical Analysis and Simulation. *SIAM J. Sci. Comput.*, 38(5):A2725–A2755, 2016.

[DPZ14] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic equations in infinite dimensions*, volume 152 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2nd edition, 2014.

[DS86] Ralph B. D'Agostino and Michael A. Stephens, editors. *Goodness-of-fit techniques*, volume 68 of *Statistics: Textbooks and Monographs*. Marcel Dekker, Inc., New York, 1986.

[EEHJ95] Kenneth Eriksson, Don Estep, Peter Hansbo, and Claes Johnson. Introduction to adaptive methods for differential equations. In *Acta numerica, 1995*, Acta Numer., pages 105–158. Cambridge Univ. Press, Cambridge, 1995.

[EEHJ96] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson. *Computational differential equations*. Cambridge University Press, Cambridge, 1996.

[Efr79] B. Efron. Bootstrap methods: another look at the jackknife. *Ann. Statist.*, 7(1):1–26, 1979.

[Fri64] Avner Friedman. *Partial differential equations of parabolic type*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964.

[Gal09] M. Galassi. *GNU Scientific Library Reference Manual - Third Edition*, 2009.

[GH93] Bo Ling Guo and Min Chun Hong. The Landau-Lifshitz equation of the ferromagnetic spin chain and harmonic maps. *Calc. Var. Partial Differential Equations*, 1(3):311–334, 1993.

[GHPS12] P. Goldenits, G. Hrkac, D. Praetorius, and D. Suess. An effective integrator for the landau-lifshitz-gilbert equation. *IFAC Proceedings Volumes*, 45(2):493–497, 2012.

[Gil04] T. L. Gilbert. A phenomenological theory of damping in ferromagnetic materials. *IEEE Transactions on Magnetics*, 40(6):3443–3449, 2004.

[GKKW02] László Györfi, Michael Kohler, Adam Krzyżak, and Harro Walk. *A distribution-free theory of nonparametric regression.* Springer Series in Statistics. Springer-Verlag, New York, 2002.

[GL97] J. G. Gaines and T. J. Lyons. Variable step size control in the numerical solution of stochastic differential equations. *SIAM J. Appl. Math.*, 57(5):1455–1484, 1997.

[Gla04] Paul Glasserman. *Monte Carlo methods in financial engineering*, volume 53 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 2004. Stochastic Modelling and Applied Probability.

[GLW05] Emmanuel Gobet, Jean-Philippe Lemor, and Xavier Warin. A regression-based Monte Carlo method to solve backward stochastic differential equations. *Ann. Appl. Probab.*, 15(3):2172–2202, 2005.

[GN88] Alan George and Esmond Ng. On the complexity of sparse QR and LU factorization of finite-element matrices. *SIAM J. Sci. Statist. Comput.*, 9(5):849–861, 1988.

[GO80] Louis Gordon and Richard A. Olshen. Consistent nonparametric regression from recursive partitioning schemes. *J. Multivariate Anal.*, 10(4):611–627, 1980.

[GO84] Louis Gordon and Richard A. Olshen. Almost surely consistent nonparametric regression from recursive partitioning schemes. *J. Multivariate Anal.*, 15(2):147–163, 1984.

[GPL98] José Luis García-Palacios and Francisco J. Lázaro. Langevin-dynamics study of the dynamical properties of small magnetic particles. *Phys. Rev. B*, 58:14937–14958, 1998.

[GR15] Eugene C. Gartland, Jr. and Alison Ramage. A renormalized Newton method for liquid crystal director modeling. *SIAM J. Numer. Anal.*, 53(1):251–278, 2015.

[GS02] Alison L. Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *Internat. Stat. Rev.*, 70(3):419–435, 2002.

[HMGR01] Norbert Hofmann, Thomas Müller-Gronbach, and Klaus Ritter. The optimal discretization of stochastic differential equations. *J. Complexity*, 17(1):117–153, 2001.

[Hoc15] Antoine Hocquet. *The Landau-Lifshitz-Gilbert equation driven by Gaussian noise.* Theses, Ecole Polytechnique, December 2015.

