Adaptive concepts for high-dimensional stochastic differential equations

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

vorgelegt von
Fabian Merle
aus Hechingen

Tübingen
2022

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

Tag der mündlichen Qualifikation:	31.10.2022
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Dr. Andreas Prohl
2. Berichterstatter:	Prof. Dr. Michael V. Tretyakov

Abstract

The objective of this thesis is the efficient approximation of high-dimensional stochastic differential equations (SDE's) via newly developed, theoretical-based adaptive methods. The thesis is split into two parts, which motivate and discuss the (temporal) approximation of high-dimensional SDE's from different aspects. Conceptually, the derivation of the corresponding adaptive methods follows the same principle: finding an appropriate scheme for the approximation of the underlying SDE, derivation of a (weak) *a posteriori* error estimate, and an implementation of an adaptive method based on it.

In the first part of this thesis we mainly consider SDE systems emerging from a spatial discretization of a given semilinear stochastic partial differential equation (SPDE). The corresponding adaptive method consists of the semi-implicit Euler scheme and a local refinement/coarsening strategy of the temporal mesh based on a computable error estimator, and generates time step sizes as well as iterates, such that the resulting (weak) error is always less or equal than a prescribed tolerance. The (computable) error estimator directly comes from the related *a posteriori* error estimate, which is derived by means of the *Kolmogorov equation*. In this regard, we (globally) bound derivatives of the solution of *Kolmogorov's equation* via (probabilistic) variation equations independently of the dimension and in terms of derivatives of the underlying test function. At this juncture, the use of the *Clark-Ocone formula* reduces the complexity of the derivatives to be bounded. Furthermore, the approximation via the semi-implicit Euler scheme allows for stability bounds which are independent of the dimension, and which, in particular, contribute to bound the error estimator. The combination of the above concepts enables an error analysis of the *a posteriori* estimate resp. the estimator, which is independent of the dimension, and, in particular, is the key for convergence of the adaptive method, as well as its applicability in high dimensions. Computational experiments compare adaptive meshes with uniform meshes and show a considerable gain in efficiency of the adaptive method.

The second part can conceptually be regarded as an extension of the first one and considers SDE systems, which arise from the probabilistic reformulation of an underlying boundary value problem, *i.e.*, of an elliptic/parabolic partial differential equation (PDE) on a bounded domain. Opposed to the setting in the first part, the solution of the SDE here takes values in a bounded domain, which, in particular, involves a convenient exposure to stopping in an approximative framework when the (approximated) solution process is about to leave the domain. To this end, we use an already existing scheme in the literature (slightly modified), which, among other things, replaces unbounded Wiener increments in the generation of (explicit) Euler iterates by bounded ones having the same distribution, and which thus allows to properly control the dynamics of the (approximated) solution process up to the boundary of the domain. Based on this scheme, we derive an *a posteriori* error estimate from which three error estimators emerge, where each of them captures different dynamics concerning the distance of the approximated process to the boundary. These dynamics are especially reflected in the choice of the local time step size selection (up to the boundary) of the adaptive method, which approximates the solution of the underlying boundary value problem at a fixed point. The choice of the local time step sizes is complemented by a suitable temporal weight factor within the related refinement/coarsening strategy, which, aside from stability results concerning stopping dynamics, ensures the (optimal) convergence of the method with respect to a given tolerance parameter. Computational experiments illustrate a stable application of the method even for violated data requirements, and a substantial gain in efficiency through adaptive (time) mesh generation.

Zusammenfassung

Ziel dieser Arbeit ist es, hochdimensionale stochastische Differentialgleichungen (SDE's) effizient mithilfe von neu entwickelten, theoretisch basierten adaptiven Verfahren zu approximieren. Die vorliegende Dissertation ist in zwei wesentliche Teile untergliedert, welche das (zeitliche) Approximieren von (hochdimensionalen) SDE's aus jeweils unterschiedlichen Aspekten motivieren und diskutieren. Die Herleitung der zugehörigen adaptiven Verfahren in beiden Teilen folgt konzeptionell dem gleichen Prinzip: Aufstellen eines geeigneten Schemas für die Approximation der zugrundeliegenden SDE, Herleitung einer (schwachen) *a posteriori* Fehlerabschätzung und Realisierung eines darauf basierenden adaptiven Verfahrens.

Im ersten Teil dieser Arbeit werden hauptsächlich SDE Systeme betrachtet, welche im Zuge einer Ortsdiskretisierung einer gegebenen (semiliniearen) stochastischen partiellen Differentialgleichung (SPDE) hervorgehen. Das zugehörige adaptive Verfahren, bestehend aus dem semi-impliziten Euler Schema und einer lokalen Verfeinerungs- bzw. Vergröberungsstrategie des zeitlichen Gitters, welche auf einem berechenbaren Fehlerschätzer basiert, generiert Zeitschrittweiten sowie Iterierte derart, dass der resultierende (schwache) Fehler stets kleiner gleich einer vorgegebenen Toleranz ist. Der (berechenbare) Fehlerschätzer resultiert unmittelbar aus der zugehörigen *a posteriori* Fehlerabschätzung, welche mithilfe der *Kolmogorov-Gleichung* hergeleitet wird. In diesem Zusammenhang werden auftretende Ableitungen der Lösung der *Kolmogorov-Gleichung* mittels (probabilistischen) Variations-Gleichungen dimensionsunabhängig und in Form von Ableitungen der zugrundeliegenden Testfunktion beschränkt. Hierbei verringert der Einsatz der *Clark-Ocone Formel* die Komplexität der Ordnung der zu beschränkenden Ableitungen. Des Weiteren erlaubt die Approximation durch das semi-implizite Euler Schema dimensionsunabhängige Stabilitätsabschätzungen, mit deren Hilfe der Fehlerschätzer beschränkt werden kann. Die Kombination dieser obigen Konzepte ermöglicht eine dimensionsunabhängige Analyse der Fehlerabschätzung bzw. des Fehlerschätzers und ist somit insbesondere der Schlüssel für die Konvergenz des adaptiven Verfahrens sowie für dessen Anwendbarkeit in hohen Raumdimensionen. Numerische Studien vergleichen adaptive Gitterwahl mit uniformer, und zeigen eine beachtliche Effizienzsteigerung des adaptiven Verfahrens auf.

Der zweite Teil kann konzeptionell als Erweiterung des ersten angesehen werden und betrachtet SDE Systeme, welche aus der probabilistischen Lösungsdarstellung eines zugrundeliegenden Randwertproblems, d.h. einer elliptischen/parabolischen partiellen Differentialgleichung (PDE), hervorgehen. Anders als im ersten Teil nimmt der Lösungsprozess der SDE hier nur Werte in einem beschränkten Gebiet an, was insbesondere einen vernünftigen Umgang mit der zugehörigen Randproblematik in einem geeigneten approximativen Rahmen erfordert. Dazu wird ein in der Literatur bereits existierendes Schema (leicht modifiziert) verwendet, welches unter anderem unbeschränkte Wiener Inkremente in der Generierung von (expliziten) Euler-Iterierten durch „in Verteilung gleiche“ beschränkte Inkremente ersetzt und somit eine exakte Kontrolle des approximierten Prozesses bis einschließlich zum Rand des Gebietes erlaubt. Basierend hierauf wird eine *a posteriori* Fehlerabschätzung hergeleitet, aus welcher drei Fehlerschätzer hervorgehen, die jeweils unterschiedliche Dynamiken (bzgl. des Abstand zum Randes) des approximierten Prozesses erfassen. Diese Dynamiken widerspiegeln sich insbesondere in der flexiblen, lokalen Zeitschrittwahl des darauf aufbauenden adaptiven Verfahrens, welches punktweise die Lösung des vorliegenden Randwertproblems approximiert. Ergänzt wird diese Zeitschrittwahl durch ein geeignetes Zeitgewicht in der zugehörigen (lokalen) Verfeinerungs- bzw. Vergröberungsstrategie, welches neben Stabilitätsresultaten hinsichtlich der Randproblematik die (optimale) Konvergenz des Verfahrens bzgl. einer gegebenen Toleranz sicherstellt. Numerische Studien belegen eine robuste Anwendbarkeit des Verfahrens in hohen Dimensionen und bei verletzten Datenanforderungen sowie eine beachtliche Effizienzsteigerung durch adaptive (Zeit-)Gitterwahl.

List of publications and contributions

The present work is based on the following list of publications:

[58] F. Merle, A. Prohl, *An adaptive time-stepping method based on a posteriori weak error analysis for large SDE systems*, Numer. Math. **149**, pp. 417-462 (2021).

[59] F. Merle, A. Prohl, *A posteriori error analysis and adaptivity for high-dimensional elliptic and parabolic boundary value problems*, Numer. Math. (to appear).

The first part of this thesis is based on [58] and the second part is based on [59]. The theoretical results are equal contributions of the authors. The implementation of the numerical experiments is due to the author of the present thesis.

Acknowledgments

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

There are some people deserving a special note of thanks.

I am extremely grateful to my supervisor, Prof. Dr. Andreas Prohl, for giving me the opportunity to write my thesis in the interesting and important field of adaptivity for stochastic differential equations. I thank him for his scholastic guidance, his inspiration, his valuable suggestions, his patience, and his overall constant support.
I also thank him and Prof. Dr. Michael V. Tretyakov for being the two evaluators of this thesis.

I would also like to thank Prof. Dr. Martin Zerner for a helpful discussion concerning some results in part two of this thesis.

I thank my former colleague Patrick König, who gave me programming advice, especially at the beginning of my thesis.

Much gratitude to my colleague, Dr. Akash Ashirbad Panda, who has also become a good friend of mine, for reading parts of this thesis and giving me feedback. I also thank him for many helpful discussions.

Special thanks to all my friends for their constant moral support, especially over the last three years, which have been a difficult time for me due to the pandemics and personal setbacks. In particular, my deep gratitude goes to Maik, Matze, Hennrik and Sissi.

Finally, I would like to thank my parents Heidrun and Ludwig, and my sister Isabelle, who have always been supporting and encouraging me my entire life.

Contents

List of Figures

List of Tables

Introduction

The mathematical modeling of 'real-world problems' in order to describe them and/or to make predictions, is one of the most important — if not the most important — subject of (applied) mathematics. Many 'real-world problems' such as *e.g.* the *motion of masses subjected to forces* (mechanics), *radioactive decay of materials* (modern physics), *games and multi-stage processes* (statistical systems), *chemical reactions* (chemistry), *epidemic models for diseases* (biology), dynamics of the population of predator and prey species (ecology), *option pricing* (finance), *firing activity of a single neuron* (neuronal activity), *coordination of human movement* (experimental psychology), *traffic reconstruction using autonomous vehicles* (autonomous driving), concern relationships between changing quantities and can thus be described via differential equations; see *e.g.* [81, 52, 63]. Once having found an appropriate differential equation to describe the dynamics of a given 'real-world problem', one typically needs to find a corresponding solution. However, 'real-world problems' are complex in general and corresponding models are often complicated, *i.e.*, involve perturbations ('noise'), complex data structures, and may be high-dimensional. Consequently, solving (such) a given differential equation is usually hard and in fact, is only possible in very few cases, which is why their numerical approximation becomes enormously important. The (efficient) numerical approximation of different types of differential equations such as *e.g.* ordinary differential equations (ODE's), partial differential equations (PDE's), stochastic differential equations (SDE's) or stochastic partial differential equations (SPDE's) has been extensively studied within the last decades and is still a very active field of research; see *e.g.* [42, 52, 73, 6, 46, 82] and references therein. The main properties and goals in this respect for a corresponding numerical method to have are always: *convergence* to the (unknown) solution, *stability/termination*, *flexibility*, *i.e.*, is the method applicable for general data settings, and *efficiency*, *i.e.*, how long does it take to finish the approximation and how large are storage resources. Typically, numerical methods for ODE's and SDE's only involve time discretizations to approximate the corresponding solution at some given time points — opposed to methods for PDE's and SPDE's, which usually require a discretization of the underlying state space (in addition). Generally, corresponding time- and space discretization parameters, *i.e.*, the time step size $\Delta t > 0$ and the spatial mesh size $h > 0$ are constant and convergence results with respect to these parameters are proposed on uniform spatial resp. temporal meshes. However, a uniform time discretization with a fixed Δt for instance may neglect temporal dynamics of the underlying ODE/SDE system and might lead to an unnecessary fine resp. too coarse resolving of the temporal mesh in certain areas, where corresponding dynamics remain unchanged resp. vary a lot, and consequently causes a loss in efficiency. To overcome this 'flaw' and to boost the efficiency of the approximation, *adaptive* methods are proposed, which automatically generate flexible mesh sizes to capture the involved dynamics

of the system more adequately.
While there is rich literature on adaptive concepts for ODE's and PDE's; see *e.g.* [16, 29, 30, 48] and Chapter 7, as well as references therein, there are only a few works on adaptivity for SDE's and SPDE's; see Chapter1 and *e.g.* [72, 57]. A reason for this is the involved 'randomness' in the stochastic models, which makes an extension of adaptive concepts (as well as their theoretical analysis) for ODE's (resp. PDE's) to SDE's (resp. SPDE's) much more difficult. Most of the adaptive strategies above rely on the error between the (unknown) solution and its numerical approximation. In the framework of SDE's and SPDE's, this view is referred to the concept of *strong approximation*, *i.e.*, pathwise approximation. However, many applications such as *e.g.* the approximation of expected functionals or the probabilistic representation (of the solution) of PDE's ('*Feynman-Kac formula*') involve the concept of *weak approximation*, *i.e.*, where one is only interested in the error between distributions of the (unknown) solution of a given SDE resp. SPDE and its numerical approximation. So far, proposed adaptive time-stepping strategies for SDE's based on the concept of *weak approximation* are limited to small dimensions [76, 64] or either lack a (fully) theoretical backup [72]. Inspired by these limitations, and as far as the expressions '*flexibility*' and '*efficiency*' are concerned; this is precisely the point where the newly developed concepts presented in this thesis come into play; see also [58, 59].

Let $L, K \in \mathbb{N}$, and $T > 0$. In the first part of this thesis, which consists of Chapters 1 – 5, we study a new adaptive time-stepping strategy to efficiently approximate the $\mathbb{R}^L$-valued solution $\mathbf{X} \equiv \{\mathbf{X}_t;\, t \in [0,T]\}$ of the SDE

$$\mathrm{d}\mathbf{X}_t = \Big(-\mathscr{A}\mathbf{X}_t + \mathbf{f}(\mathbf{X}_t)\Big)\,\mathrm{d}t + \sum_{k=1}^{K} \boldsymbol{\sigma}_k(\mathbf{X}_t)\,\mathrm{d}\beta_k(t) \quad \forall\, t \in [0,T]\,, \qquad \mathbf{X}_0 = \mathbf{y} \in \mathbb{R}^L\,, \tag{0.1}$$

where $\{\beta_k(t)\,;\, t \in [0,T]\}$, $k = 1, ..., K$ are independent $\mathbb{R}$-valued Wiener processes on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$, and $\mathscr{A} \in \mathbb{R}^{L\times L}$ is invertible and positive definite. We refer to Chapter 2, where proper settings for data $\mathscr{A}, \mathbf{f}, \{\boldsymbol{\sigma}_k\}_k$, are given. Problem (0.1) may be motivated from a spatial discretization of the semilinear stochastic partial differential equation (SPDE) on a bounded domain $\boldsymbol{\mathcal{D}} \subset \mathbb{R}^d$,

$$\mathrm{d}X_t = \Big(\varepsilon\Delta X_t + [\boldsymbol{\beta}\cdot\nabla]X_t + F(X_t)\Big)\,\mathrm{d}t + \sum_{k=1}^{K} \Sigma_k(X_t)\,\mathrm{d}\beta_k(t) \quad \forall\, t \in [0,T]\,, \qquad X_0 = y \in \mathbb{H}\,, \tag{0.2}$$

for given $\varepsilon > 0$, $\boldsymbol{\beta} : \boldsymbol{\mathcal{D}} \to \mathbb{R}^d$ constant for simplicity, and $\mathbb{H}$ a Hilbert space; see Chapter 5 for further details.
Our aim is an adaptive mesh strategy for the semi-implicit Euler method applied to (0.1), which, for every $j \in \mathbb{N}_0$, automatically selects the new step size $\tau^{j+1} = t_{j+1} - t_j$, and then determines the $\mathbb{R}^L$-valued random variable $\mathbf{Y}^{j+1}$ from $(j \in \mathbb{N}_0)$

$$\mathbf{Y}^{j+1} = \mathbf{Y}^j + \tau^{j+1}\Big(-\mathscr{A}\mathbf{Y}^{j+1} + \mathbf{f}(\mathbf{Y}^j)\Big) + \sum_{k=1}^{K} \boldsymbol{\sigma}_k(\mathbf{Y}^j)\Delta_{j+1}\beta_k\,, \qquad \mathbf{Y}^0 = \mathbf{y}\,, \tag{0.3}$$

for $\Delta_{j+1}\beta_k := \beta_k(t_{j+1}) - \beta_k(t_j)$, to approximate the solution $\mathbf{X}_{t_{j+1}}$ from (0.1) at time t_{j+1}. Conceptually, we base this local step size selection strategy on a (computable) *a posteriori* weak error estimator $\mathfrak{G}$ in each step, *i.e.*,

$$\max_{0\leq j\leq J}\left|\mathbb{E}\Big[\phi(\mathbf{X}_{t_j})\Big] - \mathbb{E}\Big[\phi(\mathbf{Y}^j)\Big]\right| \leq \sum_{j=0}^{J-1} \tau^{j+1}\mathfrak{G}\Big(\phi;\tau^{j+1},\mathbf{Y}^j\Big)\,, \tag{0.4}$$

for $\phi \in \mathcal{C}^3(\mathbb{R}^L)$ with globally bounded first, second and third derivatives. A criterion may then be set up to select a new, large τ^{j+1} in every step, such that the right-hand side of (0.4) stays below a chosen tolerance `Tol` > 0. For the derivation of (0.4) we benefit from [77], where an expansion of the weak approximation error for uniform deterministic time steps (and originally for the explicit Euler method) was obtained via *Kolmogorov's backward equation*:

$$\partial_t u(t,\mathbf{x}) + \mathcal{L}u(t,\mathbf{x}) = 0 \qquad \forall\,(t,\mathbf{x}) \in [0,T)\times\mathbb{R}^L\,, \tag{0.5a}$$

$$u(T,\mathbf{x}) = \phi(\mathbf{x}) \qquad \forall\,\mathbf{x} \in \mathbb{R}^L\,, \tag{0.5b}$$

where $\mathcal{L} \equiv \mathcal{L}_{\mathbf{X}}$ is the generator of the Markovian semigroup from $\mathbf{X} \equiv \{\mathbf{X}_t;\, t \in [0,T]\}$ in (0.1),

$$\begin{aligned}\mathcal{L}u(t,\mathbf{x}) &= \Big\langle -\mathscr{A}\mathbf{x} + \mathbf{f}(\mathbf{x}), D_{\mathbf{x}}u(t,\mathbf{x})\Big\rangle_{\mathbb{R}^L} + \frac{1}{2}\sum_{k=1}^{K}\mathrm{Tr}\Big(\boldsymbol{\sigma}_k(\mathbf{x})\boldsymbol{\sigma}_k^\top(\mathbf{x})D^2_{\mathbf{x}}u(t,\mathbf{x})\Big)\\ &= \Big\langle -\mathscr{A}\mathbf{x} + \mathbf{f}(\mathbf{x}), D_{\mathbf{x}}u(t,\mathbf{x})\Big\rangle_{\mathbb{R}^L} + \frac{1}{2}\mathrm{Tr}\Big(\boldsymbol{\sigma}(\mathbf{x})\boldsymbol{\sigma}^\top(\mathbf{x})D^2_{\mathbf{x}}u(t,\mathbf{x})\Big)\,,\end{aligned}$$

with $\boldsymbol{\sigma}(\cdot) \equiv \Big[\boldsymbol{\sigma}_1(\cdot),...,\boldsymbol{\sigma}_K(\cdot)\Big] \in \mathbb{R}^{L\times K}$. Under proper assumptions, such as for instance those stated in Chapter 2, the function $u(t,\mathbf{x}) = \mathbb{E}\Big[\phi(\mathbf{X}_T^{t,\mathbf{x}})\Big]$ is the unique solution of (0.5); see *e.g.* [49, p. 366*ff*.]. As usual we denote by $\mathbf{X}^{t,\mathbf{x}} \equiv \Big\{\mathbf{X}_s^{t,\mathbf{x}};\, s \in [t,T]\Big\}$ the $\mathbb{R}^L$-valued process which starts at time $t \in [0,T]$ in $\mathbf{x} \in \mathbb{R}^L$.
As already mentioned, we are motivated by (a spatial discretization of) SPDE (0.2), which is why we aim for adaptive methods, which are applicable to SDE (0.1) with $L \gg 1$ large; in this respect, we prefer deterministic (rather than random) meshes $\{t_j\}_{j\geq 0} \subset [0,T]$ to avoid requirements for too large storage resources, or time-consuming post-processing tasks to synchronize data, such as interpolation, or projection. This approach lends itself to a vectorized implementation (see Algorithm 4.1) and is of advantage over procedurally generated meshes (such as those in [32, 50, 51]) where the efficient implementation as a vectorized algorithm is an unsolved problem.

The following example illustrates local mesh refinement and coarsening by the adaptive Algorithm 4.1, which is detailed in Chapters 4 and 5.

Example 0.1. Let $L = 25$. Consider SDE system (0.1) with $K = 5$, which results from a finite element discretization (with spatial mesh size $h = \frac{1}{L+1}$) of SPDE (0.2) with $\varepsilon = 1$,

$\boldsymbol{\beta} \equiv \mathbf{0}$, and

$$F\Big(X_t(x)\Big) = \tfrac{1}{5}\sin\Big(\pi X_t(x)\Big), \quad \Sigma_k\Big(X_t(x)\Big) = \frac{1}{2k}\sin(\pi k x)X_t(x), \quad y(x) = \sin(\pi x)$$

for $x \in (0,1)$; see Section 5.1 for details. For the test function $\phi(\mathbf{x}) = \sqrt{h}\|\mathbf{x}\|_{\mathbb{R}^L}$ to approximate the $\mathbb{L}^2$-norm, and an initial step of size 0.1, we observe an instantaneous refinement via Algorithm 4.1 to $\tau^1 \approx 10^{-4}$; the mesh size then rapidly increases to values close to 10^{-1} at times $t \approx 0.5$, reflecting (spatial) smoothing dynamics; see Figure 0.1(B). Figure 0.1(A) shows a typical trajectory, where the buckling is caused by the driving noise. Figure 0.1(C) compares related errors for (0.3) on uniform vs. adaptive time meshes through Algorithm 4.1. Here, $\mathbb{E}_{\mathtt{M}}[\phi(\mathbf{Y}^j)] := \frac{1}{\mathtt{M}}\sum_{m=1}^{\mathtt{M}}\phi\Big(\mathbf{Y}^j(\omega_m)\Big)$ denotes the empirical mean to approximate $\mathbb{E}[\phi(\mathbf{Y}^j)]$, where we choose $\mathtt{M} = 10^4$ Monte-Carlo simulations. For the tolerance parameter $\mathtt{Tol} = 0.1$, Algorithm 4.1 generates an adaptively refined mesh with $J = 501$ time steps to stay below the given error threshold $\mathtt{Tol}$. In contrast, a uniform mesh needs $J = 2000$ time steps to perform equally well. In Figure 0.1(D), the evolution of the *a posteriori* error estimator $j \mapsto \mathfrak{G}^{(\mathtt{M})}\Big(\phi;\tau^{j+1},\mathbf{Y}^j\Big)$ is displayed, that approximately takes values between $\mathtt{Tol}$ and $\frac{\mathtt{Tol}}{2}$; this is in accordance with the tolerance criterion of Algorithm 4.1, indicating an efficient selection of variable step sizes. See Section 5.1 for more details.

Furthermore, since our adaptive mesh strategy is related to the concept of weak approximation, it can also be used to efficiently approximate the solution of (the high-dimensional) PDE (0.5) (in a pointwise manner), which is another motivation for problem (0.1). In fact, the numerical approximation of high-dimensional (parabolic) PDE's on the whole space is of great relevance due to their ubiquitous occurrence in nature. Typical examples are:

- The *Schrödinger equation* in quantum mechanics

$$\mathbf{i}\partial_t u(t,\mathbf{x}) + \Delta u(t,\mathbf{x}) - V(\mathbf{x})u(t,\mathbf{x}) = 0 \qquad \forall\,(t,\mathbf{x}) \in [0,T)\times\mathbb{R}^L\,, \tag{SE}$$

  to describe the state (represented by the wave function $u(t,\mathbf{x})$) of a quantum mechanical system and its dynamical changes in time, where $\mathbf{i}$ is the imaginary unit and $V:\mathbb{R}^L \to \mathbb{R}$ is a potential. Here, the dimension L is three times the number of particles considered in the system.

- The *Black-Scholes equation* in finance

$$\partial_t u(t,\mathbf{x}) + \frac{1}{2}\sigma^2\sum_{i=1}^{L} x_i^2\partial_{x_i}\partial_{x_i}u(t,\mathbf{x}) + r\Big\langle\mathbf{x}, D_{\mathbf{x}}u(t,\mathbf{x})\Big\rangle_{\mathbb{R}^L} - ru(t,\mathbf{x}) = 0 \qquad \forall\,(t,\mathbf{x}) \in [0,T)\times\mathbb{R}^L\,, \tag{BSE}$$

  for pricing financial derivatives, where $u(t,\mathbf{x})$ is the value of an option depending on time t and asset prices $\mathbf{x} = (x_1,\dots,x_L)^\top$, σ denotes the volatility, and r is the (risk-free) interest rate. Here, the dimension L is the number of the underlying financial assets.

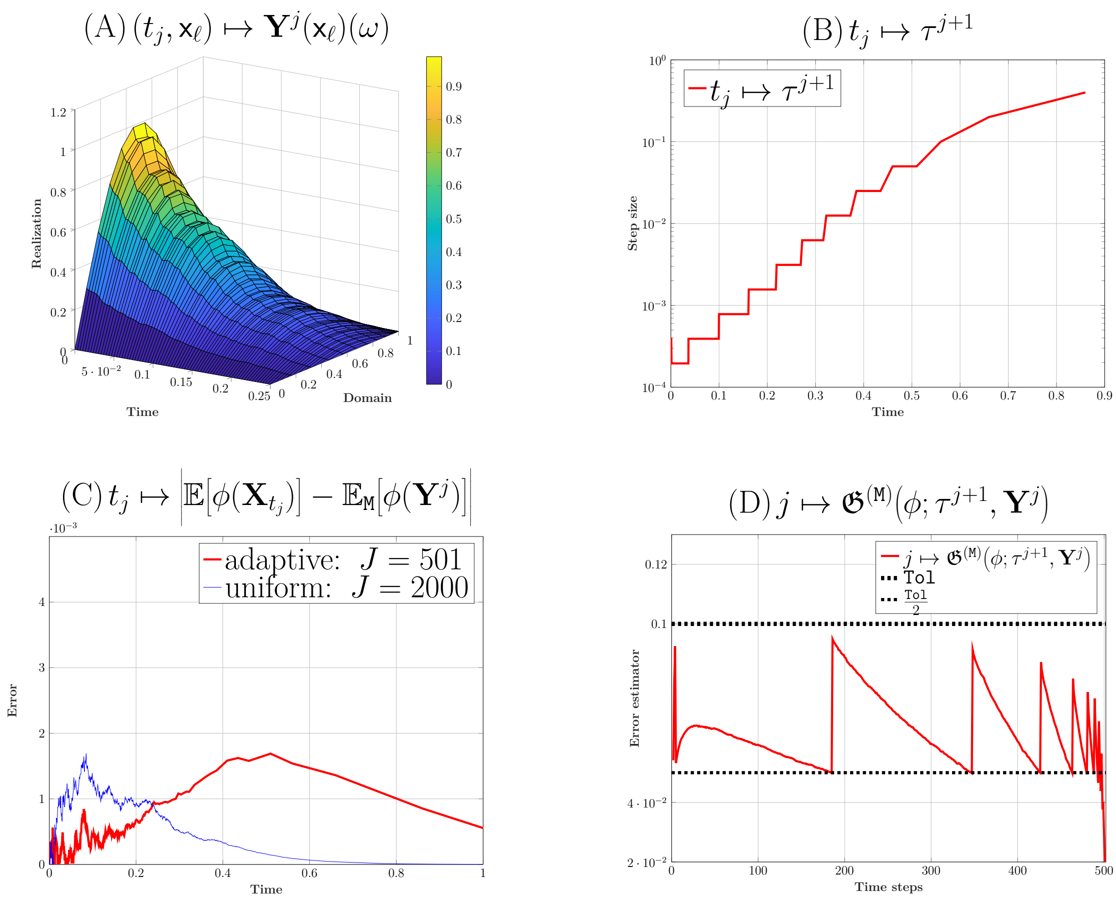


Figure 0.1.: (Example 0.1 for $\mathtt{Tol} = 0.1$, $\mathtt{M} = 10^4$, $T = 1$) (A) Contour plot of the solution for a single realization ω up to time $t = 0.25$. (B) Semi-Log-Plot of the corresponding adaptive time step size. (C) Error for uniform (—) vs. adaptive (—) time meshes via Algorithm 4.1. (D) Plot of the (empirical) *a posteriori* weak error estimator $\mathfrak{G}^{(\mathtt{M})}\big(\phi; \tau^{j+1}, \mathbf{Y}^j\big)$.

Usually, the numerical approximation of (SE) and/or (BSE) via deterministic methods is a difficult task due to the well-known '*curse of dimensionality*' (computational cost grows exponentially with the dimension). However, we are convinced that the structure of (0.5) (and hence of (0.1)) can be extended in such a way that the (efficient) approximation of (SE) and/or (BSE) via (a modified version of) Algorithm 4.1 is possible, which, in particular, avoids the *curse of dimensionality* due to the probabilistic ansatz.
We refer to Chapter 1 for a continuing discussion of the concepts presented so far.

Let now $\boldsymbol{\mathcal{D}} \subset \mathbb{R}^L$ be a bounded, smooth domain. In the second part of this thesis, which consists of Chapters 6 – 11, we derive an *a posteriori* error estimate and an adaptive time-stepping strategy based on it, for a discretization which is based on the probabilistic representation of the elliptic PDE with Dirichlet condition

$$\frac{1}{2}\mathrm{Tr}\Big(\boldsymbol{\sigma}(\mathbf{x})\boldsymbol{\sigma}^\top(\mathbf{x})D^2_{\mathbf{x}}u(\mathbf{x})\Big) + \Big\langle \mathbf{b}(\mathbf{x}), D_{\mathbf{x}}u(\mathbf{x})\Big\rangle_{\mathbb{R}^L} + c(\mathbf{x})u(\mathbf{x}) + g(\mathbf{x}) = 0 \quad \forall\, \mathbf{x} \in \boldsymbol{\mathcal{D}}, \qquad (0.6a)$$

$$u(\mathbf{x}) = \phi(\mathbf{x}) \quad \forall \mathbf{x} \in \boldsymbol{\partial\mathcal{D}}, \qquad (0.6b)$$

where $\mathbf{b} : \overline{\boldsymbol{\mathcal{D}}} \to \mathbb{R}^L$, $\boldsymbol{\sigma} : \overline{\boldsymbol{\mathcal{D}}} \to \mathbb{R}^{L\times L}$, $c : \overline{\boldsymbol{\mathcal{D}}} \to \mathbb{R}_0^-$, $g : \overline{\boldsymbol{\mathcal{D}}} \to \mathbb{R}$ and $\phi : \overline{\boldsymbol{\mathcal{D}}} \to \mathbb{R}$ are given. For proper settings of data such as stated in Section 8.1, there exists a unique classical solution $u : \overline{\boldsymbol{\mathcal{D}}} \to \mathbb{R}$ of problem (0.6), given by the probabilistic representation

$$u(\mathbf{x}) = \mathbb{E}\Big[\phi\Big(\mathbf{X}^{\mathbf{x}}_{\boldsymbol{\tau}^{\mathbf{x}}}\Big)V^{\mathbf{x}}_{\boldsymbol{\tau}^{\mathbf{x}}} + Z^{\mathbf{x}}_{\boldsymbol{\tau}^{\mathbf{x}}}\Big] \qquad \forall \mathbf{x} \in \boldsymbol{\mathcal{D}}\,, \qquad (0.7)$$

see *e.g.* [62, p. 366], where

1) $\mathbf{X}^{\mathbf{x}} \equiv \{\mathbf{X}^{\mathbf{x}}_t;\, t \geq 0\}$ denotes the $\mathbb{R}^L$-valued solution of the stochastic differential equation (SDE)

$$\mathrm{d}\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)\mathrm{d}t + \boldsymbol{\sigma}(\mathbf{X}_t)\mathrm{d}\mathbf{W}_t \quad \forall t > 0\,, \qquad \mathbf{X}_0 = \mathbf{x} \in \boldsymbol{\mathcal{D}} \subset \mathbb{R}^L\,, \qquad (0.8)$$

starting in $\mathbf{x} \in \boldsymbol{\mathcal{D}}$, where $\mathbf{W} \equiv \{\mathbf{W}_t;\, t \geq 0\}$ is an $\mathbb{R}^L$-valued Wiener process on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$, and the first exit time of $\mathbf{X}^{\mathbf{x}}$ from $\boldsymbol{\mathcal{D}}$ is

$$\boldsymbol{\tau}^{\mathbf{x}} := \inf\Big\{t > 0 \,:\, \mathbf{X}^{\mathbf{x}}_t \notin \boldsymbol{\mathcal{D}}\Big\}\,. \qquad (0.9)$$

2) $V^{\mathbf{x}} \equiv \{V^{\mathbf{x}}_t;\, t \geq 0\}$ resp. $Z^{\mathbf{x}} \equiv \{Z^{\mathbf{x}}_t;\, t \geq 0\}$ denote the $\mathbb{R}$-valued solutions of the random ordinary differential equations (ODE)

$$\mathrm{d}V_t = c(\mathbf{X}^{\mathbf{x}}_t)V_t\mathrm{d}t \quad \forall t > 0\,, \qquad V_0 = 1\,, \quad \text{and} \qquad (0.10)$$

$$\mathrm{d}Z_t = g(\mathbf{X}^{\mathbf{x}}_t)V_t\mathrm{d}t \quad \forall t > 0\,, \qquad Z_0 = 0\,. \qquad (0.11)$$

To numerically solve (0.6), deterministic schemes based on finite differences, finite volumes, or finite elements are well-known, which are complemented by rigorous *a priori* and *a posteriori* error analysis. However, these methods all suffer from the ‘curse of dimensionality’, which restricts their implementations to small values $1 \leq L \leq 4$ in practice. To simulate the boundary value problem (0.6) for $L \gg 4$, other (deterministic) mesh-based methods are available, such as sparse grids, or methods that rely on tensor-structured data and (structure-inheriting) compatible operators; see also Chapter 7 for more details.
In this work the probabilistic interpretation of (0.6) is taken to approximate $u(\mathbf{x})$, $\mathbf{x} \in \boldsymbol{\mathcal{D}}$, for high-dimensional problems, *i.e.*, $L \gg 4$, and free from (restrictive) constraints on data in (0.6): specifically, the first goal is an *a posteriori* error analysis for discretization Scheme 2 (see Section 8.3) based on [62, p. 365 *ff.*, Sec. 6.3] to bound the approximation error for $u(\mathbf{x})$, $\mathbf{x} \in \boldsymbol{\mathcal{D}}$ in terms of the computed solution $\{\mathbf{Y}^j_{\mathbf{X}}\}_{j\geq 0}$. A distinct feature of Scheme 2 is the use of a scaled random walk instead of *unbounded* Wiener increments to rigorously derive the *a posteriori* error bound (0.12) below.

To our knowledge, the first *a posteriori* (weak) error analysis for the Euler method with Wiener increments (unbounded) (9.57) to solve the *Kolmogorov* PDE on $\boldsymbol{\mathcal{D}} = \mathbb{R}^L$ goes back to [76], whose application is restricted to low dimensions L. These techniques are later extended to the related (parabolic) boundary value problem in [28], where 'stopping' is realized when corresponding iterates have come 'close' to $\boldsymbol{\partial\mathcal{D}}$ rather than onto $\boldsymbol{\partial\mathcal{D}}$; however, the *a posteriori* error analysis still suffers from same restrictions; see Chapter 7.

In this work, for $\mathbf{x} \in \boldsymbol{\mathcal{D}}$ fixed, and a given mesh $\{t_j\}_{j\geq 0} \subset [0,\infty)$ with local mesh sizes $\{\tau^{j+1}\}_{j\geq 0}$, we verify the following *a posteriori* (weak) error estimate for iterates $\{(\mathbf{Y}^j_{\mathbf{X}}, Y^j_V, Y^j_Z)\}_{j=0}^{J^*}$ from Scheme 2

$$\left|u(\mathbf{x}) - \mathbb{E}\Big[\phi(\mathbf{Y}^{J^*}_{\mathbf{X}})Y^{J^*}_V + Y^{J^*}_Z\Big]\right| \leq \boldsymbol{C}(\phi, g) \cdot \sum_{j=0}^{\infty} \tau^{j+1}\Big\{\mathfrak{G}^{(j)}_1 + \mathfrak{G}^{(j)}_2 + \mathfrak{G}^{(j)}_3\Big\}\,, \tag{0.12}$$

with $\boldsymbol{C}(\phi, g) > 0$, (computable) *a posteriori* error estimators $\{\mathfrak{G}^{(\cdot)}_\ell\}_{\ell=1}^3$ in terms of the discrete solution, and $J^* \equiv J^*(\mathbf{x}) \in \mathbb{N}_0$ the stopping index. Main achievements in our work are then

(*i*) its construction, which is based on *Taylor's formula* (rather than *Itô's formula*) to properly address the use of scaled random walk; see Theorem 9.1. At time t_j, the functional $\mathfrak{G}^{(j)}_1$ is assembled from those states that realize in the interior of $\boldsymbol{\mathcal{D}}$, while the remaining two are assembled from those in a $\mathcal{O}(\sqrt{\tau^{j+1}})-$neighborhood of the boundary, addressing possible bouncing back/stopping.

(*ii*) Stability results in Section 8.3 concerning 'discrete stopping' ensure that the sum in (0.12) is in fact finite, and, besides the 'stopping'-mechanism in Scheme 2, they are the key to verify optimal first order of convergence for (0.12) on families of (time-)meshes with maximum mesh size $\tau^{max} > 0$, when $\tau^{max} \searrow 0$; see Theorem 9.6.

(*iii*) Estimate (0.12) will be used in Chapter 10 to construct an adaptive time stepping algorithm (see Algorithm 10.1) for which we prove local, as well as global termination, and optimal convergence behaviour in terms of the tolerance ($\texttt{Tol} > 0$); see Chapter 10.

We refer to Chapter 6, where we directly take up this discussion. Moreover, we present similar concepts for the approximation of parabolic boundary value problems.

Part I.

An adaptive time-stepping method based on a posteriori weak error analysis for large SDE systems

1. Introduction

Different adaptive methods to solve SDE (0.1) may be found in the literature, addressing diverse numerical goals: in [55], an adaptive time-meshing concept is combined with the Euler-Maruyama method to foster discrete stability of the explicit time-stepping scheme in cases where the local Lipschitz drift only satisfies a 'one-sided Lipschitz condition'. Automatic mesh refinement (resp. coarsening) for each realization $\omega \in \Omega$ is applied if rapid (resp. slow) changes in the drift at two subsequent states are observed, where a maximum mesh size $\Delta_{\texttt{max}}$ bounds local (random) mesh sizes $\{\tau^j(\omega)\}_{j\geq 0} \subset [0, \Delta_{\texttt{max}}]$ to conclude asymptotic strong convergence. Another work which uses adaptive random meshes as well to strengthen the stability of the underlying explicit discretization of the above mentioned class of SDE's (0.1) is [32]: here, a 'discrete one-sided Lipschitz condition' is used to generate random mesh sizes $\{\tau_j(\omega)\}_{j\geq 0}$, which are then further constrained to lie in $[\Delta_{\min}, \Delta_{\max}]$. The main result in [32] is the derivation of an optimal convergence rate $\mathcal{O}(\Delta_{\texttt{max}}^{1/2})$ on variable random meshes of size between $\Delta_{\min}$ and $\Delta_{\texttt{max}}$. Close to the goals and applied tools in this work are [50, 51], where, again, to set parameters $\Delta_{\min}$ and $\Delta_{\max}$ requires some a priori knowledge, and the complexity in the worst case of the method may depend on $\Delta_{\min}^{-1}$, and the dimension L of the problem due to the explicit character of the discretization that affects relevant discrete solution bounds; see also [43] in this respect.

A different line of research derives *a posteriori error estimates* (such as (0.4)) to judge the quality of the current approximation, and uses it then as a 'steering tool' to initiate an automatic remeshing strategy. While this conceptual idea to design adaptive methods has been well-known in the context of (certain) ODE's and PDE's before, it has first been introduced in [76, 64] for SDE's in the contents of *weak* approximation of SDE solutions (here again via Euler-Maruyama discretization). In these works, an (asymptotic weak) *a posteriori error expansion*

$$\Big|\mathbb{E}\big[\phi(\mathbf{X}_T)\big] - \mathbb{E}\big[\phi(\mathbf{Y}_T)\big]\Big| = \mathbb{E}\left[\sum_{j=0}^{J-1} \rho_{j+1} \cdot \left(\tau^{j+1}\right)^2\right] + \text{'higher order terms'}\ , \qquad (1.1)$$

with computable $\{\rho_j\}_{j=1}^J$ has first been obtained. Its derivation in [76, 64] rests on the weak error expansion of Talay and Tubaro [77] via *Kolmogorov's backward equation* (0.5), and numerically approximates derivatives of the solution u of the PDE (0.5), whose simulation is limited to *small* dimensions L. Then, (random) time meshes are generated automatically based on the *computable part* of the right-hand side of (1.1) — with *no* minimum or maximum mesh sizes to be set, but *only* the parameter `Tol` (also serving as convergence parameter) to bound the leading error term on the right-hand side of (1.1). The iterative generation of an adapted time mesh requires the repeated computation of (approximations of) the *global* problem (0.5) — opposed to determining *local* time steps τ^{j+1} based on 'so far' computed solutions $\{\mathbf{Y}^\ell\}_{\ell=0}^{j+1}$ only. From an analytical viewpoint, the results in [76, 64] crucially rest on the assumed *boundedness* of the involved drift and diffusion functions to circumvent the deficiency of 'discrete stability' of the governing (explicit) Euler-Maruyama scheme; the advantage, however, is a theoretical backup for this weak adaptive algorithm in terms of termination at optimal rate, and the asymptotic weak *a posteriori* error estimate (1.1).

Conceptually, the derivation of the adaptive Algorithm 4.1 below is close to [76, 64] and uses a related weak error representation (see (1.3) below) for the semi-implicit Euler scheme (0.3) with the help of the solution u of (0.5) — but differs in some relevant aspects: the first is the use of the semi-implicit Euler scheme (0.3), which allows for $L-$independent (higher moment) stability bounds for its solution in case data satisfy **(A1)** – **(A3)** in Section 2.1; see Lemma 2.6. These stability bounds for (0.3) in Lemma 2.6 are the relevant property to show optimal order of weak convergence of the *a posteriori* weak error estimator proposed in Theorem 3.1 on given meshes; *cf.* Theorem 3.5.
A second difference to [76, 64] is that we bound derivatives of u that appear in (0.5) in the weak error representation (1.4) by *a priori* bounds (see (1.5) below) in terms of derivatives of ϕ, which removes the necessity to numerically approximate derivatives of the solution of (0.5) — and thus enables the applicability of Algorithm 4.1 to large SDE systems, as they *e.g.* come from SPDE (0.2) via spatial discretization (in Example 0.1).

To further detail relevant steps in our program, we start with the continuified process $\boldsymbol{\mathcal{Y}} \equiv \{\boldsymbol{\mathcal{Y}}_t;\, t \in [0,T]\}$ of the sequence of random variables $\{\mathbf{Y}^j\}_{j\geq 0}$ which solves (0.3). We easily observe in Chapter 3 that

$$\begin{aligned}\boldsymbol{\mathcal{Y}}_t := \mathbf{Y}^j &+ \Big(\Big(\mathbb{I}+\tau^{j+1}\mathscr{A}\Big)^{-1}\mathbf{f}(\mathbf{Y}^j) - \mathscr{A}\Big(\mathbb{I}+\tau^{j+1}\mathscr{A}\Big)^{-1}\mathbf{Y}^j\Big)(t-t_j) \\ &+ \sum_{k=1}^{K}\Big(\mathbb{I}+\tau^{j+1}\mathscr{A}\Big)^{-1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big(\beta_k(t)-\beta_k(t_j)\big) \qquad \forall\, t\in[t_j,t_{j+1}] \qquad (1.2)\end{aligned}$$

interpolates $\{\mathbf{Y}^j\}_{j\geq 0}$ at $\{t_j\}_{j\geq 0}$, and is $\{\mathcal{F}_t\}_{t\geq 0}$-adapted. Now assume $0 = t_0 < t_1 < ... < t_J = T$ and fix $n = 0, ..., J-1$; considering (0.5) on $[0, t_{n+1}] \times \mathbb{R}^L$, a standard argument then leads to (see Lemma 3.2)

$$\begin{aligned}\Big|\mathbb{E}\Big[\phi(\mathbf{X}_{t_{n+1}})\Big]-\mathbb{E}\Big[\phi(\mathbf{Y}^{n+1})\Big]\Big| &= \Big|\mathbb{E}\Big[u(0,\mathbf{y})-u(t_{n+1},\mathbf{Y}^{n+1})\Big]\Big| \\ &\leq \sum_{j=0}^{n}\Big|\mathbb{E}\Big[u(t_{j+1},\mathbf{Y}^{j+1})-u(t_j,\mathbf{Y}^j)\Big]\Big|. \qquad (1.3)\end{aligned}$$

We may now use *Itô's formula* with u from (0.5) to transform $\boldsymbol{\mathcal{Y}}$ on each time interval $[t_j, t_{j+1}]$ to represent each increment $u(t_{j+1}, \mathbf{Y}^{j+1}) - u(t_j, \mathbf{Y}^j)$ in the last sum, and employ (0.5) to deduce

$$\begin{aligned}&\mathbb{E}\big[u(t_{j+1},\mathbf{Y}^{j+1})-u(t_j,\mathbf{Y}^j)\big] \\ &= \int_{t_j}^{t_{j+1}}\mathbb{E}\Big[\Big\langle\underbrace{\big(\mathbb{I}+\tau^{j+1}\mathscr{A}\big)^{-1}\mathbf{f}(\mathbf{Y}^j)-\mathscr{A}\big(\mathbb{I}+\tau^{j+1}\mathscr{A}\big)^{-1}\mathbf{Y}^j-\mathbf{f}(\boldsymbol{\mathcal{Y}}_s)+\mathscr{A}\boldsymbol{\mathcal{Y}}_s}_{\text{'error indicator (drift)'}},\underbrace{D_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)}_{\text{'weight'}}\Big\rangle_{\mathbb{R}^L}\Big]\,\mathrm{d}s\end{aligned}$$

$$+\frac{1}{2}\int_{t_j}^{t_{j+1}} \mathbb{E}\Big[\mathrm{Tr}\Big(\underbrace{\Big\{\big(\mathbb{I}+\tau^{j+1}\mathscr{A}\big)^{-1}\boldsymbol{\sigma}(\mathbf{Y}^j)\big[\big(\mathbb{I}+\tau^{j+1}\mathscr{A}\big)^{-1}\boldsymbol{\sigma}(\mathbf{Y}^j)\big]^\top-\boldsymbol{\sigma}(\boldsymbol{\mathcal{Y}}_s)\boldsymbol{\sigma}^\top(\boldsymbol{\mathcal{Y}}_s)\Big\}}_{\text{'error indicator (diffusion)'}}$$

$$\cdot \underbrace{D^2_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)}_{\text{'weight'}}\Big)\Big]\,\mathrm{d}s\,; \tag{1.4}$$

see Lemma 3.2 for the justification of this identity. Conceptionally, the right-hand side of (1.4) uses the continuified process $\boldsymbol{\mathcal{Y}}$ built from the iterates $\mathbf{Y}^j$ and $\mathbf{Y}^{j+1}$ — and first and second derivatives of the solution u from (0.5) to transform $\{\boldsymbol{\mathcal{Y}}_t;\, t\in[t_j,t_{j+1}]\}$. An interpretation of (the right-hand side of) (1.4) in a corresponding setting in [76, 64] is its view as products of (local) error indicators for the drift and diffusion, and weights $D_{\mathbf{x}}u$, $D^2_{\mathbf{x}}u$, which 'encode' the chosen test function ϕ.

In the next step, we use the first to third *variation equations* for (0.1), see (2.2a) – (2.4) in Section 2.3, to deduce bounds

$$\sup_{(t,\mathbf{x})\in[0,T]\times\mathbb{R}^L}\|D^\ell_{\mathbf{x}}u(t,\mathbf{x})\|_{\mathcal{L}^\ell}\le\sum_{i=1}^{\ell}C_{\ell,i}\sup_{\mathbf{x}\in\mathbb{R}^L}\|D^i\phi(\mathbf{x})\|_{\mathcal{L}^i}\qquad(\ell\in\{1,2,3\})\,, \tag{1.5}$$

for derivatives of the solution u of (0.5), where constants $C_{\ell,i}>0$ do not depend on the dimension L. Note that only *derivatives of* u are involved in (1.4), so their estimation with the help of (1.5) suggests to choose test functions $\phi:\mathbb{R}^L\to\mathbb{R}$ in (0.4) whose *derivatives* are uniformly bounded on $\mathbb{R}^L$ — such as norms; see also Example 0.1.

While the derivation of estimate (1.5) is known for a general class of SDE's, see *e.g.* [19, Sec. 1.3], we calculate the constants $\{C_{\ell,i};\,1\le i\le\ell,\ 1\le\ell\le3\}$ under the assumptions **(A1)** – **(A3)**, which are needed in the *a posteriori* error estimate (0.4); see Lemma 2.5. A further tool to derive (0.4) then is the use of common Malliavin calculus techniques such as the *Clark-Ocone formula*, to avoid that the error estimator uses the interpolated process $\boldsymbol{\mathcal{Y}}$ in (1.2) rather than computable iterates $\mathbf{Y}^j$ from (0.3). Here, we benefitted from similar ideas and concepts, which were used in [24] in the context of *a priori* weak error analysis of SPDE's of form (0.2) with $\boldsymbol{\beta}\equiv\mathbf{0}$; see Remark 3.2 for further details. For a fixed $\phi\in\mathcal{C}^3(\mathbb{R}^L)$, our first main result in this work then is the weak *a posteriori* error estimate (0.4) (see also Theorem 3.1), giving quantitative error bounds for iterates $\{\mathbf{Y}^j\}_{j\ge0}$ solving (0.3) on a given a mesh $\{t_j\}_{j\ge0}$ covering $[0,T]$ with the help of computable (local) weak error estimators $\Big\{\mathfrak{G}\Big(\phi;\tau^{j+1},\mathbf{Y}^j\Big)\Big\}_{j\ge0}$.

In Chapter 4, we use the *a posteriori* error estimation (0.4) for iterates $\{\mathbf{Y}^j\}_{j\ge0}$ of (0.3) to automatically steer the computation of local mesh sizes τ^{j+1} via the adaptive Algorithm 4.1, yielding tuples $\Big\{(\tau^{j+1},\mathbf{Y}^{j+1})\Big\}_{j\ge0}$. Given $j\ge0$, the guiding criterion for *admissibility* of a new tuple $(\tau^{j+1},\mathbf{Y}^{j+1})$ is that the evaluation with the local error estimator yields $\mathfrak{G}\Big(\phi;\tau^{j+1},\mathbf{Y}^j\Big)\le\frac{\mathtt{Tol}}{T}$, where the tolerance $\mathtt{Tol}>0$ is provided by the user. We generate such an admissible tuple by successively halving the previous time step, and thus generating a sequence $\{\tau^{j+1,\ell}\}_{\ell\ge0}\subset\mathbb{R}^+$ with $\tau^{j+1,\ell}=\frac{\tau^{j+1,0}}{2^\ell}$ and $\tau^{j+1,0}\equiv\tau^j$, until admissibility of a tuple for some $\tau^{j+1,\ell^*_{j+1}}$ is attained; see Figure 1.1. This sequence of steps precedes a single potential step of coarsening; see Algorithm 4.1 for further details. As a result, we obtain an adaptive method where only $\texttt{Tol} > 0$ needs be set, and where admissible tuple $\big\{(\tau^{j+1}, \mathbf{Y}^{j+1})\big\}_{j\geq 0}$ satisfy (0.4), where the right-hand side is now bounded by $\texttt{Tol} > 0$. The second main result in this work then is Theorem 4.2, which ensures computation of each new time step τ^{j+1} in Algorithm 4.1 after no more than $\ell^*_{j+1} = \mathcal{O}\big(\log(\texttt{Tol}^{-1})\big)$ many iterations (indexed by j; local determination), and at most $J = \mathcal{O}(\texttt{Tol}^{-1})$ many steps to reach T (global termination); its proof again rests on the stability bounds given in Lemma 2.6 and yields the existence of a lower bound of the step sizes generated via Algorithm 4.1; see also Figure 1.1. We remark that this local construction of the new mesh size τ^{j+1} with the help of only $\mathbf{Y}^j$ differs from the strategy in [76, 64], where admissible meshes are obtained by iterative computation of *global* problems ('approximate *Kolmogorov equation*'), and where again the assumed *boundedness* of drift and diffusion is crucial to conclude optimality of attained meshes.

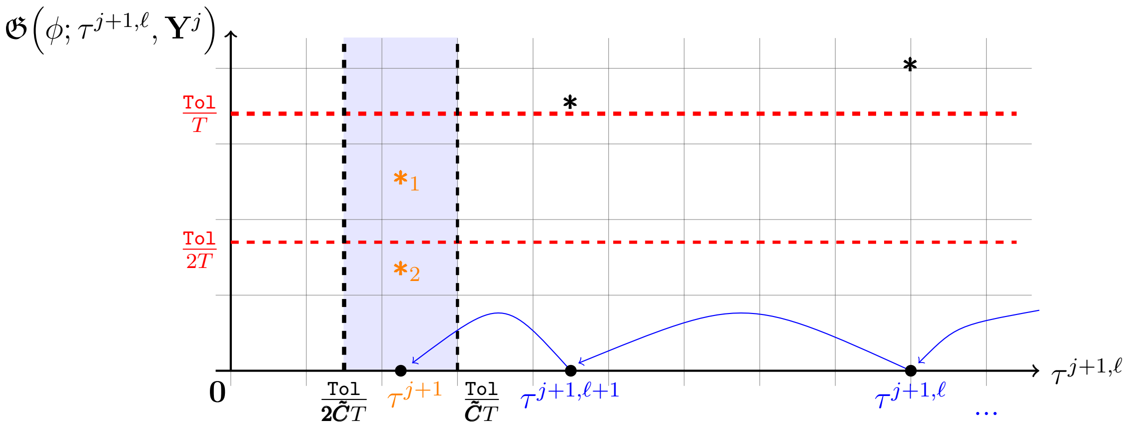


Figure 1.1.: Illustration of the local (for fixed j) and global termination argument given in Theorem 4.2, yielding a maximum of $\ell^*_{j+1} = \mathcal{O}\big(\log(\texttt{Tol}^{-1})\big)$ many refinement steps (↶); see also (4.2), within the loop of $\{\tau^{j+1,\ell}\}_{\ell\geq 0}$, to accept the new step size $\tau^{j+1} = \tau^{j+1,\ell^*_{j+1}}$, and the existence of a lower bound of step sizes generated via Algorithm 4.1, such that either **(2)** ($*_1$) or **(3)** ($*_2$) is met. Note that due to the choice of the initial mesh size τ^1 for the generation of $\mathbf{Y}^1$ and the setup of Algorithm 4.1, it is not possible that $\tau^{j+1,\ell} < \frac{\texttt{Tol}}{2\tilde{C}T}$, $\ell \geq 0$.

Chapter 5 then reports on computational studies for different SPDE's (0.2) after finite element discretization with the help of the adaptive Algorithm 4.1: we specify the corresponding *a posteriori* error estimator, and pinpoint those computable expressions $\{\mathbf{E}_{\ell}\}_{\ell\geq 1}$ involved in the error estimator in Example 0.1, which are mainly responsible for local mesh adjustments. For the different examples, including one which is convection dominated, the results evidence efficiency in comparison with uniform meshing, and accuracy of the weak adaptive Algorithm 4.1.

This part is organized as follows: Chapter 2 collects the assumptions needed for the data $\mathscr{A}, \mathbf{f}, \{\boldsymbol{\sigma}_k\}_k$ of (0.1) and recalls relevant tools from Malliavin calculus; moreover, *variation* *equations* for (0.1) are recalled to verify the bounds (1.5) and stability bounds for iterates $\{\mathbf{Y}^j\}_{j\geq 0}$ from (0.3) are presented. The *a posteriori* error analysis for (0.3) is given in Chapter 3. The related weak adaptive method is proposed and analyzed in Chapter 4, and corresponding computational studies are reported in Chapter 5.

2. Assumptions and Tools

Section 2.1 lists basic requirements on data $\mathscr{A}, \mathbf{f}, \boldsymbol{\sigma} \equiv \Big[\boldsymbol{\sigma}_1, ..., \boldsymbol{\sigma}_K\Big], \mathbf{y}$ in (0.1) throughout this work. Section 2.2 shortly recalls needed tools from Malliavin calculus. In Section 2.3, we derive explicit bounds for $\{D^\ell_{\mathbf{x}} u\}_{\ell=1}^3$ from *Kolmogorov's backward equation* (0.5) under Assumptions **(A1)** − **(A2)**. Stability bounds for $\{\mathbf{Y}^j\}_{j\geq 0}$ from (0.3) are given in Section 2.4 provided **(A1)** − **(A3)** are valid.