[Hoc16] Antoine Hocquet. Struwe-like solutions for the stochastic harmonic map flow. 2016.

[Hör67] Lars Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.

[IW89] Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo, second edition, 1989.

[JN11] Volker John and Julia Novo. Error analysis of the SUPG finite element discretization of evolutionary convection-diffusion-reaction equations. *SIAM J. Numer. Anal.*, 49(3):1149–1176, 2011.

[JNP84] Claes Johnson, Uno Nävert, and Juhani Pitkäranta. Finite element methods for linear hyperbolic problems. *Comput. Methods Appl. Mech. Engrg.*, 45(1-3):285–312, 1984.

[JS08] Volker John and Ellen Schmeyer. Finite element methods for time-dependent convection-diffusion-reaction equations with small diffusion. *Comput. Methods Appl. Mech. Engrg.*, 198(3-4):475–494, 2008.

[KP92] Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[KS91] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[Lam03] H. Lamba. An adaptive timestepping algorithm for stochastic differential equations. *J. Comput. Appl. Math.*, 161(2):417–430, 2003.

[LL35] Lev Davidovich Landau and E Lifshitz. On the theory of the dispersion of magnetic permeability in ferromagnetic bodies. *Phys. Z. Sowjet.*, 8:153–169, 1935.

[LMS07] H. Lamba, J. C. Mattingly, and A. M. Stuart. An adaptive Euler-Maruyama scheme for SDEs: convergence and stability. *IMA J. Numer. Anal.*, 27(3):479–506, 2007.

[LN96] Gábor Lugosi and Andrew Nobel. Consistency of data-driven histogram methods for density estimation and classification. *Ann. Statist.*, 24(2):687–706, 1996.

[LR05] E. L. Lehmann and Joseph P. Romano. *Testing statistical hypotheses*. Springer Texts in Statistics. Springer, New York, third edition, 2005.

[LYW16] Dangna Li, Kun Yang, and Wing Hung Wong. Density estimation via discrepancy based adaptive sequential partition. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pages 1099–1007. Curran Associates Inc., 2016.

[Mil95] G. N. Milstein. *Numerical integration of stochastic differential equations*, volume 313 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1995. Translated and revised from the 1988 Russian original.

[MSTZ03] Kyoung-Sook Moon, Anders Szepessy, Raúl Tempone, and Georgios E. Zouraris. A variational principle for adaptive approximation of ordinary differential equations. *Numer. Math.*, 96(1):131–152, 2003.

[MSTZ05] Kyoung-Sook Moon, Anders Szepessy, Raúl Tempone, and Georgios E. Zouraris. Convergence rates for adaptive weak approximation of stochastic differential equations. *Stoch. Anal. Appl.*, 23(3):511–558, 2005.

[MW01] J. M. Melenk and B. I. Wohlmuth. On residual-based a posteriori error estimation in *hp*-FEM. *Adv. Comput. Math.*, 15(1-4):311–331 (2002), 2001. A posteriori error estimation and adaptive computational methods.

[N46] Louis Néel. Bases d'une nouvelle théorie générale du champ coercitif. *Ann. Inst. Fourier (Grenoble)*, 22:299–343, 1946.

[Nob96] Andrew Nobel. Histogram regression estimation using data-dependent partitions. *Ann. Statist.*, 24(3):1084–1105, 1996.

[OBSC00] Atsuyuki Okabe, Barry Boots, Kokichi Sugihara, and Sung Nok Chiu. *Spatial tessellations: concepts and applications of Voronoi diagrams*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, second edition, 2000. With a foreword by D. G. Kendall.

[Pit79] E. J. G. Pitman. *Some basic theory for statistical inference*. Chapman and Hall, London; A Halsted Press Book, John Wiley & Sons, New York, 1979. Monographs on Applied Probability and Statistics.

[PK80] Arthur V. Peterson, Jr. and Richard A. Kronmal. A representation for discrete distributions by equiprobable mixtures. *J. Appl. Probab.*, 17(1):102–111, 1980.