2.1. Assumptions

Throughout this work, $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ is a given filtered probability space with natural filtration of the Wiener processes in (0.1). Below, we use positive constants $C_{D^\ell \mathbf{f}}$, $C_{\mathbf{f}}^{(\ell-1)}$, $C_{D^\ell \boldsymbol{\sigma}}$, $C_{\boldsymbol{\sigma}}^{(\ell-1)}$, $C_{\mathbf{y}}^{(\ell-1)}$ $(1 \leq \ell \leq 3)$, and $\lambda_{\mathscr{A}}$ to specify dependence on data $\mathscr{A}, \mathbf{f}, \boldsymbol{\sigma} \equiv \Big[\boldsymbol{\sigma}_1, ..., \boldsymbol{\sigma}_K\Big]$ in (0.1); none of these constants depend on L. For a sufficiently smooth $\mathbf{g} \in \mathcal{C}(\mathbb{R}^L; \mathbb{R}^n)$, corresponding (matrix) operator norms are given as follows ($n, L \in \mathbb{N}$, $\mathbf{x} \in \mathbb{R}^L$):

$$\|D^\ell \mathbf{g}(\mathbf{x})\|_{\mathcal{L}\Big(\underbrace{\mathbb{R}^L \times ... \times \mathbb{R}^L}_{\ell-\text{times}};\mathbb{R}^n\Big)} := \sup_{\|\mathbf{v}_i\|_{\mathbb{R}^L}=1} \|D^\ell \mathbf{g}(\mathbf{x})(\mathbf{v}_1, ..., \mathbf{v}_\ell)\|_{\mathbb{R}^n} \qquad (\ell \in \mathbb{N}),$$

where $\|\cdot\|_{\mathbb{R}^n}$ denotes the (Euclidean) vector norm of a $\mathbb{R}^n$-valued vector. If $n = L$, we write $\mathcal{L}^\ell \equiv \mathcal{L}\Big(\mathbb{R}^L \times ... \times \mathbb{R}^L; \mathbb{R}^L\Big)$. If $n = 1$, $D \equiv D_{\mathbf{x}}$ denotes the gradient and $D^2 \equiv D^2_{\mathbf{x}}$ the Hessian matrix of $\mathbf{g}$, and we also write $\mathcal{L}^\ell \equiv \mathcal{L}\Big(\mathbb{R}^L \times ... \times \mathbb{R}^L; \mathbb{R}\Big)$. Moreover, $\|D_{\mathbf{x}} \mathbf{g}(\mathbf{x})\|_{\mathcal{L}^1} = \|D_{\mathbf{x}} \mathbf{g}(\mathbf{x})\|_{\mathbb{R}^L}$, $\|D^2_{\mathbf{x}} \mathbf{g}(\mathbf{x})\|_{\mathcal{L}^2} = \|D^2_{\mathbf{x}} \mathbf{g}(\mathbf{x})\|_{\mathbb{R}^{L\times L}}$, where $\|\cdot\|_{\mathbb{R}^{L\times L}}$ denotes the spectral (matrix) norm.

(A1) **(a)** The matrix $\mathscr{A} \in \mathbb{R}^{L\times L}$ is invertible and positive definite, *i.e.*, there exists a constant $\lambda_{\mathscr{A}} > 0$, *s.t.*

$$\langle \mathscr{A}\mathbf{x}, \mathbf{x}\rangle_{\mathbb{R}^L} \geq \lambda_{\mathscr{A}} \|\mathbf{x}\|^2_{\mathbb{R}^L} \qquad \forall\, \mathbf{x} \in \mathbb{R}^L.$$

(b) The map $\mathbf{f} \in \mathcal{C}^3(\mathbb{R}^L; \mathbb{R}^L)$, and there exist constants $\{C_{D^\ell \mathbf{f}}\}_{\ell=1}^3$, *s.t.*

$$\sup_{\mathbf{x}\in\mathbb{R}^L} \|D^\ell \mathbf{f}(\mathbf{x})\|_{\mathcal{L}^\ell} \leq C_{D^\ell \mathbf{f}} \qquad (1 \leq \ell \leq 3);$$

moreover, there exist constants $\{C_{\mathbf{f}}^{(\ell)}\}_{\ell=0}^2$, *s.t.*

$$\|\mathscr{A}^\ell \mathbf{f}(\mathbf{x})\|_{\mathbb{R}^L} \leq C_{\mathbf{f}}^{(\ell)}\Big(1 + \|\mathscr{A}^\ell \mathbf{x}\|_{\mathbb{R}^L}\Big) \quad \forall\, \mathbf{x} \in \mathbb{R}^L \qquad (0 \leq \ell \leq 2).$$

(A2) The maps $\boldsymbol{\sigma}_k \in \mathcal{C}^3(\mathbb{R}^L;\mathbb{R}^L)$ for every $k = 1, ..., K$, and there exist constants $\{C_{D^\ell \boldsymbol{\sigma}}\}_{\ell=1}^3$, *s.t.*

$$\sum_{k=1}^K \sup_{\mathbf{x}\in\mathbb{R}^L} \|D^\ell \boldsymbol{\sigma}_k(\mathbf{x})\|_{\mathcal{L}^\ell} \leq C_{D^\ell \boldsymbol{\sigma}} \qquad (1 \leq \ell \leq 3)\,;$$

moreover, there exist constants $\{C_{\boldsymbol{\sigma}}^{(\ell)}\}_{\ell=0}^2$, *s.t.* for every $k = 1, ..., K$

$$\|\mathscr{A}^\ell \boldsymbol{\sigma}_k(\mathbf{x})\|_{\mathbb{R}^L} \leq C_{\boldsymbol{\sigma}}^{(\ell)}\Big(1 + \|\mathscr{A}^\ell \mathbf{x}\|_{\mathbb{R}^L}\Big) \quad \forall\, \mathbf{x} \in \mathbb{R}^L \qquad (0 \leq \ell \leq 2)\,.$$

(A3) For $0 \leq \ell \leq 2$, there exists $C_{\mathbf{y}}^{(\ell)}$, *s.t.* the initial datum $\mathbf{y} \in \mathbb{R}^L$ in (0.1) satisfies

$$\|\mathscr{A}^\ell \mathbf{y}\|_{\mathbb{R}^L} \leq C_{\mathbf{y}}^{(\ell)}\,.$$

Throughout this work, we admit test functions $\phi \in \mathcal{C}^3(\mathbb{R}^L)$ with globally bounded first, second and third derivatives.

2.2. Malliavin calculus

We briefly recall the Malliavin derivative, recall the chain rule for Malliavin derivatives and state the *Clark-Ocone formula.* For further details, we refer to [68].— We denote by $\mathcal{C}_{\mathbf{p}}^\infty(\mathbb{R}^L)$ the space of all smooth functions $g : \mathbb{R}^L \to \mathbb{R}$, such that g and all of its partial derivatives have polynomial growth. Let $\mathfrak{P}$ the set of $\mathbb{R}$-valued random variables of the form

$$\mathrm{F} = g\Big(W(h_1), ..., W(h_L)\Big)$$

for some $g \in \mathcal{C}_{\mathbf{p}}^\infty(\mathbb{R}^L)$ and $h_1, ..., h_L \in L^2(0,T)$. Here, $W : L^2(0,T) \to L^2_{\mathcal{F}_T}(\Omega)$ is defined by

$$W(h) = \int_0^T h(t)\,\mathrm{d}\beta(t)\,.$$

We further define for any $\mathrm{F} \in \mathfrak{P}$ its $\mathbb{R}$-valued Malliavin derivative process $\mathrm{DF} := \{\mathrm{D}_t\mathrm{F};\ 0 \leq t \leq T\}$ via

$$\mathrm{D}_t\mathrm{F} = \sum_{i=1}^L \partial_{x_i} g\Big(W(h_1), ..., W(h_L)\Big) h_i(t)\,. \tag{2.1}$$

For any $p \geq 1$, let $\mathbb{D}^{1,p}$ denote the closure of the class of smooth random variables with respect to the norm

$$\|\mathrm{F}\|_{\mathbb{D}^{1,p}} = \Big(\mathbb{E}[|\mathrm{F}|^p] + \mathbb{E}\Big[\|\mathrm{DF}\|^p_{L^2(0,T)}\Big]\Big)^{\frac{1}{p}}.$$

Next, we recall the chain rule for Malliavin derivatives; see [68, p. 28, Prop. 1.2.3].
Let $\varphi : \mathbb{R}^L \to \mathbb{R}$ be a continuously differentiable function with bounded partial derivatives of order 1, and $p \geq 1$ be fixed. Let further $\mathbf{F} = \Big(\mathrm{F}^1, ..., \mathrm{F}^L\Big)^\top$ be a random vector whose components belong to the space $\mathbb{D}^{1,p}$. Then $\varphi(\mathbf{F}) \in \mathbb{D}^{1,p}$, and

$$\mathrm{D}\Big(\varphi(\mathbf{F})\Big) = \sum_{i=1}^L \partial_{x_i}\varphi(\mathbf{F})\mathrm{DF}^i\,.$$

Finally, we recall the *Clark-Ocone representation formula*; see [68, p. 46, Prop. 1.3.14].

Lemma 2.1. Let $\mathrm{F} \in \mathbb{D}^{1,2}$, and β be a one-dimensional Wiener process. Then

$$\mathrm{F} = \mathbb{E}[\mathrm{F}] + \int_0^T \mathbb{E}\Big[\mathrm{D}_t\mathrm{F}|\mathcal{F}_t\Big]\,\mathrm{d}\beta(t)\,.$$

2.3. Variation equations for (0.1) and a priori bounds for $\{D^\ell_{\mathbf{x}} u\}_{\ell=1}^3$ of (0.5)

Kolmogorov's backward equation (0.5) has a unique solution $[0,T]\times\mathbb{R}^L \ni (t,\mathbf{x}) \mapsto u(t,\mathbf{x}) = \mathbb{E}[\phi(\mathbf{X}_T^{t,\mathbf{x}})]$, whenever assumptions **(A1)** − **(A2)** hold; see *e.g.* [49, p. 366*ff.*]. The derivation in Chapter 3 requires explicit bounds for derivatives u, which are uniform in L, in particular. To this end, we use the *variation equations* corresponding to (0.1) and derive these results here.

For $\mathbf{y} \in \mathbb{R}^L$ fixed, we denote by $\mathbf{X} \equiv \mathbf{X}^{0,\mathbf{y}}$ the solution of (0.1). Let $\mathbf{h} \in \mathbb{R}^L$. Following [19, p. 37*ff.*], we recall the *first variation equation* corresponding to (0.1),

$$\mathrm{d}\boldsymbol{\eta}_t^{\mathbf{h}} = \Big(-\mathscr{A} + D\mathbf{f}(\mathbf{X}_t)\Big)\cdot\boldsymbol{\eta}_t^{\mathbf{h}}\mathrm{d}t + \sum_{k=1}^K D\boldsymbol{\sigma}_k(\mathbf{X}_t)\cdot\boldsymbol{\eta}_t^{\mathbf{h}}\mathrm{d}\beta_k(t) \qquad \forall\, t\in[0,T]\,, \tag{2.2a}$$

$$\boldsymbol{\eta}_0^{\mathbf{h}} = \mathbf{h}\,. \tag{2.2b}$$

Since assumptions **(A1)** − **(A2)** are valid, there exists a unique solution $\boldsymbol{\eta}^{\mathbf{h}} \equiv \{\boldsymbol{\eta}_t^{\mathbf{h}};\, t\in[0,T]\}$; it is equal to $D_{\mathbf{y}}\mathbf{X}^{0,\mathbf{y}}\cdot\mathbf{h}$, the derivative *w.r.t.* the initial datum $\mathbf{y}\in\mathbb{R}^L$ of the map $\mathbf{y}\mapsto\mathbf{X}^{0,\mathbf{y}}$, along the direction $\mathbf{h}\in\mathbb{R}^L$.

Lemma 2.2. Assume **(A1)** − **(A2)** in (2.2a). Then, for every $\mathbf{h}\in\mathbb{R}^L$ and $p\geq 1$,

$$\sup_{t\in[0,T]} \mathbb{E}\Big[\|\boldsymbol{\eta}_t^{\mathbf{h}}\|^p_{\mathbb{R}^L}\Big] \leq V_p^{(1)}\cdot\|\mathbf{h}\|^p_{\mathbb{R}^L}\,,$$

where $V_p^{(1)} := e^{pT\max\{-\lambda_{\mathscr{A}}+C_{D\mathbf{f}}+\frac{p-1}{2}C^2_{D\boldsymbol{\sigma}},0\}}$ $(p>1)$, and $V_1^{(1)} := e^{T\max\{-\lambda_{\mathscr{A}}+C_{D\mathbf{f}}+\frac{1}{2}C^2_{D\boldsymbol{\sigma}},0\}}$.

Proof. We prove the statement for $p\in\mathbb{N}$. By *Jensen's inequality*, the result also holds true for noninteger $p\geq 1$.
a) $1<p\in\mathbb{N}$: Let $t\in[0,T]$, $\mathbf{h}\in\mathbb{R}^L$ and $1<p\in\mathbb{N}$. By *Itô's formula*,

$$\begin{aligned}\frac{1}{p}\mathbb{E}\Big[\|\boldsymbol{\eta}_t^{\mathbf{h}}\|^p_{\mathbb{R}^L}\Big] \leq \frac{\|\mathbf{h}\|^p_{\mathbb{R}^L}}{p} + \mathbb{E}\Bigg[\int_0^t \Big\langle\Big(-\mathscr{A}+D\mathbf{f}(\mathbf{X}_s)\Big)\cdot\boldsymbol{\eta}_s^{\mathbf{h}},\boldsymbol{\eta}_s^{\mathbf{h}}\Big\rangle_{\mathbb{R}^L}\cdot\|\boldsymbol{\eta}_s^{\mathbf{h}}\|^{p-2}_{\mathbb{R}^L} \\ + \frac{1}{2}\sum_{k=1}^K \mathrm{Tr}\Big(D\boldsymbol{\sigma}_k(\mathbf{X}_s)\boldsymbol{\eta}_s^{\mathbf{h}}[D\boldsymbol{\sigma}_k(\mathbf{X}_s)\boldsymbol{\eta}_s^{\mathbf{h}}]^\top\Big)\cdot(p-1)\|\boldsymbol{\eta}_s^{\mathbf{h}}\|^{p-2}_{\mathbb{R}^L}\,\mathrm{d}s\Bigg].\end{aligned}$$

Using assumptions **(A1)** − **(A2)** leads to

$$\frac{1}{p}\mathbb{E}\Big[\|\boldsymbol{\eta}_t^{\mathbf{h}}\|^p_{\mathbb{R}^L}\Big]$$

$$\leq \frac{\|\mathbf{h}\|_{\mathbb{R}^L}^p}{p} + \mathbb{E}\Big[\int_0^t \Big\{\Big\langle\Big(-\mathscr{A} + D\mathbf{f}(\mathbf{X}_s)\Big)\cdot \boldsymbol{\eta}_s^{\mathbf{h}}, \boldsymbol{\eta}_s^{\mathbf{h}}\Big\rangle_{\mathbb{R}^L} + \tfrac{p-1}{2}\sum_{k=1}^K \|D\boldsymbol{\sigma}_k(\mathbf{X}_s)\boldsymbol{\eta}_s^{\mathbf{h}}\|_{\mathbb{R}^L}^2\Big\}\|\boldsymbol{\eta}_s^{\mathbf{h}}\|_{\mathbb{R}^L}^{p-2}\,\mathrm{d}s\Big]$$

$$\leq \frac{\|\mathbf{h}\|_{\mathbb{R}^L}^p}{p} + \max\Big\{-\lambda_{\mathscr{A}} + C_{D\mathbf{f}} + \tfrac{p-1}{2}C_{D\boldsymbol{\sigma}}^2, 0\Big\}\int_0^t \mathbb{E}\Big[\|\boldsymbol{\eta}_s^{\mathbf{h}}\|_{\mathbb{R}^L}^p\Big]\,\mathrm{d}s.$$

Applying *Gronwall's inequality* leads to the first result.
b) $p = 1$: This follows by using *Jensen's inequality*.

□

Next, following [19, p. 39*ff.*], we consider the *second variation equation* corresponding to (0.1), that is, for $\mathbf{h}, \mathbf{w} \in \mathbb{R}^L$,

$$\mathrm{d}\boldsymbol{\zeta}_t^{\mathbf{h},\mathbf{w}} = \Big(\Big(-\mathscr{A} + D\mathbf{f}(\mathbf{X}_t)\Big)\cdot \boldsymbol{\zeta}_t^{\mathbf{h},\mathbf{w}} + D^2\mathbf{f}(\mathbf{X}_t)\cdot(\boldsymbol{\eta}_t^{\mathbf{h}}, \boldsymbol{\eta}_t^{\mathbf{w}})\Big)\mathrm{d}t$$
$$+ \sum_{k=1}^K \Big(D\boldsymbol{\sigma}_k(\mathbf{X}_t)\cdot \boldsymbol{\zeta}_t^{\mathbf{h},\mathbf{w}} + D^2\boldsymbol{\sigma}_k(\mathbf{X}_t)\cdot(\boldsymbol{\eta}_t^{\mathbf{h}}, \boldsymbol{\eta}_t^{\mathbf{w}})\Big)\mathrm{d}\beta_k(t) \qquad \forall\, t \in [0,T]\,, \tag{2.3a}$$
$$\boldsymbol{\zeta}_0^{\mathbf{h},\mathbf{w}} = \mathbf{0} \in \mathbb{R}^L\,. \tag{2.3b}$$

Since assumptions **(A1)** – **(A2)** are valid, there exists a unique solution $\boldsymbol{\zeta}^{\mathbf{h},\mathbf{w}} \equiv \{\boldsymbol{\zeta}_t^{\mathbf{h},\mathbf{w}};\, t \in [0,T]\}$; it is equal to $D_{\mathbf{y}}^2\mathbf{X}^{0,\mathbf{y}}\cdot(\mathbf{h},\mathbf{w})$, the second derivative *w.r.t.* the initial datum $\mathbf{y} \in \mathbb{R}^L$ of the map $\mathbf{y} \mapsto \mathbf{X}^{0,\mathbf{y}}$ along the directions $\mathbf{h}, \mathbf{w} \in \mathbb{R}^L$.

Lemma 2.3. Assume **(A1) – (A2)** in (2.3a). Then, for every $\mathbf{h}, \mathbf{w} \in \mathbb{R}^L$, $\varepsilon_1, \varepsilon_2 > 0$ and $p \geq 1$,

$$\sup_{t\in[0,T]} \mathbb{E}\Big[\|\boldsymbol{\zeta}_t^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^p\Big] \leq V_{p,\varepsilon_1,\varepsilon_2}^{(2)}\cdot \|\mathbf{h}\|_{\mathbb{R}^L}^p\|\mathbf{w}\|_{\mathbb{R}^L}^p\,,$$

where $V_{1,\varepsilon_1,\varepsilon_2}^{(2)} := \sqrt{V_{2,\varepsilon_1,\varepsilon_2}^{(2)}}$ for $p = 1$, and

$$V_{p,\varepsilon_1,\varepsilon_2}^{(2)} := T\Bigg(\frac{1}{\varepsilon_1^{p-1}}C_{D^2\mathbf{f}}^p + \frac{2}{\varepsilon_2^{(p-2)/2}}(p-1)^{\frac{p}{2}}C_{D^2\boldsymbol{\sigma}}^p\Bigg)V_{2p}^{(1)}$$
$$\cdot\, e^{pT\max\big\{-\lambda_{\mathscr{A}} + C_{D\mathbf{f}} + (p-1)C_{D\boldsymbol{\sigma}}^2 + \frac{(p-1)}{p}\varepsilon_1 + \frac{(p-2)}{p}\varepsilon_2, 0\big\}} \qquad (p \geq 2)\,.$$

Proof. The proof follows the steps outlined in the proof of Lemma 2.2: we first prove the statement for $p \in \mathbb{N}$. By *Jensen's inequality*, the result also holds true for noninteger $p \geq 1$.
a) $2 \leq p \in \mathbb{N}$: Let $t \in [0,T]$, $\mathbf{h}, \mathbf{w} \in \mathbb{R}^L$ and $2 \leq p \in \mathbb{N}$. By *Itô's formula*,

$$\frac{1}{p}\mathbb{E}\Big[\|\boldsymbol{\zeta}_t^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^p\Big]$$
$$\leq \mathbb{E}\Big[\int_0^t \Big\{\Big\langle\Big(-\mathscr{A} + D\mathbf{f}(\mathbf{X}_s)\Big)\cdot \boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}}, \boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}}\Big\rangle_{\mathbb{R}^L} + \Big\langle D^2\mathbf{f}(\mathbf{X}_s)\cdot(\boldsymbol{\eta}_s^{\mathbf{h}}, \boldsymbol{\eta}_s^{\mathbf{w}}), \boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}}\Big\rangle_{\mathbb{R}^L}\Big\}\cdot \|\boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^{p-2}$$
$$+ \frac{1}{2}\sum_{k=1}^K \|D\boldsymbol{\sigma}_k(\mathbf{X}_s)\cdot \boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}} + D^2\boldsymbol{\sigma}_k(\mathbf{X}_s)\cdot(\boldsymbol{\eta}_s^{\mathbf{h}}, \boldsymbol{\eta}_s^{\mathbf{w}})\|_{\mathbb{R}^L}^2\cdot(p-1)\cdot\|\boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^{p-2}\,\mathrm{d}s\Big]\,.$$

Using assumptions **(A1)** − **(A2)** leads to

$$\begin{aligned}\frac{1}{p}\mathbb{E}\Big[\|\boldsymbol{\zeta}_t^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^p\Big] \leq \mathbb{E}\Big[\int_0^t &\{-\lambda_{\mathscr{A}} + C_{D\mathbf{f}}\}\cdot\|\boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^p + C_{D^2\mathbf{f}}\|\boldsymbol{\eta}_s^{\mathbf{h}}\|_{\mathbb{R}^L}\|\boldsymbol{\eta}_s^{\mathbf{w}}\|_{\mathbb{R}^L}\|\boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^{p-1}\\ &+\sum_{k=1}^K\|D\boldsymbol{\sigma}_k(\mathbf{X}_s)\cdot\boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^2\cdot(p-1)\cdot\|\boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^{p-2}\\ &+\sum_{k=1}^K\|D^2\boldsymbol{\sigma}_k(\mathbf{X}_s)\cdot(\boldsymbol{\eta}_s^{\mathbf{h}},\boldsymbol{\eta}_s^{\mathbf{w}})\|_{\mathbb{R}^L}^2\cdot(p-1)\cdot\|\boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^{p-2}\,\mathrm{d}s\Big]\\ \leq \mathbb{E}\Big[\int_0^t\Big\{-\lambda_{\mathscr{A}} + C_{D\mathbf{f}} + (p-1)C_{D\boldsymbol{\sigma}}^2\Big\}&\cdot\|\boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^p + \underbrace{C_{D^2\mathbf{f}}\|\boldsymbol{\eta}_s^{\mathbf{h}}\|_{\mathbb{R}^L}\|\boldsymbol{\eta}_s^{\mathbf{w}}\|_{\mathbb{R}^L}\|\boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^{p-1}}_{=:\mathsf{I}}\\ &+\underbrace{C_{D^2\boldsymbol{\sigma}}^2\|\boldsymbol{\eta}_s^{\mathbf{h}}\|_{\mathbb{R}^L}^2\|\boldsymbol{\eta}_s^{\mathbf{w}}\|_{\mathbb{R}^L}^2(p-1)\|\boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^{p-2}}_{=:\mathsf{II}}\,\mathrm{d}s\Big]\,.\end{aligned}$$

Applying *Young's inequality* (in generalized from) to I and II leads to

$$\begin{aligned}\mathsf{I} &\leq \tfrac{(p-1)\varepsilon_1}{p}\|\boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^p + \tfrac{1}{p\varepsilon_1^{p-1}}C_{D^2\mathbf{f}}^p\|\boldsymbol{\eta}_s^{\mathbf{h}}\|_{\mathbb{R}^L}^p\|\boldsymbol{\eta}_s^{\mathbf{w}}\|_{\mathbb{R}^L}^p \qquad (\varepsilon_1>0)\,,\\ \mathsf{II} &\leq \tfrac{(p-2)\varepsilon_2}{p}\|\boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^p + \frac{2}{p\varepsilon_2^{(p-2)/p}}(p-1)^{\frac{p}{2}}C_{D^2\boldsymbol{\sigma}}^p\|\boldsymbol{\eta}_s^{\mathbf{h}}\|_{\mathbb{R}^L}^p\|\boldsymbol{\eta}_s^{\mathbf{w}}\|_{\mathbb{R}^L}^p \qquad (\varepsilon_2>0)\,,\end{aligned}$$

and hence

$$\begin{aligned}\frac{1}{p}\mathbb{E}\Big[\|\boldsymbol{\zeta}_t^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^p\Big] &\leq \mathbb{E}\Big[\int_0^t\Big\{-\lambda_{\mathscr{A}} + C_{D\mathbf{f}} + (p-1)C_{D\boldsymbol{\sigma}}^2 + \tfrac{(p-1)}{p}\varepsilon_1 + \tfrac{(p-2)}{p}\varepsilon_2\Big\}\cdot\|\boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^p\,\mathrm{d}s\Big]\\ &+\Big(\tfrac{1}{p\varepsilon_1^{p-1}}C_{D^2\mathbf{f}}^p + \tfrac{2}{p\varepsilon_2^{(p-2)/p}}(p-1)^{\frac{p}{2}}C_{D^2\boldsymbol{\sigma}}^p\Big)\int_0^t\sqrt{\mathbb{E}\Big[\|\boldsymbol{\eta}_s^{\mathbf{h}}\|_{\mathbb{R}^L}^{2p}\Big]}\sqrt{\mathbb{E}\Big[\|\boldsymbol{\eta}_s^{\mathbf{w}}\|_{\mathbb{R}^L}^{2p}\Big]}\,\mathrm{d}s\,.\end{aligned}$$

Applying Lemma 2.2 in combination with the *Cauchy-Schwarz inequality* further leads to

$$\begin{aligned}\mathbb{E}\Big[\|\boldsymbol{\zeta}_t^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^p\Big] &\leq p\max\Big\{-\lambda_{\mathscr{A}} + C_{D\mathbf{f}} + (p-1)C_{D\boldsymbol{\sigma}}^2 + \tfrac{p-1}{p}\varepsilon_1 + \tfrac{p-2}{p}\varepsilon_2, 0\Big\}\int_0^t\mathbb{E}\Big[\|\boldsymbol{\zeta}_s^{\mathbf{h},\mathbf{w}}\|_{\mathbb{R}^L}^p\Big]\,\mathrm{d}s\\ &+T\Big(\tfrac{1}{\varepsilon_1^{p-1}}C_{D^2\mathbf{f}}^p + \tfrac{2}{\varepsilon_2^{(p-2)/p}}(p-1)^{\frac{p}{2}}C_{D^2\boldsymbol{\sigma}}^p\Big)V_{2p}^{(1)}\|\mathbf{h}\|_{\mathbb{R}^L}^p\|\mathbf{w}\|_{\mathbb{R}^L}^p\,.\end{aligned}$$

Applying *Gronwall's inequality* finally yields the first result.
b) $p=1$: This follows by using *Jensen's inequality.*

□

Now, let $\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}\in\mathbb{R}^L$. Following [19, p. 43*ff.*], we recall the *third variation equation* corresponding to (0.1),

$$\begin{aligned}\mathrm{d}\boldsymbol{\Theta}_t^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}} = \Big(\Big(-\mathscr{A} + D\mathbf{f}(\mathbf{X}_t)\Big)\cdot\boldsymbol{\Theta}_t^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}} &+ \frac{1}{4}\sum_{\pi\in\mathsf{S}_3}D^2\mathbf{f}(\mathbf{X}_t)\cdot(\boldsymbol{\eta}_t^{\mathbf{h}_{\pi(1)}},\boldsymbol{\zeta}_t^{\mathbf{h}_{\pi(2)},\mathbf{h}_{\pi(3)}})\\ &+D^3\mathbf{f}(\mathbf{X}_t)\cdot(\boldsymbol{\eta}_t^{\mathbf{h_1}},\boldsymbol{\eta}_t^{\mathbf{h_2}},\boldsymbol{\eta}_t^{\mathbf{h_3}})\Big)\mathrm{d}t\end{aligned}$$

$$+\sum_{k=1}^{K}\Big(D\boldsymbol{\sigma}_k(\mathbf{X}_t)\cdot\boldsymbol{\Theta}_t^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}+\frac{1}{4}\sum_{\pi\in\mathtt{S}_3}D^2\boldsymbol{\sigma}_k(\mathbf{X}_t)\cdot(\boldsymbol{\eta}_t^{\mathbf{h}_{\pi(1)}},\boldsymbol{\zeta}_t^{\mathbf{h}_{\pi(2)},\mathbf{h}_{\pi(3)}})$$
$$+D^3\boldsymbol{\sigma}_k(\mathbf{X}_t)\cdot(\boldsymbol{\eta}_t^{\mathbf{h_1}},\boldsymbol{\eta}_t^{\mathbf{h_2}},\boldsymbol{\eta}_t^{\mathbf{h_3}})\Big)\mathrm{d}\beta_k(t)\qquad\forall\, t\in[0,T]\,,\tag{2.4a}$$
$$\boldsymbol{\Theta}_0^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}=\mathbf{0}\in\mathbb{R}^L\,.\tag{2.4b}$$

Here, $\mathtt{S}_3$ denotes the set of all permutations of a set of three elements. Since **(A1)** − **(A2)** are valid, there exists a unique solution $\boldsymbol{\Theta}^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\equiv\{\boldsymbol{\Theta}_t^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}};\, t\in[0,T]\}$; it is equal to $D^3_{\mathbf{y}}\mathbf{X}^{0,\mathbf{y}}\cdot(\mathbf{h_1},\mathbf{h_2},\mathbf{h_3})$, the third derivative *w.r.t.* the initial datum $\mathbf{y}\in\mathbb{R}^L$ of the map $\mathbf{y}\mapsto\mathbf{X}^{0,\mathbf{y}}$ along the directions $\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}\in\mathbb{R}^L$.
Moment bounds for the solution of (2.4) are obtained as in Lemmas 2.2 and 2.3. In view of Lemma 2.5 below, it suffices to consider only second moment bounds in the following lemma.

Lemma 2.4. Assume **(A1)** − **(A2)** in (2.4). Then, for every $\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}\in\mathbb{R}^L$ and $\varepsilon_3>0$,

$$\sup_{t\in[0,T]}\mathbb{E}\Big[\|\boldsymbol{\Theta}_t^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\|^2_{\mathbb{R}^L}\Big]\le V^{(3)}_{2,\varepsilon_3}\cdot\|\mathbf{h_1}\|^2_{\mathbb{R}^L}\|\mathbf{h_2}\|^2_{\mathbb{R}^L}\|\mathbf{h_3}\|^2_{\mathbb{R}^L}\,,$$

where

$$V^{(3)}_{2,\varepsilon_3}:=T\sqrt{V_4^{(1)}}\Bigg(\sqrt{V^{(2)}_{4,\varepsilon_1,\varepsilon_2}}\Big(\tfrac{9}{4\varepsilon_3}C^2_{D^2\mathbf{f}}+\tfrac{27}{4}C^2_{D^2\boldsymbol{\sigma}}\Big)+\sqrt{V_8^{(1)}}\Big(\tfrac{1}{\varepsilon_3}C^2_{D^3\mathbf{f}}+3C^2_{D^3\boldsymbol{\sigma}}\Big)\Bigg)$$
$$\cdot\, e^{2T\max\left\{-\lambda_{\mathscr{A}}+C_{D\mathbf{f}}+\frac{3}{2}C^2_{D\boldsymbol{\sigma}}+\varepsilon_3,0\right\}}\,.$$

Moreover for $V^{(3)}_{1,\varepsilon_3}:=\sqrt{V^{(3)}_{2,\varepsilon_3}}$,

$$\sup_{t\in[0,T]}\mathbb{E}\Big[\|\boldsymbol{\Theta}_t^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\|_{\mathbb{R}^L}\Big]\le V^{(3)}_{1,\varepsilon_3}\cdot\|\mathbf{h_1}\|_{\mathbb{R}^L}\|\mathbf{h_2}\|_{\mathbb{R}^L}\|\mathbf{h_3}\|_{\mathbb{R}^L}\,.$$

Proof. **a)** Let $t\in[0,T]$, and $\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}\in\mathbb{R}^L$. By *Itô's formula,*

$$\frac{1}{2}\mathbb{E}[\|\boldsymbol{\Theta}_t^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\|^2_{\mathbb{R}^L}]$$
$$\le\mathbb{E}\Big[\int_0^t\big\langle\big(-\mathscr{A}+D\mathbf{f}(\mathbf{X}_s)\big)\cdot\boldsymbol{\Theta}_s^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}},\boldsymbol{\Theta}_s^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\big\rangle_{\mathbb{R}^L}$$
$$+\frac{1}{4}\sum_{\pi\in\mathtt{S}_3}\big\langle D^2\mathbf{f}(\mathbf{X}_s)\cdot(\boldsymbol{\eta}_s^{\mathbf{h}_{\pi(1)}},\boldsymbol{\zeta}_s^{\mathbf{h}_{\pi(2)},\mathbf{h}_{\pi(3)}}),\boldsymbol{\Theta}_s^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\big\rangle_{\mathbb{R}^L}+\big\langle D^3\mathbf{f}(\mathbf{X}_s)\cdot(\boldsymbol{\eta}_s^{\mathbf{h_1}},\boldsymbol{\eta}_s^{\mathbf{h_2}},\boldsymbol{\eta}_s^{\mathbf{h_3}}),\boldsymbol{\Theta}_s^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\big\rangle_{\mathbb{R}^L}$$
$$+\frac{1}{2}\sum_{k=1}^{K}\Big\|D\boldsymbol{\sigma}_k(\mathbf{X}_s)\cdot\boldsymbol{\Theta}_s^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}+\frac{1}{4}\sum_{\pi\in\mathtt{S}_3}D^2\boldsymbol{\sigma}_k(\mathbf{X}_s)\cdot(\boldsymbol{\eta}_s^{\mathbf{h}_{\pi(1)}},\boldsymbol{\zeta}_s^{\mathbf{h}_{\pi(2)},\mathbf{h}_{\pi(3)}})$$
$$+D^3\boldsymbol{\sigma}_k(\mathbf{X}_s)\cdot(\boldsymbol{\eta}_s^{\mathbf{h_1}},\boldsymbol{\eta}_s^{\mathbf{h_2}},\boldsymbol{\eta}_s^{\mathbf{h_3}})\Big\|^2_{\mathbb{R}^L}\,\mathrm{d}s\Big]\,.$$

Using assumptions **(A1)** − **(A2)** leads to

$$\frac{1}{2}\mathbb{E}[\|\boldsymbol{\Theta}_t^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\|^2_{\mathbb{R}^L}]$$

$$
\begin{aligned}
&\le \mathbb{E}\Big[\int_0^t \big\{-\lambda_{\mathscr{A}} + C_{D\mathbf{f}}\big\} \cdot \|\boldsymbol{\Theta}_s^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\|_{\mathbb{R}^L}^2 + \frac{1}{4} C_{D^2\mathbf{f}} \sum_{\pi\in\mathsf{S}_3} \|\boldsymbol{\eta}_s^{\mathbf{h}_{\pi(1)}}\|_{\mathbb{R}^L} \|\boldsymbol{\zeta}_s^{\mathbf{h}_{\pi(2)},\mathbf{h}_{\pi(3)}}\|_{\mathbb{R}^L} \|\boldsymbol{\Theta}_s^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\|_{\mathbb{R}^L} \\
&\quad + C_{D^3\mathbf{f}} \|\boldsymbol{\eta}_s^{\mathbf{h_1}}\|_{\mathbb{R}^L} \|\boldsymbol{\eta}_s^{\mathbf{h_2}}\|_{\mathbb{R}^L} \|\boldsymbol{\eta}_s^{\mathbf{h_3}}\|_{\mathbb{R}^L} \|_{\mathbb{R}^L} \|\boldsymbol{\Theta}_s^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\|_{\mathbb{R}^L} + \frac{3}{2} \sum_{k=1}^K \|D\boldsymbol{\sigma}_k(\mathbf{X}_s) \cdot \boldsymbol{\Theta}_s^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\|_{\mathbb{R}^L}^2 \\
&\quad + \frac{3}{2\cdot 4^2} \sum_{k=1}^K \Big\| \sum_{\pi\in\mathsf{S}_3} D^2\boldsymbol{\sigma}_k(\mathbf{X}_s) \cdot (\boldsymbol{\eta}_s^{\mathbf{h}_{\pi(1)}}, \boldsymbol{\zeta}_s^{\mathbf{h}_{\pi(2)},\mathbf{h}_{\pi(3)}}) \Big\|_{\mathbb{R}^L}^2 + \frac{3}{2} \sum_{k=1}^K \|D^3\boldsymbol{\sigma}_k(\mathbf{X}_s) \cdot (\boldsymbol{\eta}_s^{\mathbf{h_1}}, \boldsymbol{\eta}_s^{\mathbf{h_2}}, \boldsymbol{\eta}_s^{\mathbf{h_3}})\|_{\mathbb{R}^L}^2 \,\mathrm{d}s\Big] \\
&\le \mathbb{E}\Big[\int_0^t \big\{-\lambda_{\mathscr{A}} + C_{D\mathbf{f}} + \tfrac{3}{2} C_{D\boldsymbol{\sigma}}^2\big\} \cdot \|\boldsymbol{\Theta}_s^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\|_{\mathbb{R}^L}^2 \\
&\quad + \underbrace{\|\boldsymbol{\Theta}_s^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\|_{\mathbb{R}^L} \cdot \Big\{ C_{D^3\mathbf{f}} \|\boldsymbol{\eta}_s^{\mathbf{h_1}}\|_{\mathbb{R}^L} \|\boldsymbol{\eta}_s^{\mathbf{h_2}}\|_{\mathbb{R}^L} \|\boldsymbol{\eta}_s^{\mathbf{h_3}}\|_{\mathbb{R}^L} + \frac{1}{4} C_{D^2\mathbf{f}} \sum_{\pi\in\mathsf{S}_3} \|\boldsymbol{\eta}_s^{\mathbf{h}_{\pi(1)}}\|_{\mathbb{R}^L} \|\boldsymbol{\zeta}_s^{\mathbf{h}_{\pi(2)},\mathbf{h}_{\pi(3)}}\|_{\mathbb{R}^L} \Big\}}_{=:\mathsf{I}} \\
&\quad + \frac{3\cdot 6}{2\cdot 4^2} C_{D^2\boldsymbol{\sigma}}^2 \sum_{\pi\in\mathsf{S}_3} \|\boldsymbol{\eta}_s^{\mathbf{h}_{\pi(1)}}\|_{\mathbb{R}^L}^2 \|\boldsymbol{\zeta}_s^{\mathbf{h}_{\pi(2)},\mathbf{h}_{\pi(3)}}\|_{\mathbb{R}^L}^2 + \frac{3}{2} C_{D^3\boldsymbol{\sigma}}^2 \|\boldsymbol{\eta}_s^{\mathbf{h_1}}\|_{\mathbb{R}^L} \|\boldsymbol{\eta}_s^{\mathbf{h_2}}\|_{\mathbb{R}^L} \|\boldsymbol{\eta}_s^{\mathbf{h_3}}\|_{\mathbb{R}^L} \,\mathrm{d}s\Big] .
\end{aligned}
$$

Applying *Young's inequality* (in generalized form) to I leads to

$$
\begin{aligned}
\mathsf{I} &\le \varepsilon_3 \|\boldsymbol{\Theta}_s^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\|_{\mathbb{R}^L}^2 \\
&\quad + \frac{1}{4\varepsilon_3} \Big\{ C_{D^3\mathbf{f}} \|\boldsymbol{\eta}_s^{\mathbf{h_1}}\|_{\mathbb{R}^L} \|\boldsymbol{\eta}_s^{\mathbf{h_2}}\|_{\mathbb{R}^L} \|\boldsymbol{\eta}_s^{\mathbf{h_3}}\|_{\mathbb{R}^L} + \frac{1}{4} C_{D^2\mathbf{f}} \sum_{\pi\in\mathsf{S}_3} \|\boldsymbol{\eta}_s^{\mathbf{h}_{\pi(1)}}\|_{\mathbb{R}^L} \|\boldsymbol{\zeta}_s^{\mathbf{h}_{\pi(2)},\mathbf{h}_{\pi(3)}}\|_{\mathbb{R}^L} \Big\}^2 \quad (\varepsilon_3 > 0) ,
\end{aligned}
$$

and hence

$$
\begin{aligned}
&\frac{1}{2}\mathbb{E}\big[\|\boldsymbol{\Theta}_t^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\|_{\mathbb{R}^L}^2\big] \\
&\quad \le \mathbb{E}\Big[\int_0^t \big\{-\lambda_{\mathscr{A}} + C_{D\mathbf{f}} + \tfrac{3}{2} C_{D\boldsymbol{\sigma}}^2 + \varepsilon_3\big\} \cdot \|\boldsymbol{\Theta}_s^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\|_{\mathbb{R}^L}^2 \,\mathrm{d}s\Big] \\
&\quad\quad + \mathbb{E}\Big[\int_0^t \frac{1}{4\varepsilon_3} \Big\{ C_{D^3\mathbf{f}} \|\boldsymbol{\eta}_s^{\mathbf{h_1}}\|_{\mathbb{R}^L} \|\boldsymbol{\eta}_s^{\mathbf{h_2}}\|_{\mathbb{R}^L} \|\boldsymbol{\eta}_s^{\mathbf{h_3}}\|_{\mathbb{R}^L} + \frac{1}{4} C_{D^2\mathbf{f}} \sum_{\pi\in\mathsf{S}_3} \|\boldsymbol{\eta}_s^{\mathbf{h}_{\pi(1)}}\|_{\mathbb{R}^L} \|\boldsymbol{\zeta}_s^{\mathbf{h}_{\pi(2)},\mathbf{h}_{\pi(3)}}\|_{\mathbb{R}^L} \Big\}^2 \\
&\quad\quad\quad + \frac{9}{16} C_{D^2\boldsymbol{\sigma}}^2 \sum_{\pi\in\mathsf{S}_3} \|\boldsymbol{\eta}_s^{\mathbf{h}_{\pi(1)}}\|_{\mathbb{R}^L}^2 \|\boldsymbol{\zeta}_s^{\mathbf{h}_{\pi(2)},\mathbf{h}_{\pi(3)}}\|_{\mathbb{R}^L}^2 + \frac{3}{2} C_{D^3\boldsymbol{\sigma}}^2 \|\boldsymbol{\eta}_s^{\mathbf{h_1}}\|_{\mathbb{R}^L} \|\boldsymbol{\eta}_s^{\mathbf{h_2}}\|_{\mathbb{R}^L} \|\boldsymbol{\eta}_s^{\mathbf{h_3}}\|_{\mathbb{R}^L} \,\mathrm{d}s\Big] \\
&\quad \le \max\big\{-\lambda_{\mathscr{A}} + C_{D\mathbf{f}} + \tfrac{3}{2} C_{D\boldsymbol{\sigma}}^2 + \varepsilon_3, 0\big\} \int_0^t \mathbb{E}\big[\|\boldsymbol{\Theta}_s^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}\|_{\mathbb{R}^L}^2\big] \,\mathrm{d}s \\
&\quad\quad + \underbrace{\Big(\tfrac{1}{2\varepsilon_3} C_{D^3\mathbf{f}}^2 + \tfrac{3}{2} C_{D^3\boldsymbol{\sigma}}^2\Big) \int_0^t \mathbb{E}\big[\|\boldsymbol{\eta}_s^{\mathbf{h_1}}\|_{\mathbb{R}^L} \|\boldsymbol{\eta}_s^{\mathbf{h_2}}\|_{\mathbb{R}^L} \|\boldsymbol{\eta}_s^{\mathbf{h_3}}\|_{\mathbb{R}^L}\big] \,\mathrm{d}s}_{=:\mathsf{II}} \\
&\quad\quad + \underbrace{\Big(\tfrac{3}{16\varepsilon_3} C_{D^2\mathbf{f}}^2 + \tfrac{9}{16} C_{D^2\boldsymbol{\sigma}}^2\Big) \int_0^t \mathbb{E}\Big[\sum_{\pi\in\mathsf{S}_3} \|\boldsymbol{\eta}_s^{\mathbf{h}_{\pi(1)}}\|_{\mathbb{R}^L}^2 \|\boldsymbol{\zeta}_s^{\mathbf{h}_{\pi(2)},\mathbf{h}_{\pi(3)}}\|_{\mathbb{R}^L}^2\Big] \,\mathrm{d}s}_{=:\mathsf{III}} .
\end{aligned}
$$

Applying Lemmas 2.2 and 2.3 in combination with the *Cauchy-Schwarz inequality* to II and III further leads to

$$
\mathsf{II} \le T\Big(\tfrac{1}{2\varepsilon_3} C_{D^3\mathbf{f}}^2 + \tfrac{3}{2} C_{D^3\boldsymbol{\sigma}}^2\Big) \sqrt{V_4^{(1)}} \sqrt{V_8^{(1)}} \|\mathbf{h_1}\|_{\mathbb{R}^L}^2 \|\mathbf{h_2}\|_{\mathbb{R}^L}^2 \|\mathbf{h_3}\|_{\mathbb{R}^L}^2 ,
$$

$$\mathsf{III} \le T\Big(\tfrac{9}{8\varepsilon_3}C^2_{D^2\mathbf{f}} + \tfrac{27}{8}C^2_{D^2\boldsymbol{\sigma}}\Big)\sqrt{V_4^{(1)}}\sqrt{V^{(2)}_{4,\varepsilon_1,\varepsilon_2}}\|\mathbf{h_1}\|^2_{\mathbb{R}^L}\|\mathbf{h_2}\|^2_{\mathbb{R}^L}\|\mathbf{h_3}\|^2_{\mathbb{R}^L}\,,$$

and consequently

$$\begin{aligned}\mathbb{E}\Big[\|\boldsymbol{\Theta}^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}_t\|^2_{\mathbb{R}^L}\Big] &\le 2\max\Big\{-\lambda_{\mathscr{A}} + C_{D\mathbf{f}} + \tfrac{3}{2}C^2_{D\boldsymbol{\sigma}} + \varepsilon_3, 0\Big\}\int_0^t \mathbb{E}\Big[\|\boldsymbol{\Theta}^{\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}}_s\|^2_{\mathbb{R}^L}\Big]\,\mathrm{d}s\\ &\quad + T\sqrt{V_4^{(1)}}\bigg(\sqrt{V^{(2)}_{4,\varepsilon_1,\varepsilon_2}}\Big(\tfrac{9}{4\varepsilon_3}C^2_{D^2\mathbf{f}} + \tfrac{27}{4}C^2_{D^2\boldsymbol{\sigma}}\Big) + \sqrt{V_8^{(1)}}\Big(\tfrac{1}{\varepsilon_3}C^2_{D^3\mathbf{f}} + 3C^2_{D^3\boldsymbol{\sigma}}\Big)\bigg)\\ &\qquad\cdot\|\mathbf{h_1}\|^2_{\mathbb{R}^L}\|\mathbf{h_2}\|^2_{\mathbb{R}^L}\|\mathbf{h_3}\|^2_{\mathbb{R}^L}\,.\end{aligned}$$

Applying *Gronwall's inequality* finally yields the first result.
b) The second result immediatly follows by using *Jensen's inequality.*

□

Let $t \in [0,T]$ and $\mathbf{x} \in \mathbb{R}^L$. In order to obtain global bounds for the first and higher derivatives of u in (0.5), we use the following identities, which can be found in *e.g.* [19, p. 94] in connection with *Kolmogorov's forward equation*, which is just a time reversal $t \mapsto T-t$ and a change of the terminal condition into an initial condition in (0.5).

$$\Big\langle D_{\mathbf{x}}u(t,\mathbf{x}),\mathbf{h}\Big\rangle_{\mathbb{R}^L} = \mathbb{E}\Big[\Big\langle D_{\mathbf{x}}\phi(\mathbf{X}^{t,\mathbf{x}}_T), D_{\mathbf{x}}\mathbf{X}^{t,\mathbf{x}}_T\cdot\mathbf{h}\Big\rangle_{\mathbb{R}^L}\Big] \qquad \forall\,\mathbf{h}\in\mathbb{R}^L, \tag{2.5}$$

$$\begin{aligned}\Big\langle D^2_{\mathbf{x}}u(t,\mathbf{x})\mathbf{h},\mathbf{w}\Big\rangle_{\mathbb{R}^L} &= \mathbb{E}\Big[\Big\langle D^2_{\mathbf{x}}\phi(\mathbf{X}^{t,\mathbf{x}}_T)D_{\mathbf{x}}\mathbf{X}^{t,\mathbf{x}}_T\cdot\mathbf{h}, D_{\mathbf{x}}\mathbf{X}^{t,\mathbf{x}}_T\cdot\mathbf{w}\Big\rangle_{\mathbb{R}^L}\Big]\\ &\quad + \mathbb{E}\Big[\Big\langle D_{\mathbf{x}}\phi(\mathbf{X}^{t,\mathbf{x}}_T), D^2_{\mathbf{x}}\mathbf{X}^{t,\mathbf{x}}_T\cdot(\mathbf{h},\mathbf{w})\Big\rangle_{\mathbb{R}^L}\Big] \qquad \forall\,\mathbf{h},\mathbf{w}\in\mathbb{R}^L\end{aligned} \tag{2.6}$$

and

$$\begin{aligned}\Big\langle D_{\mathbf{x}}\Big\langle D^2_{\mathbf{x}}u(t,\mathbf{x})\mathbf{h_1},\mathbf{h_2}\Big\rangle_{\mathbb{R}^L},\mathbf{h_3}\Big\rangle_{\mathbb{R}^L} &= \mathbb{E}\Big[\Big\langle D^3_{\mathbf{x}}\phi(\mathbf{X}^{t,\mathbf{x}}_T)D_{\mathbf{x}}\mathbf{X}^{t,\mathbf{x}}_T\cdot\mathbf{h_1}D_{\mathbf{x}}\mathbf{X}^{t,\mathbf{x}}_T\cdot\mathbf{h_2}, D_{\mathbf{x}}\mathbf{X}^{t,\mathbf{x}}_T\cdot\mathbf{h_3}\Big\rangle_{\mathbb{R}^L}\Big]\\ &\quad + \mathbb{E}\Big[\Big\langle D^2_{\mathbf{x}}\phi(\mathbf{X}^{t,\mathbf{x}}_T)D^2_{\mathbf{x}}\mathbf{X}^{t,\mathbf{x}}_T\cdot(\mathbf{h_1},\mathbf{h_3}), D_{\mathbf{x}}\mathbf{X}^{t,\mathbf{x}}_T\cdot\mathbf{h_2}\Big\rangle_{\mathbb{R}^L}\Big]\\ &\quad + \mathbb{E}\Big[\Big\langle D^2_{\mathbf{x}}\phi(\mathbf{X}^{t,\mathbf{x}}_T)D_{\mathbf{x}}\mathbf{X}^{t,\mathbf{x}}_T\cdot\mathbf{h_1}, D^2_{\mathbf{x}}\mathbf{X}^{t,\mathbf{x}}_T\cdot(\mathbf{h_2},\mathbf{h_3})\Big\rangle_{\mathbb{R}^L}\Big]\\ &\quad + \mathbb{E}\Big[\Big\langle D^2_{\mathbf{x}}\phi(\mathbf{X}^{t,\mathbf{x}}_T)D^2_{\mathbf{x}}\mathbf{X}^{t,\mathbf{x}}_T\cdot(\mathbf{h_1},\mathbf{h_2}), D_{\mathbf{x}}\mathbf{X}^{t,\mathbf{x}}_T\cdot\mathbf{h_3}\Big\rangle_{\mathbb{R}^L}\Big]\\ &\quad + \mathbb{E}\Big[\Big\langle D_{\mathbf{x}}\phi(\mathbf{X}^{t,\mathbf{x}}_T), D^3_{\mathbf{x}}\mathbf{X}^{t,\mathbf{x}}_T\cdot(\mathbf{h_1},\mathbf{h_2},\mathbf{h_3})\Big\rangle_{\mathbb{R}^L}\Big]\\ &\qquad\qquad \forall\,\mathbf{h_1},\mathbf{h_2},\mathbf{h_3}\in\mathbb{R}^L\,.\end{aligned} \tag{2.7}$$

In the following Lemma 2.5, based upon (2.5)–(2.7) and Lemmas 2.2, 2.3 and 2.4, we derive global bounds for the first, second and third derivatives of the solution u of (0.5) in terms of *derivatives* of ϕ, which are independent of the dimension L.

Lemma 2.5. Assume **(A1)** − **(A2)**, and let $\{D^{\ell}_{\mathbf{x}} u\}_{\ell=1}^{3}$ be from (0.5). Then, for all $\varepsilon_1, \varepsilon_2, \varepsilon_3 > 0$,

(i) $$\sup_{(t,\mathbf{x})\in[0,T]\times\mathbb{R}^L} \|D_{\mathbf{x}} u(t,\mathbf{x})\|_{\mathbb{R}^L} \leq V_1^{(1)} \cdot \sup_{\mathbf{x}\in\mathbb{R}^L} \|D\phi(\mathbf{x})\|_{\mathbb{R}^L} \,,$$

(ii) $$\sup_{(t,\mathbf{x})\in[0,T]\times\mathbb{R}^L} \|D^2_{\mathbf{x}} u(t,\mathbf{x})\|_{\mathbb{R}^{L\times L}} \leq V_{1,\varepsilon_1,\varepsilon_2}^{(2)} \cdot \sup_{\mathbf{x}\in\mathbb{R}^L} \|D\phi(\mathbf{x})\|_{\mathbb{R}^L} + V_2^{(1)} \cdot \sup_{\mathbf{x}\in\mathbb{R}^L} \|D^2\phi(\mathbf{x})\|_{\mathbb{R}^{L\times L}} \,,$$

(iii) $$\begin{aligned}\sup_{(t,\mathbf{x})\in[0,T]\times\mathbb{R}^L} \|D^3_{\mathbf{x}} u(t,\mathbf{x})\|_{\mathcal{L}^3} &\leq V_{1,\varepsilon_3}^{(3)} \cdot \sup_{\mathbf{x}\in\mathbb{R}^L} \|D\phi(\mathbf{x})\|_{\mathbb{R}^L} + 3V_1^{(1)} V_{1,\varepsilon_1,\varepsilon_2}^{(2)} \cdot \sup_{\mathbf{x}\in\mathbb{R}^L} \|D^2\phi(\mathbf{x})\|_{\mathbb{R}^{L\times L}} \\ &\quad + V_1^{(1)}\sqrt{V_4^{(1)}} \cdot \sup_{\mathbf{x}\in\mathbb{R}^L} \|D^3\phi(\mathbf{x})\|_{\mathcal{L}^3} \,,\end{aligned}$$

where $V_1^{(1)}, V_2^{(1)}, V_4^{(1)}, V_{1,\varepsilon_1,\varepsilon_2}^{(2)}, V_{1,\varepsilon_3}^{(3)}$ are given in Lemmas 2.2, 2.3 and 2.4.

Proof. **(i):** Let $\mathbf{x} \in \mathbb{R}^L$ and $\mathbf{0} \neq \mathbf{h} \in \mathbb{R}^L$. We apply the *Cauchy-Schwarz inequality* to the identity (2.5) and use Lemma 2.2 to get

$$\left|\left\langle D_{\mathbf{x}} u(t,\mathbf{x}), \mathbf{h} \right\rangle_{\mathbf{R}^L}\right| \leq V_1^{(1)} \cdot \sup_{\mathbf{z}\in\mathbb{R}^L} \|D\phi(\mathbf{z})\|_{\mathbb{R}^L} \cdot \|\mathbf{h}\|_{\mathbb{R}^L} \,.$$

Thus, taking $\mathbf{h} = D_{\mathbf{x}} u(t,\mathbf{x})$ immediately yields the assertion.
(ii): Let $\mathbf{x} \in \mathbb{R}^L$, $\mathbf{0} \neq \mathbf{h}, \mathbf{w} \in \mathbb{R}^L$ and $\varepsilon_1, \varepsilon_2 > 0$. Similar to **(i)** we obtain, using Lemmas 2.2 and 2.3,

$$\left|\left\langle D^2_{\mathbf{x}} u(t,\mathbf{x})\mathbf{h}, \mathbf{w} \right\rangle_{\mathbb{R}^L}\right| \leq \left(V_{1,\varepsilon_1,\varepsilon_2}^{(2)} \cdot \sup_{\mathbf{z}\in\mathbb{R}^L} \|D\phi(\mathbf{z})\|_{\mathbb{R}^L} + V_2^{(1)} \cdot \sup_{\mathbf{z}\in\mathbb{R}^L} \|D^2\phi(\mathbf{z})\|_{\mathbb{R}^{L\times L}} \right) \cdot \|\mathbf{h}\|_{\mathbb{R}^L} \|\mathbf{w}\|_{\mathbb{R}^L} \,.$$

Taking $\mathbf{w} = D^2_{\mathbf{x}} u(t,\mathbf{x})\mathbf{h}$, we further obtain

$$\frac{\|D^2_{\mathbf{x}} u(t,\mathbf{x})\mathbf{h}\|_{\mathbb{R}^L}}{\|\mathbf{h}\|_{\mathbb{R}^L}} \leq V_{1,\varepsilon_1,\varepsilon_2}^{(2)} \cdot \sup_{\mathbf{z}\in\mathbb{R}^L} \|D\phi(\mathbf{z})\|_{\mathbb{R}^L} + V_2^{(1)} \cdot \sup_{\mathbf{z}\in\mathbb{R}^L} \|D^2\phi(\mathbf{z})\|_{\mathbb{R}^{L\times L}} \,.$$

The assertion now follows since

$$\|D^2_{\mathbf{x}} u(t,\mathbf{x})\|_{\mathbb{R}^{L\times L}} := \sup_{\|\mathbf{h}\|_{\mathbb{R}^L}=1} \|D^2_{\mathbf{x}} u(t,\mathbf{x})\mathbf{h}\|_{\mathbb{R}^L} \,.$$

(iii): Let $\mathbf{x} \in \mathbb{R}^L$ and $\mathbf{0} \neq \mathbf{h_1}, \mathbf{h_2}, \mathbf{h_3} \in \mathbb{R}^L$. Similar to **(i)** and **(ii)**, the verification of assertion **(iii)** follows by means of identity (2.7), the *Cauchy-Schwarz inequality* and Lemmas 2.2, 2.3 and 2.4.

□

2.4. Stability bounds for iterates $\{\mathbf{Y}^j\}_{j\geq 0}$ from (0.3)

We derive L-independent stability bounds for iterates $\{\mathbf{Y}^j\}_{j\geq 0}$ from (0.3), provided **(A1)** – **(A3)** are valid.

Lemma 2.6. Assume **(A1)** – **(A3)**. Consider a mesh $\{t_j\}_{j=0}^J \subset [0,T]$. Let $p \in \mathbb{N}$ and let $\{\mathbf{Y}^j\}_{j\geq 0}$ solve (0.3). Then, for $\ell = 0, 1, 2$, we have

$$\sup_{j\geq 0} \mathbb{E}\Big[\|\mathscr{A}^\ell \mathbf{Y}^j\|_{\mathbb{R}^L}^{2p}\Big] \leq \boldsymbol{C}_{1,p}^{(\ell)} ,$$

where constants $\boldsymbol{C}_{1,p}^{(\ell)} > 0$ are independent of L.

We give the proof for $p = 1, 2$, and the proof for general $p > 2$ then follows inductively.

Proof. **a)** $p = 1$: Fix $j \geq 0$ and $\ell = 0, 1, 2$. Let $\mathbf{Z}^j := \mathscr{A}^\ell \mathbf{Y}^j$. We multiply (0.3) by $(\mathscr{A}^\ell)^\top \mathbf{Z}^{j+1}$ and use the *binomial formula*, as well as *Young's inequality* ($\delta_1, \delta_2 \geq 1$) to estimate

$$\begin{aligned}
&\frac{1}{2}\Big(\|\mathbf{Z}^{j+1}\|_{\mathbb{R}^L}^2 - \|\mathbf{Z}^j\|_{\mathbb{R}^L}^2\Big) + \Big(\frac{1}{2} - \frac{1}{4\delta_1} - \frac{1}{4\delta_2}\Big)\|\mathbf{Z}^{j+1} - \mathbf{Z}^j\|_{\mathbb{R}^L}^2 + \tau^{j+1}\Big\langle \mathscr{A}\mathbf{Z}^{j+1}, \mathbf{Z}^{j+1}\Big\rangle_{\mathbb{R}^L} \\
&\quad \leq \delta_1 (\tau^{j+1})^2 \|\mathscr{A}^\ell \mathbf{f}(\mathbf{Y}^j)\|_{\mathbb{R}^L}^2 + \tau^{j+1}\|\mathscr{A}^\ell \mathbf{f}(\mathbf{Y}^j)\|_{\mathbb{R}^L}\|\mathbf{Z}^j\|_{\mathbb{R}^L} \\
&\qquad + \delta_2 \sum_{k=1}^K \|\mathscr{A}^\ell \boldsymbol{\sigma}_k(\mathbf{Y}^j)\Delta_{j+1}\beta_k\|_{\mathbb{R}^L}^2 + \sum_{k=1}^K \Big\langle \mathscr{A}^\ell \boldsymbol{\sigma}_k(\mathbf{Y}^j)\Delta_{j+1}\beta_k, \mathbf{Z}^j\Big\rangle_{\mathbb{R}^L} .
\end{aligned} \tag{2.8}$$

Note that the last term vanishes if $\mathbb{E}[\cdot]$ is applied. By **(A1)** – **(A2)**, the tower property for expectations, and the identity $\mathbb{E}\Big[|\Delta_{j+1}\beta_k|^2\Big] = \tau^{j+1}$, we further conclude that

$$\begin{aligned}
\frac{1}{2}\mathbb{E}\Big[\|\mathbf{Z}^{j+1}\|_{\mathbb{R}^L}^2\Big] &\leq \Big(2\delta_1\Big(C_{\mathbf{f}}^{(\ell)}\Big)^2 T + 2C_{\mathbf{f}}^{(\ell)} + 2\delta_2 K\Big(C_{\boldsymbol{\sigma}}^{(\ell)}\Big)^2\Big) \cdot \mathbb{E}\Big[\|\mathbf{Z}^j\|_{\mathbb{R}^L}^2\Big] \cdot \tau^{j+1} \\
&\quad + \Big(2\delta_1\Big(C_{\mathbf{f}}^{(\ell)}\Big)^2 T + 2C_{\mathbf{f}}^{(\ell)} + 2\delta_2 K\Big(C_{\boldsymbol{\sigma}}^{(\ell)}\Big)^2\Big) \cdot \tau^{j+1} .
\end{aligned}$$

We set

$$\mathbf{C} := \Big(2\delta_1\Big(C_{\mathbf{f}}^{(\ell)}\Big)^2 T + 2C_{\mathbf{f}}^{(\ell)} + 2\delta_2 K\Big(C_{\boldsymbol{\sigma}}^{(\ell)}\Big)^2\Big) .$$

Summation over all iteration steps, and using **(A3)** then lead to

$$\mathbb{E}\Big[\|\mathbf{Z}^{j^*}\|_{\mathbb{R}^L}^2\Big] \leq \Big(C_{\mathbf{y}}^{(\ell)}\Big)^2 + 2\mathbf{C}t_{j^*} + 2\mathbf{C}\sum_{j=0}^{j^*-1} \tau^{j+1}\mathbb{E}\Big[\|\mathbf{Z}^j\|_{\mathbb{R}^L}^2\Big] \qquad (j^* \geq 1) .$$

Now, the *discrete Gronwall inequality* yields the assertion.
b) $p = 2$: Multiply (2.8) with $\|\mathbf{Z}^{j+1}\|_{\mathbb{R}^L}^2$ and use the *binomial formula* to get the estimate

$$\begin{aligned}
&\frac{1}{4}\Big(\|\mathbf{Z}^{j+1}\|_{\mathbb{R}^L}^4 - \|\mathbf{Z}^j\|_{\mathbb{R}^L}^4\Big) + \frac{1}{4}\Big(\|\mathbf{Z}^{j+1}\|_{\mathbb{R}^L}^2 - \|\mathbf{Z}^j\|_{\mathbb{R}^L}^2\Big)^2 + \Big(\frac{1}{2} - \frac{1}{4\delta_1} - \frac{1}{4\delta_2}\Big)\|\mathbf{Z}^{j+1} - \mathbf{Z}^j\|_{\mathbb{R}^L}^2\|\mathbf{Z}^{j+1}\|_{\mathbb{R}^L}^2 \\
&\quad + \tau^{j+1}\langle \mathscr{A}\mathbf{Z}^{j+1}, \mathbf{Z}^{j+1}\rangle_{\mathbb{R}^L}\|\mathbf{Z}^{j+1}\|_{\mathbb{R}^L}^2 \\
&\quad \leq \delta_1(\tau^{j+1})^2\|\mathscr{A}^\ell \mathbf{f}(\mathbf{Y}^j)\|_{\mathbb{R}^L}^2\|\mathbf{Z}^{j+1}\|_{\mathbb{R}^L}^2 + \tau^{j+1}\|\mathscr{A}^\ell \mathbf{f}(\mathbf{Y}^j)\|_{\mathbb{R}^L}\|\mathbf{Z}^j\|_{\mathbb{R}^L}\|\mathbf{Z}^{j+1}\|_{\mathbb{R}^L}^2
\end{aligned}$$

$$+ \delta_2 \sum_{k=1}^{K} \|\mathscr{A}^\ell \boldsymbol{\sigma}_k(\mathbf{Y}^j)\Delta_{j+1}\beta_k\|_{\mathbb{R}^L}^2 \|\mathbf{Z}^{j+1}\|_{\mathbb{R}^L}^2 + \sum_{k=1}^{K} \big\langle \mathscr{A}^\ell \boldsymbol{\sigma}_k(\mathbf{Y}^j)\Delta_{j+1}\beta_k, \mathbf{Z}^j \big\rangle_{\mathbb{R}^L} \|\mathbf{Z}^{j+1}\|_{\mathbb{R}^L}^2 .$$

We now add and substract $\|\mathbf{Z}^j\|_{\mathbb{R}^L}^2$ in the two terms on the right-hand side which involve random increments, to then absorb part of it to the second term on the left-hand side. For $\tilde{\delta}_1, \tilde{\delta}_2, \tilde{\delta}_3, \tilde{\delta}_4 > 0$, thanks to **(A1)** − **(A2)**, taking expectations and using the tower property leads to

$$\begin{aligned}
&\frac{1}{4}\Big(\mathbb{E}\Big[\|\mathbf{Z}^{j+1}\|_{\mathbb{R}^L}^4\Big] - \mathbb{E}\Big[\|\mathbf{Z}^j\|_{\mathbb{R}^L}^4\Big]\Big) + \Big(\frac{1}{4} - \frac{1}{4\tilde{\delta}_1} - \frac{1}{4\tilde{\delta}_2} - \frac{1}{4\tilde{\delta}_3} - \frac{1}{4\tilde{\delta}_4}\Big)\mathbb{E}\Big[\Big|\|\mathbf{Z}^{j+1}\|_{\mathbb{R}^L}^2 - \|\mathbf{Z}^j\|_{\mathbb{R}^L}^2\Big|^2\Big] \\
&+ \Big(\frac{1}{2} - \frac{1}{4\delta_1} - \frac{1}{4\delta_2}\Big)\mathbb{E}\Big[\|\mathbf{Z}^{j+1} - \mathbf{Z}^j\|_{\mathbb{R}^L}^2 \|\mathbf{Z}^{j+1}\|_{\mathbb{R}^L}^2\Big] + \tau^{j+1}\mathbb{E}\Big[\Big\langle \mathscr{A}\mathbf{Z}^{j+1}, \mathbf{Z}^{j+1}\Big\rangle_{\mathbb{R}^L} \|\mathbf{Z}^{j+1}\|_{\mathbb{R}^L}^2\Big] \\
&\leq \widetilde{\mathbf{C}}\Big(1 + \mathbb{E}\Big[\|\mathbf{Z}^j\|_{\mathbb{R}^L}^4\Big]\Big) \cdot \tau^{j+1} ,
\end{aligned}$$

with

$$\begin{aligned}
\widetilde{\mathbf{C}} := {}& 8\tilde{\delta}_1\delta_1^2 T^3\Big(C_{\mathbf{f}}^{(\ell)}\Big)^4 + 4T\delta_1\Big(C_{\mathbf{f}}^{(\ell)}\Big)^2 + 4T\tilde{\delta}_2\Big(C_{\mathbf{f}}^{(\ell)}\Big)^2 + 8C_{\mathbf{f}}^{(\ell)} + 24T\tilde{\delta}_3 K^2\delta_2^2\Big(C_{\boldsymbol{\sigma}}^{(\ell)}\Big)^4 \\
&+ 4K\delta_2\Big(C_{\boldsymbol{\sigma}}^{(\ell)}\Big)^2 + 4K^2\tilde{\delta}_4\Big(C_{\boldsymbol{\sigma}}^{(\ell)}\Big)^2.
\end{aligned}$$

Choosing $\delta_1, \delta_2 \geq 1$, $\tilde{\delta}_1, \tilde{\delta}_2, \tilde{\delta}_3, \tilde{\delta}_4 \geq 4$ and using **(A1)** then leads to

$$\frac{1}{4}\Big(\mathbb{E}\Big[\|\mathbf{Z}^{j+1}\|_{\mathbb{R}^L}^4\Big] - \mathbb{E}\Big[\|\mathbf{Z}^j\|_{\mathbb{R}^L}^4\Big]\Big) \leq \widetilde{\mathbf{C}} \cdot \tau^{j+1} + \widetilde{\mathbf{C}}\mathbb{E}\Big[\|\mathbf{Z}^j\|_{\mathbb{R}^L}^4\Big] \cdot \tau^{j+1}.$$

Now, similar arguments as in **b)** yield the assertion.

□

3. A posteriori weak error analysis

In Theorem 3.1, we derive an *a posteriori* error estimate for iterates $\{\mathbf{Y}^j\}_{j=0}^J$ of scheme (0.3). It is shown in Theorem 3.5 that this error estimator converges with optimal order on uniform meshes, recovering a corresponding result in [24] on *a priori* weak error analysis for a corresponding time discretization of (0.2); the relevant tools to verify Theorem 3.5 are the (discrete) stability properties of (0.3) in Lemma 2.6.

3.1. A posteriori weak error estimation: Derivation and Properties

We bound the weak approximation error $\max_{0\leq j\leq J}\left|\mathbb{E}\left[\phi(\mathbf{X}_{t_j})\right]-\mathbb{E}\left[\phi(\mathbf{Y}^j)\right]\right|$ in *a posteriori* form in Theorem 3.1. For this purpose, we employ the (data-dependent) estimates in Lemma 2.5, and therefore define (*cf.* also (1.5))

$$\begin{aligned}
\boldsymbol{C}_D(\phi) &:= \underbrace{V_1^{(1)}}_{C_{1,1}}\cdot\sup_{\mathbf{x}\in\mathbb{R}^L}\|D\phi(\mathbf{x})\|_{\mathbb{R}^L},\\
\boldsymbol{C}_{D^2}(\phi) &:= \underbrace{V_{1,\varepsilon_1,\varepsilon_2}^{(2)}}_{C_{2,1}}\cdot\sup_{\mathbf{x}\in\mathbb{R}^L}\|D\phi(\mathbf{x})\|_{\mathbb{R}^L}+\underbrace{V_2^{(1)}}_{C_{2,2}}\cdot\sup_{\mathbf{x}\in\mathbb{R}^L}\|D^2\phi(\mathbf{x})\|_{\mathbb{R}^{L\times L}},\\
\boldsymbol{C}_{D^3}(\phi) &:= \underbrace{V_{1,\varepsilon_3}^{(3)}}_{C_{3,1}}\cdot\sup_{\mathbf{x}\in\mathbb{R}^L}\|D\phi(\mathbf{x})\|_{\mathbb{R}^L}+\underbrace{3V_1^{(1)}V_{1,\varepsilon_1,\varepsilon_2}^{(2)}}_{C_{3,2}}\cdot\sup_{\mathbf{x}\in\mathbb{R}^L}\|D^2\phi(\mathbf{x})\|_{\mathbb{R}^{L\times L}}\\
&\qquad+\underbrace{V_1^{(1)}\sqrt{V_4^{(1)}}}_{C_{3,3}}\cdot\sup_{\mathbf{x}\in\mathbb{R}^L}\|D^3\phi(\mathbf{x})\|_{\mathcal{L}^3},\qquad(\varepsilon_1,\varepsilon_2,\varepsilon_3>0).
\end{aligned}$$

The following result estimates the weak error caused by $\{\mathbf{Y}^j\}_{j=0}^J$ from (0.3) on a mesh of local mesh sizes $\{\tau^{j+1}\}_{j=0}^{J-1}$ in terms of a computable *a posteriori* error estimator $\mathfrak{G}\equiv\{\mathfrak{G}\left(\phi;\tau^{j+1},\mathbf{Y}^j\right)\}_{j=0}^{J-1}$.

Theorem 3.1. Assume **(A1)** – **(A3)**. Let $\{t_j\}_{j=0}^J\subset[0,T]$ be a mesh with local sizes $\{\tau^{j+1}\}_{j=0}^{J-1}$. Let $\{\mathbf{Y}^j\}_{j=0}^J$ solve (0.3). Then, we have

$$\max_{0\leq j\leq J}\left|\mathbb{E}\left[\phi(\mathbf{X}_{t_j})\right]-\mathbb{E}\left[\phi(\mathbf{Y}^j)\right]\right|\leq\sum_{j=0}^{J-1}\tau^{j+1}\mathfrak{G}\left(\phi;\tau^{j+1},\mathbf{Y}^j\right),\tag{3.1}$$

where the *a posteriori* error estimator $\mathfrak{G}\Big(\phi;\tau^{j+1},\mathbf{Y}^j\Big)$ is given by

$$\mathfrak{G}\Big(\phi;\tau^{j+1},\mathbf{Y}^j\Big) := \Bigg\{\tfrac{3\boldsymbol{C}_D(\phi)}{2}\cdot\mathbf{E_1}(\mathbf{Y}^j) + \tfrac{\boldsymbol{C}_{D^2}(\phi)}{2}\cdot\mathbf{E_2}(\mathbf{Y}^j) + \tfrac{\boldsymbol{C}_D(\phi)}{2}\cdot\mathbf{E_3}(\mathbf{Y}^j) + \tfrac{\boldsymbol{C}_D(\phi)}{4}\cdot\mathbf{E_4}(\mathbf{Y}^j)$$
$$+ \tfrac{\boldsymbol{C}_{D^2}(\phi)}{2}\cdot\mathbf{E_5}(\mathbf{Y}^j) + \frac{\boldsymbol{C}_{D^2}(\phi)}{2}\cdot\mathbf{E_6}(\mathbf{Y}^j) + \tfrac{\boldsymbol{C}_{D^2}(\phi)}{4}\cdot\mathbf{E_7}(\mathbf{Y}^j)$$
$$+ \frac{\boldsymbol{C}_{D^3}(\phi)}{2}\cdot\mathbf{E_8}(\mathbf{Y}^j) + \boldsymbol{C}_{D^2}(\phi)\cdot\mathbf{E_9}(\mathbf{Y}^j) + \tfrac{\boldsymbol{C}_{D^2}(\phi)C^2_{D\boldsymbol{\sigma}}}{4}\cdot\mathbf{E_{12}}(\mathbf{Y}^j)\Bigg\}\cdot\tau^{j+1}$$
$$+ \Bigg\{\Bigg\{\boldsymbol{C}_D(\phi)C_{D^2\mathbf{f}}\cdot\sqrt{\mathbf{E_{10}}(\mathbf{Y}^j)} + \Big[\tfrac{\boldsymbol{C}_D(\phi)C_{D^3\mathbf{f}}}{2} + \boldsymbol{C}_{D^2}(\phi)C_{D^2\mathbf{f}}\Big]\cdot\sqrt{\mathbf{E_{11}}(\mathbf{Y}^j)}$$
$$+ \boldsymbol{C}_{D^2}(\phi)C_{D^2\boldsymbol{\sigma}}\cdot\sqrt{\mathbf{E_{13}}(\mathbf{Y}^j)} + \Big[\frac{\boldsymbol{C}_{D^2}(\phi)C_{D^3\boldsymbol{\sigma}}}{2} + \boldsymbol{C}_{D^3}(\phi)C_{D^2\boldsymbol{\sigma}}\Big]$$
$$\cdot\sqrt{\mathbf{E_{14}}(\mathbf{Y}^j)}\Bigg\}\cdot\sqrt{\tfrac{\tau^{j+1}}{15}\cdot\mathbf{E_{10}}(\mathbf{Y}^j) + \tfrac{1}{8}\cdot\mathbf{E_{12}}(\mathbf{Y}^j)}\Bigg\}\cdot\Big(\tau^{j+1}\Big)^{1.5}$$
$$+ \Bigg\{\tfrac{\boldsymbol{C}_{D^2}(\phi)C^2_{D\boldsymbol{\sigma}}}{6}\cdot\mathbf{E_{10}}(\mathbf{Y}^j) + \frac{\boldsymbol{C}_{D^2}(\phi)}{2}\cdot\mathbf{E_{15}}(\mathbf{Y}^j)\Bigg\}\cdot\Big(\tau^{j+1}\Big)^2,$$

with $\bar{\boldsymbol{\mathscr{A}}}^{j+1} := \Big(\mathbb{I}+\tau^{j+1}\mathscr{A}\Big)^{-1}$, and computable terms

1. $\mathbf{E_1}(\mathbf{Y}^j) = \mathbb{E}\Big[\Big\|\mathscr{A}^2\bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{Y}^j - \mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{f}(\mathbf{Y}^j)\Big\|_{\mathbb{R}^L}\Big]$,
2. $\mathbf{E_2}(\mathbf{Y}^j) := \mathbb{E}\Big[\sum_{k=1}^{K}\Big\|\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big\|_{\mathbb{R}^L}\Big\|\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big\|_{\mathbb{R}^L}\Big]$,
3. $\mathbf{E_3}(\mathbf{Y}^j) := \mathbb{E}\Big[\Big\|\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{Y}^j - \bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{f}(\mathbf{Y}^j)\Big\|_{\mathbb{R}^L}\|D\mathbf{f}(\mathbf{Y}^j)\|_{\mathcal{L}}\Big]$,
4. $\mathbf{E_4}(\mathbf{Y}^j) := \mathbb{E}\Big[\|D^2\mathbf{f}(\mathbf{Y}^j)\|_{\mathcal{L}^2}\cdot\sum_{k=1}^{K}\Big\|\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big\|^2_{\mathbb{R}^L}\Big]$,
5. $\mathbf{E_5}(\mathbf{Y}^j) := \mathbb{E}\Big[\|D\mathbf{f}(\mathbf{Y}^j)\|_{\mathcal{L}}\cdot\sum_{k=1}^{K}\Big\|\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big\|^2_{\mathbb{R}^L}\Big]$,
6. $\mathbf{E_6}(\mathbf{Y}^j) := \mathbb{E}\Big[\Big\|\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{Y}^j - \bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{f}(\mathbf{Y}^j)\Big\|_{\mathbb{R}^L}\cdot\sum_{k=1}^{K}\|\boldsymbol{\sigma}_k(\mathbf{Y}^j)\|_{\mathbb{R}^L}\|D\boldsymbol{\sigma}_k(\mathbf{Y}^j)\|_{\mathcal{L}}\Big]$,
7. $\mathbf{E_7}(\mathbf{Y}^j) := \mathbb{E}\Big[\sum_{k=1}^{K}\|\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\|^2_{\mathbb{R}^L}\cdot\sum_{k=1}^{K}\|\boldsymbol{\sigma}_k(\mathbf{Y}^j)\|_{\mathbb{R}^L}\|D^2\boldsymbol{\sigma}_k(\mathbf{Y}^j)\|_{\mathcal{L}^2}\Big]$,
8. $\mathbf{E_8}(\mathbf{Y}^j) := \mathbb{E}\Big[\sum_{k=1}^{K}\|\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\|^2_{\mathbb{R}^L}\cdot\sum_{k=1}^{K}\|\boldsymbol{\sigma}_k(\mathbf{Y}^j)\|_{\mathbb{R}^L}\|D\boldsymbol{\sigma}_k(\mathbf{Y}^j)\|_{\mathcal{L}}\Big]$,

9. $\mathbf{E_9}(\mathbf{Y}^j) := \mathbb{E}\Big[\sum_{k=1}^{K} \Big\|\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big\|_{\mathbb{R}^L} \|\boldsymbol{\sigma}_k(\mathbf{Y}^j)\|_{\mathbb{R}^L}\Big]$,

10. $\mathbf{E_{10}}(\mathbf{Y}^j) := \mathbb{E}\Big[\Big\|\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{Y}^j - \bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{f}(\mathbf{Y}^j)\Big\|^2_{\mathbb{R}^L}\Big]$,

11. $\mathbf{E_{11}}(\mathbf{Y}^j) := \mathbb{E}\Big[\Big|\sum_{k=1}^{K} \Big\|\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big\|^2_{\mathbb{R}^L}\Big|^2\Big]$,

12. $\mathbf{E_{12}}(\mathbf{Y}^j) := \mathbb{E}\Big[\sum_{k=1}^{K} \Big\|\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big\|^2_{\mathbb{R}^L}\Big]$,

13. $\mathbf{E_{13}}(\mathbf{Y}^j) := \mathbb{E}\Big[\Big\|\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{Y}^j - \bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{f}(\mathbf{Y}^j)\Big\|^2_{\mathbb{R}^L}\Big|\sum_{k=1}^{K}\|\boldsymbol{\sigma}_k(\mathbf{Y}^j)\|_{\mathbb{R}^L}\Big|^2\Big]$,

14. $\mathbf{E_{14}}(\mathbf{Y}^j) := \mathbb{E}\Big[\Big|\sum_{k=1}^{K} \Big\|\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big\|^2_{\mathbb{R}^L}\Big|^2\Big|\sum_{k=1}^{K}\|\boldsymbol{\sigma}_k(\mathbf{Y}^j)\|_{\mathbb{R}^L}\Big|^2\Big]$,

15. $\mathbf{E_{15}}(\mathbf{Y}^j) := \mathbb{E}\Big[\sum_{k=1}^{K} \Big\|\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big\|^2_{\mathbb{R}^L}\Big]$.

Remark 3.1. **1.** For $\mathbf{f} \equiv 0$ and/or $\boldsymbol{\sigma}_k$, $k = 1, ..., K$ constant, the estimator $\mathfrak{G}$ simplifies considerably; also, Theorem 3.1 remains valid for ODE systems, *i.e.*, for $\boldsymbol{\sigma}_k \equiv \mathbf{0}$ $(k = 1, ..., K)$, where only terms $\mathbf{E_1}(\cdot)$, $\mathbf{E_3}(\cdot)$ and $\mathbf{E_{10}}(\cdot)$ constitute $\mathfrak{G}$.
For ODE systems (0.1) with $\boldsymbol{\sigma}_k \equiv \mathbf{0}$ $(k = 1, ..., K)$, a different approach to derive a (residual-based) *a posteriori* estimate on a mesh $\{t_j\}_{j=0}^J \subset [0, T]$ for $\|\mathbf{X}_T - \mathbf{Y}^J\|_{\mathbb{R}^L}$ is via duality methods [29], which exploit (strong stability properties of) the related adjoint equation; see also **2.** below. Another variational approach here that avoids duality methods is [67], where an inherited '(discrete) energy dissipation' property of the implicit discretization of (0.1) is used to bound $\max_{1\le j\le J}\|\mathbf{X}_{t_j} - \mathbf{Y}^j\|_{\mathbb{R}^L}$ for cases where the drift operator in (0.1) is the gradient of a convex functional. We also mention [80, Ch. 6], where (residual-based) *a posteriori* estimates are derived by variational methods for space-time discretizations of the more general (0.2) with $\Sigma_k \equiv 0$ $(k = 1, ..., K)$, where the drift operator need not be the gradient of a convex functional.
2. For finite element based discretizations of (linear elliptic, parabolic) PDEs $A(u) = f$, residual-based *a posteriori estimates* are obtained in [29], where dual/adjoint problems are the relevant tool; their (*global*) stability properties may then be exploited to bound the error in terms of the residual $\rho(u_h) = A(u_h) - f$ of the computed solution u_h, times a related stability constant. In later works, dual problems involve functionals ϕ, and its solution z is computed approximately to then enter as *local* weights $\omega(z_h)$ in the 'duality-based weighted residual' estimator of the form

$$|\phi(u) - \phi(u_h)| \le |\big(\rho(u_h), \omega(z_h)\big)| + \text{'higher order terms'}$$

to sharpen computable error bounds; see the surveys [7, 36].
The derivation of *a posteriori* error estimate (3.1) for iterates of (0.3) uses the (backward) *Kolmogorov equation* (0.5) on $[0, t_{n+1}] \times \mathbb{R}^L$ for the transform $u(t, \mathbf{x}) = \mathbb{E}\Big[\phi(\mathbf{X}_{t_{n+1}}^{t,\mathbf{x}})\Big]$ — instead of an adjoint evolutionary problem on $(0, t_{n+1})$ that is motivated from optimal control: the works [76, 64] approximate derivatives of u to build *local* weights contained in $\{\rho_j\}_{j=1}^J$ in (1.1), which is possible for small L; in this work, we use the *global* stability estimate (1.5) that leads to the *a posteriori* error estimator in (3.1) for iterates from (0.3), which is applicable to SDE (0.1) for *arbitrary L*.
3. In [76, 64], (asymptotic) *a posteriori* error expansions for the *terminal time T* are given for both, random and deterministic meshes, while (3.1) bounds the error uniformly in time. The proof of Theorem (3.1) exploits that meshes $\{t_j\}_{j=0}^J \subset [0,T]$ are deterministic, *e.g.*, by repeated use of *Itô's formula*.
4. The *weak Euler method* replaces Wiener increments $\{\Delta_{j+1}\beta_k\}_{j=0}^{J-1}$ in (0.3) by bounded, discrete random variables $\{\widetilde{\xi}_k^{j+1}\sqrt{\tau^{j+1}}\}_{j=0}^{J-1}$ with approximate moments, see *e.g.* [52, p. 458]: for example, $\mathbb{P}\Big[\widetilde{\xi}_k^{j+1} = \pm 1\Big] = \frac{1}{2}$ leads to iterates $\{\widetilde{\mathbf{Y}}^j\}_{j=0}^J$, and their 'continuification' $\widetilde{\boldsymbol{\mathcal{Y}}} \equiv \{\widetilde{\boldsymbol{\mathcal{Y}}}_t;\, t \in [0,T]\}$, given by

$$\widetilde{\boldsymbol{\mathcal{Y}}}_t = \widetilde{\mathbf{Y}}^j + \Big(\bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{f}(\widetilde{\mathbf{Y}}^j) - \mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\widetilde{\mathbf{Y}}^j\Big)(t - t_j) + \sum_{k=1}^K \bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\widetilde{\mathbf{Y}}^j)\widetilde{\xi}_k^{j+1}\sqrt{t - t_j} \qquad \forall\, t \in [t_j, t_{j+1}]\,. \tag{3.2}$$

The *a posteriori* weak error analysis now starts again with (1.3), but lacks *Itô's formula* in (1.4), and thus proceeds with the *mean value theorem* and *Taylor's formula*,

$$\begin{aligned}\mathbb{E}\Big[u(t_{j+1}, \widetilde{\mathbf{Y}}^{j+1}) - u(t_j, \widetilde{\mathbf{Y}}^j)\Big] &= \mathbb{E}\Big[u(t_{j+1}, \widetilde{\mathbf{Y}}^{j+1}) - u(t_j, \widetilde{\mathbf{Y}}^{j+1}) + u(t_j, \widetilde{\mathbf{Y}}^{j+1}) - u(t_j, \widetilde{\mathbf{Y}}^j)\Big] \\ &= \mathbb{E}\Big[\partial_t u(t^*, \widetilde{\mathbf{Y}}^{j+1}) \cdot \tau^{j+1} + \Big\langle D_{\mathbf{x}} u(t_j, \widetilde{\mathbf{Y}}^j), \widetilde{\mathbf{Y}}^{j+1} - \widetilde{\mathbf{Y}}^j \Big\rangle_{\mathbb{R}^L} \\ &\quad + \frac{1}{2}\mathrm{Tr}\Big(D_{\mathbf{x}}^2 u(t_j, \widetilde{\mathbf{Y}}^*)\Big(\widetilde{\mathbf{Y}}^{j+1} - \widetilde{\mathbf{Y}}^j\Big)\Big(\widetilde{\mathbf{Y}}^{j+1} - \widetilde{\mathbf{Y}}^j\Big)^\top\Big)\Big],\end{aligned}$$

where $t^* \in (t_j, t_{j+1})$ and $\widetilde{\mathbf{Y}}^* := \widetilde{\mathbf{Y}}^j + \Theta(\widetilde{\mathbf{Y}}^{j+1} - \widetilde{\mathbf{Y}}^j)$ for some $\Theta \in [0,1]$. A repeated use of (0.5) and (0.3) (in modified form) (*cf.* also (3.2)) then causes changes to the proof of Theorem 3.1: no Malliavin calculus is needed any more for the *weak Euler method* — as is needed to handle terms (3.9) and (3.29); also, higher derivatives of u appear, and further computable terms constitute $\{\widetilde{\mathfrak{G}}\Big(\phi; \tau^{j+1}, \widetilde{\mathbf{Y}}^j, \widetilde{\mathbf{Y}}^{j+1}\Big)\}_{j\geq 0}$ for (3.1).
5. In practice, the terms $\{\mathbf{E}_{\boldsymbol{\ell}}(\cdot)\}_{\boldsymbol{\ell}=1,\ldots,15}$ may be approximated by Monte-Carlo method; see Chapter 5 for more details.

The representation of the *a posteriori* weak error estimator $\mathfrak{G} \equiv \{\mathfrak{G}\Big(\phi; \tau^{j+1}, \mathbf{Y}^j\Big)\}_{j=0}^{J-1}$ in (3.1) involves the bounds (global in time and space) $\boldsymbol{C}_{D^\ell}(\phi)$ ($\ell = 1,2,3$; *cf.* (1.5)), as well as computable error terms $\{\mathbf{E}_{\boldsymbol{\ell}}(\mathbf{Y}^j)\}_{\boldsymbol{\ell},j}$. The matrix $\bar{\boldsymbol{\mathscr{A}}}^{j+1}$, which also arises in the representations of $\{\mathbf{E}_{\boldsymbol{\ell}}(\mathbf{Y}^j)\}_{\boldsymbol{\ell}=1,\ldots,15}$ results from the use of the semi-implicit Euler scheme (0.3).

The proof of Theorem 3.1 consists of several steps: Lemma 3.2 is based on the inequality (1.3) in the introduction, where we represent the weak approximation error via (0.5). Lemma 3.3 then examines the first expectation on the right-hand side of (3.3), where only the drift term of (0.3) appears; and in a similar manner, Lemma 3.4 examines the second expectation on the right-hand side of (3.3), where only the diffusion term is involved. The proof of Theorem 3.1 then follows by combining these lemmas. Note that similar concepts for the investigation of the weak approximation error with the explicit Euler scheme have been proposed in *e.g.* [77], which here are adapted to the implicit scheme (0.3); see also [24], where weak *a priori* error estimates are obtained for a time-implicit discretization of SPDE (0.2) with $\boldsymbol{\beta} \equiv \mathbf{0}$, and Remark 3.2.

Lemma 3.2. Assume **(A1)** – **(A3)**. Let $\{t_j\}_{j=0}^J \subset [0,T]$, with local mesh sizes $\{\tau^{j+1}\}_{j=0}^{J-1}$, and $\{\mathbf{Y}^j\}_{j=0}^J$ solves (0.3). Then, for every $n = 0, ..., J-1$, we have

$$
\begin{aligned}
&\Big|\mathbb{E}\Big[\phi(\mathbf{X}_{t_{n+1}})\Big] - \mathbb{E}\Big[\phi(\mathbf{Y}^{n+1})\Big]\Big| \\
&\leq \sum_{j=0}^{n} \Bigg\{ \Bigg|\mathbb{E}\Bigg[\int_{t_j}^{t_{j+1}} \Big\langle \bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{f}(\mathbf{Y}^j) - \mathbf{f}(\boldsymbol{\mathcal{Y}}_s) + \mathscr{A}\boldsymbol{\mathcal{Y}}_s - \mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{Y}^j, D_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big\rangle_{\mathbb{R}^L}\,\mathrm{d}s\Bigg]\Bigg| \\
&\quad + \frac{1}{2}\Bigg|\sum_{k=1}^{K}\mathbb{E}\Bigg[\int_{t_j}^{t_{j+1}} \mathrm{Tr}\Big(\Big[\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big[\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big]^\top - \boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_s)\boldsymbol{\sigma}_k^\top(\boldsymbol{\mathcal{Y}}_s)\Big] D_{\mathbf{x}}^2 u(s,\boldsymbol{\mathcal{Y}}_s)\Big)\mathrm{d}s\Bigg]\Bigg| \Bigg\}.
\end{aligned} \tag{3.3}
$$

Proof. Fix $n = 0, ..., J-1$ and consider (0.5), where we replace T by t_{n+1}. Note that under assumptions **(A1)** – **(A3)**, the function $u \in \mathcal{C}^{1,3}\Big([0,t_{n+1}] \times \mathbb{R}^L; \mathbb{R}\Big)$ with bounded continuous derivatives *w.r.t.* the state and continuous derivative *w.r.t.* the time, and given by $u(t,\mathbf{x}) = \mathbb{E}\Big[\phi(\mathbf{X}_{t_{n+1}}^{t,\mathbf{x}})\Big]$ is the unique solution of (0.5); see *e.g.* [49, p. 366*ff.*]. Thus, putting $\mathbf{x} = \mathbf{Y}^{n+1}(\omega)$ in the second equation in (0.5) on $[0,t_{n+1}] \times \mathbb{R}^L$, we immediately conclude that

$$
\mathbb{E}\Big[\phi(\mathbf{X}_{t_{n+1}})\Big] = u(0,\mathbf{y}) \quad \text{and} \quad \mathbb{E}\Big[\phi(\mathbf{Y}^{n+1})\Big] = \mathbb{E}\Big[u(t_{n+1},\mathbf{Y}^{n+1})\Big]. \tag{3.4}
$$

Hence, applying (3.4), a first calculation yields (1.3). Since u is the unique solution of (0.5) on $[0,t_{n+1}] \times \mathbb{R}^L$, we use *Itô's formula* with u to (1.2) in (1.3) to deduce

$$
\begin{aligned}
&\mathbb{E}\Big[u(t_{j+1},\mathbf{Y}^{j+1}) - u(t_j,\mathbf{Y}^j)\Big] \\
&\qquad = \mathbb{E}\Bigg[\int_{t_j}^{t_{j+1}} \partial_s u(s,\boldsymbol{\mathcal{Y}}_s) + \Big\langle -\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{Y}^j + \bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{f}(\mathbf{Y}^j), D_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big\rangle_{\mathbb{R}^L} \\
&\qquad\quad + \frac{1}{2}\sum_{k=1}^{K}\mathrm{Tr}\Big(\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big[\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big]^\top D_{\mathbf{x}}^2 u(s,\boldsymbol{\mathcal{Y}}_s)\Big)\,\mathrm{d}s\Bigg].
\end{aligned} \tag{3.5}
$$

Using (0.5) on $[0,t_{n+1}] \times \mathbb{R}^L$ to eliminate $\partial_s u(s,\boldsymbol{\mathcal{Y}}_s)$ in (3.5) further leads to (1.4). Finally, combining (1.3) and (1.4) yields the assertion.

□

The next lemma examines the first expectation appearing on the right-hand side of (1.4).

Lemma 3.3. Suppose the setting in Lemma 3.2. Then we have

$$\begin{aligned}&\bigg|\mathbb{E}\Big[\int_{t_j}^{t_{j+1}}\Big\langle\bar{\mathscr{A}}^{j+1}\mathbf{f}(\mathbf{Y}^j)-\mathbf{f}(\boldsymbol{\mathcal{Y}}_s)+\mathscr{A}\boldsymbol{\mathcal{Y}}_s-\mathscr{A}\bar{\mathscr{A}}^{j+1}\mathbf{Y}^j,D_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big\rangle_{\mathbb{R}^L}\,\mathrm{d}s\Big]\bigg|\\&\leq\bigg\{\tfrac{3\boldsymbol{C}_D(\phi)}{2}\cdot\mathbf{E_1}(\mathbf{Y}^j)+\tfrac{\boldsymbol{C}_{D^2}(\phi)}{2}\cdot\mathbf{E_2}(\mathbf{Y}^j)+\tfrac{\boldsymbol{C}_D(\phi)}{2}\cdot\mathbf{E_3}(\mathbf{Y}^j)+\tfrac{\boldsymbol{C}_D(\phi)}{4}\cdot\mathbf{E_4}(\mathbf{Y}^j)\\&\qquad+\tfrac{\boldsymbol{C}_{D^2}(\phi)}{2}\cdot\mathbf{E_5}(\mathbf{Y}^j)\bigg\}\cdot\Big(\tau^{j+1}\Big)^2\\&\qquad+\bigg\{\Big\{\boldsymbol{C}_D(\phi)C_{D^2\mathbf{f}}\cdot\sqrt{\mathbf{E_{10}}(\mathbf{Y}^j)}+\Big[\tfrac{\boldsymbol{C}_D(\phi)C_{D^3\mathbf{f}}}{2}+\boldsymbol{C}_{D^2}(\phi)C_{D^2\mathbf{f}}\Big]\cdot\sqrt{\mathbf{E_{11}}(\mathbf{Y}^j)}\Big\}\\&\qquad\qquad\cdot\sqrt{\tfrac{\tau^{j+1}}{15}\cdot\mathbf{E_{10}}(\mathbf{Y}^j)+\tfrac{1}{8}\cdot\mathbf{E_{12}}(\mathbf{Y}^j)}\bigg\}\cdot\Big(\tau^{j+1}\Big)^{2.5}.\end{aligned}\tag{3.6}$$

In the proof of Lemma 3.3, we use Lemma 2.5 to (globally) bound the first and second derivative of u by $\boldsymbol{C}_D(\phi)$ and $\boldsymbol{C}_{D^2}(\phi)$, respectively. Besides some standard arguments, we also use Malliavin calculus techniques to validate $\mathcal{O}(|\tau^{j+1}|^2)$ on the right hand-side of (3.6).

Proof. Using the fact that for any $\mathbf{v}\in\mathbb{R}^L$ it holds

$$\bar{\mathscr{A}}^{j+1}\mathbf{v}=\mathbf{v}-\tau^{j+1}\mathscr{A}\bar{\mathscr{A}}^{j+1}\mathbf{v},$$

we obtain in a first calculation

$$\begin{aligned}&\Big\langle\bar{\mathscr{A}}^{j+1}\mathbf{f}(\mathbf{Y}^j)-\mathbf{f}(\boldsymbol{\mathcal{Y}}_s)+\mathscr{A}\boldsymbol{\mathcal{Y}}_s-\mathscr{A}\bar{\mathscr{A}}^{j+1}\mathbf{Y}^j,D_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big\rangle_{\mathbb{R}^L}\\&=\tau^{j+1}\cdot\Big\langle\mathscr{A}^2\bar{\mathscr{A}}^{j+1}\mathbf{Y}^j-\mathscr{A}\bar{\mathscr{A}}^{j+1}\mathbf{f}(\mathbf{Y}^j),D_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big\rangle_{\mathbb{R}^L}+\Big\langle\mathscr{A}\boldsymbol{\mathcal{Y}}_s-\mathscr{A}\mathbf{Y}^j,D_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big\rangle_{\mathbb{R}^L}\\&\quad+\Big\langle\mathbf{f}(\mathbf{Y}^j)-\mathbf{f}(\boldsymbol{\mathcal{Y}}_s),D_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big\rangle_{\mathbb{R}^L}.\end{aligned}\tag{3.7}$$

We use this identity and Lemma 2.5 to bound the left-hand side of (3.6),

$$\begin{aligned}&\bigg|\mathbb{E}\Big[\int_{t_j}^{t_{j+1}}\Big\langle\bar{\mathscr{A}}^{j+1}\mathbf{f}(\mathbf{Y}^j)-\mathbf{f}(\boldsymbol{\mathcal{Y}}_s)+\mathscr{A}\boldsymbol{\mathcal{Y}}_s-\mathscr{A}\bar{\mathscr{A}}^{j+1}\mathbf{Y}^j,D_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big\rangle_{\mathbb{R}^L}\mathrm{d}s\Big]\bigg|\\&\leq\boldsymbol{C}_D(\phi)\cdot\mathbf{E_1}(\mathbf{Y}^j)\cdot\Big(\tau^{j+1}\Big)^2+\bigg|\mathbb{E}\Big[\int_{t_j}^{t_{j+1}}\Big\langle\mathscr{A}\boldsymbol{\mathcal{Y}}_s-\mathscr{A}\mathbf{Y}^j,D_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big\rangle_{\mathbb{R}^L}\,\mathrm{d}s\Big]\bigg|\\&\quad+\bigg|\mathbb{E}\Big[\int_{t_j}^{t_{j+1}}\Big\langle\mathbf{f}(\boldsymbol{\mathcal{Y}}_s)-\mathbf{f}(\mathbf{Y}^j),D_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big\rangle_{\mathbb{R}^L}\mathrm{d}s\Big]\bigg|\\&=:\boldsymbol{I}+\boldsymbol{II}+\boldsymbol{III}.\end{aligned}\tag{3.8}$$

We estimate the terms in (3.8) independently, starting with
Step **1:** (Estimation of $\boldsymbol{III}$) **a)** We apply *Itô's formula* with $\{f_i;\,1\leq i\leq L\}$ to (1.2) to get

$$\bigg|\mathbb{E}\Big[\int_{t_j}^{t_{j+1}}\Big\langle\mathbf{f}(\boldsymbol{\mathcal{Y}}_s)-\mathbf{f}(\mathbf{Y}^j),D_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big\rangle_{\mathbb{R}^L}\mathrm{d}s\Big]\bigg|$$

$$= \Bigg| \sum_{i=1}^{L} \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \Big\{ f_i(\boldsymbol{\mathcal{Y}}_s) - f_i(\mathbf{Y}^j) \Big\} \partial_{x_i} u(s, \boldsymbol{\mathcal{Y}}_s) \, \mathrm{d}s \Big] \Bigg|$$

$$\leq \Bigg| \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \int_{t_j}^{s} \Big\langle D\mathbf{f}(\boldsymbol{\mathcal{Y}}_r) \Big(-\mathscr{A} \bar{\boldsymbol{\mathscr{A}}}^{j+1} \mathbf{Y}^j + \bar{\boldsymbol{\mathscr{A}}}^{j+1} \mathbf{f}(\mathbf{Y}^j) \Big), D_{\mathbf{x}} u(s, \boldsymbol{\mathcal{Y}}_s)) \Big\rangle_{\mathbb{R}^L} \mathrm{d}r \, \mathrm{d}s \Big] \Bigg|$$

$$+ \frac{1}{2} \sum_{k=1}^{K} \Bigg| \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \int_{t_j}^{s} \Big\langle D^2\mathbf{f}(\boldsymbol{\mathcal{Y}}_r) \bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_k(\mathbf{Y}^j) \bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_k(\mathbf{Y}^j), D_{\mathbf{x}} u(s, \boldsymbol{\mathcal{Y}}_s)) \Big\rangle_{\mathbb{R}^L} \mathrm{d}r \, \mathrm{d}s \Big] \Bigg| + \boldsymbol{M_1},$$

where

$$\boldsymbol{M_1} := \Bigg| \sum_{i=1}^{L} \sum_{k=1}^{K} \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \int_{t_j}^{s} \Big\langle Df_i(\boldsymbol{\mathcal{Y}}_r), \bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_k(\mathbf{Y}^j) \, \mathrm{d}\beta_k(r) \Big\rangle_{\mathbb{R}^L} \mathrm{d}s \cdot \partial_{x_i} u(s, \boldsymbol{\mathcal{Y}}_s) \Big] \Bigg| := \Bigg| \sum_{i=1}^{L} M_{1,i} \Bigg|. \tag{3.9}$$

We add and substract $D\mathbf{f}(\mathbf{Y}^j)$ and $D^2\mathbf{f}(\mathbf{Y}^j)$ as integrands in order to get closer to computable optimal terms. This step leads to the additional terms

$$\boldsymbol{K_1} := \Bigg| \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \int_{t_j}^{s} \Big\langle \Big\{ D\mathbf{f}(\boldsymbol{\mathcal{Y}}_r) - D\mathbf{f}(\mathbf{Y}^j) \Big\} \Big(-\mathscr{A} \bar{\boldsymbol{\mathscr{A}}}^{j+1} \mathbf{Y}^j + \bar{\boldsymbol{\mathscr{A}}}^{j+1} \mathbf{f}(\mathbf{Y}^j) \Big), D_{\mathbf{x}} u(s, \boldsymbol{\mathcal{Y}}_s)) \Big\rangle_{\mathbb{R}^L} \mathrm{d}r \, \mathrm{d}s \Big] \Bigg|$$

and

$$\boldsymbol{K_2} := \sum_{k=1}^{K} \Bigg| \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \int_{t_j}^{s} \langle \{ D^2\mathbf{f}(\boldsymbol{\mathcal{Y}}_r) - D^2\mathbf{f}(\mathbf{Y}^j) \} \bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_k(\mathbf{Y}^j) \bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_k(\mathbf{Y}^j), D_{\mathbf{x}} u(s, \boldsymbol{\mathcal{Y}}_s)) \rangle_{\mathbb{R}^L} \mathrm{d}r \, \mathrm{d}s \Big] \Bigg|,$$

which will be estimated in **c)** below. Thus, we obtain

$$\Bigg| \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \Big\langle \mathbf{f}(\boldsymbol{\mathcal{Y}}_s) - \mathbf{f}(\mathbf{Y}^j), D_{\mathbf{x}} u(s, \boldsymbol{\mathcal{Y}}_s) \Big\rangle_{\mathbb{R}^L} \mathrm{d}s \Big] \Bigg|$$

$$\leq \Bigg| \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \int_{t_j}^{s} \Big\langle D\mathbf{f}(\mathbf{Y}^j) \Big(-\mathscr{A} \bar{\boldsymbol{\mathscr{A}}}^{j+1} \mathbf{Y}^j + \bar{\boldsymbol{\mathscr{A}}}^{j+1} \mathbf{f}(\mathbf{Y}^j) \Big), D_{\mathbf{x}} u(s, \boldsymbol{\mathcal{Y}}_s) \Big\rangle_{\mathbb{R}^L} \mathrm{d}r \, \mathrm{d}s \Big] \Bigg|$$

$$+ \frac{1}{2} \sum_{k=1}^{K} \Bigg| \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \int_{t_j}^{s} \Big\langle D^2\mathbf{f}(\mathbf{Y}^j) \bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_k(\mathbf{Y}^j) \bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_k(\mathbf{Y}^j), D_{\mathbf{x}} u(s, \boldsymbol{\mathcal{Y}}_s) \Big\rangle_{\mathbb{R}^L} \mathrm{d}r \, \mathrm{d}s \Big] \Bigg|$$

$$+ \boldsymbol{M_1} + \boldsymbol{K_1} + \frac{1}{2} \boldsymbol{K_2} \, .$$

We apply the *Cauchy-Schwarz inequality*, Lemma 2.5 (i), and some standard calculations to obtain

$$\Bigg| \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \Big\langle \mathbf{f}(\boldsymbol{\mathcal{Y}}_s) - \mathbf{f}(\mathbf{Y}^j), D_{\mathbf{x}} u(s, \boldsymbol{\mathcal{Y}}_s) \Big\rangle_{\mathbb{R}^L} \mathrm{d}s \Big] \Bigg|$$

$$\leq \Bigg\{ \frac{\boldsymbol{C}_D(\phi)}{2} \cdot \mathbf{E_3}(\mathbf{Y}^j) + \frac{\boldsymbol{C}_D(\phi)}{4} \cdot \mathbf{E_4}(\mathbf{Y}^j) \Bigg\} \cdot \Big(\tau^{j+1} \Big)^2 + \boldsymbol{M_1} + \boldsymbol{K_1} + \frac{1}{2} \boldsymbol{K_2}. \tag{3.10}$$

We estimate the terms $\boldsymbol{M_1}, \boldsymbol{K_1}, \boldsymbol{K_2}$ independently in parts **b)** and **c)**.
b) We consider $\boldsymbol{M_1}$ in (3.9): for its successul treatment, we use tools from Malliavin calculus.

For $i = 1, ..., L$, we have

$$M_{1,i} = \sum_{k=1}^{K}\sum_{l=1}^{L}\mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \partial_{x_i}u(s,\boldsymbol{\mathcal{Y}}_s)\int_{t_j}^{s}\partial_{x_l}f_i(\boldsymbol{\mathcal{Y}}_r)\Big(\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big)_l \mathrm{d}\beta_k(r)\,\mathrm{d}s\Big] \tag{3.11}$$

Since $\partial_{x_i}u(s,\boldsymbol{\mathcal{Y}}_s) \in \mathbb{D}^{1,2}$ (see *e.g.* [1, 68]), we apply the *Clark-Ocone formula* in Lemma 2.1 to $\partial_{x_i}u(s,\boldsymbol{\mathcal{Y}}_s)$, to get

$$\partial_{x_i}u(s,\boldsymbol{\mathcal{Y}}_s) = \mathbb{E}\Big[\partial_{x_i}u(s,\boldsymbol{\mathcal{Y}}_s)\Big] + \int_{t_j}^{s}\mathbb{E}\Big[\mathrm{D}_r^{(k)}\Big(\partial_{x_i}u(s,\boldsymbol{\mathcal{Y}}_s)\Big)|\mathcal{F}_r\Big]\,\mathrm{d}\beta_k(r), \tag{3.12}$$

where $\mathrm{D}_r^{(k)}$ denotes the Malliavin derivative *w.r.t.* $\beta_k(r)$. Applying (2.1) to $\mathrm{D}_r^{(k)}\Big(\partial_{x_i}u(s,\boldsymbol{\mathcal{Y}}_s)\Big)$ further leads to

$$\mathrm{D}_r^{(k)}\Big(\partial_{x_i}u(s,\boldsymbol{\mathcal{Y}}_s)\Big) = \sum_{m=1}^{L}\partial_{x_m}\Big(\partial_{x_i}u(s,\boldsymbol{\mathcal{Y}}_s)\Big)\Big(\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big)_m. \tag{3.13}$$

Now, inserting (3.12) into (3.11) and using the fact that the expectation of a stochastic integral *w.r.t.* the Wiener process is zero, we conclude

$$\begin{aligned} &M_{1,i} \\ &= \sum_{k=1}^{K}\sum_{l=1}^{L}\mathbb{E}\Big[\int_{t_j}^{t_{j+1}}\int_{t_j}^{s}\mathbb{E}\Big[\mathrm{D}_r^{(k)}\Big(\partial_{x_i}u(s,\boldsymbol{\mathcal{Y}}_s)\Big)|\mathcal{F}_r\Big]\,\mathrm{d}\beta_k(r)\int_{t_j}^{s}\partial_{x_l}f_i(\boldsymbol{\mathcal{Y}}_r)\Big(\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big)_l\mathrm{d}\beta_k(r)\,\mathrm{d}s\Big] \\ &= \sum_{k=1}^{K}\sum_{l=1}^{L}\int_{t_j}^{t_{j+1}}\mathbb{E}\Big[\int_{t_j}^{s}\mathbb{E}\Big[\mathrm{D}_r^{(k)}\Big(\partial_{x_i}u(s,\boldsymbol{\mathcal{Y}}_s)\Big)|\mathcal{F}_r\Big]\partial_{x_l}f_i(\boldsymbol{\mathcal{Y}}_r)\Big(\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big)_l\mathrm{d}r\Big]\,\mathrm{d}s\,, \end{aligned} \tag{3.14}$$

where in the last step we used a (generalized) Itô isometry argument; see *e.g.* [37, p. 135, Thm. 4.2.3]. An application of the tower property (law of total expectation) in (3.14), and (3.13) then lead to

$$\begin{aligned} \boldsymbol{M_1} &= \Big|\sum_{k=1}^{K}\sum_{i,\ell=1}^{L}\int_{t_j}^{t_{j+1}}\int_{t_j}^{s}\mathbb{E}\Big[\Big(\sum_{m=1}^{L}\partial_{x_m}\big(\partial_{x_i}u(s,\boldsymbol{\mathcal{Y}}_s)\big)\big(\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big)_m\Big)\partial_{x_l}f_i(\boldsymbol{\mathcal{Y}}_r)\big(\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big)_l\Big]\,\mathrm{d}r\,\mathrm{d}s\Big| \\ &= \Big|\sum_{k=1}^{K}\mathbb{E}\Big[\int_{t_j}^{t_{j+1}}\int_{t_j}^{s}\mathrm{Tr}\Big(D_{\mathbf{x}}^2u(s,\boldsymbol{\mathcal{Y}}_s)D\mathbf{f}(\boldsymbol{\mathcal{Y}}_r)\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big[\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big]^\top\Big)\,\mathrm{d}r\,\mathrm{d}s\Big]\Big|. \end{aligned}$$

In the next step, we add and substract in the second argument $D\mathbf{f}(\mathbf{Y}^j)$ as well, to then obtain

$$\begin{aligned} \boldsymbol{M_1} \leq &\Big|\sum_{k=1}^{K}\mathbb{E}\Big[\int_{t_j}^{t_{j+1}}\int_{t_j}^{s}\mathrm{Tr}\Big(D_{\mathbf{x}}^2u(s,\boldsymbol{\mathcal{Y}}_s)D\mathbf{f}(\mathbf{Y}^j)\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big[\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big]^\top\Big)\,\mathrm{d}r\,\mathrm{d}s\Big]\Big| \\ &+ \boldsymbol{K_3} \end{aligned}$$

where

$$\boldsymbol{K_3} := \Big|\sum_{k=1}^{K}\mathbb{E}\Big[\int_{t_j}^{t_{j+1}}\int_{t_j}^{s}\mathrm{Tr}\Big(D_{\mathbf{x}}^2u(s,\boldsymbol{\mathcal{Y}}_s)\{D\mathbf{f}(\boldsymbol{\mathcal{Y}}_r) - D\mathbf{f}(\mathbf{Y}^j)\}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big[\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big]^\top\Big)\,\mathrm{d}r\,\mathrm{d}s\Big]\Big|.$$

Using Lemma 2.5 (ii), and $\mathrm{Tr}(\mathbf{B}\mathbf{v}\mathbf{w}^\top) \leq \|\mathbf{B}\|_{\mathbb{R}^{L\times L}}\|\mathbf{v}\|_{\mathbb{R}^L}\|\mathbf{w}\|_{\mathbb{R}^L}$ for any $\mathbf{B} \in \mathbb{R}^{L\times L}$, $\mathbf{v}, \mathbf{w} \in \mathbb{R}^L$, consequently leads to

$$\boldsymbol{M_1} \leq \frac{\boldsymbol{C}_{D^2(\phi)}}{2} \cdot \mathsf{E}_{\mathbf{5}}(\mathbf{Y}^j) \cdot \left(\tau^{j+1}\right)^2 + \boldsymbol{K_3}\,. \tag{3.15}$$

c) We show that the terms $\boldsymbol{K_1}, \boldsymbol{K_2}, \boldsymbol{K_3}$ are the higher order terms in (3.5), which account for the difference between $\boldsymbol{\mathcal{Y}}$ and $\{\mathbf{Y}^j\}_{j\geq 0}$. Since the treatment of $\boldsymbol{K_1}, \boldsymbol{K_2}$ and $\boldsymbol{K_3}$ is similar, we only consider $\boldsymbol{K_1}$ in detail. We start with a standard calculation, that is for $s \in [t_j, t_{j+1}]$ and $r \in [t_j, s]$ we have, considering the continuified process (1.2) and using some standard calculations

$$\mathbb{E}\Big[\|\boldsymbol{\mathcal{Y}}_r - \mathbf{Y}^j\|^2_{\mathbb{R}^L}\Big] \leq \mathsf{E}_{\mathbf{10}}(\mathbf{Y}^j) \cdot (r - t_j)^2 + \mathsf{E}_{\mathbf{12}}(\mathbf{Y}^j) \cdot (r - t_j)\,. \tag{3.16}$$

Using

$$\|D\mathbf{f}(\boldsymbol{\mathcal{Y}}_r) - D\mathbf{f}(\mathbf{Y}^j)\|_{\mathcal{L}} \leq C_{D^2\mathbf{f}} \cdot \|\boldsymbol{\mathcal{Y}}_r - \mathbf{Y}^j\|_{\mathbb{R}^L}\,,$$

the *Cauchy-Schwarz inequality* and Lemma 2.5 (i), we get

$$\boldsymbol{K_1} \leq \boldsymbol{C}_D(\phi) C_{D^2\mathbf{f}} \cdot \sqrt{\mathsf{E}_{\mathbf{10}}(\mathbf{Y}^j)} \cdot \left(\tau^{j+1} \int_{t_j}^{t_{j+1}} (s - t_j) \int_{t_j}^{s} \mathbb{E}\Big[\|\boldsymbol{\mathcal{Y}}_r - \mathbf{Y}^j\|^2_{\mathbb{R}^L}\Big]\,\mathrm{d}r\,\mathrm{d}s\right)^{\frac{1}{2}}.$$

Using (3.16) consequently leads to

$$\boldsymbol{K_1} \leq \boldsymbol{C}_D(\phi) C_{D^2\mathbf{f}} \cdot \sqrt{\mathsf{E}_{\mathbf{10}}(\mathbf{Y}^j)} \cdot \sqrt{\tfrac{\tau^{j+1}}{15} \cdot \mathsf{E}_{\mathbf{10}}(\mathbf{Y}^j) + \tfrac{1}{8} \cdot \mathsf{E}_{\mathbf{12}}(\mathbf{Y}^j)} \cdot \left(\tau^{j+1}\right)^{2.5}. \tag{3.17}$$

In a similar way, we obtain

$$\boldsymbol{K_2} \leq \boldsymbol{C}_D(\phi) C_{D^3\mathbf{f}} \cdot \sqrt{\mathsf{E}_{\mathbf{11}}(\mathbf{Y}^j)} \cdot \sqrt{\tfrac{\tau^{j+1}}{15} \cdot \mathsf{E}_{\mathbf{10}}(\mathbf{Y}^j) + \tfrac{1}{8} \cdot \mathsf{E}_{\mathbf{12}}(\mathbf{Y}^j)} \cdot \left(\tau^{j+1}\right)^{2.5} \tag{3.18}$$

and

$$\boldsymbol{K_3} \leq \boldsymbol{C}_{D^2}(\phi) C_{D^2\mathbf{f}} \cdot \sqrt{\mathsf{E}_{\mathbf{11}}(\mathbf{Y}^j)} \cdot \sqrt{\tfrac{\tau^{j+1}}{15} \cdot \mathsf{E}_{\mathbf{10}}(\mathbf{Y}^j) + \tfrac{1}{8} \cdot \mathsf{E}_{\mathbf{12}}(\mathbf{Y}^j)} \cdot \left(\tau^{j+1}\right)^{2.5}. \tag{3.19}$$

Step **2:** (Estimation of $\boldsymbol{II}$) Similar arguments as used for $\boldsymbol{III}$ in Step **1** give the bound

$$\boldsymbol{II} \leq \left\{\frac{\boldsymbol{C}_{D(\phi)}}{2} \cdot \mathsf{E}_{\mathbf{1}}(\mathbf{Y}^j) + \frac{\boldsymbol{C}_{D^2(\phi)}}{2} \cdot \mathsf{E}_{\mathbf{2}}(\mathbf{Y}^j)\right\} \cdot \left(\tau^{j+1}\right)^2. \tag{3.20}$$

Step **3:** (Finishing the proof) We combine (3.17), (3.18), (3.19) and (3.15) with (3.10) and plug the resulting expression as well as (3.20) into (3.8), which proves the assertion.