[PS18] Andreas Prohl and Christian Schellnegger. Adaptive concepts for stochastic partial differential equations. submitted, 2018.

[Rod94] Rodolfo Rodríguez. Some remarks on Zienkiewicz-Zhu estimator. *Numer. Methods Partial Differential Equations*, 10(5):625–635, 1994.

[RST08] Hans-Görg Roos, Martin Stynes, and Lutz Tobiska. *Robust numerical methods for singularly perturbed differential equations*, volume 24 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2008. Convection-diffusion-reaction and flow problems.

[Sha73] L. F. Shampine. Local extrapolation in the solution of ordinary differential equations. *Math. Comp.*, 27:91–97, 1973.

[Ske82] Robert D. Skeel. A theoretical framework for proving accuracy results for deferred corrections. *SIAM J. Numer. Anal.*, 19(1):171–196, 1982.

[SS05] Alfred Schmidt and Kunibert G. Siebert. *Design of adaptive finite element software*, volume 42 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin, 2005.

[Sto74] M. Stone. Cross-validatory choice and assessment of statistical predictions. *J. Roy. Statist. Soc. Ser. B*, 36:111–147, 1974. With discussion by G. A. Barnard, A. C. Atkinson, L. K. Chan, A. P. Dawid, F. Downton, J. Dickey, A. G. Baker, O. Barndorff-Nielsen, D. R. Cox, S. Giesser, D. Hinkley, R. R. Hocking, and A. S. Young, and with a reply by the authors.

[Sto85] Charles J. Stone. An asymptotically optimal histogram selection rule. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer,*

Vol. II (Berkeley, Calif., 1983), Wadsworth Statist./Probab. Ser., pages 513–520. Wadsworth, Belmont, CA, 1985.

[Str96] Michael Struwe. Geometric evolution problems. In *Nonlinear partial differential equations in differential geometry (Park City, UT, 1992)*, volume 2 of *IAS/Park City Math. Ser.*, pages 257–339. Amer. Math. Soc., Providence, RI, 1996.

[STZ01] Anders Szepessy, Raúl Tempone, and Georgios E. Zouraris. Adaptive weak approximation of stochastic differential equations. *Comm. Pure Appl. Math.*, 54(10):1169–1214, 2001.

[TT90] Denis Talay and Luciano Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4):483–509 (1991), 1990.

[VC15] V. N. Vapnik and A. Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. In *Measures of complexity*, pages 11–30. Springer, Cham, 2015. Reprint of Theor. Probability Appl. **16** (1971), 264–280.

[vdBW13] Jan Bouwe van den Berg and J. F. Williams. (In-)stability of singular equivariant solutions to the Landau-Lifshitz-Gilbert equation. *European J. Appl. Math.*, 24(6):921–948, 2013.

[Ver94] R. Verfürth. A posteriori error estimation and adaptive mesh-refinement techniques. In *Proceedings of the Fifth International Congress on Computational and Applied Mathematics (Leuven, 1992)*, volume 50, pages 67–83, 1994.

[vK81] N. G. van Kampen. *Stochastic processes in physics and chemistry*, volume 888 of *Lecture Notes in Mathematics*. North-Holland Publishing Co., Amsterdam-New York, 1981.

[Vos91] Michael D. Vose. A linear algorithm for generating random numbers with a given distribution. *IEEE Trans. Software Engrg.*, 17(9):972–975, 1991.

[Wal77] Alastair J. Walker. An efficient method for generating discrete random variables with general distributions. *ACM Trans. Math. Softw.*, 3(3):253–256, 1977.

[ZKC90] L. C. Zhao, P. R. Krishnaiah, and X. R. Chen. Almost sure L_r-norm convergence for data-based histogram density estimates. *Teor. Veroyatnost. i Primenen.*, 35(2):391–397, 1990.

[ZZ92] O. C. Zienkiewicz and J. Z. Zhu. The superconvergent patch recovery and a posteriori error estimates. I. The recovery technique. *Internat. J. Numer. Methods Engrg.*, 33(7):1331–1364, 1992.