□

We now bound the last sum on the right hand-side of (3.3).

Lemma 3.4. Suppose the setting in Lemma 3.2. Then we have

$$\begin{aligned}
&\Bigg|\sum_{k=1}^{K}\mathbb{E}\bigg[\int_{t_j}^{t_{j+1}}\mathrm{Tr}\Big(\Big[\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big[\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big]^\top-\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_s)\boldsymbol{\sigma}_k^\top(\boldsymbol{\mathcal{Y}}_s)\Big]D_{\mathbf{x}}^2u(s,\boldsymbol{\mathcal{Y}}_s)\Big)\mathrm{d}s\bigg]\Bigg|\\
&\leq\bigg\{\boldsymbol{C}_{D^2}(\phi)\cdot\mathbf{E_6}(\mathbf{Y}^j)+\frac{\boldsymbol{C}_{D^2}(\phi)}{2}\cdot\mathbf{E_7}(\mathbf{Y}^j)+\boldsymbol{C}_{D^3}(\phi)\cdot\mathbf{E_8}(\mathbf{Y}^j)\\
&\qquad+2\boldsymbol{C}_{D^2}(\phi)\cdot\mathbf{E_9}(\mathbf{Y}^j)+\boldsymbol{C}_{D^2}(\phi)C_{D\boldsymbol{\sigma}}^2\cdot\Big[\frac{\tau^{j+1}}{3}\cdot\mathbf{E_{10}}(\mathbf{Y}^j)+\frac{1}{2}\cdot\mathbf{E_{12}}(\mathbf{Y}^j)\Big]\bigg\}\cdot\big(\tau^{j+1}\big)^2\\
&\quad+\bigg\{\Big\{2\boldsymbol{C}_{D^2}(\phi)C_{D^2\boldsymbol{\sigma}}\cdot\sqrt{\mathbf{E_{13}}(\mathbf{Y}^j)}+\Big[\boldsymbol{C}_{D^2}(\phi)C_{D^3\boldsymbol{\sigma}}+2\boldsymbol{C}_{D^3}(\phi)C_{D^2\boldsymbol{\sigma}}\Big]\cdot\sqrt{\mathbf{E_{14}}(\mathbf{Y}^j)}\Big\}\\
&\qquad\qquad\cdot\sqrt{\frac{\tau^{j+1}}{15}\cdot\mathbf{E_{10}}(\mathbf{Y}^j)+\frac{1}{8}\cdot\mathbf{E_{12}}(\mathbf{Y}^j)}\bigg\}\cdot\big(\tau^{j+1}\big)^{2.5}\\
&\qquad+\boldsymbol{C}_{D^2}(\phi)\cdot\mathbf{E_{15}}(\mathbf{Y}^j)\cdot\big(\tau^{j+1}\big)^3.
\end{aligned}\tag{3.21}$$

Proof. Similar as in (3.7), we start with a straightforward calculation. For $k=1,...,K$, we have

$$\begin{aligned}
&\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big[\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big]^\top-\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_s)\boldsymbol{\sigma}_k^\top(\boldsymbol{\mathcal{Y}}_s)\\
&\quad=\boldsymbol{\sigma}_k(\mathbf{Y}^j)\boldsymbol{\sigma}_k^\top(\mathbf{Y}^j)-\tau^{j+1}\cdot\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big[\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big]^\top-\tau^{j+1}\cdot\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\boldsymbol{\sigma}_k^\top(\mathbf{Y}^j)\\
&\qquad+\big(\tau^{j+1}\big)^2\cdot\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big[\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big]^\top-\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_s)\boldsymbol{\sigma}_k^\top(\boldsymbol{\mathcal{Y}}_s)\,.
\end{aligned}$$

This yields

$$\begin{aligned}
&\mathrm{Tr}\Big(\Big[\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big[\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big]^\top-\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_s)\boldsymbol{\sigma}_k^\top(\boldsymbol{\mathcal{Y}}_s)\Big]D_{\mathbf{x}}^2u(s,\boldsymbol{\mathcal{Y}}_s)\Big)\\
&\quad=\mathrm{Tr}\Big(\Big[\boldsymbol{\sigma}_k(\mathbf{Y}^j)\boldsymbol{\sigma}_k^\top(\mathbf{Y}^j)-\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_s)\boldsymbol{\sigma}_k^\top(\boldsymbol{\mathcal{Y}}_s)\Big]D_{\mathbf{x}}^2u(s,\boldsymbol{\mathcal{Y}}_s)\Big)\\
&\qquad-\bigg(\mathrm{Tr}\Big(\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big[\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big]^\top D_{\mathbf{x}}^2u(s,\boldsymbol{\mathcal{Y}}_s)\Big)\\
&\qquad\qquad+\mathrm{Tr}\Big(\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\boldsymbol{\sigma}_k^\top(\mathbf{Y}^j)D_{\mathbf{x}}^2u(s,\boldsymbol{\mathcal{Y}}_s)\Big)\bigg)\cdot\tau^{j+1}\\
&\qquad+\mathrm{Tr}\Big(\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big[\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\big]^\top D_{\mathbf{x}}^2u(s,\boldsymbol{\mathcal{Y}}_s)\Big)\cdot\big(\tau^{j+1}\big)^2.
\end{aligned}\tag{3.22}$$

We set

$$\boldsymbol{T_1}:=\Bigg|\sum_{k=1}^{K}\mathbb{E}\bigg[\int_{t_j}^{t_{j+1}}\mathrm{Tr}\Big(\Big[\boldsymbol{\sigma}_k(\mathbf{Y}^j)\boldsymbol{\sigma}_k^\top(\mathbf{Y}^j)-\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_s)\boldsymbol{\sigma}_k^\top(\boldsymbol{\mathcal{Y}}_s)\Big]D_{\mathbf{x}}^2u(s,\boldsymbol{\mathcal{Y}}_s)\Big)\,\mathrm{d}s\bigg]\Bigg|$$

and plug (3.22) into the left-hand side of (3.21) to deduce

$$\begin{aligned}
&\Bigg|\sum_{k=1}^{K}\mathbb{E}\Bigg[\int_{t_j}^{t_{j+1}}\mathrm{Tr}\Big(\Big[\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big[\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big]^\top-\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_s)\boldsymbol{\sigma}_k^\top(\boldsymbol{\mathcal{Y}}_s)\Big]D^2_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big)\mathrm{d}s\Bigg]\Bigg|\\
&\leq\Bigg|\sum_{k=1}^{K}\mathbb{E}\Bigg[\int_{t_j}^{t_{j+1}}\mathrm{Tr}\Big(\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big[\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big]^\top D^2_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\\
&\qquad\qquad+\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\boldsymbol{\sigma}_k^\top(\mathbf{Y}^j)D^2_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big)\,\mathrm{d}s\Bigg]\Bigg|\cdot\tau^{j+1}\\
&\quad+\Bigg|\sum_{k=1}^{K}\mathbb{E}\Bigg[\int_{t_j}^{t_{j+1}}\mathrm{Tr}\Big(\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big[\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big]^\top D^2_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big)\,\mathrm{d}s\Bigg]\Bigg|\cdot\Big(\tau^{j+1}\Big)^2+\boldsymbol{T_1}\\
&=:\boldsymbol{IV}+\boldsymbol{V}+\boldsymbol{T_1}.
\end{aligned}$$

Step **1:** (Estimation of $\boldsymbol{IV}$) Some standard calculations and Lemma 2.5 (ii) lead to

$$\boldsymbol{IV}\leq 2\boldsymbol{C}_{D^2}(\phi)\cdot\mathbf{E_9}(\mathbf{Y}^j)\cdot\Big(\tau^{j+1}\Big)^2. \tag{3.23}$$

Step **2:** (Estimation of $\boldsymbol{V}$) Again, standard calculations and Lemma 2.5 (ii), lead to

$$\boldsymbol{V}\leq\boldsymbol{C}_{D^2}(\phi)\cdot\mathbf{E_{15}}(\mathbf{Y}^j)\cdot\Big(\tau^{j+1}\Big)^3. \tag{3.24}$$

Step **3:** (Estimation of $\boldsymbol{T_1}$) **a)** For $k=1,...,K$, we add and substract $\boldsymbol{\sigma}_k(\mathbf{Y}^j)\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_s)$ and use $\mathrm{Tr}(\mathbf{B})=\mathrm{Tr}(\mathbf{B}^\top)$ for $\mathbf{B}\in\mathbb{R}^{L\times L}$ to obtain

$$\begin{aligned}
&\mathrm{Tr}\Big(\Big[\boldsymbol{\sigma}_k(\mathbf{Y}^j)\boldsymbol{\sigma}_k^\top(\mathbf{Y}^j)-\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_s)\boldsymbol{\sigma}_k^\top(\boldsymbol{\mathcal{Y}}_s)\Big]D^2_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big)\\
&\quad=2\cdot\mathrm{Tr}\Big(\Big[\boldsymbol{\sigma}_k(\mathbf{Y}^j)-\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_s)\Big]\boldsymbol{\sigma}_k^\top(\mathbf{Y}^j)D^2_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big)\\
&\qquad-\mathrm{Tr}\Big(\Big[\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_s)-\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big]\Big[\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_s)-\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big]^\top D^2_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big).
\end{aligned}\tag{3.25}$$

Plugging (3.25) into $\boldsymbol{T_1}$ immediately leads to

$$\boldsymbol{T_1}\leq 2\cdot\boldsymbol{T_{1,a}}+\boldsymbol{T_{1,b}}\,, \tag{3.26}$$

where

$$\boldsymbol{T_{1,a}}:=\Bigg|\sum_{k=1}^{K}\mathbb{E}\Bigg[\int_{t_j}^{t_{j+1}}\mathrm{Tr}\Big(\Big[\boldsymbol{\sigma}_k(\mathbf{Y}^j)-\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_s)\Big]\boldsymbol{\sigma}_k^\top(\mathbf{Y}^j)D^2_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big)\,\mathrm{d}s\Bigg]\Bigg|$$

and

$$\boldsymbol{T_{1,b}}:=\Bigg|\sum_{k=1}^{K}\mathbb{E}\Bigg[\int_{t_j}^{t_{j+1}}\mathrm{Tr}\Big(\Big[\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_s)-\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big]\Big[\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_s)-\boldsymbol{\sigma}_k(\mathbf{Y}^j)\Big]^\top D^2_{\mathbf{x}}u(s,\boldsymbol{\mathcal{Y}}_s)\Big)\,\mathrm{d}s\Bigg]\Bigg|.$$

b) We estimate $\boldsymbol{T_{1,b}}$ with the help of Lemma 2.5 (ii),

$$\boldsymbol{T_{1,b}}\leq\boldsymbol{C}_{D^2}(\phi)\cdot\mathbb{E}\Bigg[\int_{t_j}^{t_{j+1}}\sum_{k=1}^{K}\|\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_s)-\boldsymbol{\sigma}_k(\mathbf{Y}^j)\|^2_{\mathbb{R}^L}\,\mathrm{d}s\Bigg].$$

Assumption **(A2)** and (3.16) then lead to

$$\boldsymbol{T_{1,b}} \leq \boldsymbol{C}_{D^2}(\phi) C^2_{D\boldsymbol{\sigma}} \cdot \left\{ \tfrac{\tau^{j+1}}{3} \cdot \mathbf{E_{10}}(\mathbf{Y}^j) + \tfrac{1}{2} \cdot \mathbf{E_{12}}(\mathbf{Y}^j) \right\} \cdot \left(\tau^{j+1}\right)^2. \tag{3.27}$$

c) Next, we consider the expression $\boldsymbol{T_{1,a}}$. We apply *Itô's formula* with $\{\sigma_k^{(i)};\, 1 \leq k \leq K, 1 \leq i \leq L\}$ to (1.2) and use standard arguments to get

$$\begin{aligned} \boldsymbol{T_{1,a}} &= \Bigg| \sum_{k=1}^{K} \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \sum_{i=1}^{L} \Big\langle D_{\mathbf{x}} \partial_{x_i} u(s, \boldsymbol{\mathcal{Y}}_s), \boldsymbol{\sigma}_k(\mathbf{Y}^j) \Big\rangle_{\mathbb{R}^L} \Big\{ \sigma_k^{(i)}(\boldsymbol{\mathcal{Y}}_s) - \sigma_k^{(i)}(\mathbf{Y}^j) \Big\} \,\mathrm{d}s \Big] \Bigg| \\ &\leq \boldsymbol{T_{1,a,1}} + \boldsymbol{T_{1,a,2}} + \boldsymbol{M_2}\,, \end{aligned} \tag{3.28}$$

where

$$\begin{aligned} \boldsymbol{T_{1,a,1}} :=& \Bigg| \sum_{k=1}^{K} \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \sum_{i=1}^{L} \Big\langle D_{\mathbf{x}} \partial_{x_i} u(s, \boldsymbol{\mathcal{Y}}_s), \boldsymbol{\sigma}_k(\mathbf{Y}^j) \Big\rangle_{\mathbb{R}^L} \\ & \cdot \int_{t_j}^{s} \Big\langle \bar{\boldsymbol{\mathscr{A}}}^{j+1} \mathbf{f}(\mathbf{Y}^j) - \mathscr{A} \bar{\boldsymbol{\mathscr{A}}}^{j+1} \mathbf{Y}^j, D\sigma_k^{(i)}(\boldsymbol{\mathcal{Y}}_r) \Big\rangle_{\mathbb{R}^L} \,\mathrm{d}r \,\mathrm{d}s \Big] \Bigg| \\ =& \Bigg| \sum_{k=1}^{K} \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \int_{t_j}^{s} \Big\langle D^2_{\mathbf{x}} u(s, \boldsymbol{\mathcal{Y}}_s) \boldsymbol{\sigma}_k(\mathbf{Y}^j), D\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_r) \Big(\mathscr{A} \bar{\boldsymbol{\mathscr{A}}}^{j+1} \mathbf{Y}^j - \bar{\boldsymbol{\mathscr{A}}}^{j+1} \mathbf{f}(\mathbf{Y}^j) \Big) \Big\rangle_{\mathbb{R}^L} \,\mathrm{d}r \,\mathrm{d}s \Big] \Bigg|, \end{aligned}$$

$$\begin{aligned} \boldsymbol{T_{1,a,2}} :=& \frac{1}{2} \Bigg| \sum_{k,l=1}^{K} \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \sum_{i=1}^{L} \Big\langle D_{\mathbf{x}} \partial_{x_i} u(s, \boldsymbol{\mathcal{Y}}_s), \boldsymbol{\sigma}_k(\mathbf{Y}^j) \Big\rangle_{\mathbb{R}^L} \\ & \cdot \int_{t_j}^{s} \mathrm{Tr}\Big(\bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_l(\mathbf{Y}^j) \Big[\bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_l(\mathbf{Y}^j) \Big]^\top D^2 \sigma_k^{(i)}(\boldsymbol{\mathcal{Y}}_r) \Big) \,\mathrm{d}r \,\mathrm{d}s \Big] \Bigg| \\ =& \frac{1}{2} \Bigg| \sum_{k,l=1}^{K} \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \int_{t_j}^{s} \Big\langle D^2_{\mathbf{x}} u(s, \boldsymbol{\mathcal{Y}}_s) \boldsymbol{\sigma}_k(\mathbf{Y}^j), D^2 \boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_r) \bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_l(\mathbf{Y}^j) \bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_l(\mathbf{Y}^j) \Big\rangle_{\mathbb{R}^L} \,\mathrm{d}r \,\mathrm{d}s \Big] \Bigg| \end{aligned}$$

and

$$\begin{aligned} & \boldsymbol{M_2} \\ :=& \Bigg| \sum_{k,l=1}^{K} \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \sum_{i=1}^{L} \Big\langle D_{\mathbf{x}} \partial_{x_i} u(s, \boldsymbol{\mathcal{Y}}_s), \boldsymbol{\sigma}_k(\mathbf{Y}^j) \Big\rangle_{\mathbb{R}^L} \int_{t_j}^{s} \Big\langle D\sigma_k^{(i)}(\boldsymbol{\mathcal{Y}}_r), \bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_l(\mathbf{Y}^j) \,\mathrm{d}\beta_l(r) \Big\rangle_{\mathbb{R}^L} \,\mathrm{d}s \Big] \Bigg| \\ =& \Bigg| \sum_{k,l=1}^{K} \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \int_{t_j}^{s} \Big\langle D^2_{\mathbf{x}} u(s, \boldsymbol{\mathcal{Y}}_s) \boldsymbol{\sigma}_k(\mathbf{Y}^j), D\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_r) \bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_l(\mathbf{Y}^j) \,\mathrm{d}\beta_l(r) \Big\rangle_{\mathbb{R}^L} \,\mathrm{d}s \Big] \Bigg|. \end{aligned} \tag{3.29}$$

Almost the same arguments as we used for the treatment of (3.10) in Lemma 3.3, that is generating additional higher order terms to get closer to computable terms gives

$$\boldsymbol{T_{1,a,1}} \leq \tfrac{\boldsymbol{C}_{D^2(\phi)}}{2} \cdot \mathbf{E_6}(\mathbf{Y}^j) \cdot \left(\tau^{j+1}\right)^2 + \boldsymbol{K_4} \tag{3.30}$$

and

$$\boldsymbol{T_{1,a,2}} \leq \tfrac{C_{D^2(\phi)}}{4} \cdot \mathbf{E_7}(\mathbf{Y}^j) \cdot \left(\tau^{j+1}\right)^2 + \boldsymbol{K_5}\,, \tag{3.31}$$

where

$$\boldsymbol{K_4} := \Bigg| \sum_{k=1}^{K} \mathbb{E}\Bigg[\int_{t_j}^{t_{j+1}} \int_{t_j}^{s} \Big\langle D_{\mathbf{x}}^2 u(s, \boldsymbol{\mathcal{Y}}_s) \boldsymbol{\sigma}_k(\mathbf{Y}^j), \Big(D\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_r) - D\boldsymbol{\sigma}_k(\mathbf{Y}^j) \Big) \Big(\mathscr{A} \bar{\boldsymbol{\mathscr{A}}}^{j+1} \mathbf{Y}^j - \bar{\boldsymbol{\mathscr{A}}}^{j+1} \mathbf{f}(\mathbf{Y}^j) \Big) \Big\rangle_{\mathbb{R}^L} \,\mathrm{d}r\,\mathrm{d}s \Bigg] \Bigg|$$

and

$$\boldsymbol{K_5} := \frac{1}{2} \Bigg| \sum_{k,l=1}^{K} \mathbb{E}\Bigg[\int_{t_j}^{t_{j+1}} \int_{t_j}^{s} \Big\langle D_{\mathbf{x}}^2 u(s, \boldsymbol{\mathcal{Y}}_s) \boldsymbol{\sigma}_k(\mathbf{Y}^j), \Big(D^2\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_r) - D^2\boldsymbol{\sigma}_k(\mathbf{Y}^j) \Big) \bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_l(\mathbf{Y}^j) \bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_l(\mathbf{Y}^j) \Big\rangle_{\mathbb{R}^L} \,\mathrm{d}r\,\mathrm{d}s \Bigg] \Bigg|$$

are higher order terms.

d) Next, we consider the expression $\boldsymbol{M_2}$. Here, our approach is very similar to that in Lemma 3.3, where we used tools from Malliavin calculus for an appropriate treatment of $\boldsymbol{M_1}$, and therefore skip most of the details here. We obtain

$$\begin{aligned} \boldsymbol{M_2} &= \Bigg| \sum_{k,l=1}^{K} \mathbb{E}\Bigg[\int_{t_j}^{t_{j+1}} \int_{t_j}^{s} \mathrm{Tr}\Big(D_{\mathbf{x}}^3 u(s, \boldsymbol{\mathcal{Y}}_s) \boldsymbol{\sigma}_k(\mathbf{Y}^j) D\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_r) \bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_l(\mathbf{Y}^j) \Big[\bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_l(\mathbf{Y}^j) \Big]^\top \Big) \mathrm{d}r\,\mathrm{d}s \Bigg] \Bigg| \\ &\leq \Bigg| \sum_{k,l=1}^{K} \mathbb{E}\Bigg[\int_{t_j}^{t_{j+1}} \int_{t_j}^{s} \mathrm{Tr}\Big(D_{\mathbf{x}}^3 u(s, \boldsymbol{\mathcal{Y}}_s) \boldsymbol{\sigma}_k(\mathbf{Y}^j) D\boldsymbol{\sigma}_k(\mathbf{Y}^j) \bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_l(\mathbf{Y}^j) \Big[\bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_l(\mathbf{Y}^j) \Big]^\top \Big) \mathrm{d}r\,\mathrm{d}s \Bigg] \Bigg| \\ &\quad + \boldsymbol{K_6}\,, \end{aligned}$$

where

$$\boldsymbol{K_6} := \Bigg| \sum_{k,l=1}^{K} \mathbb{E}\Bigg[\int_{t_j}^{t_{j+1}} \int_{t_j}^{s} \mathrm{Tr}\Big(D_{\mathbf{x}}^3 u(s, \boldsymbol{\mathcal{Y}}_s) \boldsymbol{\sigma}_k(\mathbf{Y}^j) \{ D\boldsymbol{\sigma}_k(\boldsymbol{\mathcal{Y}}_r) - D\boldsymbol{\sigma}_k(\mathbf{Y}^j) \} \bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_l(\mathbf{Y}^j) \Big[\bar{\boldsymbol{\mathscr{A}}}^{j+1} \boldsymbol{\sigma}_l(\mathbf{Y}^j) \Big]^\top \Big) \mathrm{d}r\,\mathrm{d}s \Bigg] \Bigg|$$

is again an additional higher order term, which results from adding and substracting $D\boldsymbol{\sigma}_k(\mathbf{Y}^j)$, $k = 1, ..., K$, in order to obtain an almost fully computable leading order term.
Similar calculations as we used before and using Lemma 2.5 (iii), yields

$$\boldsymbol{M_2} \leq \tfrac{C_{D^3(\phi)}}{2} \cdot \mathbf{E_8}(\mathbf{Y}^j) \cdot \left(\tau^{j+1}\right)^2 + \boldsymbol{K_6}\,. \tag{3.32}$$

e) Hence, plugging (3.30), (3.31) and (3.32) into (3.28) yields

$$\boldsymbol{T_{1,a}} \leq \left\{ \tfrac{C_{D^2(\phi)}}{2} \cdot \mathbf{E_6}(\mathbf{Y}^j) + \tfrac{C_{D^2(\phi)}}{4} \cdot \mathbf{E_7}(\mathbf{Y}^j) + \tfrac{C_{D^3(\phi)}}{2} \cdot \mathbf{E_8}(\mathbf{Y}^j) \right\} \cdot \left(\tau^{j+1}\right)^2 + \boldsymbol{K_4} + \boldsymbol{K_5} + \boldsymbol{K_6}\,. \tag{3.33}$$

f) Plugging further (3.27) and (3.33) into (3.26) yields

$$\boldsymbol{T_1} \leq \Bigg\{\boldsymbol{C}_{D^2}(\phi)\cdot \mathbf{E_6}(\mathbf{Y}^j) + \tfrac{\boldsymbol{C}_{D^2}(\phi)}{2}\cdot \mathbf{E_7}(\mathbf{Y}^j) + \boldsymbol{C}_{D^3}(\phi)\cdot \mathbf{E_8}(\mathbf{Y}^j)$$
$$+ \boldsymbol{C}_{D^2}(\phi) C^2_{D\boldsymbol{\sigma}} \cdot \Big\{\tfrac{\tau^{j+1}}{3}\cdot \mathbf{E_{10}}(\mathbf{Y}^j) + \tfrac{1}{2}\cdot \mathbf{E_{12}}(\mathbf{Y}^j)\Big\}\Bigg\}\cdot \Big(\tau^{j+1}\Big)^2 + 2\boldsymbol{K_4} + 2\boldsymbol{K_5} + 2\boldsymbol{K_6}\,. \tag{3.34}$$

g) It remains to examine the terms $\boldsymbol{K_4}, \boldsymbol{K_5}$ and $\boldsymbol{K_6}$ and to show that these terms are indeed higher order terms. Again, a similar treatment as we did for the higher order terms $\boldsymbol{K_1}, \boldsymbol{K_2}, \boldsymbol{K_3}$ in Lemma 3.3 yields

$$\boldsymbol{K_4} \leq \boldsymbol{C}_{D^2}(\phi) C_{D^2\boldsymbol{\sigma}} \cdot \sqrt{\mathbf{E_{13}}(\mathbf{Y}^j)}\cdot \sqrt{\tfrac{\tau^{j+1}}{15}\cdot \mathbf{E_{10}}(\mathbf{Y}^j) + \tfrac{1}{8}\cdot \mathbf{E_{12}}(\mathbf{Y}^j)}\cdot \Big(\tau^{j+1}\Big)^{2.5}, \tag{3.35}$$

$$\boldsymbol{K_5} \leq \tfrac{\boldsymbol{C}_{D^2}(\phi) C_{D^3\boldsymbol{\sigma}}}{2} \cdot \sqrt{\mathbf{E_{14}}(\mathbf{Y}^j)}\cdot \sqrt{\tfrac{\tau^{j+1}}{15}\cdot \mathbf{E_{10}}(\mathbf{Y}^j) + \tfrac{1}{8}\cdot \mathbf{E_{12}}(\mathbf{Y}^j)}\cdot \Big(\tau^{j+1}\Big)^{2.5} \tag{3.36}$$

and

$$\boldsymbol{K_6} \leq \boldsymbol{C}_{D^3}(\phi) C_{D^2\boldsymbol{\sigma}} \cdot \sqrt{\mathbf{E_{14}}(\mathbf{Y}^j)}\cdot \sqrt{\tfrac{\tau^{j+1}}{15}\cdot \mathbf{E_{10}}(\mathbf{Y}^j) + \tfrac{1}{8}\cdot \mathbf{E_{12}}(\mathbf{Y}^j)}\cdot \Big(\tau^{j+1}\Big)^{2.5}. \tag{3.37}$$

Step **4:** (Finishing the proof) We combine (3.35), (3.36) and (3.37) with (3.34), and then combine the resulting expression with (3.23) and (3.24), which proves the assertion.

□

Next, we show convergence with optimal weak order $\mathcal{O}(\tau)$ for the *a posteriori* error estimator in (3.1) on a mesh with maximum mesh size $\tau > 0$ with the help of Lemma 2.6 — and hence of the weak error of $\{\mathbf{Y}^j\}_{j=0}^J$ from (0.3) thanks to Theorem 3.1.

Theorem 3.5. Assume **(A1)** – **(A3)**. Let $\{\mathbf{Y}^j\}_{j=0}^J$ solve (0.3) on a mesh $\{t_j\}_{j=0}^J \subset [0,T]$ with local mesh sizes $\{\tau^{j+1}\}_{j=0}^{J-1}$ and maximum mesh size $\tau = \max_j \tau^{j+1}$. Then, there exists $\boldsymbol{C} \equiv \boldsymbol{C}(\phi) > 0$ independent of L, such that

$$\sum_{j=0}^{J-1} \tau^{j+1}\mathfrak{G}\Big(\phi; \tau^{j+1}, \mathbf{Y}^j\Big) \leq \boldsymbol{C}\cdot\tau\,.$$

Remark 3.2. **1.** The work [25] derives a weak *a priori* error estimate for the linear stochastic heat equation with additive noise, where the analysis exploits the representation formula for the mild solution, and a transformation of it to another process which solves a further SPDE without drift term and additive noise. In contrast, the weak a priori error analysis in [24] for SPDE (0.2) with $\boldsymbol{\beta} = \mathbf{0}$ requires Malliavin calculus to efficiently estimate additionally appearing stochastic integral terms due to the nonlinearities F and Σ — which are of similar type as $\boldsymbol{M_1}$ in (3.9) and $\boldsymbol{M_2}$ in (3.29) appearing here. The *a posteriori* error analysis to verify Theorem 3.1 with estimators $\{\mathfrak{G}\Big(\phi; \tau^{j+1}, \mathbf{Y}^j\Big)\}_{j\geq 0}$ also exploits Malliavin calculus,

and eventually enables Theorem 3.5. We remark that the tools from Malliavin calculus used here slightly differ from those in [24]: while [24] utilizes an *integration by parts formula* (*cf.* [24, Lemma 2.1]), we are making use of the *Clark-Ocone formula* (*cf.* Lemma 2.1).
2. The work [24] uses less regular initial data, in particular, and exploits the regularizing effect of the involved semigroup. In this work, we assume **(A3)** to verify Theorem 3.5; however, we believe a corresponding result to hold for less 'regular' initial data, by using a modification of Lemma 2.6 that involves temporal weights in the functional to handle less regular initial data, and mimic the regularizing effect in the present context of arguments.

Proof. We independently bound $\{\mathbf{E}_{\boldsymbol{\ell}}(\mathbf{Y}^j)\}_{\boldsymbol{\ell}=1,\dots,15,\, j=0,\dots,J-1}$ in $\mathfrak{G} \equiv \{\mathfrak{G}\big(\phi; \tau^{j+1}, \mathbf{Y}^j\big)\}_{j=0}^{J-1}$ in (3.1) with the help of Lemma 2.6
a) Second moment bounds for $\mathbf{E}_{\boldsymbol{\ell}}(\mathbf{Y}^j)$, $\boldsymbol{\ell} = 1,2,3,4,5,6,9,10,12,15$, $j = 0,\dots,J-1$: We show for $\ell = 0,1,2$ that

$$(\mathbf{i}) \quad \max_{j=0,\dots,J-1} \mathbb{E}\Big[\|\mathscr{A}^\ell \bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{Y}^j\|^2_{\mathbb{R}^L}\Big] \leq \boldsymbol{C}^{(\ell)}_{1,1}\,,$$

$$(\mathbf{ii}) \quad \max_{j=0,\dots,J-1} \mathbb{E}\Big[\|\mathscr{A}^\ell \bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{f}(\mathbf{Y}^j)\|^2_{\mathbb{R}^L}\Big] \leq 2\Big(C^{(\ell)}_{\mathbf{f}}\Big)^2\Big(1+\boldsymbol{C}^{(\ell)}_{1,1}\Big)\,,$$

$$(\mathbf{iii}) \quad \max_{j=0,\dots,J-1} \sum_{k=1}^{K} \mathbb{E}\Big[\|\mathscr{A}^\ell \bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\|^2_{\mathbb{R}^L}\Big] \leq 2K\Big(C^{(\ell)}_{\boldsymbol{\sigma}}\Big)^2\Big(1+\boldsymbol{C}^{(\ell)}_{1,1}\Big)\,,$$

$$(\mathbf{iv}) \quad \max_{j=0,\dots,J-1} \sum_{k=1}^{K} \mathbb{E}\Big[\|\boldsymbol{\sigma}_k(\mathbf{Y}^j)\|^2_{\mathbb{R}^L}\Big] \leq 2K\Big(C^{(0)}_{\boldsymbol{\sigma}}\Big)^2\Big(1+\boldsymbol{C}^{(0)}_{1,2}\Big)\,,$$

where $\boldsymbol{C}^{(\ell)}_{1,1} > 0$, $\ell = 0,1,2$, are the constants from Lemma 2.6, and $\bar{\boldsymbol{\mathscr{A}}}^{j+1} = \Big(\mathbb{I} + \tau^{j+1}\mathscr{A}\Big)^{-1}$.
Let $\ell = 0,1,2$ and $j = 0,\dots,J-1$. Since $\|\bar{\boldsymbol{\mathscr{A}}}^{j+1}\|_{\mathbb{R}^{L\times L}} \leq 1$ for every $\tau^{j+1} > 0$, we immediately obtain

$$\|\mathscr{A}^\ell \bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{Y}^j\|^2_{\mathbb{R}^L} \leq \|\mathscr{A}^\ell\mathbf{Y}^j\|^2_{\mathbb{R}^L} \cdot \|\bar{\boldsymbol{\mathscr{A}}}^{j+1}\|_{\mathbb{R}^{L\times L}} \leq \|\mathscr{A}^\ell\mathbf{Y}^j\|^2_{\mathbb{R}^L}\,.$$

Hence, taking the expectation and using Lemma 2.6, assertion (**i**) follows. In almost the same way, on using **(A1) (b)**, we obtain

$$\|\mathscr{A}^\ell \bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{f}(\mathbf{Y}^j)\|^2_{\mathbb{R}^L} \leq \|\mathscr{A}^\ell\mathbf{f}(\mathbf{Y}^j)\|^2_{\mathbb{R}^L} \leq 2\Big(C^{(\ell)}_{\mathbf{f}}\Big)^2\Big(1+\|\mathscr{A}^\ell\mathbf{Y}^j\|^2_{\mathbb{R}^L}\Big)\,.$$

Again, taking the expectation and applying Lemma 2.6, assertion (**ii**) follows. Statement (**iii**) follows by the same argumentation and (**iv**) immediately follows from **(A2)** and Lemma 2.6.
The bounds **(i)** – **(iv)** now yield these for $\mathbf{E}_{\boldsymbol{\ell}}(\mathbf{Y}^j)$, $\boldsymbol{\ell} = 1,2,3,4,5,6,9,10,12,15$, $j = 0,\dots,J-1$.
b) Fourth moment bounds for $\mathbf{E}_{\boldsymbol{\ell}}(\mathbf{Y}^j)$, $\boldsymbol{\ell} = 7,8,11,13,14$, $j = 0,\dots,J-1$: Similar as in **a)**, by Lemma 2.6, we obtain

(v) $\max_{j=0,\ldots,J-1} \mathbb{E}\Big[\|\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{Y}^j\|_{\mathbb{R}^L}^4\Big] \leq \boldsymbol{C}_{1,2}^{(1)}\,,$

(vi) $\max_{j=0,\ldots,J-1} \mathbb{E}\Big[\|\mathscr{A}\bar{\boldsymbol{\mathscr{A}}}^{j+1}\mathbf{f}(\mathbf{Y}^j)\|_{\mathbb{R}^L}^4\Big] \leq 8\Big(C_{\mathbf{f}}^{(0)}\Big)^4\Big(1+\boldsymbol{C}_{1,2}^{(0)}\Big)\,,$

(vii) $\max_{j=0,\ldots,J-1} \sum_{k=1}^{K} \mathbb{E}\Big[\|\bar{\boldsymbol{\mathscr{A}}}^{j+1}\boldsymbol{\sigma}_k(\mathbf{Y}^j)\|_{\mathbb{R}^L}^4\Big] \leq 8K\Big(C_{\boldsymbol{\sigma}}^{(0)})^4\Big(1+\boldsymbol{C}_{1,2}^{(0)}\Big)\,,$

(viii) $\max_{j=0,\ldots,J-1} \sum_{k=1}^{K} \mathbb{E}\Big[\|\boldsymbol{\sigma}_k(\mathbf{Y}^j)\|_{\mathbb{R}^L}^4\Big] \leq 8K\Big(C_{\boldsymbol{\sigma}}^{(0)})^4\Big(1+\boldsymbol{C}_{1,2}^{(0)}\Big)\,,$

where $\boldsymbol{C}_{1,2}^{(\ell)} > 0$, $\ell = 0, 1$, are the constants from Lemma 2.6. Again, the bounds **(v)** – **(viii)** yield these for $\mathbf{E}_{\boldsymbol{\ell}}(\mathbf{Y}^j)$, $\boldsymbol{\ell} = 7, 8, 11, 13, 14$, $j = 0, ..., J-1$.
c): By means of **a)** and **b)**, we can find a constant $\tilde{\boldsymbol{C}} \geq 1$ independent of L and j such that for all $j \geq 0$

$$\mathfrak{G}\Big(\phi; \tau^{j+1}, \mathbf{Y}^j\Big) \leq \tilde{\boldsymbol{C}} \cdot \tau^{j+1}\,. \tag{3.38}$$

Hence, plugging (3.38) into (3.1), using $\tau^{j+1} \leq \tau$ for every $j \geq 0$, and setting $\boldsymbol{C} := \tilde{\boldsymbol{C}} \cdot T$ yields the assertion.

□

In the next chapter, we base an adaptive method on the *a posteriori* error estimate (3.1) to automatically select local step sizes. For every $j \geq 0$, we show that the adaptive method selects a new time step τ^{j+1} within finitely many steps, and that the algorithm reaches the terminal time $T > 0$ after finitely many steps as well (global termination).

4. Weak adaptive approximation: Algorithm and Convergence

By Theorem 3.1, the weak error caused by scheme (0.3) on a *given* partition $\{t_j\}_{j=0}^J \subset [0,T]$ is controllable via the *a posteriori* error estimate (0.4). In this chapter, we use this result for an adaptive method that automatically steers local mesh size selection. For this purpose, we check if the criterion $\mathfrak{G}\Big(\phi;\tau^{j+1},\mathbf{Y}^j\Big) \leq \frac{\mathtt{Tol}}{T}$ is met or not: in the first case, τ^{j+1} is admissible, bounding the new local error in such a way that the overall error will be bounded by `Tol` through Theorem 3.1; in the latter case, τ^{j+1} will be replaced by the refined mesh size $\widetilde{\tau}^{j+1} := \frac{\tau^{j+1}}{2}$, and the criterion will be checked again. The following algorithm contains a refinement step **(1)** to generate $\{\tau^{j+1,\ell}\}_{\ell\geq 0}$ and — if $\mathfrak{G}\Big(\phi;\tau^{j+1,\ell},\mathbf{Y}^j\Big)$ is 'too small' — a final coarsening step **(3)** in the loop of generating the subsequent iterate from (0.3), after accepting the possible underestimation of $\mathfrak{G}\Big(\phi;\tau^{j+1,\ell},\mathbf{Y}^j\Big)$.

Algorithm 4.1. Fix $\mathtt{Tol} > 0$ and $\tau^1 \geq \frac{\mathtt{Tol}}{T}$. Let $(\tau^j,\mathbf{Y}^j)$ be given for some $j \geq 1$. Define $\tau^{j+1,0} := \tau^j$. — For $\ell = 0,1,2,\ldots$ compute $\mathfrak{G}\Big(\phi;\tau^{j+1,\ell},\mathbf{Y}^j\Big)$ and decide:

(1) If $\mathfrak{G}\Big(\phi;\tau^{j+1,\ell},\mathbf{Y}^j\Big) > \frac{\mathtt{Tol}}{T}$, set $\tau^{j+1,\ell+1} := \frac{\tau^{j+1,\ell}}{2}$, and $\ell \hookrightarrow \ell+1$.

(2) If $\frac{\mathtt{Tol}}{2T} \leq \mathfrak{G}\Big(\phi;\tau^{j+1,\ell},\mathbf{Y}^j\Big) \leq \frac{\mathtt{Tol}}{T}$, set $\tau^{j+1} := \tau^{j+1,\ell}$, $t_{j+1} := t_j + \tau^{j+1}$, compute $\Delta_{j+1,\ell}\beta_k := \beta_k\Big(t_j+\tau^{j+1,\ell}\Big) - \beta_k(t_j)$, for $k = 1,\ldots,K$, then solve (0.3) for $\mathbf{Y}^{j+1}$, and $j \hookrightarrow j+1$.

(3) If $\mathfrak{G}\Big(\phi;\tau^{j+1,\ell},\mathbf{Y}^j\Big) < \frac{\mathtt{Tol}}{2T}$, set $\tau^{j+1} := \tau^{j+1,\ell}$, $t_{j+1} := t_j + \tau^{j+1}$, compute $\mathbf{Y}^{j+1}$ via (0.3) with τ^{j+1} and $\{\Delta_{j+1,\ell}\beta_k,\, k=1,\ldots,K\}$. Then set $\tau^{j+1} := 2\tau^{j+1}$ and $j \hookrightarrow j+1$.

Stop, if $t_j \geq T$ for some j and set $J := j$.

This sequence of refinement steps, which is succeeded by possibly one coarsening step prevents infinite loops of refinement and coarsening, and enables a flexible re-meshing to capture local dynamics. The following theorem validates termination of the adaptive method, consisting of (0.3) and Algorithm 4.1.

Theorem 4.2. Let $\mathtt{Tol} > 0$. Suppose **(A1)** – **(A3)**. Then, the adaptive method consisting of (0.3) and Algorithm 4.1 generates each of the local step sizes of $\{\tau^{j+1}\}_{j\geq 0}$ after

$\mathcal{O}\Big(\log(\mathtt{Tol}^{-1})\Big)$ many iterations and the algorithm reaches the terminal time $T > 0$ within $J = \mathcal{O}\Big(\mathtt{Tol}^{-1}\Big)$ time steps. Furthermore, admissible tuple $\{(\tau^{j+1}, \mathbf{Y}^{j+1})\}_{j=0}^{J-1}$ satisfy

$$\max_{0\leq j\leq J} \Big|\mathbb{E}\Big[\phi(\mathbf{X}_{t_j})\Big] - \mathbb{E}\Big[\phi(\mathbf{Y}^j)\Big]\Big| \leq \mathtt{Tol}\,. \tag{4.1}$$

Remark 4.1. **1.** An adaptive method based on a strong *a posteriori* error estimate to steer automatic spatio-temporal remeshing for a discretization of (0.2) with additive noise is proposed in [57], for which termination of an iterative strategy to select new local mesh parameters for a *fixed* index $j \in \mathbb{N}$ is shown. [57] conceptionally follows ideas in [20] for the heat equation (*i.e.*, $\Sigma_k \equiv 0$ $(1 \leq k \leq K)$ in (0.2)), where a local approximation argument in step $j \in \mathbb{N}$ settles the existence of a value $\tau_*^{j+1} > 0$, *s.t.* values $\tau^{j+1,\ell} \leq \tau_*^{j+1}$ meet the stopping criterion; this argument, however, does not exclude selected τ_*^{j+1} in [57, 20] to crucially depend on j, leaving open *global* termination. This deficiency has been overcome in [53] for a modified version of the adaptive algorithm in [20], which, in particular, exploits a *discrete stability* property of the underlying discretization to herewith establish $\inf_j \tau_*^{j+1} \geq \tau_* > 0$. — We here proceed analogously to settle convergence of Algorithm 4.1 with the help of Lemma 2.6. **2.** Automatic mesh refinement in [64] is based on the computable *leading-order* term in the weak *a posteriori* estimator (1.1) that has been derived in [76], provided that the involved drift and diffusion functions are *bounded*. For a *sufficiently fine* initial mesh $\mathcal{I}_{J^0} := \{t_j\}_{j=0}^{J^0} \subset [0,T]$ and given $\mathcal{I}_{J^\ell}$, the new mesh $\mathcal{I}_{J^{\ell+1}} \supset \mathcal{I}_{J^\ell}$ refines those intervals $[t_{j^\ell}, t_{(j+1)^\ell}]$, where $\rho_{(j+1)^\ell}|\tau^{(j+1)^\ell}|^2$ overshoots $\frac{\mathtt{Tol}}{J^\ell}$. Note that this iterative strategy requires the *global* re-computation of $\{\rho_{j^\ell}\}_{j^\ell}$ for every $\ell \geq 0$: termination after $\ell^* < \infty$ iterations, with $J^{\ell^*} = \mathcal{O}(\mathtt{Tol}^{-1})$ is then shown in [64].

Proof. **a) Termination for each** $j \geq 0$**:** Fix $j \geq 0$, and recall (3.38) in the proof of Theorem 3.5. Since the constant $\tilde{\boldsymbol{C}} \geq 1$ appearing there does not depend on j, we generate a finite sequence $\{\tau^{j+1,\ell}\}_{\ell=0}^{\ell^*_{j+1}}$ with $\tau^{j+1,\ell} = \frac{\tau^{j+1,0}}{2^\ell}$, $\ell = 0, ..., \ell^*_{j+1}$, according to the refinement mechanism **(1)** of Algorithm 4.1, until either **(2)** or **(3)** is met. In view of (3.38), we find out that $\ell = \left\lceil \log\left(\frac{\tau^{j+1,0}\tilde{\boldsymbol{C}}T}{\mathtt{Tol}}\right)/\log(2)\right\rceil$ is the smallest natural number such that

$$\mathfrak{G}\Big(\phi; \tau^{j+1,\ell}, \mathbf{Y}^j\Big) \leq \tilde{\boldsymbol{C}} \cdot \tau^{j+1,\ell} = \tilde{\boldsymbol{C}} \cdot \frac{\tau^{j+1,0}}{2^\ell} \overset{!}{\leq} \frac{\mathtt{Tol}}{T}\,.$$

Consequently, we have

$$0 \leq \ell^*_{j+1} \leq \left\lceil \frac{\log\left(\frac{\tau^{j+1,0}\tilde{\boldsymbol{C}}T}{\mathtt{Tol}}\right)}{\log(2)} \right\rceil, \tag{4.2}$$

which yields a maximum of $\mathcal{O}\Big(\log(\mathtt{Tol}^{-1})\Big)$ (refinement) steps to accept the local step size $\tau^{j+1} := \tau^{j+1,\ell^*_{j+1}} = \frac{\tau^{j+1,0}}{2^{\ell^*_{j+1}}}$.

b) Global termination rate: We show by induction that

$$\tau^{j+1} \geq \frac{\mathtt{Tol}}{2\tilde{\boldsymbol{C}}T} \quad (j \geq 0)\,,$$

where $\tilde{\boldsymbol{C}} \geq 1$ is the constant in (3.38). In particular, this means that T is reached within $J = \mathcal{O}\left(\mathtt{Tol}^{-1}\right)$ many time steps.
The base case follows by the choice of the initial mesh size $\tau^1 \geq \frac{\mathtt{Tol}}{T}$. Now suppose that we have generated $\mathbf{Y}^j$ with step size $\tau^j \geq \frac{\mathtt{Tol}}{2\tilde{\boldsymbol{C}}T}$; see also Figure 1.1. In order to successfully compute $\mathbf{Y}^{j+1}$, we set $\tau^{j+1,0} := \tau^j$ (if **(2)** occurred in the generation of τ^j), or $\tau^{j+1,0} := 2\tau^j$ (if **(3)** occurred in the generation of τ^j). In both cases, $\tau^{j+1,0} \geq \frac{\mathtt{Tol}}{2\tilde{\boldsymbol{C}}T}$. Via **a)**, we generate a finite sequence $\{\tau^{j+1,\ell}\}_{\ell=0}^{\ell^*_{j+1}}$ until either **(2)** or **(3)** is met, and then generate $\mathbf{Y}^{j+1}$ with step size $\tau^{j+1} := \tau^{j+1,\ell^*_{j+1}} = \frac{\tau^{j+1,0}}{2^{\ell^*_{j+1}}}$. Since $\lceil x \rceil < 1 + x$, $x \in \mathbb{R}$, we conclude by means of (4.2)

$$\tau^{j+1} := \tau^{j+1,\ell^*_{j+1}} = \frac{\tau^{j+1,0}}{2^{\ell^*_{j+1}}} \geq \frac{\mathtt{Tol}}{2\tilde{\boldsymbol{C}}T}\,.$$

c) Estimate (4.1) immediately follows from (3.1) and part **(2)** of Algorithm 4.1.

□

5. Computational Experiments

The simulations of iterates $\{\mathbf{Y}^j\}_{j\geq 0}$ from (0.3) use independent standard normally distributed pseudo-random numbers (‘`randn`’) in `MATLAB` (version: 2017a). By *Kolmogorov's extension theorem*; see *e.g.* [69, p. 11, Thm. 2.1.5], a family of probability measures $\{\boldsymbol{\mathcal{N}}_{\mathbf{0},t\mathbb{I}};\, 0 \leq t \leq T\}$ on $(\mathbb{R}^K, \mathcal{B}(\mathbb{R}^K))$ yields the existence of a (filtered) probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ and Wiener processes $\{\beta_k(t)\,;\, t \in [0,T]\}$, $k = 1, ..., K$ on it; we consider them to be the ones with which (0.1) and (0.3) are defined.

Let $\mathtt{Tol} > 0$. We use (0.3) in combination with the adaptive Algorithm 4.1 for different examples, which result from a finite element spatial discretization of a SPDE (0.2). We show how the involved *a posteriori* error estimate (0.4) serves to estimate related weak errors, and that adaptive remeshing substantially reduces the amount of needed steps to overcome the interval $[0,T]$. For the sake of computations, we therefore use sufficiently large $\mathtt{M}$-samples to suppress additional statistical errors in the Monte-Carlo method due to approximating appearing expectations $\mathbb{E}[\cdot]$ by $\mathbb{E}_{\mathtt{M}}[\cdot]$ — as in (0.4), which then takes the form

$$\max_{0\leq j\leq J} \Big|\mathbb{E}\Big[\phi(\mathbf{X}_{t_j})\Big] - \mathbb{E}_{\mathtt{M}}\Big[\phi(\mathbf{Y}^j)\Big]\Big| \leq 2\mathtt{Tol}\,, \tag{5.1}$$

and which now holds with high probability. In order to obtain (5.1), we first add and substract $\mathbb{E}\Big[\phi(\mathbf{Y}^j)\Big]$ and use Theorem 3.1, which yields

$$\max_{0\leq j\leq J} \Big|\mathbb{E}\Big[\phi(\mathbf{X}_{t_j})\Big] - \mathbb{E}_{\mathtt{M}}\Big[\phi(\mathbf{Y}^j)\Big]\Big| \leq \sum_{j=0}^{J-1} \tau^{j+1}\mathfrak{G}\Big(\phi; \tau^{j+1}, \mathbf{Y}^j\Big) + \max_{0\leq j\leq J} \Big|\mathbb{E}\Big[\phi(\mathbf{Y}^j)\Big] - \mathbb{E}_{\mathtt{M}}\Big[\phi(\mathbf{Y}^j)\Big]\Big|.$$

Replacing all arising expectations $\mathbf{E}_{\boldsymbol{\ell}}(\cdot)$, $\boldsymbol{\ell} = 1, ..., 15$ in the representation of the error estimator $\mathfrak{G}$ by their corresponding empirical means $\mathbf{E}_{\boldsymbol{\ell}}^{(\mathtt{M})}(\cdot)$ and writing $\mathfrak{G}^{(\mathtt{M})}$ for the related (empirical) error estimator further leads to

$$\max_{0\leq j\leq J} \Big|\mathbb{E}\Big[\phi(\mathbf{X}_{t_j})\Big] - \mathbb{E}_{\mathtt{M}}\Big[\phi(\mathbf{Y}^j)\Big]\Big| \leq \sum_{j=0}^{J-1} \tau^{j+1}\mathfrak{G}^{(\mathtt{M})}\Big(\phi; \tau^{j+1}, \mathbf{Y}^j\Big) + \mathtt{ERR}_{\mathfrak{G}}(\mathtt{M}) + \mathtt{ERR}_{\phi}(\mathtt{M})\,, \tag{5.2}$$

where $\mathtt{ERR}_{\mathfrak{G}}(\mathtt{M})$, $\mathtt{ERR}_{\phi}(\mathtt{M})$ denote the arising statistical errors resulting from the approximation of the error estimator $\mathfrak{G}$ and from the approximation of the expectation of the test function ϕ. Algorithm 4.1 controls the first expression on the right-hand side of (6.1a). In order to control the remaining statistical errors, *i.e.*, to ensure that $\mathtt{ERR}_{\mathfrak{G}}(\mathtt{M}) + \mathtt{ERR}_{\phi}(\mathtt{M}) \leq \mathtt{Tol}$ holds with high probability, and to conclude (5.1), one can (asymptotically) determine a number $\mathtt{M} \equiv \mathtt{M}(\mathtt{Tol}) \in \mathbb{N}$ of Monte-Carlo samples by means of concentration inequalities, the *central limit theorem* or other (non-)asymptotic controls; we refer to [37] for more details in this direction — In the computational studies reported below, we mostly chose $\mathtt{M} = 10^4$ for which the Monte-Carlo simulations performed stably.

5.1. The one-dimensional stochastic heat equation in Example 0.1

Let $T > 0$, $K \in \mathbb{N}$ and $\boldsymbol{\mathcal{D}} = (0,1) \subset \mathbb{R}$. We consider the SPDE (0.2) with $\varepsilon = 1$, $\boldsymbol{\beta} \equiv \mathbf{0}$, and $F, \Sigma \in \mathcal{C}^3\Big(\mathbb{R}\Big) \cap \mathbb{H}_0^1(\boldsymbol{\mathcal{D}})$, as well as homogeneous Dirichlet boundary conditions, and $y_0 \in \mathbb{H}_0^1(\boldsymbol{\mathcal{D}})$. These assumptions ensure the existence of a unique strong solution of (0.2); see *e.g.* [22, p. 197*ff.*]. We use standard notation, as *e.g.* $\mathbb{L}^2 \equiv \mathbb{L}^2(\boldsymbol{\mathcal{D}})$ and $\mathbb{H}_0^1 \equiv \mathbb{H}_0^1(\boldsymbol{\mathcal{D}})$ below; see *e.g.* [31, p. 244 *ff.*].

Let $L \in \mathbb{N}$. We consider a (uniform) triangulation of the domain $\boldsymbol{\mathcal{D}}$ such that

$$0 = \mathsf{x}_0 < \mathsf{x}_1 < ... < \mathsf{x}_L < \mathsf{x}_{L+1} = 1\,, \quad \text{with} \quad h \equiv \mathsf{x}_\ell - \mathsf{x}_{\ell-1} = \tfrac{1}{L+1} \quad (\ell = 1, ..., L).$$

Following [47], we use a finite element method based on piecewise affine functions to spatially discretize SPDE (0.2) with $\varepsilon = 1$ and $\boldsymbol{\beta} \equiv \mathbf{0}$. In combination with 'mass lumping', we obtain the following $L-$dimensional SDE system:

$$\begin{cases} \mathrm{d}\mathbf{X}_t^h &= \Big(-\mathscr{A}\mathbf{X}_t^h + \mathbf{f}(\mathbf{X}_t^h)\Big)\mathrm{d}t + \sum\limits_{k=1}^{K} \boldsymbol{\sigma}_k(\mathbf{X}_t^h)\mathrm{d}\beta_k(t) \quad \forall\, t \in [0,T]\,, \\ \mathbf{X}_0^h &= \Big(y_0(\mathsf{x}_1), ..., y_0(\mathsf{x}_L)\Big)^\top \in \mathbb{R}^L\,, \end{cases} \tag{5.3}$$

where

$$\mathbf{X}_t^h := \Big(X_t^h(\mathsf{x}_1), ..., X_t^h(\mathsf{x}_L)\Big)^\top \in \mathbb{R}^L\,, \qquad \mathbf{f}(\mathbf{X}_t^h) := \Big(F\Big(X_t^h(\mathsf{x}_1)\Big), ..., F\Big(X_t^h(\mathsf{x}_L)\Big)\Big)^\top \in \mathbb{R}^L\,,$$
$$\boldsymbol{\sigma}_k(\mathbf{X}_t^h) := \Big(\Sigma_k\Big(X_t^h(\mathsf{x}_1)\Big), ..., \Sigma_k\Big(X_t^h(\mathsf{x}_L)\Big)\Big)^\top \in \mathbb{R}^L\,, \quad \mathscr{A} := \frac{1}{h^2}\mathrm{tridiag}[-1, 2, -1] \in \mathbb{R}^{L\times L}\,.$$

Example 0.1 discusses computational studies for (5.3) via (0.3) in combination with adaptive Algorithm 4.1. We choose $\phi(\mathbf{x}) = \sqrt{h}\|\mathbf{x}\|_{\mathbb{R}^L}$ to approximate the $\mathbb{L}^2-$norm of (the finite element approximation of) $\{X_t,\ t \in [0,T]\}$ from SPDE (0.2). The related constants to compute $\mathfrak{G} \equiv \{\mathfrak{G}\Big(\phi; \tau^{j+1}, \mathbf{Y}^j\Big)\}_{j\geq 0}$ below (3.1) are:

$$\lambda_{\mathscr{A}} \approx \pi^2, \quad C_{D\mathbf{f}} = \tfrac{\pi}{5}, \quad C_{D^2\mathbf{f}} = \tfrac{\pi^2}{5}, \quad C_{D^3\mathbf{f}} = \tfrac{\pi^3}{5}, \quad C_{D\boldsymbol{\sigma}} = \tfrac{137}{120}, \quad C_{D^2\boldsymbol{\sigma}} = 0.158,$$
$$C_{D^3\boldsymbol{\sigma}} = 0.518, \quad \boldsymbol{C}_D(\phi) = \sqrt{h}, \quad \boldsymbol{C}_{D^2}(\phi) = 0, \quad \boldsymbol{C}_{D^3}(\phi) = 0 \quad \Big(\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = 6\Big).$$

Figure 5.1 below displays the contributions of the different $\mathbf{E}_{\boldsymbol{\ell}}^{(\mathtt{M})}(\cdot)$, $\boldsymbol{\ell} \in \{1, ..., 15\}\backslash\{7, 13, 14\}$ in the *a posteriori* error estimator $\mathfrak{G}$, which steers the step size selection of the adaptive Algorithm 4.1. Note that $\mathfrak{G}$ consists of leading order terms $\mathbf{E}_{\boldsymbol{\ell}}(\cdot)$, $\boldsymbol{\ell} \in \{1, 2, 3, 4, 5, 6, 8, 9, 12\}$, as well as 'higher order terms' $\mathbf{E}_{\boldsymbol{\ell}}(\cdot)$, $\boldsymbol{\ell} \in \{10, 11, 15\}$. Mainly responsible for mesh adjustments are the leading order terms — in particular $\mathbf{E}_{\mathbf{1}}(\cdot)$, which addresses higher derivatives (up to order 4) of the (approximated) solution $\{X_t, t \in [0,T]\}$ of (0.2), starting with $x \mapsto \sin(\pi x)$ as initial function. Heuristically, it might therefore be justified to neglect 'higher order terms' in order to save computational effort, and only take into account leading order terms in $\mathfrak{G}$.

For different noise parameters $K \in \{0, 1, 3, 5, 6, 8, 10\}$, Figure 5.2 below compares the total amount of time steps needed for adaptive and uniform meshes to perform equally well, indicating that: the larger K the more improved performance of the adaptive Algorithm 4.1 compared to scheme (0.3) with uniform time steps may be expected.

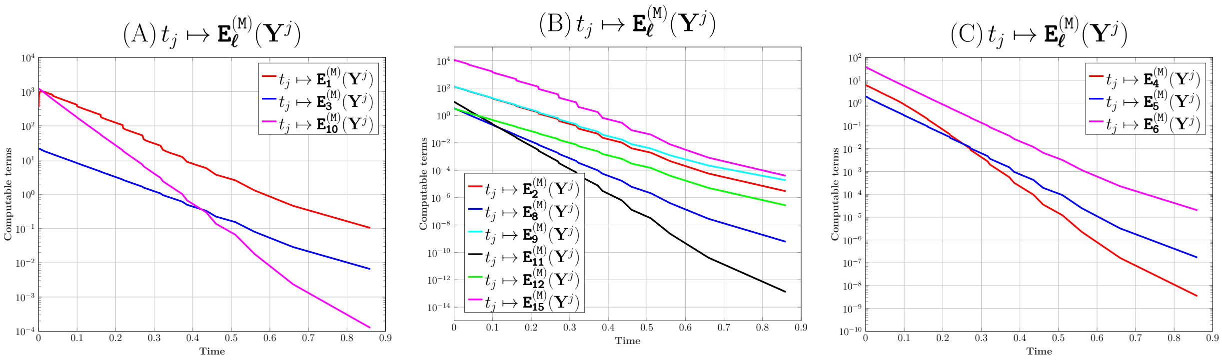


Figure 5.1.: (Error indicators of the weak *a posteriori* error estimator $\mathfrak{G}^{(\mathtt{M})}$ of Example 0.1 for $\mathtt{Tol} = 0.1$, $\mathtt{M} = 10^4$, $T = 1$) (A) Semi-Log-Plot of the corresponding computable error indicators $t_j \mapsto \mathbf{E}_{\boldsymbol{\ell}}^{(\mathtt{M})}(\mathbf{Y}^j)$, $\boldsymbol{\ell} = 1, 3, 10$, which only involve the drift term. (B) Semi-Log-Plot of the corresponding computable error indicators $t_j \mapsto \mathbf{E}_{\boldsymbol{\ell}}^{(\mathtt{M})}(\mathbf{Y}^j)$, $\boldsymbol{\ell} = 2, 8, 9, 11, 12, 15$, which only involve the diffusion term. (C) Semi-Log-Plot of the corresponding computable error indicators $t_j \mapsto \mathbf{E}_{\boldsymbol{\ell}}^{(\mathtt{M})}(\mathbf{Y}^j)$, $\boldsymbol{\ell} = 4, 5, 6$, where both the drift and diffusion are involved.

5.2. A convection dominated (stochastic) problem

Example 5.1. Consider (0.2) on $\boldsymbol{\mathcal{D}} = (0, 1)$, $T > 0$, with $\varepsilon > 0$, $\boldsymbol{\beta} \in \mathbb{R}$, $F \equiv 0$ and homogeneous Dirichlet boundary conditions. After a finite element discretization, using ‘mass lumping’, and $h = \frac{1}{L+1}$ for some $L \in \mathbb{N}$, we obtain

$$\begin{cases} \mathrm{d}\mathbf{X}_t^h &= -\mathscr{A}\mathbf{X}_t^h \mathrm{d}t + \sum\limits_{k=1}^{K} \boldsymbol{\sigma}_k(\mathbf{X}_t^h)\mathrm{d}\beta_k(t) \quad \forall\, t \in [0, T]\,, \\ \mathbf{X}_0^h &= \Big(y_0(\mathsf{x}_1), ..., y_0(\mathsf{x}_L)\Big)^\top \in \mathbb{R}^L\,, \end{cases}$$

where $\boldsymbol{\sigma}_k(\mathbf{X}_t^h)$ as in (5.3), and

$$\mathscr{A} := \frac{\varepsilon}{h^2}\mathrm{tridiag}[-1, 2, -1] - \frac{\boldsymbol{\beta}}{2h}\mathrm{tridiag}[-1, 0, 1] \in \mathbb{R}^{L\times L}\,.$$

We study three different cases given in Table 5.1: **Setup A** deals with a purely deterministic version of Example 5.1, *i.e.*, no diffusion is involved. In this context, for fixed $\boldsymbol{\beta}$ and ε, we discuss the role of the chosen test function ϕ in the adaptive method; see Figure 5.4 and Figure 5.5 below. Then, for a fixed test function, **Setup B** studies the impact of different

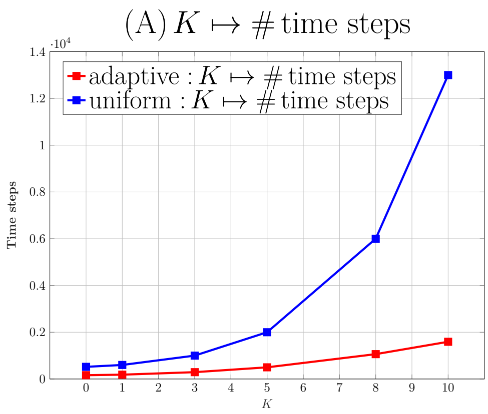


Figure 5.2.: (Adaptive vs. uniform: Influence of noise parameter K on the total amount of time steps within Example 0.1 for `Tol` = 0.1, `M` = 5000, T = 1) (A) Semi-Log-Plot of the total amount of time steps for unifom (—) vs. adaptive (—).

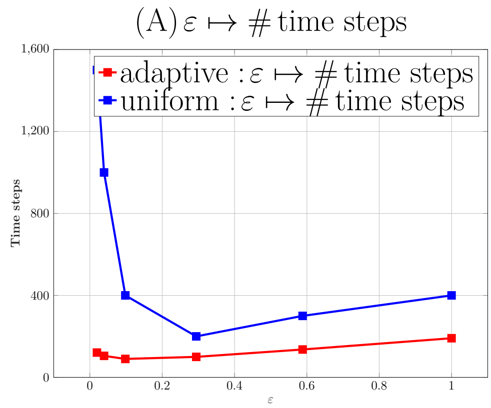


Figure 5.3.: (Adaptive vs. uniform: Influence of parameter ε on the total amount of time steps within **Setup B** for `Tol` = 0.1, `M` = 5000, $\boldsymbol{\beta}$ = 1, T = 1) (A) Semi-Log-Plot of the total amount of time steps for unifom (—) vs. adaptive (—).

choices of $\boldsymbol{\beta}$, ε on adaptive meshing; see Figures 5.6, 5.7 and 5.8 below. **Setup C** investigates Example 5.1 for a different initial function and a different type of multiplicative noise, but with fixed $\boldsymbol{\beta}$, ε and ϕ; see Figure 5.9 below.

	L	T	$y_0(x)$	K	$\Sigma_k\big(X_t(x)\big)$
Setup A	50	1	$\sin(\pi x)$	1	0
Setup B	50	1	$\sin(\pi x)$	5	$\frac{1}{2}X_t$
Setup C	50	1	$\sin(3\pi x)$	5	$\frac{1}{K+1-k}\sin(\pi kx)\big(X_t + 0.2\big)$

Table 5.1.: Different Setups for Example 5.1

As we can see in Figures 5.4 and 5.5, different choices of test functions might lead to huge changes in the amount of time steps generated via Algorithm 4.1. In Figures 5.4 and 5.5 we choose $\phi(\mathbf{x}) = \sqrt{h}\|\mathbf{x}\|_{\mathbb{R}^L}$ (resp. $\phi(\mathbf{x}) = \|\mathbf{x}\|_\infty := \max_{\ell=1,\ldots,L}|x_\ell|$) to approximate the $\mathbb{L}^2-$norm (resp. the $\mathbb{L}^\infty-$norm) of (the finite element approximation of) $\{X_t,\ t \in [0,T]\}$, from SPDE (0.2). Although both figures illustrate similar behaviours of time step size, error and *a posteriori* error plots, the amount of time steps needed to stay below the given error

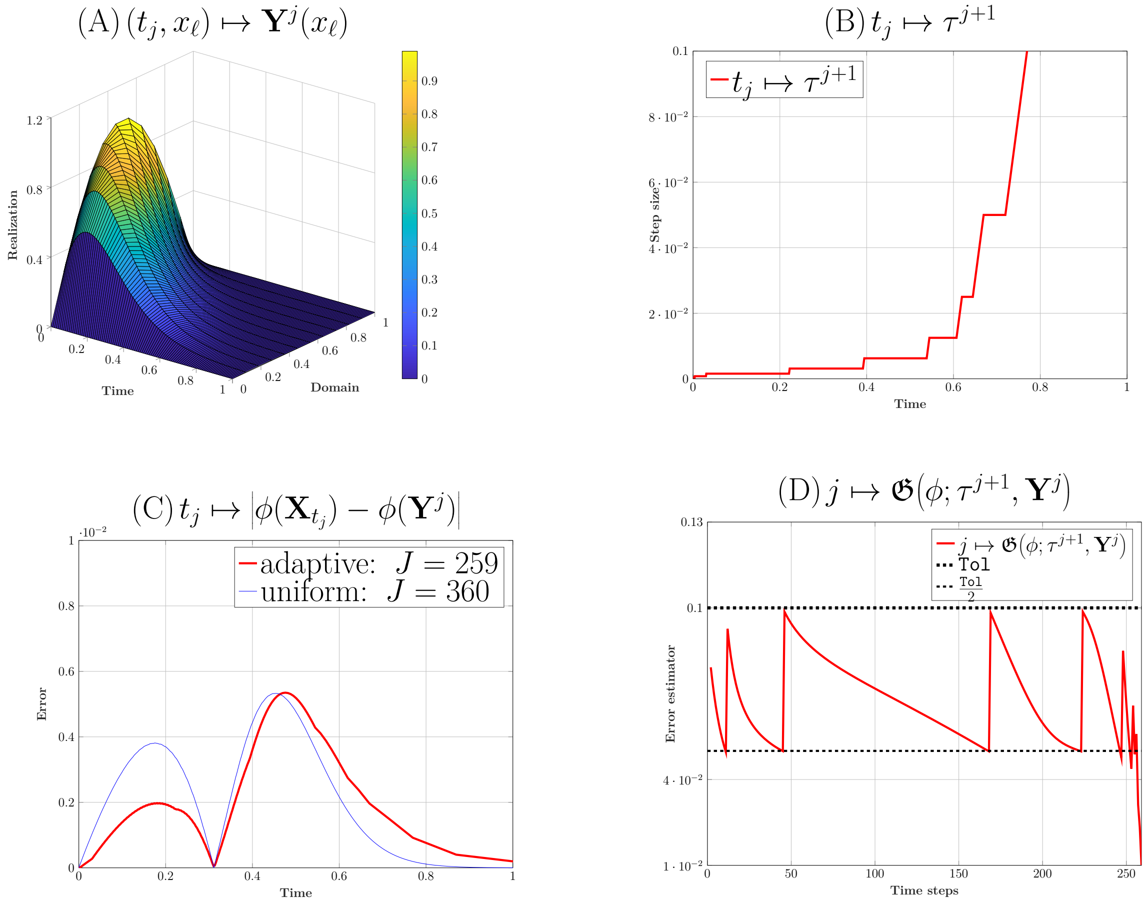


Figure 5.4.: (**Setup A** with `Tol` $= 0.1$, $\boldsymbol{\beta} = 2$, $\varepsilon = 2h$ and $\phi(\mathbf{x}) = \sqrt{h}\|\mathbf{x}\|_{\mathbb{R}^L}$) (A) Contour plot of the solution. (B) Plot of the corresponding adaptive time step size. (C) Error for uniform (—) vs. adaptive (—) time meshes via Algorithm 4.1. (D) Plot of the *a posteriori* weak error estimator $\mathfrak{G}\Big(\phi; \tau^{j+1}, \mathbf{Y}^j\Big)$.

threshold `Tol` in Figure 5.5 is larger compared to Figure 5.4, which is due to the different scalings of the considered norms.
Different choices of $\boldsymbol{\beta}$ and ε affect the amount of total steps generated via Algorithm 4.1. A larger size of $\boldsymbol{\beta}$ increases the convection effect and leads to more time steps; see Figures 5.8 and 5.4. In turn, larger values of ε reduce the transport, which requires fewer time steps, see Figure 5.7. For different parameters $\varepsilon \in \{h, 2h, 5h, 15h, 30h, 1\}$, Figure 5.3 compares the total amount of time steps needed for adaptive and uniform meshes to perform equally well, indicating that the smaller ε is, the more savings are obtained via the adaptive Algorithm 4.1, if compared to scheme (0.3) with uniform time steps.

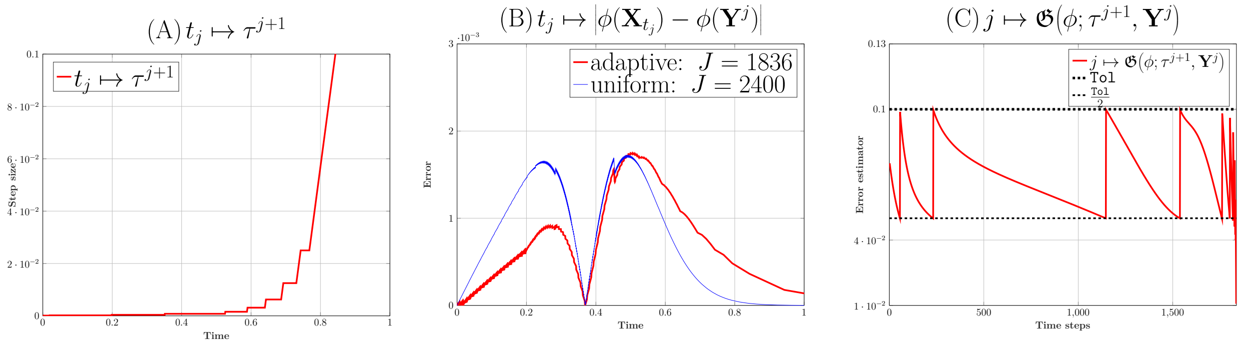


Figure 5.5.: (**Setup A** with $\texttt{Tol} = 0.1$, $\boldsymbol{\beta} = 2$, $\varepsilon = 2h$ and $\phi(\mathbf{x}) = \|\mathbf{x}\|_\infty$) (A) Plot of the corresponding adaptive time step size. (B) Error for uniform (—) vs. adaptive (—) time meshes via Algorithm 4.1. (C) Plot of the *a posteriori* weak error estimator $\mathfrak{G}\big(\phi; \tau^{j+1}, \mathbf{Y}^j\big)$.

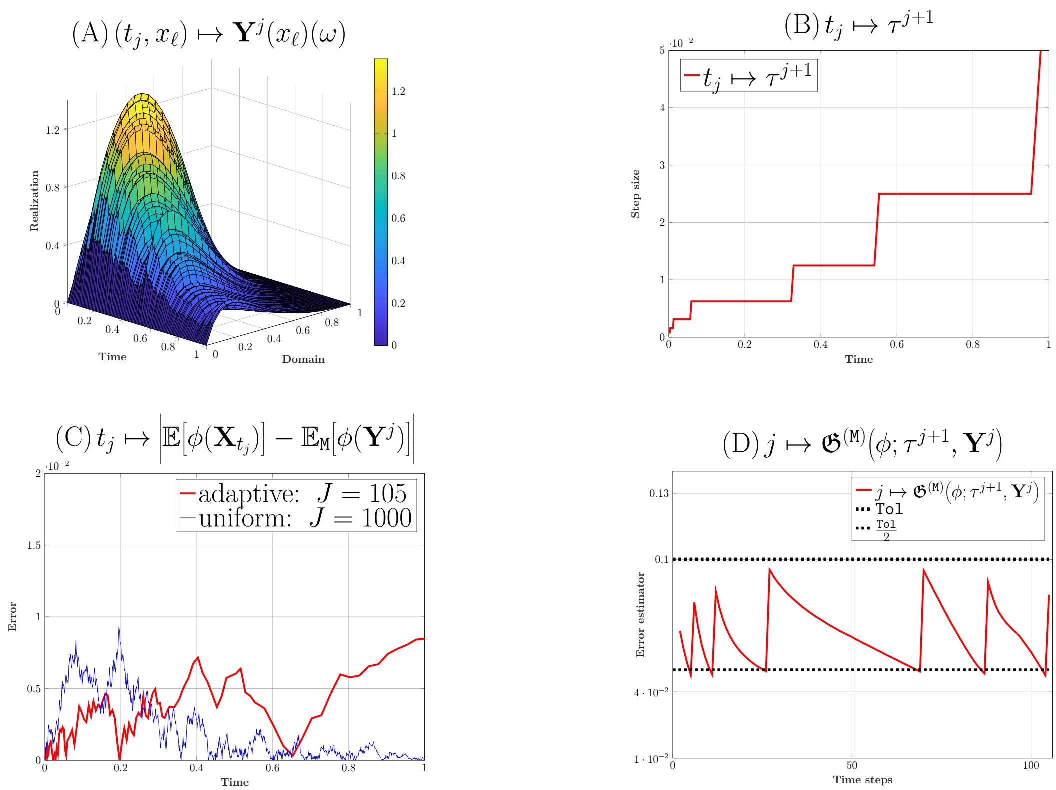


Figure 5.6.: (**Setup B** with $\texttt{Tol} = 0.1$, $\mathtt{M} = 10^4$, $\boldsymbol{\beta} = 1$, $\varepsilon = 2h$ and $\phi(\mathbf{x}) = \sqrt{h}\|\mathbf{x}\|_{\mathbb{R}^L}$) (A) Contour plot of the solution for a single realization ω. (B) Plot of the corresponding adaptive time step size. (C) Error for uniform (—) vs. adaptive (—) time meshes via Algorithm 4.1. (D) Plot of the *a posteriori* weak error estimator $\mathfrak{G}^{(\mathtt{M})}\big(\phi; \tau^{j+1}, \mathbf{Y}^j\big)$.

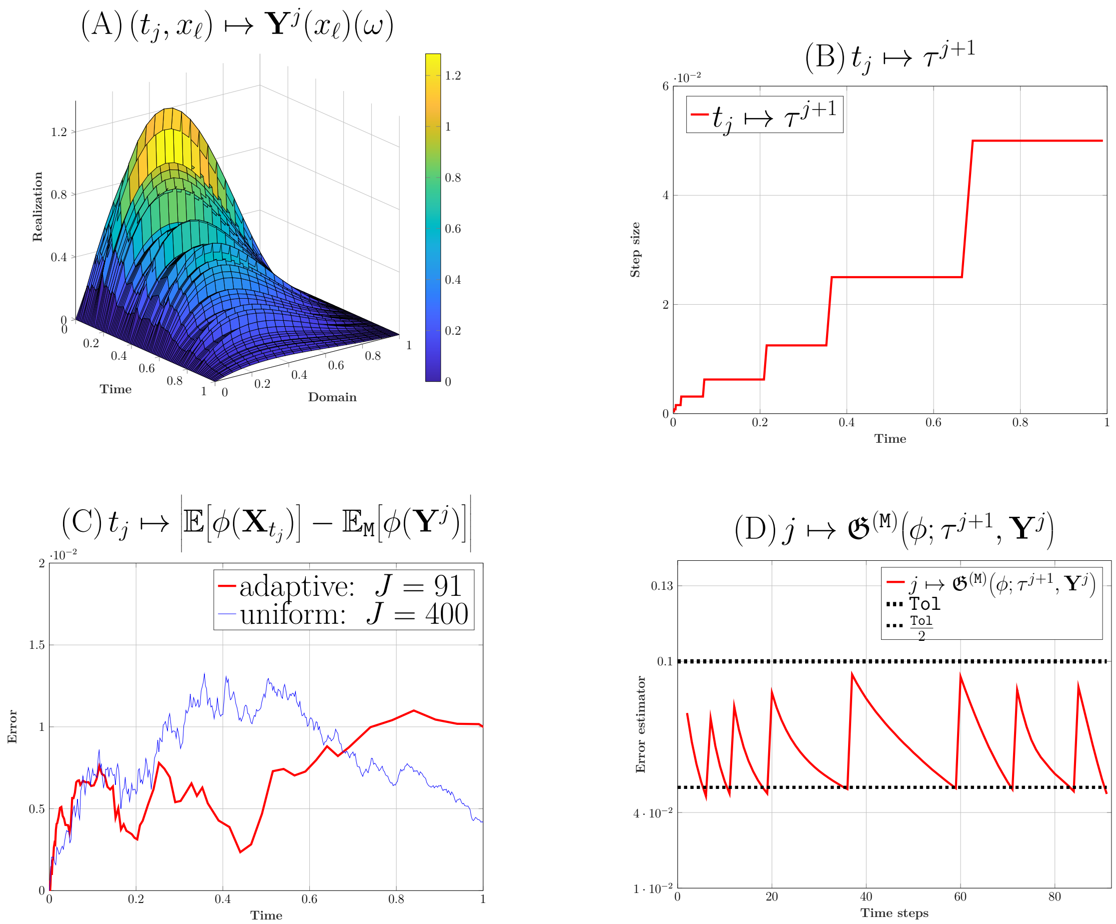


Figure 5.7.: (**Setup B** with $\texttt{Tol} = 0.1$, $\texttt{M} = 10^4$, $\boldsymbol{\beta} = 1$, $\varepsilon = 5h$ and $\phi(\mathbf{x}) = \sqrt{h}\|\mathbf{x}\|_{\mathbb{R}^L}$) (A) Contour plot of the solution for a single realization ω. (B) Plot of the corresponding adaptive time step size. (C) Error for uniform (—) vs. adaptive (—) time meshes via Algorithm 4.1. (D) Plot of the *a posteriori* weak error estimator $\mathfrak{G}^{(\texttt{M})}\big(\phi; \tau^{j+1}, \mathbf{Y}^j\big)$.

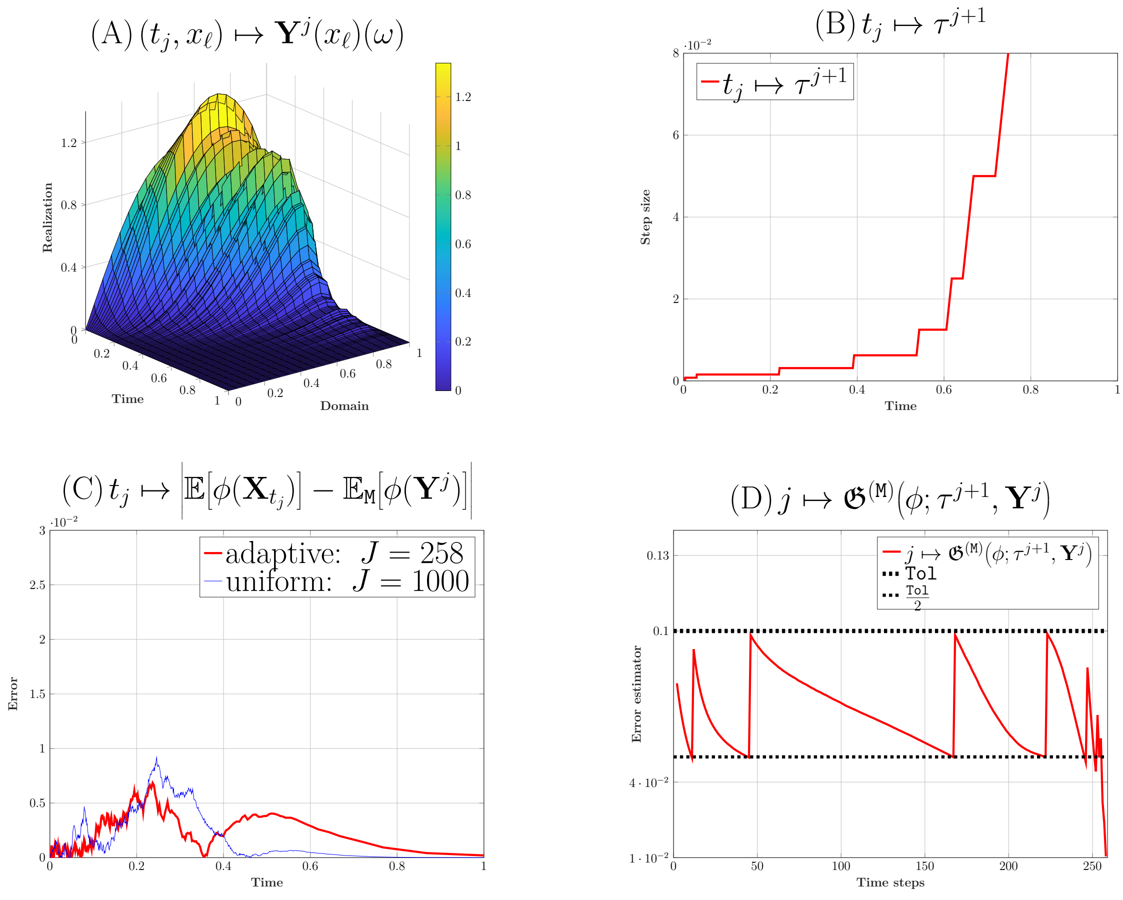


Figure 5.8.: (**Setup B** with $\mathtt{Tol} = 0.1$, $\mathtt{M} = 10^4$, $\boldsymbol{\beta} = -2$, $\varepsilon = 2h$ and $\phi(\mathbf{x}) = \sqrt{h}\|\mathbf{x}\|_{\mathbb{R}^L}$) (A) Contour plot of the solution for a single realization ω. (B) Plot of the corresponding adaptive time step size. (C) Error for uniform (—) vs. adaptive (—) time meshes via Algorithm 4.1. (D) Plot of the *a posteriori* weak error estimator $\mathfrak{G}^{(\mathtt{M})}\left(\phi; \tau^{j+1}, \mathbf{Y}^j\right)$.

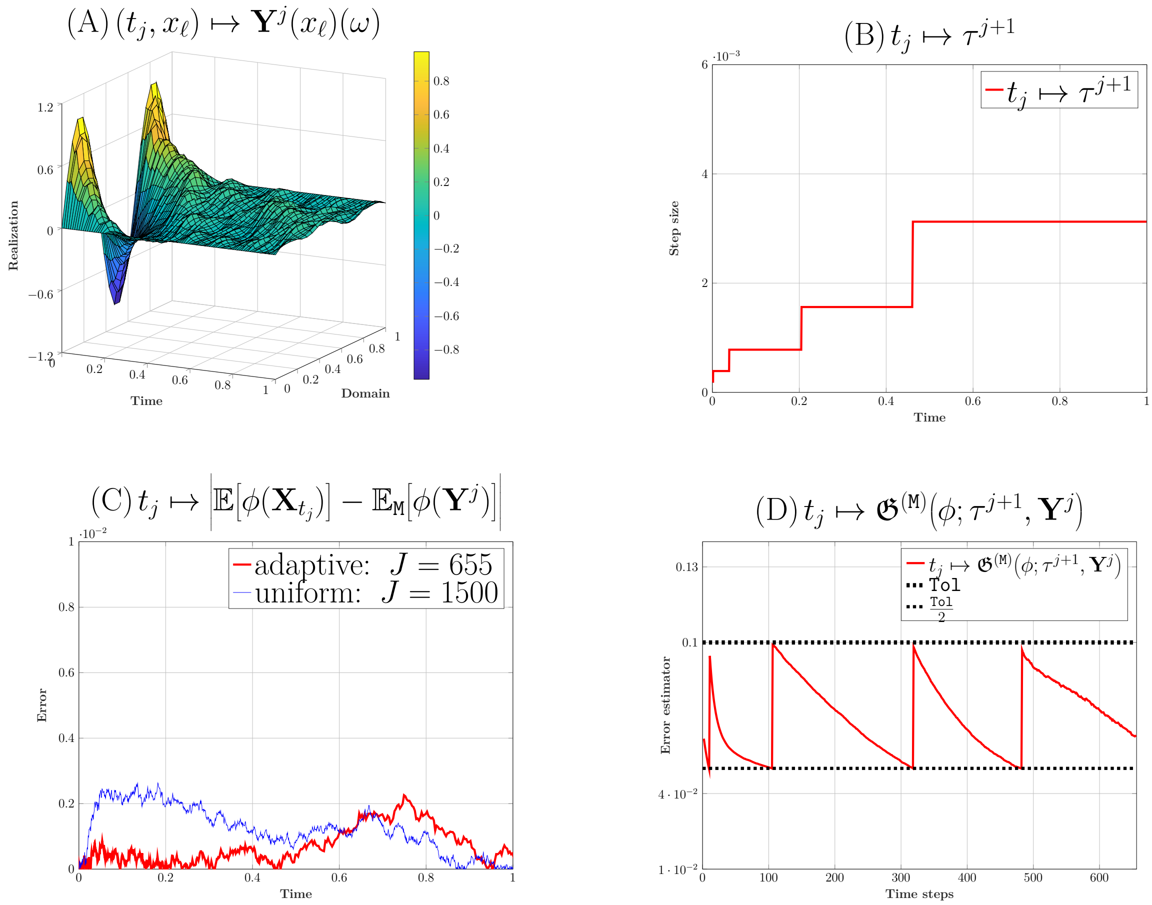


Figure 5.9.: (**Setup C** with $\mathtt{Tol} = 0.1$, $\mathtt{M} = 10^4$, $\beta = 1$, $\varepsilon = 2h$ and $\phi(\mathbf{x}) = \sqrt{h}\|\mathbf{x}\|_{\mathbb{R}^L}$) (A) Contour plot of the solution for a single realization ω. (B) Plot of the corresponding adaptive time step size. (C) Error for uniform (—) vs. adaptive (—) time meshes via Algorithm 4.1. (D) Plot of the *a posteriori* weak error estimator $\mathfrak{G}^{(\mathtt{M})}\left(\phi; \tau^{j+1}, \mathbf{Y}^j\right)$.

Part II.

A posteriori error analysis and adaptivity for high-dimensional elliptic and parabolic boundary value problems

6. Introduction

The following aspects directly tie in with items (*i*), (*ii*) and (*iii*) in the introduction in the beginning of this thesis.

ad (i). The derivation of (0.12) conceptually follows the guideline of Theorem 3.1 in the first part, where an *a posteriori* (weak) error estimate is presented for the (semi-implicit) Euler method, which uses (unbounded) Wiener increments; in fact, $\mathfrak{G}_1^{(\cdot)}$ in (0.12) is conceptually close to the estimator in (3.1). While Theorem 3.1 considers the *Kolmogorov* PDE (see (0.5), and also (6.1)) on the whole space $\boldsymbol{\mathcal{D}} = \mathbb{R}^L$, we here consider bounded domains $\boldsymbol{\mathcal{D}} \subset \mathbb{R}^L$, which requires the proper numerical approximation of the stopping time $\boldsymbol{\tau}^{\mathbf{x}}$ in (0.9) when $\mathbf{X}^{\mathbf{x}}$ crosses the boundary $\boldsymbol{\partial\mathcal{D}}$. To this end, the *weak* Euler method in Scheme 2 in combination with the corresponding 'stopping'-mechanism enables a successive (local) construction of iterates $\{\mathbf{Y}_{\mathbf{X}}^j\}_{j\geq 0}$ up to the boundary $\boldsymbol{\partial\mathcal{D}}$, where all of them lie in $\overline{\boldsymbol{\mathcal{D}}}$: in this respect, we denote by $\boldsymbol{\mathcal{S}}_{\tau^{j+1}} \subset \overline{\boldsymbol{\mathcal{D}}}$, $j \geq 0$, the set of points which are close to $\boldsymbol{\partial\mathcal{D}}$. We characterize this 'boundary strip' via the verification:

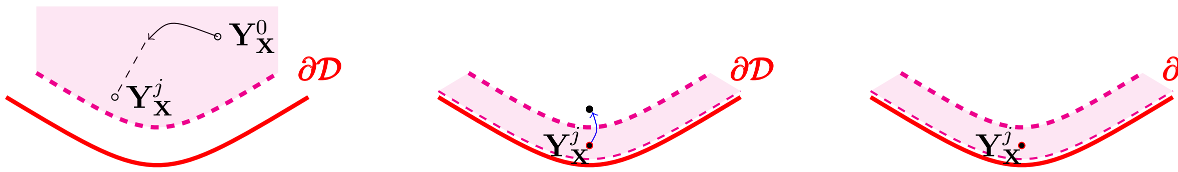
- if $\mathrm{d}(\mathbf{Y}_{\mathbf{X}}^j, \boldsymbol{\partial\mathcal{D}}) := \inf\{\|\mathbf{Y}_{\mathbf{X}}^j - \mathbf{v}\|_{\mathbb{R}^L} \,|\, \mathbf{v} \in \boldsymbol{\partial\mathcal{D}}\} \geq \lambda_j \sqrt{\tau^{j+1}}$, then $\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}$, and hence $\mathbf{Y}_{\mathbf{X}}^{j+1} \in \overline{\boldsymbol{\mathcal{D}}}$,
- if $0 < \mathrm{d}(\mathbf{Y}_{\mathbf{X}}^j, \boldsymbol{\partial\mathcal{D}}) < \lambda_j \sqrt{\tau^{j+1}}$, then $\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}$,

for a suitable number $\lambda_j > 0$; see Section 8.3 for a proper choice. Once $\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}$, it is either projected onto $\boldsymbol{\partial\mathcal{D}}$ and the procedure stops or is 'bounced back' to the interior of $\boldsymbol{\mathcal{D}}$ (with some probability). This different treatment of realizations of $\mathbf{Y}_{\mathbf{X}}^j$ via Scheme 2 is reflected in the error estimators $\{\mathfrak{G}_\ell^{(\cdot)}\}_{\ell=1}^3$ in (0.12) (see Figure 6.1 below): those which contribute to the functional $\mathfrak{G}_1^{(j)}$ take positions in $\boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}$; in contrast, $\mathfrak{G}_3^{(j)}$ accounts for those in the boundary strip $\boldsymbol{\mathcal{S}}_{\tau^{j+1}}$, while $\mathfrak{G}_2^{(j)}$ assembles the subset of those realizations, which bounce back to the interior of $\boldsymbol{\mathcal{D}}$.
We illustrate the role of the different estimators $\{\mathfrak{G}_\ell^{(\cdot)}\}_{\ell=1}^3$ in (0.12) for a prototype PDE (0.6).

Example 6.1. Let $L \in \mathbb{N}$ and $\boldsymbol{\mathcal{D}} = \Big\{\mathbf{x} \in \mathbb{R}^L \,:\, \|\mathbf{x}\|_{\mathbb{R}^L} < 1\Big\}$. Consider (0.6) with $\mathbf{b}(\mathbf{x}) \equiv \mathbf{0}$, $\boldsymbol{\sigma}(\mathbf{x}) \equiv \sqrt{\frac{2}{L}} \cdot \mathbb{I}$, where $\mathbb{I}$ denotes the $L-$dimensional identity matrix, $c(\mathbf{x}) \equiv 0$. Then, $\{\mathfrak{G}_\ell^{(j)}\}_{\ell=1}^3$ in Theorem 9.1 are $(j \geq 0)$[1]:

$$\mathfrak{G}_1^{(j)} = \frac{1}{2} \cdot \mathbb{E}\Big[\mathbf{1}_{\Big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\Big\}} \cdot \Big\|\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^j\Big\|_{\mathbb{R}^L}^2\Big],$$

[1] Here, given an event $\mathcal{A}$, $\mathbf{1}_{\mathcal{A}} : \Omega \to \{0,1\}$ denotes the indicator function of $\mathcal{A}$, *i.e.*, $\mathbf{1}_{\mathcal{A}}(\omega) = 1$ if $\omega \in \mathcal{A}$, and $\mathbf{1}_{\mathcal{A}}(\omega) = 0$ if $\omega \notin \mathcal{A}$.



(a) $\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}$

(b) $\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}$, and 'bouncing back'

(c) $\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}$

Figure 6.1.: Realizations which (a) contribute to $\mathfrak{G}_1^{(j)}$, (b) to $\mathfrak{G}_2^{(j)}$, and (c) to $\mathfrak{G}_3^{(j)}$.

$$\mathfrak{G}_2^{(j)} = \frac{1}{2} \cdot \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}} \mathbf{1}_{\left\{\overline{\mathbf{Y}}_{\mathbf{X}}^j = \mathbf{Y}_{\mathbf{X}}^j + \lambda_j \sqrt{\tau^{j+1}} \boldsymbol{n}\left(\Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\right)\right\}} \cdot \left\| \mathbf{Y}_{\mathbf{X}}^{j+1} - \overline{\mathbf{Y}}_{\mathbf{X}}^j \right\|_{\mathbb{R}^L}^2 \Big],$$

$$\mathfrak{G}_3^{(j)} = 2 \cdot \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\Big].$$

$\mathfrak{G}_1^{(j)}$ accounts for incremental changes within the interior $\boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}$. The terms in $\mathfrak{G}_3^{(j)}$ account for those iterates that have already entered $\boldsymbol{\mathcal{S}}_{\tau^{j+1}}$, and where the event of a 'projection' resp. 'bouncing back' is about to happen next. The terms in $\mathfrak{G}_2^{(j)}$ account for those realizations in $\boldsymbol{\mathcal{S}}_{\tau^{j+1}}$ which will be 'bounced back'.

ad (ii). Once the *a posteriori* error estimate has been established in Theorem 9.1, we analyze its convergence behavior along sequences of shrinking meshes with maximum mesh size $\tau^{max} > 0$. The result in Theorem 9.6 shows an optimal rate of convergence, and thus recovers the well-known *a priori* estimate for iterates $\{(\mathbf{Y}_{\mathbf{X}}^j, Y_V^j, Y_Z^j)\}_{j=0}^{J^*}$ of Scheme 2; see [62, p. 369, Thm. 3.4]. In fact, the presence of the estimator $\{\mathfrak{G}_2^{(j)}\}_{j\geq 0}$ is crucial to validate order 1; in fact, if it would be removed from the estimator, and an immediate projection onto $\partial\boldsymbol{\mathcal{D}}$ of an iterate in the boundary strip would occur, only a convergence order $\frac{1}{2}$ of the reduced *a posteriori* error estimate may be expected; this conclusion may be drawn from the *a priori* error analysis in [62, p. 370, Rem. 3.5] where this selective 'bouncing back/projection'-mechanism was conceived. Further crucial tools in the proof of Theorem 9.6 are stability results in Section 8.3 for the boundedness of visits in the boundary strips $\left\{\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}_{j\geq 0}$, and the discrete stopping time (see Lemmas 8.3 and 8.4), which generalize related stability results in [62, p. 367, Lem. 3.2] and [62, p. 367, Lem. 3.2] to non-uniform time steps; see also Remark 8.1 for further details.

ad (iii). In Chapter 10, the *a posteriori* error estimate (0.12) is used to construct an adaptive method (see Algorithm 10.1) which automatically selects deterministic (local) step sizes $\tau^{j+1} = t_{j+1} - t_j$ in every iteration step. For this purpose, given some tolerance $\texttt{Tol} > 0$ and $j \geq 0$, we check via iterated refinement/coarsening of the current step size τ^{j+1} whether the partial sum $\sum_{k=0}^{j} \tau^{k+1}\left\{\mathfrak{G}_1^{(k)} + \mathfrak{G}_2^{(k)} + \mathfrak{G}_3^{(k)}\right\}$ is below (a multiple of) a pre-assigned tolerance $\texttt{Tol} > 0$; see (10.1). If compared to Algorithm 4.1 in the first part, the main difficulty for the boundary value problem (0.6) here is to set up a thresholding criterion that properly addresses the 'discrete stopping', for which the stability results in Lemmas 8.3 and 8.4 hold. Theorem 10.2 then validates computation of each new time step τ^{j+1} in Algorithm 10.1 after *finitely* many iterations (*i.e.*, in $\mathcal{O}\Big(\log(\mathtt{Tol}^{-1})\Big)$), at most $\mathbb{E}\Big[J^*\Big] = \mathcal{O}(\mathtt{Tol}^{-1})$ many steps to (globally) terminate, and a weak error convergence order $\mathcal{O}(\mathtt{Tol})$.

The following example from [13] illustrates efficient local mesh refinement–coarsening by the adaptive Algorithm 10.1 for $L \gg 1$.

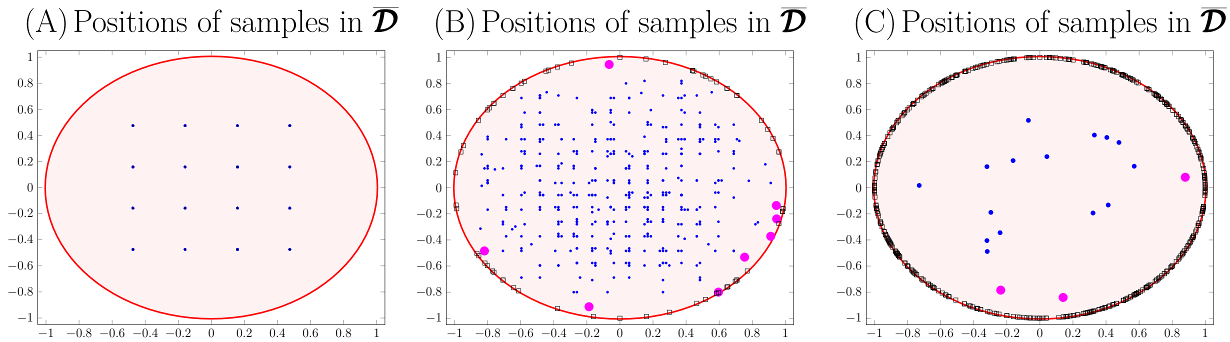


Figure 6.2.: Example 6.2 for $L = 2$: Temporal evolution of positions of samples in $\overline{\boldsymbol{\mathcal{D}}}$: • samples in the interior of $\boldsymbol{\mathcal{D}}$; • samples in the corresponding boundary strips; □ samples on $\boldsymbol{\partial\mathcal{D}}$.

Example 6.2 (see [13])**.** Let $L = 10$ and $\boldsymbol{\mathcal{D}} := \Big\{\mathbf{x} \in \mathbb{R}^L \,:\, \|\mathbf{x}\|_{\mathbb{R}^L} < 1\Big\}$. Consider (0.6) with $\boldsymbol{\sigma}(\mathbf{x}) = \mathbb{I}$, $\mathbf{b}(\mathbf{x}) \equiv \mathbf{0}$, $c(\mathbf{x}) \equiv 0$, $g(\mathbf{x}) \equiv 1$ and $\phi(\mathbf{x}) \equiv 0$. Fix $\mathbf{x} = \mathbf{0}$. We use Algorithm 10.1 (with $\mathtt{Tol} = 0.005$, $\mathtt{M} = 10^4$) to get the approximation $\mathtt{u}^{(\mathtt{M})}(\mathbf{x})$ of the solution $u(\mathbf{x}) = \frac{1}{L}\Big(1 - \|\mathbf{x}\|^2_{\mathbb{R}^L}\Big)$. Here,

$$\mathtt{u}^{(\mathtt{M})}(\mathbf{x}) := \mathbb{E}_{\mathtt{M}}\Big[\phi(\mathbf{Y}_{\mathbf{X}}^{J^*})Y_V^{J^*} + Y_Z^{J^*}\Big] := \frac{1}{\mathtt{M}}\sum_{m=1}^{\mathtt{M}} \phi\Big(\mathbf{Y}_{\mathbf{X}}^{J^*,m}\Big)Y_V^{J^*,m} + Y_Z^{J^*,m} \quad (\mathbf{x} \in \boldsymbol{\mathcal{D}})$$

denotes the empirical mean to approximate $\mathbb{E}\Big[\phi(\mathbf{Y}_{\mathbf{X}}^{J^*})Y_V^{J^*} + Y_Z^{J^*}\Big]$. The initial refinement and gradual coarsening of the step sizes ('U'-profile) in Figure 6.3 (A) is a typical consequence of Algorihm 10.1 allowing for an interaction between informations from the (empirical) error estimators $\{\mathfrak{G}_\ell^{(\cdot),(\mathtt{M})}\}_{\ell=1}^3$ and a minor weightening of 'outlier-samples' according to the shape of the distribution of the stopping time t_{J^*}; see Figure 6.3 (B), and also Figure 6.2. In a comparative consideration of Figures 6.3 (A), (B), and (C), first samples enter the boundary strips at time ≈ 0.025 and hence (possibly) get projected onto $\boldsymbol{\partial\mathcal{D}}$, which is why we observe a refinement of step sizes up to this time. Within the time interval $[0.025, 0.125]$, most of the samples hit $\boldsymbol{\partial\mathcal{D}}$ which involves fine step sizes in this region to reach a certain level of accuracy regulated by the choice of $\mathtt{Tol}$. Those samples, which have not been stopped before time 0.125 may be considered as 'outlier-samples' which most likely spoil the approximation. The mechanism in Algorithm 10.1 automatically allows a gradual coarsening of related step sizes for the generation of these leftover samples, which increases the width of their boundary strips, and hence forces their immediate projection onto $\partial\boldsymbol{\mathcal{D}}$, *i.e.*, a stopping of Algorithm 10.1. Moreover, Algorithm 10.1 is efficient to reach the same accuracy (**Error** ≈ 0.002, $\texttt{Tol} = 0.005$, $\texttt{M} = 10^4$, $\mathbf{x} = \mathbf{0}$); the needed number of steps to terminate in Algorithm 10.1 resp. the empirical mean of the stopping index J^* is $\max_{m=1,\dots,\texttt{M}} J^*(\omega_m) = 642$ (CPU time: 243 sec) resp. $\mathbb{E}_{\texttt{M}}[J^*] \approx 362$ — opposed to $\max_{m=1,\dots,\texttt{M}} J^*(\omega_m) = 3757$ (CPU time: 800 sec) resp. $\mathbb{E}_{\texttt{M}}[J^*] \approx 957$ for Scheme 2 on a uniform mesh. Hence, automatic mesh size selection which leans on *where* current states realize (*i.e.*, in the interior, where only $\mathfrak{G}_1^{(\cdot)}$ is active or close to the boundary, where $\mathfrak{G}_2^{(\cdot)}$ and $\mathfrak{G}_3^{(\cdot)}$ adjust proper scaling) highly increases the efficiency of Scheme 2.

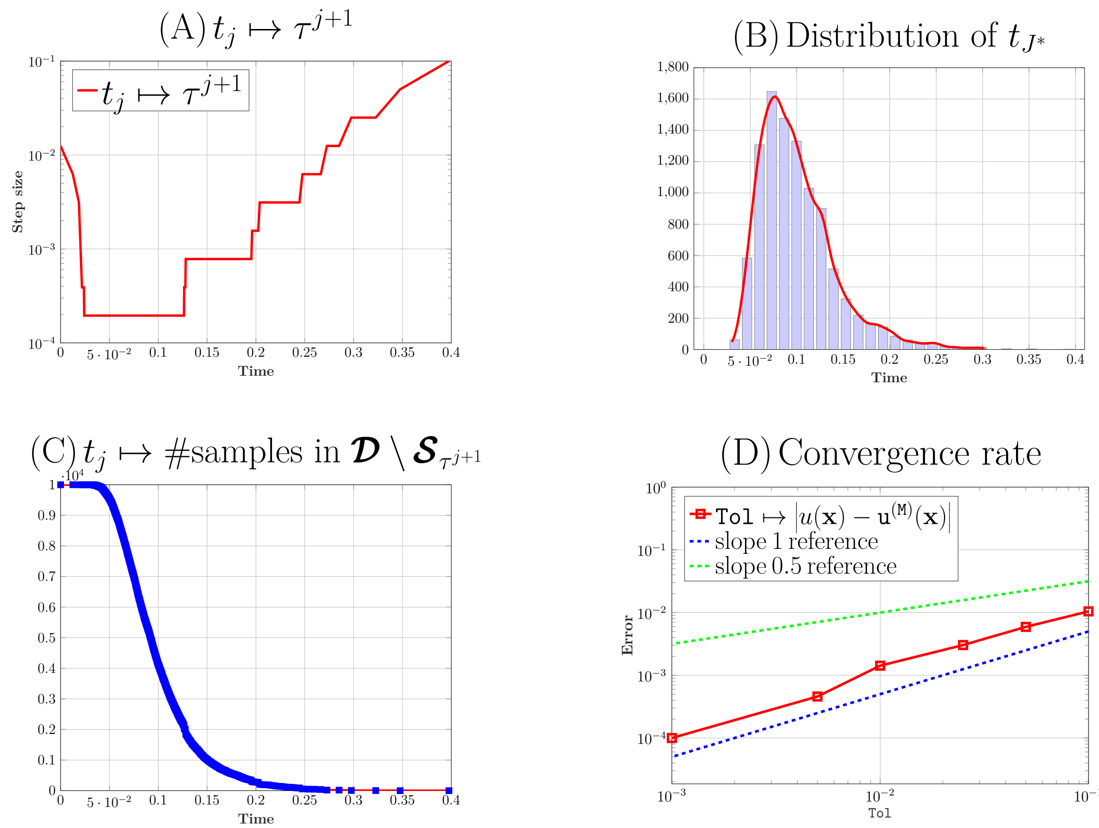


Figure 6.3.: (A) Semi-Log plot of the (adaptive) step sizes generated via Algorithm 10.1. (B) Shape of the distribution of t_{J^*} illustrated via a histogram plot. (C) Temporal evolution of (sample-)iterates in the interior of $\boldsymbol{\mathcal{D}}$. (D) Convergence rate (error) Log-log plot via Algorithm 10.1 ($\texttt{M} = 10^5$, $\mathbf{x} = \mathbf{0}$).

Secondly, we focus on the parabolic PDE with proper terminal and Dirichlet boundary data,

$$\partial_t u(t,\mathbf{x}) + \frac{1}{2}\mathrm{Tr}\Big(\boldsymbol{\sigma}(\mathbf{x})\boldsymbol{\sigma}^\top(\mathbf{x})D^2_{\mathbf{x}}u(t,\mathbf{x})\Big) + \big\langle \mathbf{b}(\mathbf{x}), D_{\mathbf{x}}u(t,\mathbf{x})\big\rangle_{\mathbb{R}^L} + g(t,\mathbf{x}) = 0 \qquad \forall\,(t,\mathbf{x}) \in [0,T)\times\boldsymbol{\mathcal{D}}\,, \tag{6.1a}$$

$$u(T,\mathbf{x}) = \phi(T,\mathbf{x}) \qquad \forall\,\mathbf{x}\in\overline{\boldsymbol{\mathcal{D}}}\,, \tag{6.1b}$$

$$u(t,\mathbf{x}) = \phi(t,\mathbf{x}) \qquad \forall (t,\mathbf{x}) \in [0,T)\times\boldsymbol{\partial\mathcal{D}}\,, \tag{6.1c}$$

where additionally $T>0$, and $g:[0,T]\times\overline{\boldsymbol{\mathcal{D}}}\to\mathbb{R}$, $\phi:[0,T]\times\overline{\boldsymbol{\mathcal{D}}}\to\mathbb{R}$. Under proper settings of data stated in Section 8.2, there exists a unique classical solution $u:[0,T]\times\overline{\boldsymbol{\mathcal{D}}}\to\mathbb{R}$ of problem (6.1), which has the following probabilistic representation; see *e.g.* [62, p. 340]:

$$u(t,\mathbf{x}) = \mathbb{E}\Big[\phi\Big(\boldsymbol{\tau}^{t,\mathbf{x}}, \mathbf{X}^{t,\mathbf{x}}_{\boldsymbol{\tau}^{t,\mathbf{x}}}\Big) + Z_{\boldsymbol{\tau}^{t,\mathbf{x}}}\Big] \qquad \forall (t,\mathbf{x}) \in [0,T)\times\boldsymbol{\mathcal{D}}\,, \tag{6.2}$$

where

1) $\mathbf{X}^{t,\mathbf{x}} \equiv \{\mathbf{X}^{t,\mathbf{x}}_s;\ s\in[t,T]\}$ denotes the $\mathbb{R}^L$-valued solution of the SDE

$$\mathrm{d}\mathbf{X}_s = \mathbf{b}(\mathbf{X}_s)\mathrm{d}s + \boldsymbol{\sigma}(\mathbf{X}_s)\mathrm{d}\mathbf{W}_s \quad \forall s\in(t,T]\,, \qquad \mathbf{X}_t = \mathbf{x}\in\boldsymbol{\mathcal{D}}\subset\mathbb{R}^L\,, \tag{6.3}$$

starting at time $t\in[0,T)$ in $\mathbf{x}\in\boldsymbol{\mathcal{D}}$, and the first exit time of $\mathbf{X}^{t,\mathbf{x}}$ from $\boldsymbol{\mathcal{D}}$ is

$$\boldsymbol{\tau}^{t,\mathbf{x}} := \inf\Big\{s>t\ :\ \mathbf{X}^{t,\mathbf{x}}_s \notin \boldsymbol{\mathcal{D}} \text{ or } s\notin(t,T)\Big\}\,. \tag{6.4}$$

2) $Z\equiv\{Z_s\,;\ s\in[t,T]\}$ denotes the $\mathbb{R}$-valued solution of the (random) (ODE)

$$\mathrm{d}Z_s = g(s,\mathbf{X}^{t,\mathbf{x}}_s)\mathrm{d}s \quad \forall s\in(t,T]\,, \qquad Z_t = 0\,. \tag{6.5}$$

If compared to deterministic numerical methods — see also Section 7.2 —, a conceptional advantage of probabilistic numerical methods which approximate (6.1) is that only *one* (temporal) discretization parameter is needed. Consequently, main structural tools which lead to *a posteriori* error estimate (0.12) for (0.6) now may easily be adopted to approximate (6.2), and evenly so for the later construction of an adaptive method; for $(t,\mathbf{x})\in[0,T)\times\boldsymbol{\mathcal{D}}$ fixed, the *a posteriori* error estimate on a given mesh $\{t_j\}_{j=0}^J\subset[t,T]$ with local mesh sizes $\{\tau^{j+1}\}_{j=0}^{J-1}$ for iterates $\{(\mathbf{Y}^j_{\mathbf{X}}, Y^j_Z)\}_{j=0}^{J^*}$ from Scheme 3 to approximate (6.2) takes again the form

$$\Big|u(t,\mathbf{x}) - \mathbb{E}\Big[\phi(t_{J^*}, \mathbf{Y}^{J^*}_{\mathbf{X}}) + Y^{J^*}_Z\Big]\Big| \le \mathfrak{C}(\phi,g)\cdot\sum_{j=0}^{J-1}\tau^{j+1}\Big\{\mathfrak{H}^{(j)}_1 + \mathfrak{H}^{(j)}_2 + \mathfrak{H}^{(j)}_3\Big\}\,, \tag{6.6}$$

with $\mathfrak{C}(\phi,g)>0$, $J\equiv J(t,\mathbf{x})\in\mathbb{N}$, (computable) *a posteriori* error estimators $\{\mathfrak{H}^{(\cdot)}_\ell\}_{\ell=1}^3$, and $0\le J^*\equiv J^*(t,\mathbf{x})\le J$ the stopping index; see Theorem 9.7.
The following example details $\{\mathfrak{H}^{(\cdot)}_\ell\}_{\ell=1}^3$ in (6.6) for a prototype PDE (6.1).

Example 6.3. Let $L \in \mathbb{N}$ and $\boldsymbol{\mathcal{D}} = \Big\{\mathbf{x} \in \mathbb{R}^L \,:\, \|\mathbf{x}\|_{\mathbb{R}^L} < 1\Big\}$. Consider (6.1) with $\mathbf{b}(\mathbf{x}) \equiv \mathbf{0}$, $\boldsymbol{\sigma}(\mathbf{x}) \equiv \sqrt{\frac{2}{L}} \cdot \mathbb{I}$, $g(t,\mathbf{x}) \equiv 0$, and ϕ smooth. Then,

$$\begin{aligned}
\mathfrak{H}_1^{(j)} &= \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}} \cdot \Big\|\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^j\Big\|_{\mathbb{R}^L}^2\Big] + 2 \cdot \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big] \cdot \tau^{j+1}\,, \\
\mathfrak{H}_2^{(j)} &= \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}} \mathbf{1}_{\big\{\overline{\mathbf{Y}}_{\mathbf{X}}^j = \mathbf{Y}_{\mathbf{X}}^j + \lambda_j \sqrt{\tau^{j+1}} \boldsymbol{n}\big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\big)\big\}} \cdot \Big\|\mathbf{Y}_{\mathbf{X}}^{j+1} - \overline{\mathbf{Y}}_{\mathbf{X}}^j\Big\|_{\mathbb{R}^L}^2\Big] \\
&\quad + 2 \cdot \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}} \mathbf{1}_{\big\{\overline{\mathbf{Y}}_{\mathbf{X}}^j = \mathbf{Y}_{\mathbf{X}}^j + \lambda_j \sqrt{\tau^{j+1}} \boldsymbol{n}\big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\big)\big\}}\Big] \cdot \tau^{j+1}\,, \\
\mathfrak{H}_3^{(j)} &= 2 \cdot \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big]\,.
\end{aligned}$$

The three estimators take similar roles as in Example 6.1 for (0.6).

The following items $(i) - (iii)$ comment on the construction of (6.6), its convergence analysis, and use to construct an adaptive method.

- (i) If compared to Scheme 2 for the elliptic problem (0.6), Scheme 3 (see Section 8.4) exploits an additional observance of 'stopping' when there is no projection onto $\boldsymbol{\partial\mathcal{D}}$ before the terminal time $T > 0$, but is similar elsewise. Consequently, the form of (6.6) is close to (0.12).
- (ii) If compared to (0.12), the convergence analysis of (6.6) along sequences of shrinking meshes with a maximum mesh size simplifies since the stopping time $\boldsymbol{\tau}^{t,\mathbf{x}}$ in (6.4) is $\mathbb{P}-a.s.$ bounded by the terminal time $T > 0$, which, in particular, avoids a related stability result concerning 'discrete-stopping'. In fact, only Lemma 8.5 is needed, which is an analogue of Lemma 8.3 in the elliptic setting.
- (iii) Similar to Algorithm 10.1 in Chapter 10, we construct an adaptive time-stepping algorithm (see Algorithm 10.3) based on (6.6), for which we prove (again) local and global termination, as well as optimal convergence in terms of a given tolerance $\mathtt{Tol} > 0$. The (successive) step size selection procedure in Algorithm 10.3 proceeds in the same way as in Algorithm 10.1: given $\mathtt{Tol} > 0$ and $j \geq 0$, the current step size τ^{j+1} is (automatically) generated (via iterated refinement/coarsening), such that the partial sum $\sum_{k=0}^{j} \tau^{k+1}\Big\{\mathfrak{H}_1^{(k)} + \mathfrak{H}_2^{(k)} + \mathfrak{H}_3^{(k)}\Big\}$ is below $\mathtt{Tol}$ times a specified 'temporal weight', which grows with t_j, but is bounded by means of the stability result in Lemma 8.5. In the fully practical implementation of Algorithm 10.3 (as well as Algorithm 10.1), where arising expectations are approximated by Monte-Carlo method, the 'temporal weight' gradually forces those leftover samples, which have not been projected onto $\boldsymbol{\partial\mathcal{D}}$ with the majority of samples, to a projection. These 'forced' projections are obtained by enlarging corresponding boundary strips through a gradual coarsening of the step sizes (see Examples 6.2 and 6.4, and also Figure 6.2). We refer to Chapters 10 and 11 for further details.

The following example from [56, Experiment 7.1] illustrates local mesh refinement-coarsening by Algorithm 10.3.

Example 6.4 (see [56, Experiment 7.1])**.** Let $T = 1$, and $\boldsymbol{\mathcal{D}} := \Big\{\mathbf{x} = (x_1, x_2)^\top \in \mathbb{R}^2 : \|\mathbf{x}\|_{\mathbb{R}^2} < 1\Big\}$. Consider (6.1) with

$$\boldsymbol{\sigma}(\mathbf{x}) \equiv \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \ \mathbf{b}(\mathbf{x}) = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}, \ g(t,\mathbf{x}) = 5\Big(1 - e^{-(T-t)}\Big) - \Big(25 - x_1^2 - x_2^2\Big)e^{-(T-t)}\,,$$
$$\phi(t,\mathbf{x}) = \Big(25 - x_1^2 - x_2^2\Big)\Big(1 - e^{-(T-t)}\Big)\,.$$

The corresponding solution is given by $u(t,\mathbf{x}) = \Big(25 - x_1^2 - x_2^2\Big)\Big(1 - e^{-(T-t)}\Big)$. We fix $(t,\mathbf{x}) = (0,\mathbf{0})$, and use Algorithm 10.3 (with $\mathtt{Tol} = 0.01$, $\mathtt{M} = 10^4$) to approximate $u(t,\mathbf{x})$ by $\mathtt{u}^{(\mathtt{M})}(t,\mathbf{x})$. Illustrated in Figure 6.4 below, the methodology of Algorithm 10.3 allowing for interactions between $\{\mathfrak{H}_\ell^{(\cdot),(\mathtt{M})}\}_{\ell=1}^3$ and a less weightening of 'outlier-samples' is conceptually similar to Algorithm 10.1: we observe a refinement of step sizes (within $[0, 0.05]$) till first samples hit $\boldsymbol{\partial\mathcal{D}}$; fine step sizes are needed within $[0.05, 0.4]$, where most samples are projected onto $\boldsymbol{\partial\mathcal{D}}$; afterwards, we observe a gradual coarsening of the step sizes (within $[0.4, 1]$) to force 'outlier-samples' to hit $\boldsymbol{\partial\mathcal{D}}$ resp. to proceed to the terminal time T as fast as possible. Furthermore, Algorithm 10.3 is (also) efficient to reach the same accuracy (**Error** ≈ 0.015, $\mathtt{Tol} = 0.02$, $\mathtt{M} = 10^4$, $(t,\mathbf{x}) = (0,\mathbf{0})$); the needed number of steps to terminate in Algorithm 10.3 resp. the empirical mean of the stopping index J^* is $\max\limits_{m=1,\dots,\mathtt{M}} J^*(\omega_m) = 709$ (CPU time: 297 sec) resp. $\mathbb{E}_{\mathtt{M}}[J^*] \approx 129$ — as opposed to $\max\limits_{m=1,\dots,\mathtt{M}} J^*(\omega_m) = 8000$ (CPU time: 1500 sec) resp. $\mathbb{E}_{\mathtt{M}}[J^*] \approx 1607$ for Scheme 3 on a uniform mesh.

This part in organized as follows: Chapter 7 provides a survey of existing (adaptive) methods for the approximation of the elliptic, as well as the parabolic PDE. Chapter 8 collects the assumptions needed for the data in (0.6) resp. (6.1), recalls *a priori* bounds for the solution of (0.6) resp. (6.1) and presents Schemes 1 to 3, as well as corresponding stability results. The *a posteriori* error estimates (0.12) and (6.6) are derived in Chapter 9, where also its optimal convergence orders are shown. The related adaptive methods are proposed and analyzed in Chapter 10. Chapter 11 presents computational studies.

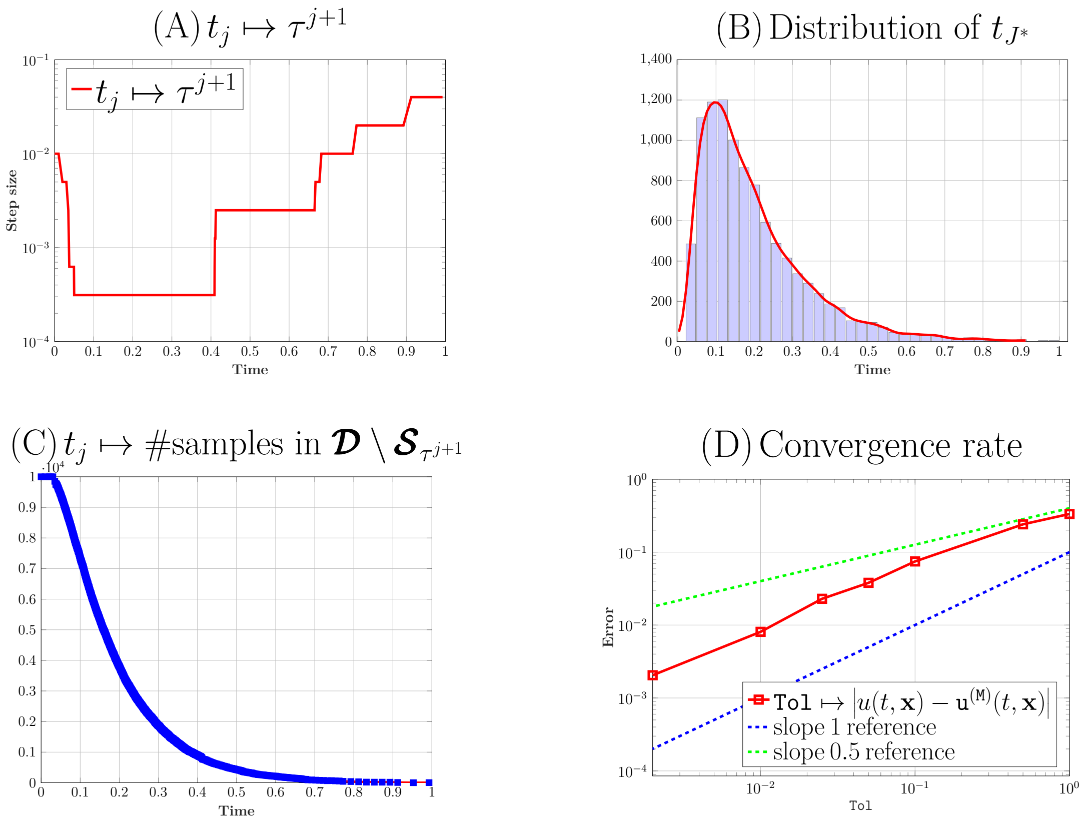


Figure 6.4.: (A) Semi-Log plot of the (adaptive) step sizes generated via Algorithm 10.3. (B) Shape of the distribution of t_{J^*} illustrated via a histogram plot. (C) Temporal evolution of (sample-)iterates in the interior of $\mathcal{D}$. (D) Convergence rate (error) Log-log plot via Algorithm 10.3 ($\mathtt{M} = 10^5$, $(t, \mathbf{x}) = (0, \mathbf{0})$).

7. A short review of A posteriori error analysis and Adaptivity

Deterministic methods to solve PDE's (0.6) and (6.1) usually employ meshes to resolve the state space, and their implementation usually is complicated. In contrast, probabilistic methods are meshless, comparatively easier to implement, and still are applicable in high dimensions L. Their efficiency increases rapidly with the recent emergence of modern (parallel) GPU architectures; see [45]. The main goal in this chapter is to survey some existing representative directions in the *a posteriori* error analysis and *adaptive numerical methods* for the (initial-)boundary value problems (0.6) and (6.1).

7.1. Probabilistic methods to discretize high-dimensional PDE's

In the literature, there exist different numerical methods for (0.7) or (6.2), which may be seen as examples of more general '*stopped diffusion*' problems. Most of them use the explicit Euler method, *i.e.*, (8.1) where '$\boldsymbol{\xi}_{j+1}\sqrt{\tau^{j+1}}$' is replaced by the Wiener increment '$\mathbf{W}_{t_{j+1}} - \mathbf{W}_{t_j}$'; see (9.57). The main difficulty then is to accurately compute the (discrete) stopping time (0.9) resp. (6.4), when the related (discrete) solution path leaves the domain $\boldsymbol{\mathcal{D}}$. This problem becomes even more prominent when the related first exit time $\tilde{\boldsymbol{\tau}}$ of the (abstract) continuified Euler process $\boldsymbol{\mathcal{Y}}^{\mathbf{X}}$ (see (9.58)) is compared in this context on an interval $[t_j, t_{j+1}]$ — where trajectories may exit $\boldsymbol{\mathcal{D}}$ even though all discrete (explicit) Euler iterates lie in $\boldsymbol{\mathcal{D}}$; see Figure 8.1 (a) below. An *a priori* error analysis therefore cuts the expectable convergence rate from 1 to $\frac{1}{2}$; see [38, Thm. 2.3]. To recover optimal order, more simulations are needed close to the boundary to accurately capture discrete stopping. First works in this direction are [38, 39], which prove optimal convergence order 1. To use the method in [39], the exit probability of $\boldsymbol{\mathcal{Y}}^{\mathbf{X}}$ leaving the domain, *i.e.*, the probability that $\tilde{\boldsymbol{\tau}}$ lies in a time interval specified by two (consecutive) grid points needs be available explicitly, which is only known for certain domains (*e.g.* when $\boldsymbol{\mathcal{D}}$ is a half-space; see *e.g.* [39]). For general underlying domains $\boldsymbol{\mathcal{D}}$, this approach needs be combined with local transformations of $\partial\boldsymbol{\mathcal{D}}$ to be successful.
In order to avoid local charts close to $\partial\boldsymbol{\mathcal{D}}$, the '*boundary shifting method*' is presented in [40] which shrinks the domain $\boldsymbol{\mathcal{D}}$ to generate more frequent exits. If compared to the '*Brownian bridge method*' above, an explicit formula for exit probabilities of $\boldsymbol{\mathcal{Y}}^{\mathbf{X}}$ is not required anymore, which broadens the applicability of the method to more general domains. The corresponding error analysis guarantees order $o(\sqrt{\Delta t})$, while computations evidence order 1. For further, different strategies to ensure accurate 'stopping', we also refer to [60, 13].

The methods that we discussed so far were supplemented by *a priori* error analysis; to our knowledge, the only work that addresses *a posteriori* error analysis in this setting is [28]. For $g \equiv 0$ in (6.1), and based on an (asymptotic) weak *a posteriori* error expansion with computable leading order term, a time-stepping method is proposed which generates global stochastic (adaptive) meshes to approximate (6.2). However, these (random) mesh generations are only based on the *computable part* of the underlying error expansion, and adaptivity here thus remains heuristic. The corresponding derivation uses computable exit probabilities (similar to [38, 39]), and is elsewise similar to the procedure in [76, 64]: the derivation rests on the weak error expansion via PDE (6.1), which practically involves numerical approximations of derivatives of the (unknown) solution u of (6.1), whose simulation is only feasible in small dimensions L. The computational experiments in this work indicate a convergence order 1, but no theoretical results are known that support these observations.
The methods above are primarily addressing the (efficient) approximation (6.1) rather than (0.6). We here mention the works [11, 12, 14] which computationally study the approximation of (0.6) by extending the ideas from [39]: there, $\boldsymbol{\tau}^{\mathbf{x}}$ in (0.9) is (accurately) approximated by sampling from a distribution, which is constructed by means of the related exit probability. Computational studies with the corresponding method evidence an improved convergence order 1 as well.
These schemes all use the (explicit) Euler method with unbounded Wiener increments. From a practical viewpoint however, the *weak* Euler method in (8.1) is an alternative option, as it uses *bounded* random variables in every iteration step (see Scheme 1 below) to avoid *overshootings* outside the domain by controlling the steps up to the boundary; in particular the work [61] verifies first order convergence for (8.1) in an associated scheme (very close to Scheme 2 resp. 3), which — except for a projection onto $\partial\boldsymbol{\mathcal{D}}$ resp. bouncing back to the inside of $\boldsymbol{\mathcal{D}}$ — does not require further adjustments of its iterates close to $\partial\boldsymbol{\mathcal{D}}$. We use this simple, fully practical method (8.1) within Schemes 2 and 3 in Chapter 9 to provide computable right-hand sides in an *a posteriori* error analysis, which may then be used to set up an adaptive time-stepping strategy based on it in Chapter 10[1].

7.2. Deterministic adaptive methods in low dimension — AFEM

Adaptive finite element methods (AFEM) base an automatic adjustment of a given mesh $\mathcal{T}_0$ covering $\boldsymbol{\mathcal{D}} \subset \mathbb{R}^L$ on an *a posteriori error estimator* $\eta^2(u_{\mathcal{T}_0}, \mathcal{T}_0) \equiv \eta^2 = \sum_{T^{m,0} \in \mathcal{T}_0} \eta^2_{T^{m,0}}$ for the computed approximation $u_{\mathcal{T}_0} : \overline{\boldsymbol{\mathcal{D}}} \to \mathbb{R}$ for (0.6) on $\mathcal{T}_0$. FEM is a general deterministic Galerkin method for (0.6) posed on an arbitrary *low-dimensional* domain $\boldsymbol{\mathcal{D}}$; in practice, its use then leads to large coupled algebraic systems to be inverted by means of advanced iterative solvers, where its performance crucially hinges on the ellipticity of (0.6). AFEM extends these concepts, by trying to optimally distribute nodal mesh points across $\boldsymbol{\mathcal{D}}$, guided

[1] For comparison, we give an *a posteriori* error analysis for the usual Euler method (9.57) in Section 9.3, where an upper error bound is given in terms of the stopping time $\tilde{\boldsymbol{\tau}}$ of the continuified Euler process $\boldsymbol{\mathcal{Y}}^{\mathbf{X}}$ (9.58) — which requires an approximation via additional computations; see Remark 9.3.

by local $\{\eta_{T^{m,0}};\, T^{m,0} \in \mathcal{T}_0\}$ where $\eta_T \equiv \eta_T(u_{\mathcal{T}_0}\big|_{T^{m,0}}, T^{m,0})$, while aiming for optimal accuracy under fixed computational costs; it is an iterative method which repeatedly refines meshes *locally* and thus generates a family of *nested* $\{\mathcal{T}_\ell\}_{\ell=1}^{\ell^*}$ — until the related approximate $u_{\mathcal{T}_{\ell^*}} : \overline{\boldsymbol{\mathcal{D}}} \to \mathbb{R}$ of (0.6) fulfills a certain threshold criterion.
Existing AFEM mainly uses Hilbert space methods to derive an (residual-based) a posteriori estimator $\eta(u_{\mathcal{T}_\ell}, \mathcal{T}_\ell)$ to upperly bound the error $u - u_{\mathcal{T}_\ell}$ in the 'energy norm', with an unknown factor (*e.g.*, Poincare's constant reflecting stability properties of (0.6), and another one which accounts for admitted triangulations; see [29, 65]). For AFEM, this estimate then suggests the following loop

$$\texttt{Solve} \longrightarrow \texttt{Estimate} \longrightarrow \texttt{Mark} \longrightarrow \texttt{Refine} \tag{7.1}$$

to automatically generate a sequence of (increasingly more) specific, (locally) refined meshes $\{\mathcal{T}_{h^\ell}\}_\ell$, starting from a coarse mesh $\mathcal{T}_0$, which all cover $\mathcal{D}$: for a given $\mathcal{T}_\ell$, we

1) (`Solve`) first compute $u_{\mathcal{T}_\ell}$ with the help of direct or indirect solvers (*e.g.*, PCG, multigrid method, GMRES, or BICG) that solves a large linear system. Then

2) (`Estimate`) the estimator $\eta(u_{\mathcal{T}_\ell}, \mathcal{T}_\ell)$ is computed to decide whether or not $u_{\mathcal{T}_\ell}$ is sufficiently accurate, and/or $\mathcal{T}_\ell$ should be refined or not. Based on the estimator alone is

3) (`Mark`) 'Dörfler's marking criterion' (see (7.2) below), which selects those elements $\widetilde{\mathcal{T}}_\ell := \{T^{m,\ell} \in \mathcal{T}_\ell\}$ which are up to refinement.

4) (`Refine`) Only mesh refinement is admitted to obtain the new nested mesh $\mathcal{T}_{\ell+1}$ — via the 'newest bisection method' that splits the marked elements in **3)**.

It remained open until [27] to show that tuple $\{(u_{\mathcal{T}_\ell}, \mathcal{T}_\ell)\}_\ell$ obtained from **1)** – **4)** meet a pre-assigned error tolerance within finite steps ℓ^*: next to the assumption of a sufficiently fine initial $\mathcal{T}_0$ and the 'one interior node'-condition in **4)**, the convergence proof for AFEM in [27] for Poisson's problem rests on 'Dörfler's strategy marking':

$$\text{Find a subset } \widetilde{\mathcal{T}}_\ell \subset \mathcal{T}_\ell : \qquad \eta(u_{\mathcal{T}_\ell}, \widetilde{\mathcal{T}}_\ell) \geq \theta \eta(u_{\mathcal{T}_\ell}, \mathcal{T}_\ell) \tag{7.2}$$

for a fixed $0 < \theta < 1$, which ensures that sufficiently many elements from $\mathcal{T}_\ell \equiv \{T^{m,\ell}\}_m$ are chosen that constitute a fixed proportion of the global error estimator [65]. The work [27] initiated a whole series of works to broaden convergence results for more general AFEM of sort (7.1), *s.t.* the relevant contraction property remains valid, which is

$$\exists\, \beta \in (0,1): \qquad \int_{\boldsymbol{\mathcal{D}}} |\nabla(u - u_{\mathcal{T}_\ell})|^2 \,\mathrm{d}x \leq C\beta^\ell \qquad (\ell \in \mathbb{N}),$$

with a generic constant $C \equiv C(\boldsymbol{\mathcal{D}}) > 0$ that depends on $\boldsymbol{\mathcal{D}}$ and admitted mesh geometries: to *e.g.* remove in [18] the (too costly) 'one interior node'-condition in **4)**, next to required sufficiently fine initial $\mathcal{T}_0$, the concept of 'total error' was central. Another direction generalizes the convergence property of AFEM to nonsymmetric linear-elliptic, and even quasi-linear problems; see *e.g.* [33, 8]. Next to the contraction property, 'mesh optimality' is a crucial property for AFEM to have, which bounds the number of degrees of freedom $N_{\ell^*} = \sharp \mathcal{T}_{\ell^*}$ in the terminating mesh $\mathcal{T}_{\ell^*}$: the first work in this direction is [10], which shows optimal convergence rates (in terms of N_{ℓ^*}; for the Poisson problem) for a certain AFEM which included a crucial coasening step; this step was later removed by a modified approach in [75]. For a further discussion of 'mesh optimality' for AFEM we refer to [66], and [17] where sufficient criteria are identified which ensure optimal convergence rates for a general AFEM; and to *e.g.* [33, 8] for more general (PDE's). We remark that the proof of 'mesh optimality' usually requires $\theta \in (0, \theta^*)$ in (7.2), for θ^* sufficiently small to bound the number of marked elements in step **3)** — whose value is not explicit for actual simulations. We refer to [26] for a further discussion.

These concepts are applied in [20, 53, 34] to construct adaptive methods based on the implicit Euler method for the heat equation, as a special example of the evolutionary PDE (6.1): for every $n \geq 0$, to (iteratively) find the new time step τ_n, and *then* the spatial mesh $\mathcal{T}_n$ to cover $\boldsymbol{\mathcal{D}}$, different error indicators are identified which subsequently (and thus independently) address these goals. These indicators are space-time *localizations* of computable terms in the *a posteriori error estimate* [79], see also [53, Thm. 3.1], whose derivation is based on the concept of weak (variational) solution for (6.1), to bound the error in the *global* Bochner norm $L^2(0,T;\mathbb{H}^1_0) \cap W^{1,2}(0,T;\mathbb{H}^{-1})$. As a consequence, given $n \geq 0$, and τ_n, the construction of a mesh $\mathcal{T}_n$ to approximate the solution of

$$\frac{1}{\tau_n} u^n + \mathcal{L} u^n = \frac{1}{\tau_n} u^{n-1} + g\,,$$

where $\mathcal{L} = -\Delta$, via the convergent AFEM strategy (7.1) is then possible. However, the subtle interplay of *different* spatial and temporal scales and the *decoupled* treatment of related error occuring in each time step makes the construction of an efficient adaptive method for (6.1) more challenging in this parabolic case (see also [34]): also, we have to make sure that

(i) τ_n may iteratively be constructed via a *finite* sequence $\{\tau_{n,\ell}\}_{\ell \geq 0}^{\ell_n^*}$, and that

(ii) the final time T is reached after *finitely* many steps, *i.e.*, there exists $N \in \mathbb{N}$: $\tau_N \geq T$

to conclude convergence of the adaptive method. The adaptive method in [20] satisfies (i) but lacks (ii); a first convergent method is given in [53], where each time-step starts with a possible coarsening of $(\tau_{n-1}, \mathcal{T}_{n-1})$, and only refinements afterwards; a *uniform* energy estimate for iterates is now employed to determine a (uniform) minimum admissible time step for each n to meet the error tolerance, and thus show termination of the adaptive method, *i.e.*, property (ii) — although with nonoptimal complexity bounds.

7.3. Deterministic methods to discretize high-dimensional PDE's — tensor sparsity

For $\boldsymbol{\mathcal{D}} = (0,1)^L$, *mesh-based* methods (such as FEM used in Section 7.2) to *e.g.* solve PDE (0.6) suffer the *curse of dimensionality*: for N the number of points on a uniform mesh per dimension, the number of related *nodal basis* functions is $\mathcal{O}(N^L)$, which grows exponentially with the dimension L. *Sparse grids* on hypercubes $\boldsymbol{\mathcal{D}} = (0,1)^L$ drastically cut down this complexity of a full mesh to $\mathcal{O}(N|\log(N)|^{L-1})$ many grid points: they discard those elements of a *hierarchical basis* in tensor product form which have small support, such that no loss of approximation power for sufficiently smooth solutions of PDE (0.6) occurs, see *e.g.* [15]; for the heat equation in (6.1) and rough initial data in tensor form, graded time meshes properly address this requirement for sparse spatial grids in [70]. As *e.g.* detailed in [15, 71], the efficient use of sparse grids for high dimensions L requires a restricted data setting $(\mathbf{b}, \boldsymbol{\sigma}, c, g, \phi, \boldsymbol{\mathcal{D}})$: for non-constant elliptic operators $\mathcal{L}$ including convection, or a domain that is not of tensor structure, as well as non-constant (Dirichlet-) boundary data partly non-trivial extensions are necessary, and those setups of data typically lower accuracy, and reachable L; see the discussion in [78]. Also, a theoretical backup for local adaptive mesh adjustments (see [15, 71]) that preserve optimal complexity *as L increases* is less developed. To approach even larger dimensions L based on tensor product representations for approximate solutions of PDE (0.6) with 'Laplacian like operator', the construction of a proper (sub-)set of basis functions will be part in the *low rank approximation* method itself; see *e.g.* [5] for a recent survey. We also mention [41, 74], where its complexity is compared with sparse grids, and smoothness of the function was again found to be crucial for the efficiency of the low rank approximation. According to [2], its efficiency crucially hinges on the differential operator in PDE (0.6) to 'have a simple tensor product structure', and that (L-dependent) ranks, whose optimal value is not evident in general [3] should be chosen properly; see also [5, Sect. 5.3]. In fact, related theoretical discussions in [23] for the high-dimensional PDE (0.6) with constant, symmetric elliptic operators $\mathcal{L} \equiv -\mathrm{div}(\boldsymbol{A}\nabla u)$ with $\boldsymbol{A} \in \mathbb{R}^{L\times L}_{\mathtt{diag}}$ conclude the transfer of tensor-sparsity from data to solutions, which motivates low-rank tensor format approximations for the solution of PDE (0.6) in those cases; but such a structural transfer may get lost in the case of stronger couplings [4] for general $\boldsymbol{A} \in \mathbb{R}^{L\times L}_{\mathtt{spd}}$, demanding higher ranks for a proper approximation.

While current research on deterministic methods for large L mainly focuses on the efficient use of 'tensor-sparsity respecting data' to fight the 'curse of dimensionality', we base the construction of easily implementable, adaptive methods to solve PDE's (0.6) resp. (6.1) on their probabilistic reformulations (0.7) resp. (6.2): general domains $\boldsymbol{\mathcal{D}} \subset \mathbb{R}^L$, and elliptic differential operators $\mathcal{L} \equiv \mathcal{L}(\mathbf{x})$ in (0.6) resp. (6.1) are admitted, which appear in physical applications in particular, for which convergence with optimal rates for the related adaptive Algorithms 10.1 and 10.3 that base on *a posteriori* error estimators will provide a theoretical backup.

8. Assumptions and Tools

Section 8.1 lists basic requirements on data $\mathbf{b}, \boldsymbol{\sigma}, c, g, \phi$ in (0.6), which guarantee the existence of a unique classical solution $u : \overline{\boldsymbol{\mathcal{D}}} \to \mathbb{R}$ of (0.6); see *e.g.* [35, Ch. 6]. Moreover, we recall bounds for $\{D^\ell_{\mathbf{x}} u\}_{\ell=1}^4$ of (0.6). In almost the same manner, Section 8.2 presents assumptions on data $\mathbf{b}, \boldsymbol{\sigma}, g, \phi$ in (6.1), which ensure the existence of a unique classical solution u of (6.1); see *e.g.* [54, p. 320, Thm. 5.2]. Moreover, we recall bounds for $\{D^\ell_{\mathbf{x}} u\}_{\ell=1}^4$ of (6.1).

For a sufficiently smooth $\varphi \in \mathcal{C}(\mathbb{R}^L; \mathbb{R}^n)$, corresponding (matrix) operator norms are given as ($n, L \in \mathbb{N}$, $\mathbf{x} \in \mathbb{R}^L$)

$$\|D^\ell \varphi(\mathbf{x})\|_{\mathcal{L}\Big(\underbrace{\mathbb{R}^L \times ... \times \mathbb{R}^L}_{\ell-\text{times}};\mathbb{R}^n\Big)} := \sup_{\|\mathbf{v}_i\|_{\mathbb{R}^L}=1} \|D^\ell \varphi(\mathbf{x})(\mathbf{v}_1, ..., \mathbf{v}_\ell)\|_{\mathbb{R}^n} \qquad (\ell \in \mathbb{N}),$$

where $\|\cdot\|_{\mathbb{R}^n}$ denotes the (Euclidean) vector norm of a $\mathbb{R}^n$-valued vector. If $n = L$, we write $\mathcal{L}^\ell \equiv \mathcal{L}\Big(\mathbb{R}^L \times ... \times \mathbb{R}^L; \mathbb{R}^L\Big)$. If $n = 1$, $D \equiv D_{\mathbf{x}}$ denotes the gradient and $D^2 \equiv D^2_{\mathbf{x}}$ the Hessian matrix of φ, and we also write $\mathcal{L}^\ell \equiv \mathcal{L}\Big(\mathbb{R}^L \times ... \times \mathbb{R}^L; \mathbb{R}\Big)$. Moreover, $\|D_{\mathbf{x}}\varphi(\mathbf{x})\|_{\mathcal{L}^1} = \|D_{\mathbf{x}}\varphi(\mathbf{x})\|_{\mathbb{R}^L}$, $\|D^2_{\mathbf{x}}\varphi(\mathbf{x})\|_{\mathcal{L}^2} = \|D^2_{\mathbf{x}}\varphi(\mathbf{x})\|_{\mathbb{R}^{L\times L}}$, where $\|\cdot\|_{\mathbb{R}^{L\times L}}$ denotes the spectral (matrix) norm.

For $k \in \mathbb{N}$ and $\beta \in (0,1)$, we denote by $\mathcal{C}^{k+\beta}(\overline{\boldsymbol{\mathcal{D}}}; \mathbb{R})$ the Banach space consisting of continuous functions v in $\boldsymbol{\mathcal{D}}$, with continuous derivatives up to order k in $\overline{\boldsymbol{\mathcal{D}}}$, such that

$$\|v\|^{(k+\beta)}_{\boldsymbol{\mathcal{D}}} := \sum_{j=0}^{k} \sum_{|\boldsymbol{j}'|=j} \sup_{\mathbf{x}\in\boldsymbol{\mathcal{D}}} |\partial^{\boldsymbol{j}'}_{\mathbf{x}} v(\mathbf{x})| + \sum_{|\boldsymbol{j}'|=k} \sup_{\mathbf{x},\mathbf{y}\in\boldsymbol{\mathcal{D}},\, \mathbf{x}\neq\mathbf{y}} \frac{|\partial^{\boldsymbol{j}'}_{\mathbf{x}} v(\mathbf{x}) - \partial^{\boldsymbol{j}'}_{\mathbf{x}} v(\mathbf{y})|}{\|\mathbf{x}-\mathbf{y}\|^\beta_{\mathbb{R}^L}} < \infty,$$

or rather

$$\|v\|^{(k,\beta)}_{\boldsymbol{\mathcal{D}}} := \sum_{j=0}^{k} \sup_{|\boldsymbol{j}'|=j} \sup_{\mathbf{x}\in\boldsymbol{\mathcal{D}}} |\partial^{\boldsymbol{j}'}_{\mathbf{x}} v(\mathbf{x})| + \sup_{|\boldsymbol{j}'|=k} \sup_{\mathbf{x},\mathbf{y}\in\boldsymbol{\mathcal{D}},\, \mathbf{x}\neq\mathbf{y}} \frac{|\partial^{\boldsymbol{j}'}_{\mathbf{x}} v(\mathbf{x}) - \partial^{\boldsymbol{j}'}_{\mathbf{x}} v(\mathbf{y})|}{\|\mathbf{x}-\mathbf{y}\|^\beta_{\mathbb{R}^L}} < \infty,$$

where the above summation is taken over all multi-index $\boldsymbol{j}'$ of length $|\boldsymbol{j}'|$.
In a similar manner, we denote by $\mathcal{C}^{(k+\beta)/2,k+\beta}([0,T] \times \overline{\boldsymbol{\mathcal{D}}}; \mathbb{R})$ the Banach space consisting of continuous functions w in $[0,T) \times \boldsymbol{\mathcal{D}}$, with continuous derivatives up to order k in $[0,T] \times \overline{\boldsymbol{\mathcal{D}}}$, such that

$$\begin{aligned}\|w\|^{(k+\beta)}_{[0,T)\times\boldsymbol{\mathcal{D}}} := &\sum_{j=0}^{k} \sum_{\substack{2r+|\boldsymbol{j}'|=j\\ r\in\mathbb{N}_0}} \sup_{(t,\mathbf{x})\in[0,T)\times\boldsymbol{\mathcal{D}}} |\partial^r_t \partial^{\boldsymbol{j}'}_{\mathbf{x}} w(t,\mathbf{x})| \\ &+ \sum_{\substack{2r+|\boldsymbol{j}'|=k\\ r\in\mathbb{N}_0}} \sup_{\substack{(t,\mathbf{x}),(t,\mathbf{y})\in[0,T]\times\overline{\boldsymbol{\mathcal{D}}}\\ \mathbf{x}\neq\mathbf{y}}} \frac{|\partial^r_t \partial^{\boldsymbol{j}'}_{\mathbf{x}} w(t,\mathbf{x}) - \partial^r_t \partial^{\boldsymbol{j}'}_{\mathbf{x}} w(t,\mathbf{y})|}{\|\mathbf{x}-\mathbf{y}\|^\beta_{\mathbb{R}^L}}\end{aligned}$$

$$+\sum_{\substack{2r+|\boldsymbol{j}'|=k\\ r\in\mathbb{N}_0}}\ \sup_{\substack{(t,\mathbf{x}),(s,\mathbf{x})\in[0,T]\times\overline{\boldsymbol{\mathcal{D}}}\\ t\neq s}} \frac{|\partial_t^r\partial_{\mathbf{x}}^{\boldsymbol{j}'}w(t,\mathbf{x})-\partial_t^r\partial_{\mathbf{x}}^{\boldsymbol{j}'}w(s,\mathbf{x})|}{|t-s|^\beta}<\infty\,,$$

see also [54, p. 2 *ff.*] for further details.

8.1. The elliptic PDE (0.6): assumptions and bounds for $\{D_{\mathbf{x}}^\ell u\}_{\ell=1}^4$

We give assumptions, under which there exists a unique classical solution $u\in\mathcal{C}^{4+\beta}\big(\overline{\boldsymbol{\mathcal{D}}};\mathbb{R}\big)$ $(0<\beta<1)$ of PDE (0.6); see [35, Ch. 6].

(A1) $\mathbf{b}:\overline{\boldsymbol{\mathcal{D}}}\to\mathbb{R}^L$, with $b_i(\cdot)\in\mathcal{C}^{2+\beta}(\overline{\boldsymbol{\mathcal{D}}};\mathbb{R}),\ i=1,...,L\,.$

(A2) $\boldsymbol{\sigma}:\overline{\boldsymbol{\mathcal{D}}}\to\mathbb{R}^{L\times L}$, with $\sigma_{ij}(\cdot)\in\mathcal{C}^{2+\beta}(\overline{\boldsymbol{\mathcal{D}}};\mathbb{R}),\ i,j=1,...,L$. Moreover, there exists a constant $\lambda_{\boldsymbol{\sigma}}>0$, *s.t.*

$$\Big\langle\mathbf{y},\boldsymbol{\sigma}(\mathbf{z})\boldsymbol{\sigma}^\top(\mathbf{z})\mathbf{y}\Big\rangle_{\mathbb{R}^L}\geq\lambda_{\boldsymbol{\sigma}}\|\mathbf{y}\|_{\mathbb{R}^L}^2\quad\forall\mathbf{z}\in\overline{\boldsymbol{\mathcal{D}}},\ \mathbf{y}\in\mathbb{R}^L\,.$$

(A3) $c\in\mathcal{C}^{2+\beta}(\overline{\boldsymbol{\mathcal{D}}};\mathbb{R}_0^-)$, $g\in\mathcal{C}^{2+\beta}(\overline{\boldsymbol{\mathcal{D}}};\mathbb{R})$ and $\phi\in\mathcal{C}^{4+\beta}\big(\overline{\boldsymbol{\mathcal{D}}};\mathbb{R}\big)$.

In the next chapter, we need to sharpen these asssumptions. Hence, we assume

(A1*) $\langle\mathbf{z},\mathbf{b}(\mathbf{z})\rangle_{\mathbb{R}^L}\geq 0\quad\forall\mathbf{z}\in\boldsymbol{\mathcal{D}}$.

(A1**) There exists a constant $C_{\mathbf{b},\boldsymbol{\sigma},L}>0$, *s.t.*

$$2\langle\mathbf{z},\mathbf{b}(\mathbf{z})\rangle_{\mathbb{R}^L}+L\lambda_{\boldsymbol{\sigma}}\geq C_{\mathbf{b},\boldsymbol{\sigma},L}\quad\forall\mathbf{z}\in\boldsymbol{\mathcal{D}}\,.$$

Lemma 8.1. Assume **(A1)** − **(A3)** in (0.6). Then, for $\ell=1,2,3,4$,

$$\boldsymbol{C}_{D^\ell}(\phi,g):=\sup_{\mathbf{z}\in\boldsymbol{\mathcal{D}}}\|D_{\mathbf{x}}^\ell u(\mathbf{z})\|_{\mathcal{L}^\ell}\leq\boldsymbol{C}(\phi,g)\,,$$

where

$$\boldsymbol{C}(\phi,g):=\mathsf{C}\bigg\{\|g\|_{\boldsymbol{\mathcal{D}}}^{(2,\beta)}+\|\phi\|_{\boldsymbol{\mathcal{D}}}^{(4,\beta)}+\sup_{\mathbf{z}\in\partial\boldsymbol{\mathcal{D}}}|\phi(\mathbf{z})|\bigg\}\,,$$

for some constant $\mathsf{C}>0$ depending on the data in (0.6), the dimension L and the domain $\boldsymbol{\mathcal{D}}$.

The proof of Lemma 8.1 is an immediate consequence of [35, p. 142, Prob. 6.2 and p. 36, Thm. 3.7].

8.2. The parabolic PDE (6.1): assumptions and bounds for $\{D^\ell_{\mathbf{x}} u\}_{\ell=1}^4$

We give assumptions, under which there exists a unique classical solution $u \in \mathcal{C}^{(2+\beta)/2+1,4+\beta}\Big([0,T]\times \overline{\boldsymbol{\mathcal{D}}};\mathbb{R}\Big)$ $(0<\beta<1)$ of PDE (6.1); see [54, p. 320, Thm. 5.2].

(B1) $\mathbf{b}:\overline{\boldsymbol{\mathcal{D}}}\to\mathbb{R}^L$, with $b_i(\cdot)\in\mathcal{C}^{2+\beta}(\overline{\boldsymbol{\mathcal{D}}};\mathbb{R})\quad i=1,...,L\,.$

(B2) $\boldsymbol{\sigma}:\overline{\boldsymbol{\mathcal{D}}}\to\mathbb{R}^{L\times L}$, with $\sigma_{ij}(\cdot)\in\mathcal{C}^{2+\beta}(\overline{\boldsymbol{\mathcal{D}}};\mathbb{R})$, $i,j=1,...,L$. Moreover, there exists a constant $\lambda_{\boldsymbol{\sigma}}>0$, *s.t.*

$$\Big\langle \mathbf{y}, \boldsymbol{\sigma}(\mathbf{z})\boldsymbol{\sigma}^\top(\mathbf{z})\mathbf{y}\Big\rangle_{\mathbb{R}^L} \geq \lambda_{\boldsymbol{\sigma}}\|\mathbf{y}\|^2_{\mathbb{R}^L}\quad \forall \mathbf{z}\in\overline{\boldsymbol{\mathcal{D}}},\ \mathbf{y}\in\mathbb{R}^L\,.$$

(B3) $g\in\mathcal{C}^{(2+\beta)/2,2+\beta}([0,T]\times\overline{\boldsymbol{\mathcal{D}}};\mathbb{R})$ and $\phi\in\mathcal{C}^{(2+\beta)/2+1,4+\beta}\Big([0,T]\times\overline{\boldsymbol{\mathcal{D}}};\mathbb{R}\Big)$, with $(j=0,1)$

$$\partial_t^{j+1}\phi(t,\mathbf{x}) + \Big\langle \mathbf{b}(\mathbf{x}), D_{\mathbf{x}}\Big(\partial_t^j\phi(t,\mathbf{x})\Big)\Big\rangle_{\mathbb{R}^L} + \frac{1}{2}\mathrm{Tr}\Big(\boldsymbol{\sigma}(\mathbf{x})\boldsymbol{\sigma}^\top(\mathbf{x})D^2_{\mathbf{x}}\Big(\partial_t^j\phi(t,\mathbf{x})\Big)\Big) + \partial_t^j g(t,\mathbf{x}) = 0 \qquad \forall\,(t,\mathbf{x})\in\{T\}\times\boldsymbol{\partial\mathcal{D}}\,.$$

Lemma 8.2. Assume **(B1)** − **(B3)** in (6.1). Then, for $\ell=1,2,3,4$,

$$\sup_{(t,\mathbf{z})\in[0,T)\times\boldsymbol{\mathcal{D}}}\|D^\ell_{\mathbf{x}}u(t,\mathbf{z})\|_{\mathcal{L}^\ell}\leq \mathfrak{C}(\phi,g)\,,$$

where

$$\mathfrak{C}(\phi,g) := \mathsf{C}\Big\{\|g\|^{(2+\beta)}_{[0,T)\times\boldsymbol{\mathcal{D}}} + \|\phi(T,\cdot)\|^{(4+\beta)}_{\boldsymbol{\mathcal{D}}} + \|\phi\|^{(4+\beta)}_{[0,T)\times\boldsymbol{\partial\mathcal{D}}}\Big\}\,,$$

for some constant $\mathsf{C}>0$ depending on the data in (6.1), the dimension L and the domain $\boldsymbol{\mathcal{D}}$.

The proof of Lemma 8.2 is an immediate consequence of [54, p. 320, Thm. 5.2].

8.3. Discretization for the elliptic PDE (0.6): Scheme and stability

Scheme 2 below will be used to approximate (0.7) from (0.6). For this purpose, we fix $\mathbf{x}\in\boldsymbol{\mathcal{D}}$ and let $\{t_j\}_{j\geq0}\subset[0,\infty)$ be a mesh with local mesh sizes $\{\tau^{j+1}\}_{j\geq0}$.

Scheme 1. Let $j\geq0$. For given $(\mathbf{Y}^j_{\mathbf{X}}, Y^j_V, Y^j_Z)$ at time t_j, find the $\mathbb{R}^L$-valued random variable $\mathbf{Y}^{j+1}_{\mathbf{X}}$ from

$$\mathbf{Y}^{j+1}_{\mathbf{X}} = \mathbf{Y}^j_{\mathbf{X}} + \mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\tau^{j+1} + \boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\sqrt{\tau^{j+1}}\,,\qquad \mathbf{Y}^0_{\mathbf{X}} = \mathbf{x}\,,\tag{8.1}$$

where $\boldsymbol{\xi}_{j+1} = \left(\xi_{j+1}^{(1)}, ..., \xi_{j+1}^{(L)}\right)^\top$ is a $\mathbb{R}^L$-valued random vector, whose entries are independent two-point distributed random variables, taking values ± 1 with probability $\frac{1}{2}$ each, as well as the $\mathbb{R}$-valued random variables Y_V^{j+1}, Y_Z^{j+1} from

$$\begin{aligned} Y_V^{j+1} &= Y_V^j + \tau^{j+1} c(\mathbf{Y}_\mathbf{X}^j) Y_V^{j+1}\,, \qquad Y_V^0 = 1\,, \qquad \text{and} && (8.2)\\ Y_Z^{j+1} &= Y_Z^j + \tau^{j+1} g(\mathbf{Y}_\mathbf{X}^j) Y_V^j\,, \qquad Y_Z^0 = 0\,, && (8.3)\end{aligned}$$

to approximate solution $\mathbf{X}_{t_{j+1}}$ from (0.8), $V_{t_{j+1}}$ from (0.10), and $Z_{t_{j+1}}$ from (0.11) at time $t_{j+1} = t_j + \tau^{j+1}$.

The iterates $\{Y_V^j\}_{j\geq 0}$ from (8.2) are computed via the *implicit* Euler method in order to ensure $0 < Y_V^j \leq 1$ $(j \geq 0)$ without additonal smallness assumptions of the corresponding step sizes $\{\tau^{j+1}\}_{j\geq 0}$.

Scheme 2 below is closely based on [62, p. 365 *ff.*, Sec. 6.3] and uses Scheme 1 to approximate (0.7) by '$\mathbb{E}\Big[\phi(\mathbf{Y}_\mathbf{X}^{J^*}) Y_V^{J^*} + Y_Z^{J^*}\Big]$', where $J^* = J^*(\omega)$ is the smallest number such that $\mathbf{Y}_\mathbf{X}^{J^*} \in \partial\boldsymbol{\mathcal{D}}$. Recalling the characterization of the boundary strip $\boldsymbol{\mathcal{S}}_{\tau^{j+1}}$ in Chapter 6 (see also Figure 8.1 (b)), we observe that $\lambda_j > 0$ has to be chosen such that $\lambda_j \sqrt{\tau^{j+1}} \geq \|\mathbf{Y}_\mathbf{X}^{j+1} - \mathbf{Y}_\mathbf{X}^j\|_{\mathbb{R}^L}$, *i.e.*, as (computable) upper bound of the distance between two consecutive iterates. Hence, choosing

$$\lambda_j := \|\mathbf{b}(\mathbf{Y}_\mathbf{X}^j)\|_{\mathbb{R}^L} \sqrt{\tau^{j+1}} + \sqrt{L} \|\boldsymbol{\sigma}(\mathbf{Y}_\mathbf{X}^j)\|_{\mathbb{R}^{L\times L}}\,, \tag{8.4}$$

is suitable. Consequently, we identify $\mathbf{Y}_\mathbf{X}^j \in \boldsymbol{\mathcal{D}}$ as being 'close' to resp. 'away' from $\partial\boldsymbol{\mathcal{D}}$, when $\mathbf{Y}_\mathbf{X}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}$ resp. $\mathbf{Y}_\mathbf{X}^j \notin \boldsymbol{\mathcal{S}}_{\tau^{j+1}}$. For the following, we denote by $\Pi_{\partial\boldsymbol{\mathcal{D}}} : \overline{\boldsymbol{\mathcal{D}}} \to \partial\boldsymbol{\mathcal{D}}$ the projection onto the boundary $\partial\boldsymbol{\mathcal{D}}$, and by $\boldsymbol{n}\Big(\Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{z})\Big)$ the unit internal normal to $\partial\boldsymbol{\mathcal{D}}$ at $\Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{z})$; see Figure 8.1 (b).

Scheme 2. Let $j \geq 0$. Let $(\mathbf{Y}_\mathbf{X}^j, Y_V^j, Y_Z^j)$ be given, and $\mathbf{Y}_\mathbf{X}^k \in \boldsymbol{\mathcal{D}}$, for $k = 0, ..., j$.

(1) (`' Localization`') If $\mathbf{Y}_\mathbf{X}^j \notin \boldsymbol{\mathcal{S}}_{\tau^{j+1}}$, set $\overline{\mathbf{Y}}_\mathbf{X}^j := \mathbf{Y}_\mathbf{X}^j$.

- **a)** If $\mathbf{Y}_\mathbf{X}^j \in \boldsymbol{\mathcal{D}}$, go to **(4)**.
- **b)** If $\mathbf{Y}_\mathbf{X}^j \notin \boldsymbol{\mathcal{D}}$, set $J^* := j$, $\mathbf{Y}_\mathbf{X}^{J^*} := \Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}_\mathbf{X}^j)$, $Y_V^{J^*} := Y_V^j$, $Y_Z^{J^*} := Y_Z^j$, and **STOP**.

(2) (`' Localization`') If $\mathbf{Y}_\mathbf{X}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}$, then either $\overline{\mathbf{Y}}_\mathbf{X}^j = \Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}_\mathbf{X}^j)$ with probability

$$p_j := \mathbb{P}\Big[\overline{\mathbf{Y}}_\mathbf{X}^j = \Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}_\mathbf{X}^j)\,|\,\mathbf{Y}_\mathbf{X}^j\Big] = \frac{\lambda_j\sqrt{\tau^{j+1}}}{\Big\|\mathbf{Y}_\mathbf{X}^j + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\Big(\Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}_\mathbf{X}^j)\Big) - \Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}_\mathbf{X}^j)\Big\|_{\mathbb{R}^L}}\,, \tag{8.5}$$

or $\overline{\mathbf{Y}}_\mathbf{X}^j = \mathbf{Y}_\mathbf{X}^j + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\Big(\Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}_\mathbf{X}^j)\Big)$ with probability $1 - p_j$.

(3) (`'Projection`') If $\overline{\mathbf{Y}}_\mathbf{X}^j = \Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}_\mathbf{X}^j)$, set $J^* := j$, $\mathbf{Y}_\mathbf{X}^{J^*} := \Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}_\mathbf{X}^j)$, $Y_V^{J^*} := Y_V^j$, $Y_Z^{J^*} := Y_Z^j$, and **STOP**.

(4) (`' Solve`') Set $\mathbf{Y}_\mathbf{X}^j := \overline{\mathbf{Y}}_\mathbf{X}^j$. Compute $\mathbf{Y}_\mathbf{X}^{j+1}$, Y_V^{j+1} and Y_Z^{j+1} via Scheme 1.

(5) Put $j := j+1$, and return to **(1)**.

For $j > J^*$, we set $(\mathbf{Y}^j_{\mathbf{X}}, Y^j_V, Y^j_Z) = (\mathbf{Y}^{J^*}_{\mathbf{X}}, Y^{J^*}_V, Y^{J^*}_Z)$.

Note that $p_j > \frac{1}{2}$ in Step **(2)** of Scheme 2, since $\mathrm{d}(\mathbf{Y}^j_{\mathbf{X}}, \partial\boldsymbol{\mathcal{D}}) = \big\|\mathbf{Y}^j_{\mathbf{X}} - \Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\big\|_{\mathbb{R}^L} < \lambda_j\sqrt{\tau^{j+1}}$, and $\mathbf{Y}^j_{\mathbf{X}} + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\Big(\Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big) \notin \boldsymbol{\mathcal{S}}_{\tau^{j+1}}$.

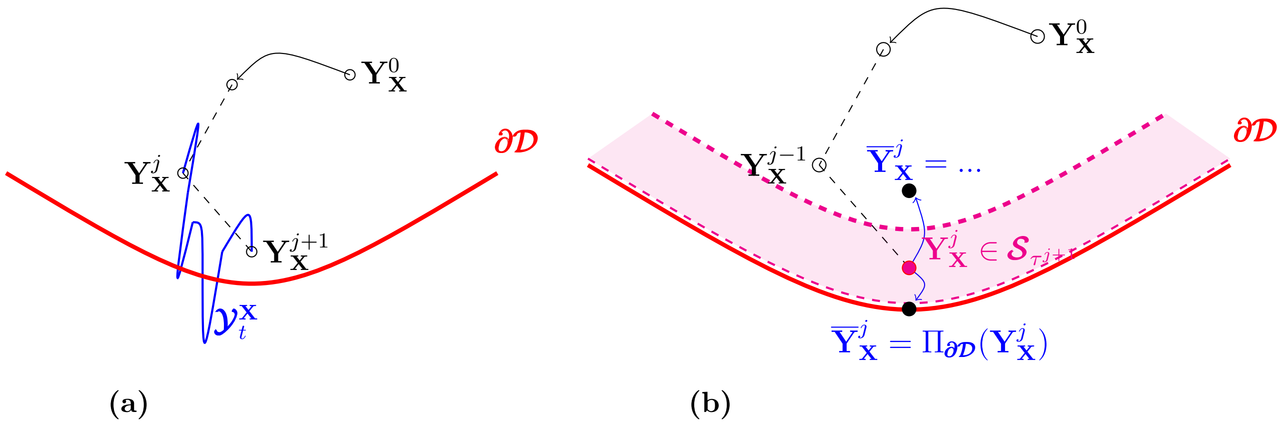


Figure 8.1.: (a) Exit of the continuified Euler process $\boldsymbol{\mathcal{Y}}^{\mathbf{X}}$ in (9.58). (b) Projection resp. bouncing back mechanism in Scheme 2.

The following lemma estimates the number of iterates $\{\mathbf{Y}^j_{\mathbf{X}}\}_{j\geq 0}$ from Scheme 2 in the boundary strips; it may be considered as a generalization of [62, p. 367, Lem. 3.2] for non-uniform time steps.

Lemma 8.3. Assume **(A1)** – **(A3)**. Fix $\mathbf{x} \in \boldsymbol{\mathcal{D}}$. Let $\{t_j\}_{j\geq 0} \subset [0,\infty)$ be a mesh with local mesh sizes $\{\tau^{j+1}\}_{j\geq 0}$. Let $\{\mathbf{Y}^j_{\mathbf{X}}\}_{j\geq 0}$ be from Scheme 2. Then

$$\sum_{j=0}^{\infty} \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}^j_{\mathbf{X}} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big] < 2\,.$$

Proof. Let $j \in \mathbb{N}$. Since the probability p_j in (8.5) is greater than $\frac{1}{2}$, we obtain

$$\begin{aligned}\mathbb{P}[J^* = j] &= \mathbb{E}\Big[\mathbb{E}\big[\mathbf{1}_{\{J^*=j\}} \,|\, \mathbf{Y}^j_{\mathbf{X}}\big]\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}^j_{\mathbf{X}} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}} \mathbf{1}_{\big\{\overline{\mathbf{Y}}^j_{\mathbf{X}} = \Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\big\}} \,|\, \mathbf{Y}^j_{\mathbf{X}}\Big]\Big] \\ &= \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}^j_{\mathbf{X}} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}} \cdot p_j\Big] > \frac{1}{2}\mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}^j_{\mathbf{X}} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big]\,.\end{aligned}$$

Consequently, we have

$$\sum_{j=0}^{\infty} \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}^j_{\mathbf{X}} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big] < 2\sum_{j=0}^{\infty} \mathbb{P}[J^* = j] = 2\,.$$

□

The following lemma yields boundedness of the expected discrete stopping time

$$t_{J^*} := \min\Big\{t_j \ : \ \mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\partial\mathcal{D}}\,,\ j \geq 0\Big\}\,,$$

which approximates (0.9).

Lemma 8.4. Assume **(A1) – (A3)**. Fix $\mathbf{x} \in \boldsymbol{\mathcal{D}}$. Let $\{t_j\}_{j\geq 0} \subset [0,\infty)$ be a mesh with local mesh sizes $\{\tau^{j+1}\}_{j\geq 0}$ and maximum mesh size $\tau^{max} := \max_j \tau^{j+1}$. Let $\{\mathbf{Y}_{\mathbf{X}}^j\}_{j\geq 0}$ be from Scheme 2. For τ^{max} either sufficiently small, or for general $\tau^{max} > 0$ if **(A1)** is complemented by **(A1*)** or **(A1**)**, we have

$$\mathbb{E}[t_{J^*}] \leq C\,,$$

where $C > 0$ depends on the dimension L, the domain $\boldsymbol{\mathcal{D}}$ and the data in (0.6), but is independent of $\mathbf{x} \in \boldsymbol{\mathcal{D}}$.

Proof. **Step 1:** (Derivation of a 'discrete Dynkin-formula') We derive a 'discrete Dynkin-formula' adapted to our setting. Let $f \in \mathcal{C}(\mathbb{R}^L)$ and $k \in \mathbb{N}$. A first calculation yields

$$\begin{aligned}
\mathbb{E}\Big[f\Big(\mathbf{Y}_{\mathbf{X}}^{J^*\wedge k}\Big) - f(\mathbf{x})\Big] &= \mathbb{E}\Bigg[\sum_{j=0}^{J^*\wedge k-1}\Big\{f(\mathbf{Y}_{\mathbf{X}}^{j+1}) - f(\mathbf{Y}_{\mathbf{X}}^j)\Big\}\Bigg] \\
&= \mathbb{E}\Bigg[\sum_{j=0}^{J^*\wedge k-1}\Big\{f(\mathbf{Y}_{\mathbf{X}}^{j+1}) - f(\overline{\mathbf{Y}}_{\mathbf{X}}^j) + f(\overline{\mathbf{Y}}_{\mathbf{X}}^j) - f(\mathbf{Y}_{\mathbf{X}}^j)\Big\}\Bigg] \\
&= \underbrace{\mathbb{E}\Bigg[\sum_{j=0}^{k-1}\mathbf{1}_{\{J^*>j\}}\cdot\Big\{f(\mathbf{Y}_{\mathbf{X}}^{j+1}) - f(\overline{\mathbf{Y}}_{\mathbf{X}}^j)\Big\}\Bigg]}_{=:\boldsymbol{T_1}} + \underbrace{\mathbb{E}\Bigg[\sum_{j=0}^{k-1}\mathbf{1}_{\{J^*>j\}}\cdot\Big\{f(\overline{\mathbf{Y}}_{\mathbf{X}}^j) - f(\mathbf{Y}_{\mathbf{X}}^j)\Big\}\Bigg]}_{=:\boldsymbol{T_2}}\,.
\end{aligned} \tag{8.6}$$

a) (Investigation of $\boldsymbol{T_1}$) According to the procedure in Scheme 2, we have

$$\boldsymbol{T_1} = \boldsymbol{T_{1,1}} + \boldsymbol{T_{1,2}}\,, \tag{8.7}$$

where

$$\begin{aligned}
\boldsymbol{T_{1,1}} &:= \mathbb{E}\Bigg[\sum_{j=0}^{k-1}\mathbf{1}_{\Big\{\mathbf{Y}_{\mathbf{X}}^j\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\Big\}}\cdot\Big\{f(\mathbf{Y}_{\mathbf{X}}^{j+1}) - f(\mathbf{Y}_{\mathbf{X}}^j)\Big\}\Bigg]\,, \\
\boldsymbol{T_{1,2}} &:= \mathbb{E}\Bigg[\sum_{j=0}^{k-1}\mathbf{1}_{\Big\{\mathbf{Y}_{\mathbf{X}}^j\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\Big\}}\mathbf{1}_{\Big\{\overline{\mathbf{Y}}_{\mathbf{X}}^j=\mathbf{Y}_{\mathbf{X}}^j+\sqrt{\tau^{j+1}}\boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\Big)\Big\}}\cdot\Big\{f(\mathbf{Y}_{\mathbf{X}}^{j+1}) - f(\overline{\mathbf{Y}}_{\mathbf{X}}^j)\Big\}\Bigg]\,.
\end{aligned}$$

b) (Investigation of $\boldsymbol{T_{1,1}}$) *Taylor's formula* yields

$$\boldsymbol{T_{1,1}} = \boldsymbol{T_{1,1,1}} + \boldsymbol{T_{1,1,2}}\,, \tag{8.8}$$

where

$$\boldsymbol{T_{1,1,1}} := \mathbb{E}\Bigg[\sum_{j=0}^{k-1}\mathbf{1}_{\Big\{\mathbf{Y}_{\mathbf{X}}^j\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\Big\}}\cdot\Big\langle D_{\mathbf{x}}f(\mathbf{Y}_{\mathbf{X}}^j), \mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^j\Big\rangle_{\mathbb{R}^L}\Bigg]\,,$$

$$\boldsymbol{T}_{\mathbf{1,1,2}} := \frac{1}{2}\mathbb{E}\Bigg[\sum_{j=0}^{k-1}\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\cdot\mathrm{Tr}\Big(D_{\mathbf{x}}^{2}f(\hat{\mathbf{Y}}_{\mathbf{X}}^{j})\Big(\mathbf{Y}_{\mathbf{X}}^{j+1}-\mathbf{Y}_{\mathbf{X}}^{j}\Big)\Big(\mathbf{Y}_{\mathbf{X}}^{j+1}-\mathbf{Y}_{\mathbf{X}}^{j}\Big)^{\top}\Big)\Bigg]\,,$$

for some $\hat{\mathbf{Y}}_{\mathbf{X}}^{j}$ between $\mathbf{Y}_{\mathbf{X}}^{j+1}$ and $\mathbf{Y}_{\mathbf{X}}^{j}$, *i.e.*, $\hat{\mathbf{Y}}_{\mathbf{X}}^{j} = \mathbf{Y}_{\mathbf{X}}^{j} + \theta\Big(\mathbf{Y}_{\mathbf{X}}^{j+1}-\mathbf{Y}_{\mathbf{X}}^{j}\Big)$ with $\theta\in(0,1)$. We use (8.1) to represent the increment '$\mathbf{Y}_{\mathbf{X}}^{j+1}-\mathbf{Y}_{\mathbf{X}}^{j}$' in $\boldsymbol{T}_{\mathbf{1,1,1}}$, as well as the tower property and independency arguments to get

$$\begin{aligned}\boldsymbol{T}_{\mathbf{1,1,1}} &= \mathbb{E}\Bigg[\sum_{j=0}^{k-1}\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\cdot\Big\langle D_{\mathbf{x}}f(\mathbf{Y}_{\mathbf{X}}^{j}),\mathbf{b}(\mathbf{Y}_{\mathbf{X}}^{j})\Big\rangle_{\mathbb{R}^{L}}\cdot\tau^{j+1}\Bigg]\\ &\quad+\sum_{j=0}^{k-1}\underbrace{\mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\cdot\Big\langle D_{\mathbf{x}}f(\mathbf{Y}_{\mathbf{X}}^{j}),\boldsymbol{\sigma}(\mathbf{Y}_{\mathbf{X}}^{j})\boldsymbol{\xi}_{j+1}\Big\rangle_{\mathbb{R}^{L}}\cdot\sqrt{\tau^{j+1}}\Big]}_{=0}\\ &= \mathbb{E}\Bigg[\sum_{j=0}^{k-1}\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\cdot\Big\langle D_{\mathbf{x}}f(\mathbf{Y}_{\mathbf{X}}^{j}),\mathbf{b}(\mathbf{Y}_{\mathbf{X}}^{j})\Big\rangle_{\mathbb{R}^{L}}\cdot\tau^{j+1}\Bigg]\,.\end{aligned}\tag{8.9}$$

Similar arguments as in (8.9), *i.e.*, representing the increment '$\mathbf{Y}_{\mathbf{X}}^{j+1}-\mathbf{Y}_{\mathbf{X}}^{j}$' in $\boldsymbol{T}_{\mathbf{1,1,2}}$ via (8.1) and using standard calculations, as well as independency arguments lead to

$$\boldsymbol{T}_{\mathbf{1,1,2}} = \frac{1}{2}\mathbb{E}\Bigg[\sum_{j=0}^{k-1}\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\cdot\mathrm{Tr}\Big(D_{\mathbf{x}}^{2}f(\mathbf{Y}_{\mathbf{X}}^{j})\boldsymbol{\sigma}(\mathbf{Y}_{\mathbf{X}}^{j})\boldsymbol{\sigma}^{\top}(\mathbf{Y}_{\mathbf{X}}^{j})\Big)\cdot\tau^{j+1}\Bigg]+\boldsymbol{T}_{\mathbf{1,1,2,1}}\,,\tag{8.10}$$

where

$$\begin{aligned}\boldsymbol{T}_{\mathbf{1,1,2,1}} &= \frac{1}{2}\mathbb{E}\Bigg[\sum_{j=0}^{k-1}\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\cdot\mathrm{Tr}\Big(D_{\mathbf{x}}^{2}f(\hat{\mathbf{Y}}_{\mathbf{X}}^{j})\mathbf{b}(\mathbf{Y}_{\mathbf{X}}^{j})\mathbf{b}^{\top}(\mathbf{Y}_{\mathbf{X}}^{j})\Big)\cdot\big(\tau^{j+1}\big)^{2}\Bigg]\\ &\quad+\frac{1}{2}\mathbb{E}\Bigg[\sum_{j=0}^{k-1}\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\cdot\mathrm{Tr}\Big(\Big\{D_{\mathbf{x}}^{2}f(\hat{\mathbf{Y}}_{\mathbf{X}}^{j})-D_{\mathbf{x}}^{2}f(\mathbf{Y}_{\mathbf{X}}^{j})\Big\}\mathbf{b}(\mathbf{Y}_{\mathbf{X}}^{j})\big(\boldsymbol{\sigma}(\mathbf{Y}_{\mathbf{X}}^{j})\boldsymbol{\xi}_{j+1}\big)^{\top}\Big)\cdot\big(\tau^{j+1}\big)^{\frac{3}{2}}\Bigg]\\ &\quad+\frac{1}{2}\mathbb{E}\Bigg[\sum_{j=0}^{k-1}\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\cdot\mathrm{Tr}\Big(\Big\{D_{\mathbf{x}}^{2}f(\hat{\mathbf{Y}}_{\mathbf{X}}^{j})-D_{\mathbf{x}}^{2}f(\mathbf{Y}_{\mathbf{X}}^{j})\Big\}\boldsymbol{\sigma}(\mathbf{Y}_{\mathbf{X}}^{j})\boldsymbol{\xi}_{j+1}\mathbf{b}^{\top}(\mathbf{Y}_{\mathbf{X}}^{j})\Big)\cdot\big(\tau^{j+1}\big)^{\frac{3}{2}}\Bigg]\\ &\quad+\frac{1}{2}\mathbb{E}\Bigg[\sum_{j=0}^{k-1}\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\cdot\mathrm{Tr}\Big(\Big\{D_{\mathbf{x}}^{2}f(\hat{\mathbf{Y}}_{\mathbf{X}}^{j})-D_{\mathbf{x}}^{2}f(\mathbf{Y}_{\mathbf{X}}^{j})\Big\}\boldsymbol{\sigma}(\mathbf{Y}_{\mathbf{X}}^{j})\boldsymbol{\xi}_{j+1}\big(\boldsymbol{\sigma}(\mathbf{Y}_{\mathbf{X}}^{j}\boldsymbol{\xi}_{j+1})\big)^{\top}\Big)\cdot\tau^{j+1}\Bigg]\,.\end{aligned}$$

Plugging (8.9) and (8.10) into (8.8) yields

$$\begin{aligned}\boldsymbol{T}_{\mathbf{1,1}} &= \mathbb{E}\Bigg[\sum_{j=0}^{k-1}\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\cdot\Big\langle D_{\mathbf{x}}f(\mathbf{Y}_{\mathbf{X}}^{j}),\mathbf{b}(\mathbf{Y}_{\mathbf{X}}^{j})\Big\rangle_{\mathbb{R}^{L}}\cdot\tau^{j+1}\Bigg]\\ &\quad+\frac{1}{2}\mathbb{E}\Bigg[\sum_{j=0}^{k-1}\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\cdot\mathrm{Tr}\Big(D_{\mathbf{x}}^{2}f(\mathbf{Y}_{\mathbf{X}}^{j})\boldsymbol{\sigma}(\mathbf{Y}_{\mathbf{X}}^{j})\boldsymbol{\sigma}^{\top}(\mathbf{Y}_{\mathbf{X}}^{j})\Big)\cdot\tau^{j+1}\Bigg]+\boldsymbol{T}_{\mathbf{1,1,2,1}}\,.\end{aligned}\tag{8.11}$$

c) (Investigation of $\boldsymbol{T}_{\mathbf{1,2}}$) Since the investigation of $\boldsymbol{T}_{\mathbf{1,2}}$ is similar to $\boldsymbol{T}_{\mathbf{1,1}}$, we obtain

$$\boldsymbol{T}_{\mathbf{1,2}} = \mathbb{E}\Bigg[\sum_{j=0}^{k-1}\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\mathbf{1}_{\left\{\overline{\mathbf{Y}}_{\mathbf{X}}^{j}=\mathbf{Y}_{\mathbf{X}}^{j}+\sqrt{\tau^{j+1}}\boldsymbol{n}\left(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^{j})\right)\right\}}\cdot\big\langle D_{\mathbf{x}}f(\overline{\mathbf{Y}}_{\mathbf{X}}^{j}),\mathbf{b}(\overline{\mathbf{Y}}_{\mathbf{X}}^{j})\big\rangle_{\mathbb{R}^{L}}\cdot\tau^{j+1}\Bigg]$$

$$+\frac{1}{2}\mathbb{E}\Big[\sum_{j=0}^{k-1}\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^j\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}_{\mathbf{X}}^j=\mathbf{Y}_{\mathbf{X}}^j+\sqrt{\tau^{j+1}}\boldsymbol{n}\big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\big)\}}\cdot\mathrm{Tr}\Big(D_{\mathbf{x}}^2 f(\overline{\mathbf{Y}}_{\mathbf{X}}^j)\boldsymbol{\sigma}(\overline{\mathbf{Y}}_{\mathbf{X}}^j)\boldsymbol{\sigma}^\top(\overline{\mathbf{Y}}_{\mathbf{X}}^j)\Big)\cdot\tau^{j+1}\Big]$$
$$+\overline{\boldsymbol{T}_{\mathbf{1,1,2,1}}}\,, \tag{8.12}$$

where $\overline{\boldsymbol{T}_{\mathbf{1,1,2,1}}}$ has the same representation as $\boldsymbol{T}_{\mathbf{1,1,2,1}}$, where every $\mathbf{Y}_{\mathbf{X}}^j$ in the trace terms is replaced by $\overline{\mathbf{Y}}_{\mathbf{X}}^j$, and $\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^j\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}$ is replaced by $\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^j\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}_{\mathbf{X}}^j=\mathbf{Y}_{\mathbf{X}}^j+\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j))\}}$.

d) (Investigation of $\boldsymbol{T_2}$) According to the procedure in Scheme 2, we obtain

$$\boldsymbol{T_2}=\mathbb{E}\Bigg[\sum_{j=0}^{k-1}\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^j\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}_{\mathbf{X}}^j=\mathbf{Y}_{\mathbf{X}}^j+\sqrt{\tau^{j+1}}\boldsymbol{n}\big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\big)\}}\cdot\Big\{f(\overline{\mathbf{Y}}_{\mathbf{X}}^j)-f(\mathbf{Y}_{\mathbf{X}}^j)\Big\}\Bigg]. \tag{8.13}$$

e) We insert (8.11) and (8.12) into (8.7), and plug the resulting expression as well as (8.13) into (8.6) to obtain

$$\mathbb{E}\big[f\big(\mathbf{Y}_{\mathbf{X}}^{J^*\wedge k}\big)\big]=f(\mathbf{x})+\mathbb{E}\Big[\sum_{j=0}^{k-1}\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^j\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\cdot\big\langle D_{\mathbf{x}}f(\mathbf{Y}_{\mathbf{X}}^j),\mathbf{b}(\mathbf{Y}_{\mathbf{X}}^j)\big\rangle_{\mathbb{R}^L}\cdot\tau^{j+1}\Big]$$
$$+\frac{1}{2}\mathbb{E}\Big[\sum_{j=0}^{k-1}\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^j\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\cdot\mathrm{Tr}\Big(D_{\mathbf{x}}^2 f(\mathbf{Y}_{\mathbf{X}}^j)\boldsymbol{\sigma}(\mathbf{Y}_{\mathbf{X}}^j)\boldsymbol{\sigma}^\top(\mathbf{Y}_{\mathbf{X}}^j)\Big)\cdot\tau^{j+1}\Big]+\boldsymbol{T}_{\mathbf{1,1,2,1}}$$
$$+\mathbb{E}\Big[\sum_{j=0}^{k-1}\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^j\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}_{\mathbf{X}}^j=\mathbf{Y}_{\mathbf{X}}^j+\sqrt{\tau^{j+1}}\boldsymbol{n}\big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\big)\}}\cdot\big\langle D_{\mathbf{x}}f(\overline{\mathbf{Y}}_{\mathbf{X}}^j),\mathbf{b}(\overline{\mathbf{Y}}_{\mathbf{X}}^j)\big\rangle_{\mathbb{R}^L}\cdot\tau^{j+1}\Big]$$
$$+\frac{1}{2}\mathbb{E}\Big[\sum_{j=0}^{k-1}\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^j\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}_{\mathbf{X}}^j=\mathbf{Y}_{\mathbf{X}}^j+\sqrt{\tau^{j+1}}\boldsymbol{n}\big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\big)\}}\cdot\mathrm{Tr}\Big(D_{\mathbf{x}}^2 f(\overline{\mathbf{Y}}_{\mathbf{X}}^j)\boldsymbol{\sigma}(\overline{\mathbf{Y}}_{\mathbf{X}}^j)\boldsymbol{\sigma}^\top(\overline{\mathbf{Y}}_{\mathbf{X}}^j)\Big)\cdot\tau^{j+1}\Big]$$
$$+\overline{\boldsymbol{T}_{\mathbf{1,1,2,1}}}+\mathbb{E}\Big[\sum_{j=0}^{k-1}\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^j\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}_{\mathbf{X}}^j=\mathbf{Y}_{\mathbf{X}}^j+\sqrt{\tau^{j+1}}\boldsymbol{n}\big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\big)\}}\cdot\big\{f(\overline{\mathbf{Y}}_{\mathbf{X}}^j)-f(\mathbf{Y}_{\mathbf{X}}^j)\big\}\Big]. \tag{8.14}$$

Step 2: (Proof of the statement for τ^{max} sufficiently small) Let $n\in\mathbb{N}$. Choose a $\mathbf{B}\in\mathbb{R}^L$ such that $\min_{\mathbf{z}\in\overline{\boldsymbol{\mathcal{D}}}}\|\mathbf{z}+\mathbf{B}\|_{\mathbb{R}^L}^{2n}\geq 1$. Set $A^2:=\max_{\mathbf{z}\in\overline{\boldsymbol{\mathcal{D}}}}\|\mathbf{z}+\mathbf{B}\|_{\mathbb{R}^L}^{2n}$ and consider (8.14) with

$$f(\mathbf{z})=A^2-\|\mathbf{z}+\mathbf{B}\|_{\mathbb{R}^L}^{2n}\qquad(\mathbf{z}\in\boldsymbol{\mathcal{D}})\,,$$
$$\partial_{x_i}f(\mathbf{z})=-2n\|\mathbf{z}+\mathbf{B}\|_{\mathbb{R}^L}^{2(n-1)}(z_i+B_i)\qquad i=1,...,L\,,$$
$$\partial_{x_i}\partial_{x_j}f(\mathbf{z})=-4n(n-1)\|\mathbf{z}+\mathbf{B}\|_{\mathbb{R}^L}^{2(n-2)}(z_i+B_i)(z_j+B_j)-2n\|\mathbf{z}+\mathbf{B}\|_{\mathbb{R}^L}^{2(n-1)}\delta_{ij}\qquad i,j=1,...,L\,.$$

We refer to [62, p. 367, Lemma 3.2] for a similar choice of a function f in a related setting. By applying **(A2)** four times, we consequently obtain for $k\in\mathbb{N}$

$$\mathbb{E}\big[f\big(\mathbf{Y}_{\mathbf{X}}^{J^*\wedge k}\big)\big]=f(\mathbf{x})-2n\mathbb{E}\Big[\sum_{j=0}^{k-1}\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^j\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\cdot\tau^{j+1}\|\mathbf{Y}_{\mathbf{X}}^j+\mathbf{B}\|_{\mathbb{R}^L}^{2(n-1)}\big\langle\mathbf{Y}_{\mathbf{X}}^j+\mathbf{B},\mathbf{b}(\mathbf{Y}_{\mathbf{X}}^j)\big\rangle_{\mathbb{R}^L}\Big]$$
$$-n\mathbb{E}\Big[\sum_{j=0}^{k-1}\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^j\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\cdot\tau^{j+1}\|\mathbf{Y}_{\mathbf{X}}^j+\mathbf{B}\|_{\mathbb{R}^L}^{2(n-1)}\underbrace{\sum_{k=1}^{L}\big(\boldsymbol{\sigma}(\mathbf{Y}_{\mathbf{X}}^j)\boldsymbol{\sigma}^\top(\mathbf{Y}_{\mathbf{X}}^j)\big)_{kk}}_{\geq\lambda_{\boldsymbol{\sigma}}L}\Big]$$

$$
\begin{aligned}
&- 2n(n-1)\mathbb{E}\Big[\sum_{j=0}^{k-1}\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\mathcal{D}\setminus\mathcal{S}_{\tau^{j+1}}\}}\cdot\tau^{j+1}\|\mathbf{Y}^j_{\mathbf{X}}+\mathbf{B}\|^{2(n-2)}_{\mathbb{R}^L}\cdot\underbrace{\big\langle\mathbf{Y}^j_{\mathbf{X}}+\mathbf{B},\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\sigma}^\top(\mathbf{Y}^j_{\mathbf{X}})(\mathbf{Y}^j_{\mathbf{X}}+\mathbf{B})\big\rangle_{\mathbb{R}^L}}_{\geq\lambda_{\boldsymbol{\sigma}}\|\mathbf{Y}^j_{\mathbf{X}}+\mathbf{B}\|^2_{\mathbb{R}^L}}\Big]\\
&+\boldsymbol{T}_{\mathbf{1,1,1,2}}\\
&- 2n\mathbb{E}\Big[\sum_{j=0}^{k-1}\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\mathcal{S}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\partial\mathcal{D}}(\mathbf{Y}^j_{\mathbf{X}}))\}}\cdot\tau^{j+1}\|\overline{\mathbf{Y}}^j_{\mathbf{X}}+\mathbf{B}\|^{2(n-1)}_{\mathbb{R}^L}\big\langle\overline{\mathbf{Y}}^j_{\mathbf{X}}+\mathbf{B},\mathbf{b}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\big\rangle_{\mathbb{R}^L}\Big]\\
&- n\mathbb{E}\Big[\sum_{j=0}^{k-1}\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\mathcal{S}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\partial\mathcal{D}}(\mathbf{Y}^j_{\mathbf{X}}))\}}\cdot\tau^{j+1}\|\overline{\mathbf{Y}}^j_{\mathbf{X}}+\mathbf{B}\|^{2(n-1)}_{\mathbb{R}^L}\underbrace{\sum_{k=1}^{L}\big(\boldsymbol{\sigma}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\boldsymbol{\sigma}^\top(\overline{\mathbf{Y}}^j_{\mathbf{X}})\big)_{kk}}_{\geq\lambda_{\boldsymbol{\sigma}}L}\Big]\\
&- 2n(n-1)\mathbb{E}\Big[\sum_{j=0}^{k-1}\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\mathcal{S}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\partial\mathcal{D}}(\mathbf{Y}^j_{\mathbf{X}}))\}}\cdot\tau^{j+1}\|\overline{\mathbf{Y}}^j_{\mathbf{X}}+\mathbf{B}\|^{2(n-2)}_{\mathbb{R}^L}\\
&\qquad\qquad\cdot\underbrace{\big\langle\overline{\mathbf{Y}}^j_{\mathbf{X}}+\mathbf{B},\boldsymbol{\sigma}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\boldsymbol{\sigma}^\top(\overline{\mathbf{Y}}^j_{\mathbf{X}})(\overline{\mathbf{Y}}^j_{\mathbf{X}}+\mathbf{B})\big\rangle_{\mathbb{R}^L}}_{\geq\lambda_{\boldsymbol{\sigma}}\|\overline{\mathbf{Y}}^j_{\mathbf{X}}+\mathbf{B}\|^2_{\mathbb{R}^L}}\Big]\\
&+\overline{\boldsymbol{T}_{\mathbf{1,1,1,2}}}+\mathbb{E}\Big[\sum_{j=0}^{k-1}\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\mathcal{S}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\partial\mathcal{D}}(\mathbf{Y}^j_{\mathbf{X}}))\}}\cdot\big\{\|\mathbf{Y}^j_{\mathbf{X}}+\mathbf{B}\|^{2n}_{\mathbb{R}^L}-\|\overline{\mathbf{Y}}^j_{\mathbf{X}}+\mathbf{B}\|^{2n}_{\mathbb{R}^L}\big\}\Big]\\
\leq{}& f(\mathbf{x})-n\mathbb{E}\Big[\sum_{j=0}^{k-1}\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\mathcal{D}\setminus\mathcal{S}_{\tau^{j+1}}\}}\cdot\tau^{j+1}\|\mathbf{Y}^j_{\mathbf{X}}+\mathbf{B}\|^{2(n-1)}_{\mathbb{R}^L}\cdot\Big\{2\big\langle\mathbf{Y}^j_{\mathbf{X}}+\mathbf{B},\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\big\rangle_{\mathbb{R}^L}+(2n-2+L)\lambda_{\boldsymbol{\sigma}}\Big\}\Big]\\
&- n\mathbb{E}\Big[\sum_{j=0}^{k-1}\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\mathcal{S}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\partial\mathcal{D}}(\mathbf{Y}^j_{\mathbf{X}}))\}}\\
&\qquad\cdot\tau^{j+1}\|\overline{\mathbf{Y}}^j_{\mathbf{X}}+\mathbf{B}\|^{2(n-1)}_{\mathbb{R}^L}\cdot\Big\{2\big\langle\overline{\mathbf{Y}}^j_{\mathbf{X}}+\mathbf{B},\mathbf{b}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\big\rangle_{\mathbb{R}^L}+(2n-2+L)\lambda_{\boldsymbol{\sigma}}\Big\}\Big]\\
&+\boldsymbol{T}_{\mathbf{1,1,1,2}}+\overline{\boldsymbol{T}_{\mathbf{1,1,1,2}}}+A^2\mathbb{E}\Big[\sum_{j=0}^{k-1}\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\mathcal{S}_{\tau^{j+1}}\}}\Big].
\end{aligned}
$$

We choose $n\in\mathbb{N}$ large enough such that

$$2\Big\langle\mathbf{z}+\mathbf{B},\mathbf{b}(\mathbf{z})\Big\rangle_{\mathbb{R}^L}+(2n-2+L)\lambda_{\boldsymbol{\sigma}}\geq 1\quad\forall\mathbf{z}\in\mathcal{D}.$$

Since $f(\mathbf{z})\geq 0$, $\mathbf{z}\in\overline{\mathcal{D}}$, $\|\mathbf{Y}^j_{\mathbf{X}}+\mathbf{B}\|^{2(n-1)}_{\mathbb{R}^L}\geq 1$, $\|\overline{\mathbf{Y}}^j_{\mathbf{X}}+\mathbf{B}\|^{2(n-1)}_{\mathbb{R}^L}\geq 1$, and $\sum_{j=0}^{k-1}\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\mathcal{S}_{\tau^{j+1}}\}}\Big]<2$ due to Lemma 8.3, we further get

$$
\begin{aligned}
0\leq{}& f(\mathbf{x})-n\cdot\mathbb{E}\Big[\sum_{j=0}^{k-1}\Big\{\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\mathcal{D}\setminus\mathcal{S}_{\tau^{j+1}}\}}+\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\mathcal{S}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\partial\mathcal{D}}(\mathbf{Y}^j_{\mathbf{X}}))\}}\Big\}\cdot\tau^{j+1}\Big]\\
&+\boldsymbol{T}_{\mathbf{1,1,1,2}}+\overline{\boldsymbol{T}_{\mathbf{1,1,1,2}}}+2A^2\\
={}& f(\mathbf{x})-n\cdot\mathbb{E}\Big[\sum_{j=0}^{k-1}\mathbf{1}_{\{J^*>j\}}\cdot\tau^{j+1}\Big]+\boldsymbol{T}_{\mathbf{1,1,1,2}}+\overline{\boldsymbol{T}_{\mathbf{1,1,1,2}}}+2A^2\\
={}& f(\mathbf{x})-n\cdot\mathbb{E}[t_{J^*\wedge k}]+\boldsymbol{T}_{\mathbf{1,1,1,2}}+\overline{\boldsymbol{T}_{\mathbf{1,1,1,2}}}+2A^2.
\end{aligned}
$$

By means of standard calculations and *Taylor's formula* one can show that

$$\boldsymbol{T}_{\mathbf{1,1,1,2}} + \overline{\boldsymbol{T}_{\mathbf{1,1,1,2}}} \leq \mathfrak{C}(n)\tau^{max}\mathbb{E}[t_{J^*\wedge k}]\,,$$

where $\mathfrak{C}(n) > 0$. For τ^{max} sufficiently small, *i.e.*, $\tau^{max} \leq \frac{n}{\mathfrak{C}(n)}$, we thus have

$$\mathbb{E}[t_{J^*\wedge k}] \leq \frac{f(\mathbf{x}) + 2A^2}{n - \mathfrak{C}(n)\tau^{max}}\,,$$

Letting $k \to \infty$ yields the assertion.
Step 3: (Proof of the statement under **(A1*****)**) Set $A^{max} := \max_{\mathbf{z}\in\overline{\boldsymbol{\mathcal{D}}}} \|\mathbf{z}\|^2_{\mathbb{R}^L}$ and consider (8.14) with

$$f(\mathbf{z}) = A^{max} - \|\mathbf{z}\|^2_{\mathbb{R}^L} \qquad (\mathbf{z} \in \boldsymbol{\mathcal{D}})\,.$$

Applying **(A2)** and **(A1*****)** and using the fact that $\boldsymbol{T}_{\mathbf{1,1,1,2}} + \overline{\boldsymbol{T}_{\mathbf{1,1,1,2}}} \leq 0$, we obtain for $k \in \mathbb{N}$

$$\mathbb{E}\Big[f\Big(\mathbf{Y}_{\mathbf{X}}^{J^*\wedge k}\Big)\Big] \leq f(\mathbf{x}) - \lambda_{\boldsymbol{\sigma}} L \cdot \mathbb{E}[t_{J^*\wedge k}] + 2A^{max}\,. \tag{8.15}$$

Since $f(\mathbf{z}) \geq 0$, $\mathbf{z} \in \overline{\boldsymbol{\mathcal{D}}}$, we obtain

$$\mathbb{E}[t_{J^*\wedge k}] \leq \frac{3A^{max} - \|\mathbf{x}\|^2_{\mathbb{R}^L}}{\lambda_{\boldsymbol{\sigma}} L}\,.$$

Hence, letting $k \to \infty$ yields the assertion.
Step 4: (Proof of the statement under **(A1******)**) The assertion immediately follows from (8.15) and **(A1******)**.

□

Remark 8.1. **1.** Lemma 8.4 generalizes [62, p. 367, Lem. 3.2] to non-uniform time steps. There, (uniform) time steps are chosen 'small enough' to ensure the statement from Lemma 8.4 without postulating additional assumptions such as **(A1*****)** or **(A1******)**.
2. For **(A1*****)** or **(A1******)**, Lemma 8.4 holds for general mesh sizes, which is needed to establish optimal convergence of the adaptive Algorithm 10.1; see Theorem 10.2. Examples 6.2, 11.1 and 11.3 satisfy **(A1*****)**.
3. For the usual Euler method in (9.57), it is possible to derive a similar result as in Lemma 8.4 *only* under **(A1)** − **(A3)** thanks to *Dynkin's-formula*.
4. The constant $C > 0$ can be explicitly identified under assumption **(A1*****)** resp. **(A1******)**. In the first case

$$\mathbb{E}[t_{J^*}] \leq \frac{1}{\lambda_{\boldsymbol{\sigma}} L}\Big(3\max_{\mathbf{z}\in\overline{\boldsymbol{\mathcal{D}}}}\|\mathbf{z}\|^2_{\mathbb{R}^L} - \|\mathbf{x}\|^2_{\mathbb{R}^L}\Big) \leq \frac{1}{\lambda_{\boldsymbol{\sigma}} L} \cdot 3\max_{\mathbf{z}\in\overline{\boldsymbol{\mathcal{D}}}}\|\mathbf{z}\|^2_{\mathbb{R}^L} =: C\,,$$

while in the second

$$\mathbb{E}[t_{J^*}] \leq \frac{1}{C_{\mathbf{b},\boldsymbol{\sigma},L}}\Big(3\max_{\mathbf{z}\in\overline{\boldsymbol{\mathcal{D}}}}\|\mathbf{z}\|^2_{\mathbb{R}^L} - \|\mathbf{x}\|^2_{\mathbb{R}^L}\Big) \leq \frac{1}{C_{\mathbf{b},\boldsymbol{\sigma},L}} \cdot 3\max_{\mathbf{z}\in\overline{\boldsymbol{\mathcal{D}}}}\|\mathbf{z}\|^2_{\mathbb{R}^L} =: C\,.$$

8.4. Discretization for the parabolic PDE (6.1): Scheme and stability

Fix $(t, \mathbf{x}) \in [0,T) \times \boldsymbol{\mathcal{D}}$ in (6.2) and let $\{t_j\}_{j=0}^{J} \subset [t,T]$ be a mesh with local step sizes $\{\tau^{j+1}\}_{j=0}^{J-1}$, where $J \equiv J(t,\mathbf{x}) \in \mathbb{N}$. We use (8.1) and (8.3) with $Y_V^j \equiv 1$, and where '$g(\mathbf{Y}_{\mathbf{X}}^j)$' is replaced by '$g(t_j, \mathbf{Y}_{\mathbf{X}}^j)$' to approximate (6.3) and (6.5). In the following, we state Scheme 3, which is closely based on [62, p. 353 *ff.*, Subsec. 6.2.1], and which can be seen as an analog to Scheme 2 in the elliptic setting to approximate (6.2) by '$\mathbb{E}\Big[\phi(t_{J^*}, \mathbf{Y}_{\mathbf{X}}^{J^*}) + Y_Z^{J^*}\Big]$'.

Scheme 3. Let $j \geq 0$. Let $(\mathbf{Y}_{\mathbf{X}}^j, Y_Z^j)$ be given with $\mathbf{Y}_{\mathbf{X}}^k \in \boldsymbol{\mathcal{D}}$, $k = 0, ..., j$.

(1) Proceed as in **(1) - (4)** in Scheme 2.

(2) (' `Stop`') If $j+1 = J$, set $J^* := j+1$, $\mathbf{Y}_{\mathbf{X}}^{J^*} := \mathbf{Y}_{\mathbf{X}}^{J}$, $Y_Z^{J^*} := Y_Z^J$, and **STOP**.

For $j > J^*$, we set $(\mathbf{Y}_{\mathbf{X}}^j, Y_V^j, Y_Z^j) = (\mathbf{Y}_{\mathbf{X}}^{J^*}, Y_V^{J^*}, Y_Z^{J^*})$.

Similar to Lemma 8.3, the following lemma can be considered as a generalization of [62, p. 356, Lem. 2.2] to non-uniform time steps, which estimates the number of iterates $\{\mathbf{Y}_{\mathbf{X}}^j\}_{j=0}^{J}$ from Scheme 3 in the boundary strips.

Lemma 8.5. Assume **(B1)** $-$ **(B3)**. Fix $(t,\mathbf{x}) \in [0,T) \times \boldsymbol{\mathcal{D}}$. Let $J \equiv J(t,\mathbf{x}) \in \mathbb{N}$ and $\{t_j\}_{j=0}^{J} \subset [t,T]$ be a mesh with local mesh sizes $\{\tau^{j+1}\}_{j=0}^{J-1}$. Let $\{\mathbf{Y}_{\mathbf{X}}^j\}_{j=0}^{J}$ be from Scheme 3. Then

$$\sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big] < 2\,.$$

9. A posteriori weak error analysis

In Section 9.1, we derive an *a posteriori* error estimate for iterates $\{(\mathbf{Y}^j_{\mathbf{X}}, Y^j_V, Y^j_Z)\}_{j\geq 0}$ of Scheme 2 within the approximative framework of the elliptic PDE (0.6); see Theorem 9.1. It is shown in Theorem 9.6 that the resulting error estimators converge with optimal order 1 on a mesh with maximum mesh size $\tau^{max} > 0$; the relevant tools for its verification are Theorem 9.1 and Lemmas 8.3 and 8.4. Corresponding results for the parabolic PDE (6.1), *cf.* Theorem 9.7 and Theorem 9.12 are derived in Section 9.2. In Section 9.3, we derive an *a posteriori* error estimate for the (usual) Euler scheme and discuss related difficulties.

9.1. A posteriori weak error estimation: Derivation and Optimality for the elliptic PDE (0.6)

The following result bounds the approximation error $\Big|u(\mathbf{x}) - \mathbb{E}\Big[\phi(\mathbf{Y}^{J^*}_{\mathbf{X}})Y^{J^*}_V + Y^{J^*}_Z\Big]\Big|$ in terms of computable *a posteriori* error estimators $\{\mathfrak{G}^{(\cdot)}_\ell\}_{\ell=1}^3$.

Theorem 9.1. Assume **(A1)** − **(A3)** in Section 8.1. Fix $\mathbf{x} \in \boldsymbol{\mathcal{D}}$. Let $\{t_j\}_{j\geq 0} \subset [0,\infty)$ be a mesh with local step sizes $\{\tau^{j+1}\}_{j\geq 0}$. Let $\{(\mathbf{Y}^j_{\mathbf{X}}, Y^j_V, Y^j_Z)\}_{j\geq 0}$ solve Scheme 2. Then we have

$$\Big|u(\mathbf{x}) - \mathbb{E}\Big[\phi(\mathbf{Y}^{J^*}_{\mathbf{X}})Y^{J^*}_V + Y^{J^*}_Z\Big]\Big| \leq \boldsymbol{C}(\phi, g) \cdot \sum_{j=0}^{\infty} \tau^{j+1}\Big\{\mathfrak{G}^{(j)}_1 + \mathfrak{G}^{(j)}_2 + \mathfrak{G}^{(j)}_3\Big\}, \tag{9.1}$$

where $\boldsymbol{C}(\phi, g) > 0$ is the constant from Lemma 8.1, and the *a posteriori* error estimators $\{\mathfrak{G}^{(j)}_\ell\}_{\ell=1}^3$, are given by

$$\begin{aligned}
\mathfrak{G}^{(j)}_1 &:= \mathsf{E_1}(\mathbf{Y}^j_{\mathbf{X}}, Y^j_V) \cdot \tau^{j+1} + \frac{1}{2} \cdot \mathsf{E_2}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^j_{\mathbf{X}}, Y^j_V) + \frac{1}{2} \cdot \mathsf{E_3}(\mathbf{Y}^j_{\mathbf{X}}, Y^j_V) \cdot \tau^{j+1} \\
&\quad + \sqrt{L} \cdot \mathsf{E_4}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^j_{\mathbf{X}}, Y^j_V) \cdot \Big(\tau^{j+1}\Big)^{\frac{1}{2}} + \frac{1}{4} L \cdot \mathsf{E_5}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^j_{\mathbf{X}}, Y^j_V) \\
&\quad + \frac{1}{2} L \cdot \mathsf{E_6}(\mathbf{Y}^j_{\mathbf{X}}, Y^j_V) \cdot \tau^{j+1} + \mathsf{E_7}(\mathbf{Y}^j_{\mathbf{X}}, Y^{j+1}_V, Y^j_V)\,, \\
\mathfrak{G}^{(j)}_2 &:= \mathsf{E_8}(\mathbf{Y}^j_{\mathbf{X}}, \overline{\mathbf{Y}}^j_{\mathbf{X}}, Y^j_V) \cdot \tau^{j+1} + \frac{1}{2} \cdot \mathsf{E_9}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^j_{\mathbf{X}}, \overline{\mathbf{Y}}^j_{\mathbf{X}}, Y^j_V) \\
&\quad + \frac{1}{2} \cdot \mathsf{E_{10}}(\mathbf{Y}^j_{\mathbf{X}}, \overline{\mathbf{Y}}^j_{\mathbf{X}}, Y^j_V) \cdot \tau^{j+1} + \sqrt{L} \cdot \mathsf{E_{11}}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^j_{\mathbf{X}}, \overline{\mathbf{Y}}^j_{\mathbf{X}}, Y^j_V) \cdot \Big(\tau^{j+1}\Big)^{\frac{1}{2}} \\
&\quad + \frac{1}{4} L \cdot \mathsf{E_{12}}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^j_{\mathbf{X}}, \overline{\mathbf{Y}}^j_{\mathbf{X}}, Y^j_V) + \frac{1}{2} L \cdot \mathsf{E_{13}}(\mathbf{Y}^j_{\mathbf{X}}, \overline{\mathbf{Y}}^j_{\mathbf{X}}, Y^j_V) \cdot \tau^{j+1}
\end{aligned}$$

$$+\,\mathbf{E_{14}}(\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}},Y_V^{j+1},Y_V^j)\,,$$
$$\mathfrak{G}_3^{(j)} := 2\cdot \mathbf{E_{15}}\Big(\mathbf{Y}^j_{\mathbf{X}},Y_V^j\Big)\,,$$

with computable terms

1. $\mathbf{E_1}(\mathbf{Y}^j_{\mathbf{X}},Y_V^j)=\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\cdot\big|\big(1-\tau^{j+1}c(\mathbf{Y}^j_{\mathbf{X}})\big)^{-1}c(\mathbf{Y}^j_{\mathbf{X}})\big|\cdot Y_V^j\cdot\big\|\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\big\|_{\mathbb{R}^L}\Big]\,,$
2. $\mathbf{E_2}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}},Y_V^j):=\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\cdot\big|\big(1-\tau^{j+1}c(\mathbf{Y}^j_{\mathbf{X}})\big)^{-1}c(\mathbf{Y}^j_{\mathbf{X}})\big|\cdot Y_V^j\cdot\big\|\mathbf{Y}^{j+1}_{\mathbf{X}}-\mathbf{Y}^j_{\mathbf{X}}\big\|^2_{\mathbb{R}^L}\Big]\,,$
3. $\mathbf{E_3}(\mathbf{Y}^j_{\mathbf{X}},Y_V^j):=\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\cdot Y_V^j\cdot\big\|\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\big\|^2_{\mathbb{R}^L}\Big]\,,$
4. $\mathbf{E_4}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}},Y_V^j):=\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\cdot Y_V^j\cdot\big\|\mathbf{Y}^{j+1}_{\mathbf{X}}-\mathbf{Y}^j_{\mathbf{X}}\big\|_{\mathbb{R}^L}\cdot\big\|\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\big\|_{\mathbb{R}^L}\cdot\big\|\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\big\|_{\mathbb{R}^{L\times L}}\Big]\,,$
5. $\mathbf{E_5}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}},Y_V^j):=\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\cdot Y_V^j\cdot\big\|\mathbf{Y}^{j+1}_{\mathbf{X}}-\mathbf{Y}^j_{\mathbf{X}}\big\|^2_{\mathbb{R}^L}\cdot\big\|\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\big\|^2_{\mathbb{R}^{L\times L}}\Big]\,,$
6. $\mathbf{E_6}(\mathbf{Y}^j_{\mathbf{X}},Y_V^j):=\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\cdot Y_V^j\cdot\big\|\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\big\|_{\mathbb{R}^L}\cdot\big\|\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\big\|^2_{\mathbb{R}^{L\times L}}\Big]\,,$
7. $\mathbf{E_7}(\mathbf{Y}^j_{\mathbf{X}},Y_V^{j+1},Y_V^j):=\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\cdot\big|c(\mathbf{Y}^j_{\mathbf{X}})\big|\cdot\big|Y_V^{j+1}-Y_V^j\big|\Big]\,,$
8. $\mathbf{E_8}(\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}},Y_V^j):=\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}}))\}}\cdot\big|\big(1-\tau^{j+1}c(\overline{\mathbf{Y}}^j_{\mathbf{X}})\big)^{-1}c(\overline{\mathbf{Y}}^j_{\mathbf{X}})\big|$
$\cdot Y_V^j\cdot\big\|\mathbf{b}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\big\|_{\mathbb{R}^L}\Big]\,,$
9. $\mathbf{E_9}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}},Y_V^j):=\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}}))\}}$
$\cdot\big|\big(1-\tau^{j+1}c(\overline{\mathbf{Y}}^j_{\mathbf{X}})\big)^{-1}c(\overline{\mathbf{Y}}^j_{\mathbf{X}})\big|\cdot Y_V^j\cdot\big\|\mathbf{Y}^{j+1}_{\mathbf{X}}-\overline{\mathbf{Y}}^j_{\mathbf{X}}\big\|^2_{\mathbb{R}^L}\Big]\,,$
10. $\mathbf{E_{10}}(\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}},Y_V^j):=\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}}))\}}\cdot Y_V^j\cdot\big\|\mathbf{b}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\big\|^2_{\mathbb{R}^L}\Big]\,,$
11. $\mathbf{E_{11}}(\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}},Y_V^j):=\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}}))\}}\cdot Y_V^j\cdot\big\|\mathbf{Y}^{j+1}_{\mathbf{X}}-\overline{\mathbf{Y}}^j_{\mathbf{X}}\big\|_{\mathbb{R}^L}$
$\cdot\big\|\mathbf{b}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\big\|_{\mathbb{R}^L}\cdot\big\|\boldsymbol{\sigma}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\big\|_{\mathbb{R}^{L\times L}}\Big]\,,$
12. $\mathbf{E_{12}}(\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}},Y_V^j):=\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}}))\}}\cdot Y_V^j\cdot\big\|\mathbf{Y}^{j+1}_{\mathbf{X}}-\overline{\mathbf{Y}}^j_{\mathbf{X}}\big\|^2_{\mathbb{R}^L}$
$\cdot\big\|\boldsymbol{\sigma}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\big\|^2_{\mathbb{R}^{L\times L}}\Big]\,,$
13. $\mathbf{E_{13}}(\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}},Y_V^j):=\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}}))\}}\cdot Y_V^j\cdot\big\|\mathbf{b}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\big\|_{\mathbb{R}^L}$
$\cdot\big\|\boldsymbol{\sigma}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\big\|^2_{\mathbb{R}^{L\times L}}\Big]\,,$
14. $\mathbf{E_{14}}(\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}},Y_V^{j+1},Y_V^j):=\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}}))\}}\cdot\big|c(\overline{\mathbf{Y}}^j_{\mathbf{X}})\big|$
$\cdot\big|Y_V^{j+1}-Y_V^j\big|\Big]\,,$
15. $\mathbf{E_{15}}\big(\mathbf{Y}^j_{\mathbf{X}},Y_V^j\big)=\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\lambda_j^2Y_V^j\Big]\,.$

Remark 9.1. **1.** For $\mathbf{b}(\mathbf{x}) \equiv \mathbf{0}$, $\boldsymbol{\sigma}(\mathbf{x}) \equiv \sqrt{2} \cdot \mathbb{I}$, where $\mathbb{I}$ denotes the $L-$dimensional identity matrix, which are data requirements in (0.6) for well-known elliptic PDE's such as the *Poisson equation* or *Helmholtz equation*, the particular error estimators $\{\mathfrak{G}_\ell^{(\cdot)}\}_{\ell=1}^3$ simplify considerably. For *Poisson's equation*, where additionally $c(\mathbf{x}) \equiv 0$ is required in (0.6), only $\mathbf{E_5}(\cdot), \mathbf{E_{12}}(\cdot)$ and $\mathbf{E_{15}}(\cdot)$ constitute $\{\mathfrak{G}_\ell^{(\cdot)}\}_{\ell=1}^3$; *cf.* Example 6.1.
2. The derivation of the *a posteriori* error estimate (9.1) crucially depends on the use of the *weak* Euler method (8.1) and the associated procedure in Scheme 2. Note that the right-hand side of (9.1) is 'computable', *i.e.*, in practice, the terms $\{\mathbf{E}_{\boldsymbol{\ell}}\big(\cdot\big)\}_{\boldsymbol{\ell}=1,\dots,15}$ may be approximated by Monte-Carlo method, which typically provides a basis for an efficient error approximation (see Chapter 11 for further details). In contrast, we present an *a posteriori* error analysis via the explicit Euler method (9.57) in Section 9.3, whose derivation is (also) close to Theorem 3.1 in the first part, and discuss upcoming difficulties, where, in particular, the computation of terms in *a posteriori* form involved there is not straightforward; *cf.* Remark 9.3.
3. The terms $\{\mathbf{E}_{\boldsymbol{\ell}}\big(\cdot\big)\}_{\boldsymbol{\ell}=1,\dots,7}$ in $\mathfrak{G}_1^{(\cdot)}$, which capture dynamics away from $\boldsymbol{\partial\mathcal{D}}$, may be related to the terms in the error estimator in (3.1). The additional terms $\{\mathbf{E}_{\boldsymbol{\ell}}\big(\cdot\big)\}_{\boldsymbol{\ell}=8,\dots,15}$ in $\mathfrak{G}_2^{(\cdot)}$ and $\mathfrak{G}_3^{(\cdot)}$ address stopping dynamics near the boundary, which, however, do not occur in the framework of Theorem 3.1.

The proof of Theorem 9.1 consists of several steps: Lemma 9.2 represents the error on the left-hand side of (9.1) with the help of the (unknown) solution u of (0.6). Lemmas 9.3, 9.4 and 9.5 estimate the expressions '$\boldsymbol{I_j}$', '$\boldsymbol{II_j}$' and '$\boldsymbol{III_j}$' emerging from Lemma 9.2 and given in (9.2), (9.3) and (9.4), respectively. The derivation of the *a posteriori* error estimate (9.1) then follows by combining these lemmas.

Lemma 9.2. Assume **(A1)** − **(A3)**. Fix $\mathbf{x} \in \boldsymbol{\mathcal{D}}$. Let $\{t_j\}_{j\geq 0} \subset [0,\infty)$ be a mesh with local step sizes $\{\tau^{j+1}\}_{j\geq 0}$. Let $\{(\mathbf{Y}_\mathbf{X}^j, Y_V^j, Y_Z^j)\}_{j\geq 0}$ solve Scheme 2. Then we have

$$\Big|u(\mathbf{x}) - \mathbb{E}\Big[\phi(\mathbf{Y}_\mathbf{X}^{J^*})Y_V^{J^*} + Y_Z^{J^*}\Big]\Big| \leq \sum_{j=0}^{\infty} \Big\{\boldsymbol{I_j} + \boldsymbol{II_j} + \boldsymbol{III_j}\Big\},$$

where

$$\boldsymbol{I_j} := \Big|\mathbb{E}\Big[\mathbf{1}_{\Big\{\mathbf{Y}_\mathbf{X}^j \in \boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\Big\}}\Big\{u(\mathbf{Y}_\mathbf{X}^{j+1})Y_V^{j+1} - u(\mathbf{Y}_\mathbf{X}^j)Y_V^j + Y_Z^{j+1} - Y_Z^j\Big\}\Big]\Big|, \tag{9.2}$$

$$\boldsymbol{II_j} := \Big|\mathbb{E}\Big[\mathbf{1}_{\Big\{\mathbf{Y}_\mathbf{X}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\Big\}}\mathbf{1}_{\Big\{\overline{\mathbf{Y}}_\mathbf{X}^j = \mathbf{Y}_\mathbf{X}^j + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_\mathbf{X}^j)\big)\Big\}} \cdot \Big\{u(\mathbf{Y}_\mathbf{X}^{j+1})Y_V^{j+1} - u(\overline{\mathbf{Y}}_\mathbf{X}^j)Y_V^j + Y_Z^{j+1} - Y_Z^j\Big\}\Big]\Big|, \tag{9.3}$$

$$\boldsymbol{III_j} := \Big|\mathbb{E}\Big[\mathbf{1}_{\Big\{\mathbf{Y}_\mathbf{X}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\Big\}}\Big\{u(\overline{\mathbf{Y}}_\mathbf{X}^j)Y_V^j - u(\mathbf{Y}_\mathbf{X}^j)Y_V^j\Big\}\Big]\Big|. \tag{9.4}$$

Proof. Considering PDE (0.6) and observing that $Y_V^0 = 1$, $Y_Z^0 = 0$, a first calculation leads to

$$\Big|u(\mathbf{x}) - \mathbb{E}\Big[\phi(\mathbf{Y}_\mathbf{X}^{J^*})Y_V^{J^*} + Y_Z^{J^*}\Big]\Big| = \Big|u(\mathbf{x})Y_V^0 + Y_Z^0 - \mathbb{E}\Big[u(\mathbf{Y}_\mathbf{X}^{J^*})Y_V^{J^*} + Y_Z^{J^*}\Big]\Big|$$

$$= \left| \mathbb{E}\Big[\sum_{j=0}^{J^*-1} u(\mathbf{Y}_{\mathbf{X}}^{j+1})Y_V^{j+1} + Y_Z^{j+1} - u(\mathbf{Y}_{\mathbf{X}}^{j})Y_V^{j} - Y_Z^{j} \Big] \right|$$
$$= \left| \sum_{j=0}^{\infty} \mathbb{E}\Big[\underbrace{u(\overline{\mathbf{Y}}_{\mathbf{X}}^{j})Y_V^{j} + Y_Z^{j} - u(\mathbf{Y}_{\mathbf{X}}^{j})Y_V^{j} - Y_Z^{j}}_{=:\boldsymbol{d}_j} + \underbrace{u(\mathbf{Y}_{\mathbf{X}}^{j+1})Y_V^{j+1} + Y_Z^{j+1} - u(\overline{\mathbf{Y}}_{\mathbf{X}}^{j})Y_V^{j} - Y_Z^{j}}_{=:\boldsymbol{d}'_j} \Big] \right|. \tag{9.5}$$

Since $\boldsymbol{d}_j \equiv 0$ on the event $\Big\{\mathbf{Y}_{\mathbf{X}}^j \notin \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\Big\}$, $\boldsymbol{d}'_j \equiv 0$ on the event $\Big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\Big\} \cap \Big\{\overline{\mathbf{Y}}_{\mathbf{X}}^j = \Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\Big\}$, $\boldsymbol{d}'_j \equiv 0$ on the event $\Big\{\mathbf{Y}_{\mathbf{X}}^j \notin \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\Big\} \cap \Big\{\mathbf{Y}_{\mathbf{X}}^j \notin \boldsymbol{\mathcal{D}}\Big\}$ and $\Big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\Big\} \cap \Big\{\mathbf{Y}_{\mathbf{X}}^j \notin \boldsymbol{\mathcal{D}}\Big\} = \emptyset$, the assertion follows from (9.5).

□

Lemma 9.3. Assume **(A1) – (A3)**. Fix $\mathbf{x} \in \boldsymbol{\mathcal{D}}$. Let $\{t_j\}_{j\geq 0} \subset [0,\infty)$ be a mesh with local step sizes $\{\tau^{j+1}\}_{j\geq 0}$. Let $\{(\mathbf{Y}_{\mathbf{X}}^j, Y_V^j, Y_Z^j)\}_{j\geq 0}$ solve Scheme 2. Then, for every $j \geq 0$, we have

$$\boldsymbol{I_j} \leq \boldsymbol{C}(\phi, g) \cdot \mathfrak{G}_1^{(j)} \cdot \tau^{j+1},$$

where $\boldsymbol{I_j}$ is given in (9.2), and $\boldsymbol{C}(\phi, g) > 0$ is from Lemma 8.1.

Proof. In the following, we write $A_j := \Big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\Big\}$ to simplify the notation. In a first step, we rewrite $\boldsymbol{I_j}$ by making use of (8.2) and (8.3) in Scheme 1.

$$\begin{aligned} \boldsymbol{I_j} &= \left| \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \Big\{ u(\mathbf{Y}_{\mathbf{X}}^{j+1})Y_V^{j+1} - u(\mathbf{Y}_{\mathbf{X}}^{j+1})Y_V^{j} + u(\mathbf{Y}_{\mathbf{X}}^{j+1})Y_V^{j} - u(\mathbf{Y}_{\mathbf{X}}^{j})Y_V^{j} + g(\mathbf{Y}_{\mathbf{X}}^{j})Y_V^{j} \cdot \tau^{j+1} \Big\} \Big] \right| \\ &= \left| \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \Big\{ u(\mathbf{Y}_{\mathbf{X}}^{j+1})\Big(1 - \tau^{j+1}c(\mathbf{Y}_{\mathbf{X}}^{j})\Big)^{-1} c(\mathbf{Y}_{\mathbf{X}}^{j})Y_V^{j} \cdot \tau^{j+1} \right. \\ &\qquad \left. + \Big[u(\mathbf{Y}_{\mathbf{X}}^{j+1}) - u(\mathbf{Y}_{\mathbf{X}}^{j})\Big]Y_V^{j} + g(\mathbf{Y}_{\mathbf{X}}^{j})Y_V^{j} \cdot \tau^{j+1} \Big\} \Big] \right|. \end{aligned} \tag{9.6}$$

Step 1: (Employing PDE (0.6)) We use *Taylor's formula* to deduce from (9.6)

$$\begin{aligned} \boldsymbol{I_j} = \Bigg| \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot Y_V^{j} \cdot \Big\{ &\Big[u(\mathbf{Y}_{\mathbf{X}}^{j+1}) - u(\mathbf{Y}_{\mathbf{X}}^{j})\Big]\Big(1 - \tau^{j+1}c(\mathbf{Y}_{\mathbf{X}}^{j})\Big)^{-1} c(\mathbf{Y}_{\mathbf{X}}^{j}) \cdot \tau^{j+1} \\ &+ u(\mathbf{Y}_{\mathbf{X}}^{j})\Big(1 - \tau^{j+1}c(\mathbf{Y}_{\mathbf{X}}^{j})\Big)^{-1} c(\mathbf{Y}_{\mathbf{X}}^{j}) \cdot \tau^{j+1} + g(\mathbf{Y}_{\mathbf{X}}^{j}) \cdot \tau^{j+1} \\ &+ \Big\langle D_{\mathbf{x}}u(\mathbf{Y}_{\mathbf{X}}^{j}), \mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^{j} \Big\rangle_{\mathbb{R}^L} + \frac{1}{2}\mathrm{Tr}\Big(D_{\mathbf{x}}^2 u(\hat{\mathbf{Y}}_{\mathbf{X}}^{j})(\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^{j})(\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^{j})^\top \Big) \Big\} \Big] \Bigg|, \end{aligned} \tag{9.7}$$

for some $\hat{\mathbf{Y}}_{\mathbf{X}}^j$ between $\mathbf{Y}_{\mathbf{X}}^{j+1}$ and $\mathbf{Y}_{\mathbf{X}}^j$, *i.e.*, $\hat{\mathbf{Y}}_{\mathbf{X}}^j = \mathbf{Y}_{\mathbf{X}}^j + \theta\Big(\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^j\Big)$ with $\theta \in (0,1)$. Now, we use the identity in (0.6) to restate $g(\mathbf{Y}_{\mathbf{X}}^j) \cdot \tau^{j+1}$ in (9.7)

$$\begin{aligned} \boldsymbol{I_j} = \Bigg| \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot Y_V^{j} \cdot \Big\{ &\Big[u(\mathbf{Y}_{\mathbf{X}}^{j+1}) - u(\mathbf{Y}_{\mathbf{X}}^{j})\Big]\Big(1 - \tau^{j+1}c(\mathbf{Y}_{\mathbf{X}}^{j})\Big)^{-1} c(\mathbf{Y}_{\mathbf{X}}^{j}) \cdot \tau^{j+1} \\ &+ u(\mathbf{Y}_{\mathbf{X}}^{j})\Big(1 - \tau^{j+1}c(\mathbf{Y}_{\mathbf{X}}^{j})\Big)^{-1} c(\mathbf{Y}_{\mathbf{X}}^{j}) \cdot \tau^{j+1} - u(\mathbf{Y}_{\mathbf{X}}^{j})c(\mathbf{Y}_{\mathbf{X}}^{j}) \cdot \tau^{j+1} \end{aligned}$$

$$
\begin{aligned}
&- \langle \mathbf{b}(\mathbf{Y}^j_{\mathbf{X}}), D_{\mathbf{x}} u(\mathbf{Y}^j_{\mathbf{X}})\rangle_{\mathbb{R}^L} \cdot \tau^{j+1} - \frac{1}{2}\mathrm{Tr}\Big(\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\sigma}^\top(\mathbf{Y}^j_{\mathbf{X}}) D^2_{\mathbf{x}} u(\mathbf{Y}^j_{\mathbf{X}})\Big) \cdot \tau^{j+1} \\
&+ \Big\langle D_{\mathbf{x}} u(\mathbf{Y}^j_{\mathbf{X}}), \mathbf{Y}^{j+1}_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}}\Big\rangle_{\mathbb{R}^L} + \frac{1}{2}\mathrm{Tr}\Big(D^2_{\mathbf{x}} u(\hat{\mathbf{Y}}^j_{\mathbf{X}})(\mathbf{Y}^{j+1}_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}})(\mathbf{Y}^{j+1}_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}})^\top\Big)\Big\}\Big]\Big|.
\end{aligned}
\tag{9.8}
$$

Next, we use (8.1) to represent $\mathbf{Y}^{j+1}_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}}$ in (9.8), and standard calculations,

$$
\boldsymbol{I_j} \leq \boldsymbol{K_1} + \boldsymbol{K_2} + \boldsymbol{K_3} + \boldsymbol{K_4} + \boldsymbol{K_5}\,, \tag{9.9}
$$

where

$$
\boldsymbol{K_1} := \Big|\mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \Big(1 - \tau^{j+1} c(\mathbf{Y}^j_{\mathbf{X}})\Big)^{-1} c(\mathbf{Y}^j_{\mathbf{X}}) Y^j_V \cdot \Big\{u(\mathbf{Y}^{j+1}_{\mathbf{X}}) - u(\mathbf{Y}^j_{\mathbf{X}})\Big\}\Big]\Big| \cdot \tau^{j+1},
$$

$$
\boldsymbol{K_2} := \frac{1}{2}\Big|\mathbb{E}\Big[\mathbf{1}_{A_j} \cdot Y^j_V \cdot \mathrm{Tr}\Big(D^2_{\mathbf{x}} u(\hat{\mathbf{Y}}^j_{\mathbf{X}})\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})^\top\Big)\Big]\Big| \cdot \Big(\tau^{j+1}\Big)^2,
$$

$$
\begin{aligned}
\boldsymbol{K_3} : = \frac{1}{2}\Big|\mathbb{E}\Big[\mathbf{1}_{A_j} \cdot Y^j_V \cdot \mathrm{Tr}\Big(D^2_{\mathbf{x}} u(\hat{\mathbf{Y}}^j_{\mathbf{X}})\Big\{\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\Big(\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\Big)^\top \\
+ \boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})^\top\Big\}\Big)\Big]\Big| \cdot \Big(\tau^{j+1}\Big)^{\frac{3}{2}},
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{K_4} : = \frac{1}{2}\Big|\mathbb{E}\Big[\mathbf{1}_{A_j} \cdot Y^j_V \cdot \Big\{\mathrm{Tr}\Big(D^2_{\mathbf{x}} u(\hat{\mathbf{Y}}^j_{\mathbf{X}})\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\Big(\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\Big)^\top\Big) \\
- \mathrm{Tr}\Big(D^2_{\mathbf{x}} u(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\sigma}^\top(\mathbf{Y}^j_{\mathbf{X}})\Big)\Big\}\Big]\Big| \cdot \tau^{j+1}\,,
\end{aligned}
$$

$$
\boldsymbol{K_5} := \Big|\mathbb{E}\Big[\mathbf{1}_{A_j} \cdot u(\mathbf{Y}^j_{\mathbf{X}}) c(\mathbf{Y}^j_{\mathbf{X}}) \cdot \Big\{Y^{j+1}_V - Y^j_V\Big\}\Big]\Big| \cdot \tau^{j+1}\,.
$$

Step 2: (Estimation of $\boldsymbol{K_1}, \boldsymbol{K_2}, \boldsymbol{K_3}, \boldsymbol{K_4}, \boldsymbol{K_5}$) We estimate the terms in (9.9) independently.
a) (Estimation of $\boldsymbol{K_1}$) We use *Taylor's formula* to get

$$
\begin{aligned}
\boldsymbol{K_1} = \Big|\mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \Big(1 - \tau^{j+1} c(\mathbf{Y}^j_{\mathbf{X}})\Big)^{-1} c(\mathbf{Y}^j_{\mathbf{X}}) Y^j_V \cdot \Big\{\Big\langle D_{\mathbf{x}} u(\mathbf{Y}^j_{\mathbf{X}}), \mathbf{Y}^{j+1}_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}}\Big\rangle_{\mathbb{R}^L} \\
+ \frac{1}{2}\mathrm{Tr}\Big(D^2_{\mathbf{x}} u(\hat{\hat{\mathbf{Y}}}^j_{\mathbf{X}})(\mathbf{Y}^{j+1}_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}})(\mathbf{Y}^{j+1}_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}})^\top\Big)\Big\}\Big]\Big| \cdot \tau^{j+1}\,,
\end{aligned}
$$

for some $\hat{\hat{\mathbf{Y}}}^j_{\mathbf{X}}$ between $\mathbf{Y}^{j+1}_{\mathbf{X}}$ and $\mathbf{Y}^j_{\mathbf{X}}$. Using again (8.1) for $\mathbf{Y}^{j+1}_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}}$, independency, the fact that $Y^j_V > 0$ (this follows from the generation of Y^j_V via the implicit Euler method (8.2)), Lemma 8.1 and standard arguments lead to

$$
\boldsymbol{K_1} \leq \boldsymbol{C}(\phi, g) \cdot \mathbf{E_1}(\mathbf{Y}^j_{\mathbf{X}}, Y^j_V) \cdot \Big(\tau^{j+1}\Big)^2 + \frac{1}{2}\boldsymbol{C}(\phi, g) \cdot \mathbf{E_2}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^j_{\mathbf{X}}, Y^j_V) \cdot \tau^{j+1}\,. \tag{9.10}
$$

b) (Estimation of $\boldsymbol{K_2}$) Lemma 8.1 and standard arguments immediately lead to

$$\boldsymbol{K_2} \leq \frac{1}{2}\boldsymbol{C}(\phi, g) \cdot \mathbf{E_3}(\mathbf{Y}^j_{\mathbf{X}}, Y^j_V) \cdot \left(\tau^{j+1}\right)^2 . \tag{9.11}$$

c) (Estimation of $\boldsymbol{K_3}$) We add and substract $D^2_{\mathbf{x}} u(\mathbf{Y}^j_{\mathbf{X}})$, use independency and Lemma 8.1 to obtain

$$\begin{aligned}
\boldsymbol{K_3} = \frac{1}{2}\bigg|\mathbb{E}\Big[\mathbf{1}_{A_j} \cdot Y^j_V \cdot \mathrm{Tr}\Big(&\Big\{D^2_{\mathbf{x}} u(\hat{\mathbf{Y}}^j_{\mathbf{X}}) - D^2_{\mathbf{x}} u(\mathbf{Y}^j_{\mathbf{X}})\Big\} \\
&\Big\{\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\Big(\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\Big)^\top + \boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})^\top\Big\}\Big)\Big]\bigg| \cdot \left(\tau^{j+1}\right)^{\frac{3}{2}} \\
\leq \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot Y^j_V \cdot \|D^2_{\mathbf{x}} u(\hat{\mathbf{Y}}^j_{\mathbf{X}}) - D^2_{\mathbf{x}} u(\mathbf{Y}^j_{\mathbf{X}})\|_{\mathbb{R}^{L\times L}} &\|\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\|_{\mathbb{R}^L} \cdot \|\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\|_{\mathbb{R}^L}\Big] \cdot \left(\tau^{j+1}\right)^{\frac{3}{2}} \\
\leq \boldsymbol{C}(\phi, g) \cdot \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot Y^j_V \cdot \|\hat{\mathbf{Y}}^j_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}}\|_{\mathbb{R}^L} &\|\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\|_{\mathbb{R}^L} \cdot \|\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\|_{\mathbb{R}^L}\Big] \cdot \left(\tau^{j+1}\right)^{\frac{3}{2}} ,
\end{aligned} \tag{9.12}$$

where we estimate $\|D^2_{\mathbf{x}} u(\hat{\mathbf{Y}}^j_{\mathbf{X}}) - D^2_{\mathbf{x}} u(\mathbf{Y}^j_{\mathbf{X}})\|_{\mathbb{R}^{L\times L}} \leq \boldsymbol{C}(\phi, g) \cdot \|\hat{\mathbf{Y}}^j_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}}\|_{\mathbb{R}^L}$. In order to estimate the term $\|\hat{\mathbf{Y}}^j_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}}\|_{\mathbb{R}^L}$ in (9.12), we recall that $\hat{\mathbf{Y}}^j_{\mathbf{X}}$ is a point between $\mathbf{Y}^j_{\mathbf{X}}$ and $\mathbf{Y}^j_{\mathbf{X}}$, *i.e.*, $\hat{\mathbf{Y}}^j_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}} = \theta\Big(\mathbf{Y}^{j+1}_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}}\Big)$ with $\theta \in (0,1)$; thus we have

$$\|\hat{\mathbf{Y}}^j_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}}\|_{\mathbb{R}^L} \leq \|\mathbf{Y}^{j+1}_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}}\|_{\mathbb{R}^L} . \tag{9.13}$$

Plugging (9.13) into (9.12) and using $\|\boldsymbol{\xi}_{j+1}\|_{\mathbb{R}^L} = \sqrt{L}$ then leads to

$$\boldsymbol{K_3} \leq \boldsymbol{C}(\phi, g)\sqrt{L} \cdot \mathbf{E_4}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^j_{\mathbf{X}}, Y^j_V) \cdot \left(\tau^{j+1}\right)^{\frac{3}{2}} . \tag{9.14}$$

d) (Estimation of $\boldsymbol{K_4}$) We start with a straightforward rewriting of $\boldsymbol{K_4}$.

$$\begin{aligned}
\boldsymbol{K_4} = \frac{1}{2}\bigg|\mathbb{E}\Big[\mathbf{1}_{A_j} \cdot Y^j_V \cdot \Big\{&\Big\langle\Big\{D^2_{\mathbf{x}} u(\hat{\mathbf{Y}}^j_{\mathbf{X}}) - D^2_{\mathbf{x}} u(\mathbf{Y}^j_{\mathbf{X}})\Big\}\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}, \boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\Big\rangle_{\mathbb{R}^L} \\
&+ \Big\langle D^2_{\mathbf{x}} u(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}, \boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\Big\rangle_{\mathbb{R}^L} \\
&- \mathrm{Tr}\Big(D^2_{\mathbf{x}} u(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\sigma}^\top(\mathbf{Y}^j_{\mathbf{X}})\Big)\Big\}\Big]\bigg| \cdot \tau^{j+1} .
\end{aligned}$$

By the independence of $\xi^{(i)}_{j+1}$ and $\xi^{(k)}_{j+1}$ (note that $\boldsymbol{\xi}_{j+1} = (\xi^{(1)}_{j+1}, ..., \xi^{(L)}_{j+1})^\top$), and Lemma 8.1, we deduce

$$\boldsymbol{K_4} \leq \boldsymbol{K_{4,1}} , \tag{9.15}$$

where

$$\begin{aligned}
\boldsymbol{K_{4,1}} : &= \frac{1}{2}\bigg|\mathbb{E}\Big[\mathbf{1}_{A_j} \cdot Y^j_V \cdot \Big\{\Big\langle\Big\{D^2_{\mathbf{x}} u(\hat{\mathbf{Y}}^j_{\mathbf{X}}) - D^2_{\mathbf{x}} u(\mathbf{Y}^j_{\mathbf{X}})\Big\}\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}, \boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\Big\rangle_{\mathbb{R}^L}\Big]\bigg| \cdot \tau^{j+1} \\
&= \frac{1}{2}\bigg|\mathbb{E}\Big[\mathbf{1}_{A_j} \cdot Y^j_V \cdot \Big\langle D^3_{\mathbf{x}} u(\mathbf{Y}^j_{\mathbf{X}})\Big\{\hat{\mathbf{Y}}^j_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}}\Big\}\boldsymbol{Z}_j, \boldsymbol{Z}_j\Big\rangle_{\mathbb{R}^L}\Big]\bigg| \cdot \tau^{j+1}
\end{aligned}$$

$$+\frac{1}{4}\Big|\mathbb{E}\Big[\mathbf{1}_{A_j}\cdot Y_V^j\cdot\Big\langle D_{\mathbf{x}}^4 u(\hat{\hat{\mathbf{Y}}}_{\mathbf{X}}^j)\Big\{\hat{\mathbf{Y}}_{\mathbf{X}}^j-\mathbf{Y}_{\mathbf{X}}^j\Big\}\Big\{\hat{\mathbf{Y}}_{\mathbf{X}}^j-\mathbf{Y}_{\mathbf{X}}^j\Big\}\boldsymbol{Z}_j,\boldsymbol{Z}_j\Big\rangle_{\mathbb{R}^L}\Big]\Big|\cdot\tau^{j+1},$$

by *Taylor's formula*, where $\boldsymbol{Z}_j := \boldsymbol{\sigma}(\mathbf{Y}_{\mathbf{X}}^j)\boldsymbol{\xi}_{j+1}$, and, for some $\hat{\hat{\mathbf{Y}}}_{\mathbf{X}}^j$ between $\mathbf{Y}_{\mathbf{X}}^j$ and $\hat{\mathbf{Y}}_{\mathbf{X}}^j$. Next, by Lemma 8.1, (9.13), since $\|\boldsymbol{\xi}_{j+1}\|_{\mathbb{R}^L}^2 = L$, we estimate

$$\boldsymbol{K_{4,1}} \leq \boldsymbol{K_{4,1,1}} + \frac{1}{4}\boldsymbol{C}(\phi,g)L\cdot\mathbf{E_5}(\mathbf{Y}_{\mathbf{X}}^{j+1},\mathbf{Y}_{\mathbf{X}}^j,Y_V^j)\cdot\tau^{j+1}, \tag{9.16}$$

where

$$\boldsymbol{K_{4,1,1}} := \frac{1}{2}\Big|\mathbb{E}\Big[\mathbf{1}_{A_j}\cdot Y_V^j\cdot\Big\langle D_{\mathbf{x}}^3 u(\mathbf{Y}_{\mathbf{X}}^j)\Big\{\hat{\mathbf{Y}}_{\mathbf{X}}^j-\mathbf{Y}_{\mathbf{X}}^j\Big\}\boldsymbol{\sigma}(\mathbf{Y}_{\mathbf{X}}^j)\boldsymbol{\xi}_{j+1},\boldsymbol{\sigma}(\mathbf{Y}_{\mathbf{X}}^j)\boldsymbol{\xi}_{j+1}\Big\rangle_{\mathbb{R}^L}\Big]\Big|\cdot\tau^{j+1}.$$

In order to estimate $\boldsymbol{K_{4,1,1}}$, we again use the representation $\hat{\mathbf{Y}}_{\mathbf{X}}^j - \mathbf{Y}_{\mathbf{X}}^j = \theta\Big(\mathbf{Y}_{\mathbf{X}}^{j+1}-\mathbf{Y}_{\mathbf{X}}^j\Big)$ with $\theta\in(0,1)$, and (8.1) to represent $\mathbf{Y}_{\mathbf{X}}^{j+1}-\mathbf{Y}_{\mathbf{X}}^j$,

$$\boldsymbol{K_{4,1,1}} \leq \frac{1}{2}\boldsymbol{C}(\phi,g)L\cdot\mathbf{E_6}(\mathbf{Y}_{\mathbf{X}}^j,Y_V^j)\cdot\Big(\tau^{j+1}\Big)^2. \tag{9.17}$$

We combine (9.17) with (9.16) and plug the resulting expression into (9.15) to obtain

$$\boldsymbol{K_4} \leq \frac{1}{4}\boldsymbol{C}(\phi,g)L\cdot\mathbf{E_5}(\mathbf{Y}_{\mathbf{X}}^{j+1},\mathbf{Y}_{\mathbf{X}}^j,Y_V^j)\cdot\tau^{j+1} + \frac{1}{2}\boldsymbol{C}(\phi,g)L\cdot\mathbf{E_6}(\mathbf{Y}_{\mathbf{X}}^j,Y_V^j)\cdot\Big(\tau^{j+1}\Big)^2. \tag{9.18}$$

e) (Estimation of $\boldsymbol{K_5}$) Lemma 8.1 and standard arguments immediately lead to

$$\boldsymbol{K_5} \leq \boldsymbol{C}(\phi,g)\cdot\mathbf{E_7}(\mathbf{Y}_{\mathbf{X}}^j,Y_V^{j+1},Y_V^j)\cdot\tau^{j+1}. \tag{9.19}$$

Step 3: Finally, combining (9.10), (9.11), (9.14), (9.18) and (9.19) with (9.9) proves the assertion.

□

The following lemma estimates $\boldsymbol{II_j}$ from (9.3). Its proof is very similar to the proof of Lemma 9.3 and is thus omitted. In fact, by replacing A_j resp. $\mathbf{Y}_{\mathbf{X}}^j$ in the proof of Lemma 9.3 by $\overline{A}_j := \mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^j\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\mathbf{1}_{\Big\{\overline{\mathbf{Y}}_{\mathbf{X}}^j=\mathbf{Y}_{\mathbf{X}}^j+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\big)\Big\}}$ resp. $\overline{\mathbf{Y}}_{\mathbf{X}}^j$, and observing that $\mathbf{Y}_{\mathbf{X}}^{j+1}$ is generated starting from $\overline{\mathbf{Y}}_{\mathbf{X}}^j$, yield the proof of Lemma 9.4.

Lemma 9.4. Assume **(A1)** – **(A3)**. Fix $\mathbf{x}\in\boldsymbol{\mathcal{D}}$. Let $\{t_j\}_{j\geq 0}\subset[0,\infty)$ be a mesh with local step sizes $\{\tau^{j+1}\}_{j\geq 0}$. Let $\{(\mathbf{Y}_{\mathbf{X}}^j,Y_V^j,Y_Z^j)\}_{j\geq 0}$ solve Scheme 2. Then, for every $j\geq 0$, we have

$$\boldsymbol{II_j} \leq \boldsymbol{C}(\phi,g)\cdot\mathfrak{G}_2^{(j)}\cdot\tau^{j+1},$$

where $\boldsymbol{II_j}$ is given in (9.3), and $\boldsymbol{C}(\phi,g)>0$ is from Lemma 8.1.

The next lemma estimates $\boldsymbol{III_j}$ from (9.4).

Lemma 9.5. Assume **(A1)** − **(A3)**. Fix $\mathbf{x} \in \boldsymbol{\mathcal{D}}$. Let $\{t_j\}_{j\geq 0} \subset [0,\infty)$ be a mesh with local step sizes $\{\tau^{j+1}\}_{j\geq 0}$. Let $\{(\mathbf{Y}^j_{\mathbf{X}}, Y^j_V, Y^j_Z)\}_{j\geq 0}$ solve Scheme 2. Then, for every $j \geq 0$, we have

$$\boldsymbol{III_j} \leq \boldsymbol{C}(\phi, g) \cdot \mathfrak{G}^{(j)}_3 \cdot \tau^{j+1},$$

where $\boldsymbol{III_j}$ is given in (9.4), and $\boldsymbol{C}(\phi, g) > 0$ is from Lemma 8.1.

Proof. We take the conditional expectation *w.r.t.* $\mathbf{Y}^j_{\mathbf{X}}$ and use measureability arguments to obtain in a first calculation

$$\begin{aligned}\boldsymbol{III_j} &= \Bigg|\mathbb{E}\Bigg[\mathbb{E}\Bigg[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}} Y^j_V\Big\{u(\mathbf{Y}^j_{\mathbf{X}}) - \mathbf{1}_{\left\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\right\}} u\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big) \\ &\qquad - \mathbf{1}_{\left\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\left(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\right)\right\}} u\Big(\mathbf{Y}^j_{\mathbf{X}} + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big)\Big)\Big\} \,\Big|\, \mathbf{Y}^j_{\mathbf{X}}\Bigg]\Bigg]\Bigg| \\ &= \Bigg|\mathbb{E}\Bigg[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}} Y^j_V\Big\{u(\mathbf{Y}^j_{\mathbf{X}}) - p_j\cdot u\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big) \\ &\qquad - (1-p_j)\cdot u\Big(\mathbf{Y}^j_{\mathbf{X}} + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big)\Big)\Big\}\Bigg]\Bigg|,\end{aligned}$$

where p_j is given in (8.5). We apply the *mean value theorem* twice to get

$$\begin{aligned}\boldsymbol{III_j} &= \Bigg|\mathbb{E}\Bigg[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}} Y^j_V\Big\{u(\mathbf{Y}^j_{\mathbf{X}}) - u(\mathbf{Y}^j_{\mathbf{X}} + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}}))) \\ &\qquad + \frac{u(\mathbf{Y}^j_{\mathbf{X}} + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})) - u(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})))}{\|\mathbf{Y}^j_{\mathbf{X}} + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})) - \Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\|_{\mathbb{R}^L}} \cdot \lambda_j\sqrt{\tau^{j+1}}\Big\}\Bigg]\Bigg| \\ &\leq \mathbb{E}\Bigg[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}} |Y^j_V|\Bigg| - \Big\langle D_{\mathbf{x}}u(\hat{\mathbf{Y}}^j_{\mathbf{X}}), \boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}}))\Big\rangle_{\mathbb{R}^L} \cdot \lambda_j\sqrt{\tau^{j+1}} \\ &\qquad + \frac{\Big\langle D_{\mathbf{x}}u(\hat{\hat{\mathbf{Y}}}^j_{\mathbf{X}}), \mathbf{Y}^j_{\mathbf{X}} + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})) - \Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big\rangle_{\mathbb{R}^L}}{\|\mathbf{Y}^j_{\mathbf{X}} + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})) - \Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\|_{\mathbb{R}^L}} \cdot \lambda_j\sqrt{\tau^{j+1}}\Bigg|\Bigg], \end{aligned} \tag{9.20}$$

for some points $\hat{\mathbf{Y}}^j_{\mathbf{X}}$, $\hat{\hat{\mathbf{Y}}}^j_{\mathbf{X}}$ between $\mathbf{Y}^j_{\mathbf{X}}$ and $\mathbf{Y}^j_{\mathbf{X}} + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big)$, and between $\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})$ and $\mathbf{Y}^j_{\mathbf{X}} + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big)$, respectively.
Next, we define $\varphi\Big(\mathbf{z}, \boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big)\Big) := \Big\langle D_{\mathbf{x}}u(\mathbf{z}), \boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big)\Big\rangle_{\mathbb{R}^L}$, $\mathbf{z} \in \boldsymbol{\mathcal{D}}$. Since

$$\boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big) = \frac{\mathbf{Y}^j_{\mathbf{X}} + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big) - \Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})}{\Big\|\mathbf{Y}^j_{\mathbf{X}} + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big) - \Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big\|_{\mathbb{R}^L}},$$

we can rewrite (9.20) as follows,

$$\boldsymbol{III_j} \leq \mathbb{E}\Bigg[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}} \lambda_j\Big|Y^j_V\Big|\Big|\varphi\Big(\hat{\hat{\mathbf{Y}}}^j_{\mathbf{X}}, \boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big)\Big) - \varphi\Big(\hat{\mathbf{Y}}^j_{\mathbf{X}}, \boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big)\Big)\Big|\Bigg] \cdot \sqrt{\tau^{j+1}}$$

$$\leq \sup_{\mathbf{z}\in\boldsymbol{\mathcal{D}}} \sup_{\|\mathbf{v}\|_{\mathbb{R}^L}=1} \|D_{\mathbf{z}}\varphi(\mathbf{z},\mathbf{v})\|_{\mathbb{R}^L} \cdot \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}} \lambda_j \big|Y^j_V\big| \|\hat{\mathbf{Y}}^j_{\mathbf{X}} - \hat{\hat{\mathbf{Y}}}^j_{\mathbf{X}}\|_{\mathbb{R}^L}\Big] \cdot \sqrt{\tau^{j+1}}\,. \tag{9.21}$$

Since $\|\hat{\mathbf{Y}}^j_{\mathbf{X}} - \hat{\hat{\mathbf{Y}}}^j_{\mathbf{X}}\|_{\mathbb{R}^L} \leq 2\lambda_j\sqrt{\tau^{j+1}}$ $\mathbb{P}-a.s.$, and $Y^j_V > 0$, $j \geq 0$, and for all $\mathbf{v} \in \mathbb{R}^L$ with $\|\mathbf{v}\|_{\mathbb{R}^L} = 1$

$$\|D_{\mathbf{z}}\varphi(\mathbf{z},\mathbf{v})\|_{\mathbb{R}^L} = \sqrt{\sum_{i=1}^L \Big|\langle D_{\mathbf{z}}\partial_{z_i} u(\mathbf{z}), \mathbf{v}\rangle_{\mathbb{R}^L}\Big|^2} \leq \boldsymbol{C}(\phi, g)\,,$$

the assertion then follows from (9.21).

□

Next, we show convergence with optimal (weak) order 1 of the *a posteriori* error estimate (9.1) on a mesh with maximum mesh size $\tau^{max} > 0$. Theorems 9.1 and 9.6 then imply the *a priori* estimate [62, p. 369, Thm. 3.4], in particular.

Theorem 9.6. Assume **(A1)** − **(A3)**. Fix $\mathbf{x} \in \boldsymbol{\mathcal{D}}$. Let $\{(\mathbf{Y}^j_{\mathbf{X}}, Y^j_V, Y^j_Z)\}_{j\geq 0}$ solve Scheme 2 on a mesh $\{t_j\}_{j\geq 0} \subset [0,\infty)$ with local mesh sizes $\{\tau^{j+1}\}_{j\geq 0}$ and maximum mesh size $\tau^{max} = \max_j \tau^{j+1} > 0$.

(i) If τ^{max} is sufficiently small, there exists $\boldsymbol{C} > 0$, independent of τ^{max}, such that

$$\boldsymbol{C}(\phi, g) \cdot \sum_{j=0}^{\infty} \tau^{j+1}\Big\{\mathfrak{G}^{(j)}_1 + \mathfrak{G}^{(j)}_2 + \mathfrak{G}^{(j)}_3\Big\} \leq \boldsymbol{C} \cdot \tau^{max}\,.$$

(ii) If **(A1*)** or **(A1**)** holds in addition, there exists $\boldsymbol{C} > 0$, independent of τ^{max}, such that

$$\boldsymbol{C}(\phi, g) \cdot \sum_{j=0}^{\infty} \tau^{j+1}\Big\{\mathfrak{G}^{(j)}_1 + \mathfrak{G}^{(j)}_2 + \mathfrak{G}^{(j)}_3\Big\} \leq \boldsymbol{C} \cdot \tau^{max}\,.$$

Proof. In the following, $\boldsymbol{C} > 0$ is a constant, which might differ from line to line, but is always independent of τ^{max}.

Step 1: We independently bound $\{\mathbf{E_k}(\cdot)\}_{k=1,\ldots,14}$ appearing in the error estimators $\{\mathfrak{G}^{(j)}_\ell\}_{j\geq 0}$, $\ell = 1, 2$, in (9.1).

a) Bounds for $\{\mathbf{E_k}(\cdot)\}_{k=1,\ldots,7}$ in $\mathfrak{G}^{(j)}_1$: Due to **(A1)** − **(A3)**, we have

$$\max_{\mathbf{y}\in\overline{\boldsymbol{\mathcal{D}}}} \|\mathbf{b}(\mathbf{y})\|_{\mathbb{R}^L} \leq \boldsymbol{C}\,, \quad \max_{\mathbf{y}\in\overline{\boldsymbol{\mathcal{D}}}} \|\boldsymbol{\sigma}(\mathbf{y})\|_{\mathbb{R}^{L\times L}} \leq \boldsymbol{C}, \quad \max_{\mathbf{y}\in\overline{\boldsymbol{\mathcal{D}}}} |c(\mathbf{y})| \leq \boldsymbol{C}\,,$$

for some constant $\boldsymbol{C} > 0$. Moreover, since $c \leq 0$, we have $0 < Y^j_V \leq 1$ for every $j \geq 0$. Let $j \geq 0$. We immediately obtain

$$\mathbf{E_k}(\mathbf{Y}^j_{\mathbf{X}}, Y^j_V) \leq \boldsymbol{C} \cdot \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\Big] \qquad \mathbf{k} = 1, 3, 6\,.$$

By means of (8.3) and standard arguments used before, we further get

$$\mathbf{E_k}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^j_{\mathbf{X}}, Y^j_V) \leq \boldsymbol{C} \cdot \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\Big] \cdot \tau^{j+1} \qquad \mathbf{k} = 2, 5\,,$$

and

$$\mathbf{E_4}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^{j}, Y_V^j) \leq \boldsymbol{C} \cdot \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big] \cdot \big(\tau^{j+1}\big)^{\frac{1}{2}}\,.$$

Moreover, due to (8.2), we obtain

$$\mathbf{E_7}(\mathbf{Y}_{\mathbf{X}}^{j}, Y_V^{j+1}, Y_V^j) \leq \boldsymbol{C} \cdot \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big] \cdot \tau^{j+1}\,.$$

Consequently, by considering the representation of $\mathfrak{G}_1^{(j)}$ in Theorem 9.1, we obtain

$$\mathfrak{G}_1^{(j)} \leq \boldsymbol{C} \cdot \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big] \cdot \tau^{j+1}\,. \tag{9.22}$$

b) Bounds for $\{\mathbf{E_k}(\cdot)\}_{k=8,\ldots,14}$ in $\mathfrak{G}_2^{(j)}$: Similar to **a)**, we obtain

$$\mathfrak{G}_2^{(j)} \leq \boldsymbol{C} \cdot \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}} \mathbf{1}_{\big\{\overline{\mathbf{Y}}_{\mathbf{X}}^j = \mathbf{Y}_{\mathbf{X}}^j + \lambda_j \sqrt{\tau^{j+1}}\boldsymbol{n}\big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\big)\big\}}\Big] \cdot \tau^{j+1}\,. \tag{9.23}$$

Step 2: Let $j \geq 0$. By means of the representation of λ_j in (8.4), and the fact that $0 < Y_V^j \leq 1$, we get

$$\mathfrak{G}_3^{(j)} \leq \boldsymbol{C} \cdot \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big]\,. \tag{9.24}$$

Step 3: We plug (9.22), (9.23) and (9.24) into (9.1), and use $\tau^{j+1} \leq \tau^{max}$, to get

$$\begin{aligned}
&\sum_{j=0}^{\infty} \tau^{j+1}\Big\{\mathfrak{G}_1^{(j)} + \mathfrak{G}_2^{(j)} + \mathfrak{G}_3^{(j)}\Big\} \\
&\quad \leq \boldsymbol{C} \cdot \tau^{max} \cdot \sum_{j=0}^{\infty} \tau^{j+1} \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}} + \mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}} \mathbf{1}_{\big\{\overline{\mathbf{Y}}_{\mathbf{X}}^j = \mathbf{Y}_{\mathbf{X}}^j + \lambda_j \sqrt{\tau^{j+1}}\boldsymbol{n}\big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\big)\big\}}\Big] \\
&\qquad + \boldsymbol{C} \cdot \tau^{max} \cdot \sum_{j=0}^{\infty} \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big] \\
&\quad = \boldsymbol{C} \cdot \tau^{max} \cdot \left\{ \sum_{j=0}^{\infty} \tau^{j+1} \mathbb{E}\Big[\mathbf{1}_{\{J^* > j\}}\Big] + \sum_{j=0}^{\infty} \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big] \right\}.
\end{aligned} \tag{9.25}$$

Due to Lemma 8.3, we have

$$\sum_{j=0}^{\infty} \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big] < 2\,, \tag{9.26}$$

and due to Lemma 8.4, we obtain in both cases **(i)** and **(ii)**

$$\sum_{j=0}^{\infty} \tau^{j+1} \mathbb{E}\Big[\mathbf{1}_{\{J^* > j\}}\Big] = \mathbb{E}[t_{J^*}] \leq C \tag{9.27}$$

for some constant $C > 0$ independent of $j \geq 0$ and τ^{max}.
Step 4: We plug (9.26) and (9.27) into (9.25), which proves assertions **(i)** and **(ii)**.

□

9.2. A posteriori weak error estimation: Derivation and Optimality for the parabolic PDE (6.1)

Close to Theorems 9.1 and 9.6 for the elliptic PDE (0.6), we derive corresponding results for PDE (6.1); see Theorems 9.7 and 9.12 below.

Theorem 9.7. Assume **(B1)** − **(B3)** in Section 8.2. Fix $(t,\mathbf{x}) \in [0,T) \times \boldsymbol{\mathcal{D}}$. Let $J \equiv J(t,\mathbf{x}) \in \mathbb{N}$ and $\{t_j\}_{j=0}^{J} \subset [t,T]$ be a mesh with local step sizes $\{\tau^{j+1}\}_{j=0}^{J-1}$. Let $\{(\mathbf{Y}_{\mathbf{X}}^j, Y_Z^j)\}_{j=0}^{J}$ solve Scheme 3. Then we have

$$\Big|u(t,\mathbf{x}) - \mathbb{E}\Big[\phi(t_{J^*}, \mathbf{Y}_{\mathbf{X}}^{J^*}) + Y_Z^{J^*}\Big]\Big| \leq \mathfrak{C}(\phi,g) \cdot \sum_{j=0}^{J-1} \tau^{j+1}\Big\{\mathfrak{H}_1^{(j)} + \mathfrak{H}_2^{(j)} + \mathfrak{H}_3^{(j)}\Big\}, \tag{9.28}$$

where the *a posteriori* error estimators $\{\mathfrak{H}_\ell^{(j)}\}_{\ell=1}^{3}$, are given by

$$\begin{aligned}
\mathfrak{H}_1^{(j)} :=& \, \mathbf{E_1}(\mathbf{Y}_{\mathbf{X}}^j) \cdot \tau^{j+1} + \frac{3}{2} \cdot \mathbf{E_2}(\mathbf{Y}_{\mathbf{X}}^j) \cdot \tau^{j+1} + \frac{1}{2} \cdot \mathbf{E_3}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j) + \mathbf{E_4}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j) + \mathbf{E_5}(\mathbf{Y}_{\mathbf{X}}^j) \cdot \tau^{j+1} \\
&+ \Big\{\frac{1}{2} C_{D^2\mathbf{b}} + \frac{1}{4} L C_{D^2\boldsymbol{\sigma}\boldsymbol{\sigma}^\top}\Big\} \cdot \mathbf{E_6}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j) + L \cdot \mathbf{E_7}(\mathbf{Y}_{\mathbf{X}}^j) \cdot \tau^{j+1} + \frac{1}{2} L \cdot \mathbf{E_8}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j) \\
&+ \frac{1}{2} L \cdot \mathbf{E_9}(\mathbf{Y}_{\mathbf{X}}^j) \cdot \tau^{j+1} + \frac{1}{2} L \cdot \mathbf{E_{10}}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j) + \frac{1}{2} L \cdot \mathbf{E_{11}}(\mathbf{Y}_{\mathbf{X}}^j) \cdot \tau^{j+1} \\
&+ \sqrt{L} \cdot \mathbf{E_{12}}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j) \cdot \sqrt{\tau^{j+1}} + \big|\mathbf{E_{13}}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j, t_j)\big| + \mathbf{E_{14}}(\mathbf{Y}_{\mathbf{X}}^j) \cdot \tau^{j+1} \\
\mathfrak{H}_2^{(j)} :=& \, \mathbf{E_{15}}(\mathbf{Y}_{\mathbf{X}}^j, \overline{\mathbf{Y}}_{\mathbf{X}}^j) \cdot \tau^{j+1} + \frac{3}{2} \cdot \mathbf{E_{16}}(\mathbf{Y}_{\mathbf{X}}^j, \overline{\mathbf{Y}}_{\mathbf{X}}^j) \cdot \tau^{j+1} + \frac{1}{2} \cdot \mathbf{E_{17}}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j, \overline{\mathbf{Y}}_{\mathbf{X}}^j) + \mathbf{E_{18}}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j, \overline{\mathbf{Y}}_{\mathbf{X}}^j) \\
&+ \mathbf{E_{19}}(\mathbf{Y}_{\mathbf{X}}^j, \overline{\mathbf{Y}}_{\mathbf{X}}^j) \cdot \tau^{j+1} + \Big\{\frac{1}{2} C_{D^2\mathbf{b}} + \frac{1}{4} L C_{D^2\boldsymbol{\sigma}\boldsymbol{\sigma}^\top}\Big\} \cdot \mathbf{E_{20}}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j, \overline{\mathbf{Y}}_{\mathbf{X}}^j) \\
&+ L \cdot \mathbf{E_{21}}(\mathbf{Y}_{\mathbf{X}}^j, \overline{\mathbf{Y}}_{\mathbf{X}}^j) \cdot \tau^{j+1} + \frac{1}{2} L \cdot \mathbf{E_{22}}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j, \overline{\mathbf{Y}}_{\mathbf{X}}^j) + \frac{1}{2} L \cdot \mathbf{E_{23}}(\mathbf{Y}_{\mathbf{X}}^j, \overline{\mathbf{Y}}_{\mathbf{X}}^j) \cdot \tau^{j+1} \\
&+ \frac{1}{2} L \cdot \mathbf{E_{24}}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j, \overline{\mathbf{Y}}_{\mathbf{X}}^j) + \frac{1}{2} L \cdot \mathbf{E_{25}}(\mathbf{Y}_{\mathbf{X}}^j, \overline{\mathbf{Y}}_{\mathbf{X}}^j) \cdot \tau^{j+1} + \sqrt{L} \cdot \mathbf{E_{26}}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j, \overline{\mathbf{Y}}_{\mathbf{X}}^j) \cdot \sqrt{\tau^{j+1}} \\
&+ \big|\mathbf{E_{27}}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j, \overline{\mathbf{Y}}_{\mathbf{X}}^j, t_j)\big| + \mathbf{E_{28}}(\mathbf{Y}_{\mathbf{X}}^j, \overline{\mathbf{Y}}_{\mathbf{X}}^j) \cdot \tau^{j+1} \\
\mathfrak{H}_3^{(j)} :=& \, 2 \cdot \mathbf{E_{29}}(\mathbf{Y}_{\mathbf{X}}^j),
\end{aligned}$$

where $C_{D^2\mathbf{b}} := \sup_{\mathbf{y}\in\overline{\boldsymbol{\mathcal{D}}}} \|D^2\mathbf{b}(\mathbf{y})\|_{\mathcal{L}^2}$, $C_{D^2\boldsymbol{\sigma}\boldsymbol{\sigma}^\top} := \sup_{\mathbf{y}\in\overline{\boldsymbol{\mathcal{D}}}} \sup_{\|\mathbf{v}_i\|_{\mathbb{R}^L}} \|D^2\boldsymbol{\sigma}(\mathbf{y})\boldsymbol{\sigma}^\top(\mathbf{y})(\mathbf{v}_1,\mathbf{v}_2)\|_{\mathbb{R}^{L\times L}}$,
and $\mathfrak{C}(\phi,g) \geq 1$ is the constant from Lemma 8.2, and with computable terms

1. $\mathbf{E_1}(\mathbf{Y}_{\mathbf{X}}^j) := \mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}} \|\mathbf{b}(\mathbf{Y}_{\mathbf{X}}^j)\|_{\mathbb{R}^L}\Big]$,
2. $\mathbf{E_2}(\mathbf{Y}_{\mathbf{X}}^j) := \mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}} \|\mathbf{b}(\mathbf{Y}_{\mathbf{X}}^j)\|_{\mathbb{R}^L}^2\Big]$,
3. $\mathbf{E_3}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j) := \mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}} \|\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^j\|_{\mathbb{R}^L}^2 \|\mathbf{b}(\mathbf{Y}_{\mathbf{X}}^j)\|_{\mathbb{R}^L}\Big]$,
4. $\mathbf{E_4}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j) := \mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}} \|\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^j\|_{\mathbb{R}^L} \|\mathbf{b}(\mathbf{Y}_{\mathbf{X}}^{j+1}) - \mathbf{b}(\mathbf{Y}_{\mathbf{X}}^j)\|_{\mathbb{R}^L}\Big]$,
5. $\mathbf{E_5}(\mathbf{Y}_{\mathbf{X}}^j) := \mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}} \|D\mathbf{b}(\mathbf{Y}_{\mathbf{X}}^j) \cdot \mathbf{b}(\mathbf{Y}_{\mathbf{X}}^j)\|_{\mathbb{R}^L}\Big]$,
6. $\mathbf{E_6}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j) := \mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}} \|\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^j\|_{\mathbb{R}^L}^2\Big]$,

7. $\mathbf{E_7}(\mathbf{Y}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\|\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\|_{\mathbb{R}^L}\|\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\|^2_{\mathbb{R}^{L\times L}}\Big]$,

8. $\mathbf{E_8}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\|\mathbf{Y}^{j+1}_{\mathbf{X}}-\mathbf{Y}^j_{\mathbf{X}}\|^2_{\mathbb{R}^L}\|\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\|^2_{\mathbb{R}^{L\times L}}\Big]$,

9. $\mathbf{E_9}(\mathbf{Y}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\|\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\|^2_{\mathbb{R}^{L\times L}}\Big]$,

10. $$\mathbf{E_{10}}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\|\mathbf{Y}^{j+1}_{\mathbf{X}}-\mathbf{Y}^j_{\mathbf{X}}\|_{\mathbb{R}^L} \cdot \|\boldsymbol{\sigma}(\mathbf{Y}^{j+1}_{\mathbf{X}})\boldsymbol{\sigma}^\top(\mathbf{Y}^{j+1}_{\mathbf{X}})-\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\sigma}^\top(\mathbf{Y}^j_{\mathbf{X}})\|_{\mathbb{R}^{L\times L}}\Big],$$

11. $\mathbf{E_{11}}(\mathbf{Y}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\|D\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\sigma}^\top(\mathbf{Y}^j_{\mathbf{X}})\cdot\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\|_{\mathbb{R}^{L\times L}}\Big]$,

12. $\mathbf{E_{12}}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\|\mathbf{Y}^{j+1}_{\mathbf{X}}-\mathbf{Y}^j_{\mathbf{X}}\|_{\mathbb{R}^L}\|\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\|_{\mathbb{R}^L}\|\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\|_{\mathbb{R}^{L\times L}}\Big]$,

13. $\mathbf{E_{13}}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}},t_j) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\big(g(t_j,\mathbf{Y}^{j+1}_{\mathbf{X}})-g(t_j,\mathbf{Y}^j_{\mathbf{X}})\big)\Big]$,

14. $\mathbf{E_{14}}(\mathbf{Y}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\Big]$,

15. $\mathbf{E_{15}}(\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\mathbf{1}_{\left\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\left(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\right)\right\}}\|\mathbf{b}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\|_{\mathbb{R}^L}\Big]$,

16. $\mathbf{E_{16}}(\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\mathbf{1}_{\left\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\left(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\right)\right\}}\|\mathbf{b}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\|^2_{\mathbb{R}^L}\Big]$,

17. $$\mathbf{E_{17}}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\mathbf{1}_{\left\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\left(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\right)\right\}}\|\mathbf{Y}^{j+1}_{\mathbf{X}}-\overline{\mathbf{Y}}^j_{\mathbf{X}}\|^2_{\mathbb{R}^L} \cdot \|\mathbf{b}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\|_{\mathbb{R}^L}\Big],$$

18. $$\mathbf{E_{18}}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\mathbf{1}_{\left\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\left(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\right)\right\}}\|\mathbf{Y}^{j+1}_{\mathbf{X}}-\overline{\mathbf{Y}}^j_{\mathbf{X}}\|_{\mathbb{R}^L}\|\mathbf{b}(\mathbf{Y}^{j+1}_{\mathbf{X}})-\mathbf{b}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\|_{\mathbb{R}^L}\Big],$$

19. $\mathbf{E_{19}}(\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\mathbf{1}_{\left\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\left(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\right)\right\}}\|D\mathbf{b}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\cdot\mathbf{b}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\|_{\mathbb{R}^L}\Big]$,

20. $\mathbf{E_{20}}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\mathbf{1}_{\left\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\left(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\right)\right\}}\|\mathbf{Y}^{j+1}_{\mathbf{X}}-\overline{\mathbf{Y}}^j_{\mathbf{X}}\|^2_{\mathbb{R}^L}\Big]$,

21. $\mathbf{E_{21}}(\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\mathbf{1}_{\left\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\left(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\right)\right\}}\|\mathbf{b}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\|_{\mathbb{R}^L}\|\boldsymbol{\sigma}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\|^2_{\mathbb{R}^{L\times L}}\Big]$,

22. $$\mathbf{E_{22}}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\mathbf{1}_{\left\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\left(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\right)\right\}}\|\mathbf{Y}^{j+1}_{\mathbf{X}}-\overline{\mathbf{Y}}^j_{\mathbf{X}}\|^2_{\mathbb{R}^L} \cdot \|\boldsymbol{\sigma}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\|^2_{\mathbb{R}^{L\times L}}\Big],$$

23. $\mathbf{E_{23}}(\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\mathbf{1}_{\left\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\left(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\right)\right\}}\|\boldsymbol{\sigma}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\|^2_{\mathbb{R}^{L\times L}}\Big]$,

24. $$\mathbf{E_{24}}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\mathbf{1}_{\left\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\left(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\right)\right\}}\|\mathbf{Y}^{j+1}_{\mathbf{X}}-\overline{\mathbf{Y}}^j_{\mathbf{X}}\|_{\mathbb{R}^L} \cdot \|\boldsymbol{\sigma}(\mathbf{Y}^{j+1}_{\mathbf{X}})\boldsymbol{\sigma}^\top(\mathbf{Y}^{j+1}_{\mathbf{X}})-\boldsymbol{\sigma}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\boldsymbol{\sigma}^\top(\overline{\mathbf{Y}}^j_{\mathbf{X}})\|_{\mathbb{R}^{L\times L}}\Big],$$

25. $$\mathbf{E_{25}}(\mathbf{Y}^j_{\mathbf{X}},\overline{\mathbf{Y}}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}}\mathbf{1}_{\left\{\overline{\mathbf{Y}}^j_{\mathbf{X}}=\mathbf{Y}^j_{\mathbf{X}}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\left(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\right)\right\}} \cdot \|D\boldsymbol{\sigma}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\boldsymbol{\sigma}^\top(\overline{\mathbf{Y}}^j_{\mathbf{X}})\cdot\mathbf{b}(\overline{\mathbf{Y}}^j_{\mathbf{X}})\|_{\mathbb{R}^{L\times L}}\Big],$$

26. $\mathbf{E_{26}}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^{j}, \overline{\mathbf{Y}}_{\mathbf{X}}^{j}) := \mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}_{\mathbf{X}}^{j}=\mathbf{Y}_{\mathbf{X}}^{j}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^{j}))\}}\|\mathbf{Y}_{\mathbf{X}}^{j+1} - \overline{\mathbf{Y}}_{\mathbf{X}}^{j}\|_{\mathbb{R}^L} \cdot \|\mathbf{b}(\overline{\mathbf{Y}}_{\mathbf{X}}^{j})\|_{\mathbb{R}^L}\|\boldsymbol{\sigma}(\overline{\mathbf{Y}}_{\mathbf{X}}^{j})\|_{\mathbb{R}^{L\times L}}\Big]$,
27. $\mathbf{E_{27}}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^{j}, \overline{\mathbf{Y}}_{\mathbf{X}}^{j}, t_j) := \mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}_{\mathbf{X}}^{j}=\mathbf{Y}_{\mathbf{X}}^{j}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^{j}))\}} \cdot \big(g(t_j, \mathbf{Y}_{\mathbf{X}}^{j+1}) - g(t_j, \overline{\mathbf{Y}}_{\mathbf{X}}^{j})\big)\Big]$,
28. $\mathbf{E_{28}}(\mathbf{Y}_{\mathbf{X}}^{j}, \overline{\mathbf{Y}}_{\mathbf{X}}^{j}) := \mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}_{\mathbf{X}}^{j}=\mathbf{Y}_{\mathbf{X}}^{j}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^{j}))\}}\Big]$,
29. $\mathbf{E_{29}}(\mathbf{Y}_{\mathbf{X}}^{j}) = \mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\lambda_j^2\Big]$.

Remark 9.2. **1.** For $\mathbf{b}(\mathbf{x}) \equiv \mathbf{0}$, $g(t,\mathbf{x}) \equiv 0$ and $\boldsymbol{\sigma}(\mathbf{x}) \equiv \sqrt{2a}\cdot\mathbb{I}$, where $a > 0$, which are data requirements in (6.1) for the *(homogeneous) heat equation*, the particular error estimators $\{\mathfrak{H}_\ell\}_{\ell=1}^3$ simplify considerably: only $\mathbf{E_6}(\cdot), \mathbf{E_8}(\cdot), \mathbf{E_{14}}(\cdot), \mathbf{E_{20}}(\cdot), \mathbf{E_{22}}(\cdot), \mathbf{E_{28}}(\cdot)$ and $\mathbf{E_{29}}(\cdot)$ constitute $\{\mathfrak{H}_\ell^{(\cdot)}\}_{\ell=1}^3$; *cf.* Example 6.3.
2. An *a posteriori* error estimate for a more general PDE (6.1) with $c(t,\mathbf{x}) \neq 0$ (*cf.* also(0.6)) is possible, but its corresponding derivation would be more complicated.

The proof of Theorem 9.7 follows the guideline of the proof of Theorem 9.1 and consists of several steps: Lemma 9.8 represents the error on the left-hand side of (9.28) with the help of the (unknown) solution u of (6.1). Lemmas 9.9, 9.10 and 9.11 estimate the expressions '$\tilde{\boldsymbol{I}}_{\boldsymbol{j}}$', '$\boldsymbol{I}\tilde{\boldsymbol{I}}_{\boldsymbol{j}}$' and '$\boldsymbol{II}\tilde{\boldsymbol{I}}_{\boldsymbol{j}}$' emerging from Lemma 9.8 and given in (9.29), (9.30) and (9.31), respectively. The derivation of the *a posteriori* error estimate (9.28) then follows by combining these lemmas.

Lemma 9.8. Assume **(B1)** $-$ **(B3)**. Fix $(t,\mathbf{x}) \in [0,T) \times \boldsymbol{\mathcal{D}}$. Let $J \equiv J(t,\mathbf{x}) \in \mathbb{N}$ and $\{t_j\}_{j\geq J} \subset [t,T]$ be a mesh with local step sizes $\{\tau^{j+1}\}_{j=0}^{J-1}$. Let $\{(\mathbf{Y}_{\mathbf{X}}^{j}, Y_Z^j)\}_{j=0}^{J}$ solve Scheme 3. Then we have

$$\Big|u(t,\mathbf{x}) - \mathbb{E}\Big[\phi(t_{J^*}, \mathbf{Y}_{\mathbf{X}}^{J^*}) + Y_Z^{J^*}\Big]\Big| \leq \sum_{j=0}^{J-1}\Big\{\tilde{\boldsymbol{I}}_{\boldsymbol{j}} + \boldsymbol{I}\tilde{\boldsymbol{I}}_{\boldsymbol{j}} + \boldsymbol{II}\tilde{\boldsymbol{I}}_{\boldsymbol{j}}\Big\},$$

where

$$\tilde{\boldsymbol{I}}_{\boldsymbol{j}} := \Big|\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\Big\{u(t_{j+1}, \mathbf{Y}_{\mathbf{X}}^{j+1}) - u(t_j, \mathbf{Y}_{\mathbf{X}}^{j}) + Y_Z^{j+1} - Y_Z^j\Big\}\Big]\Big|, \tag{9.29}$$

$$\boldsymbol{I}\tilde{\boldsymbol{I}}_{\boldsymbol{j}} := \Big|\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\mathbf{1}_{\{\overline{\mathbf{Y}}_{\mathbf{X}}^{j}=\mathbf{Y}_{\mathbf{X}}^{j}+\lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^{j}))\}} \cdot \Big\{u(t_{j+1}, \mathbf{Y}_{\mathbf{X}}^{j+1}) - u(t_j, \overline{\mathbf{Y}}_{\mathbf{X}}^{j}) + Y_Z^{j+1} - Y_Z^j\Big\}\Big]\Big|, \tag{9.30}$$

$$\boldsymbol{II}\tilde{\boldsymbol{I}}_{\boldsymbol{j}} := \Big|\mathbb{E}\Big[\mathbf{1}_{\{\mathbf{Y}_{\mathbf{X}}^{j}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\}}\Big\{u(t_j, \overline{\mathbf{Y}}_{\mathbf{X}}^{j}) - u(t_j, \mathbf{Y}_{\mathbf{X}}^{j})\Big\}\Big]\Big|. \tag{9.31}$$

Proof. Considering PDE (6.1) and observing that $Y_Z^0 = 0$, a first calculation leads to

$$\begin{aligned}\Big|u(t,\mathbf{x}) - \mathbb{E}\Big[\phi(t_{J^*}, \mathbf{Y}_{\mathbf{X}}^{J^*}) + Y_Z^{J^*}\Big]\Big| &= \Big|u(t,\mathbf{x}) + Y_Z^0 - \mathbb{E}\Big[u(t_{J^*}, \mathbf{Y}_{\mathbf{X}}^{J^*}) + Y_Z^{J^*}\Big]\Big| \\ &= \Bigg|\mathbb{E}\Bigg[\sum_{j=0}^{J^*-1} u(t_{j+1}, \mathbf{Y}_{\mathbf{X}}^{j+1}) + Y_Z^{j+1} - u(t_j, \mathbf{Y}_{\mathbf{X}}^{j}) - Y_Z^{j}\Bigg]\Bigg| \\ &= \Bigg|\sum_{j=0}^{J-1} \mathbb{E}\Big[\underbrace{u(t_j, \overline{\mathbf{Y}}_{\mathbf{X}}^{j}) + Y_Z^{j} - u(t_j, \mathbf{Y}_{\mathbf{X}}^{j}) - Y_Z^{j}}_{=:\tilde{\boldsymbol{d}}_j} + \underbrace{u(t_{j+1}, \mathbf{Y}_{\mathbf{X}}^{j+1}) + Y_Z^{j+1} - u(t_j, \overline{\mathbf{Y}}_{\mathbf{X}}^{j}) - Y_Z^{j}}_{=:\tilde{\boldsymbol{d}}_j'}\Big]\Bigg|.\end{aligned} \tag{9.32}$$

Since $\tilde{\boldsymbol{d}}_j \equiv 0$ on the event $\Big\{\mathbf{Y}_{\mathbf{X}}^{j} \notin \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\Big\}$, $\tilde{\boldsymbol{d}}_j' \equiv 0$ on the event $\Big\{\mathbf{Y}_{\mathbf{X}}^{j} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\Big\} \cap \Big\{\overline{\mathbf{Y}}_{\mathbf{X}}^{j} = \Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^{j})\Big\}$, $\tilde{\boldsymbol{d}}_j' \equiv 0$ on the event $\Big\{\mathbf{Y}_{\mathbf{X}}^{j} \notin \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\Big\} \cap \Big\{\mathbf{Y}_{\mathbf{X}}^{j} \notin \boldsymbol{\mathcal{D}}\Big\}$ and $\Big\{\mathbf{Y}_{\mathbf{X}}^{j} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\Big\} \cap \Big\{\mathbf{Y}_{\mathbf{X}}^{j} \notin \boldsymbol{\mathcal{D}}\Big\} = \emptyset$, the assertion follows from (9.32).

□

Lemma 9.9. Assume **(B1)** − **(B3)**. Fix $(t,\mathbf{x}) \in [0,T) \times \boldsymbol{\mathcal{D}}$. Let $J \equiv J(t,\mathbf{x}) \in \mathbb{N}$ and $\{t_j\}_{j\geq J} \subset [t,T]$ be a mesh with local step sizes $\{\tau^{j+1}\}_{j=0}^{J-1}$. Let $\{(\mathbf{Y}_{\mathbf{X}}^{j}, Y_Z^{j})\}_{j=0}^{J}$ solve Scheme 3. Then, for every $j \geq 0$, we have

$$\tilde{\boldsymbol{I}}_{\boldsymbol{j}} \leq \mathfrak{C}(\phi, g) \cdot \mathfrak{H}_1^{(j)} \cdot \tau^{j+1},$$

where $\tilde{\boldsymbol{I}}_{\boldsymbol{j}}$ is given in (9.29), and $\mathfrak{C}(\phi, g) > 0$ is from Lemma 8.2.

Proof. In the following, we write $A_j := \Big\{\mathbf{Y}_{\mathbf{X}}^{j} \in \boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\Big\}$ to simplify the notation. In a first step, we rewrite $\tilde{\boldsymbol{I}}_{\boldsymbol{j}}$.

$$\tilde{\boldsymbol{I}}_{\boldsymbol{j}} = \Big|\mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \Big\{u(t_{j+1}, \mathbf{Y}_{\mathbf{X}}^{j+1}) - u(t_j, \mathbf{Y}_{\mathbf{X}}^{j+1}) + u(t_j, \mathbf{Y}_{\mathbf{X}}^{j+1}) - u(t_j, \mathbf{Y}_{\mathbf{X}}^{j}) + g(t_j, \mathbf{Y}_{\mathbf{X}}^{j}) \cdot \tau^{j+1}\Big\}\Big]\Big|. \tag{9.33}$$

Step 1: (Employing PDE (6.1)) We use the *mean value theorem* and *Taylor's formula* to deduce from (9.33)

$$\begin{aligned}\tilde{\boldsymbol{I}}_{\boldsymbol{j}} = \Big|\mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \Big\{&\partial_t u(\hat{t}, \mathbf{Y}_{\mathbf{X}}^{j+1}) \cdot \tau^{j+1} + \Big\langle D_{\mathbf{x}} u(t_j, \mathbf{Y}_{\mathbf{X}}^{j}), \mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^{j}\Big\rangle_{\mathbb{R}^L} \\ &+ \frac{1}{2}\mathrm{Tr}\Big(D_{\mathbf{x}}^2 u(t_j, \hat{\mathbf{Y}}_{\mathbf{X}}^{j})(\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^{j})(\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^{j})^\top\Big) + g(t_j, \mathbf{Y}_{\mathbf{X}}^{j}) \cdot \tau^{j+1}\Big\}\Big]\Big|,\end{aligned} \tag{9.34}$$

for some $\hat{t}$ between t_j and t_{j+1}, and for some $\hat{\mathbf{Y}}_{\mathbf{X}}^{j}$ between $\mathbf{Y}_{\mathbf{X}}^{j+1}$ and $\mathbf{Y}_{\mathbf{X}}^{j}$, *i.e.*, $\hat{\mathbf{Y}}_{\mathbf{X}}^{j} = \mathbf{Y}_{\mathbf{X}}^{j} + \theta\Big(\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^{j}\Big)$ with $\theta \in (0,1)$. Now, we use the identity in (6.1) and represent the increment $\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^{j}$ in (9.34) to get

$$\begin{aligned}\tilde{\boldsymbol{I}}_{\boldsymbol{j}} = \Big|\mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \Big\{\Big[&-\big\langle \mathbf{b}(\mathbf{Y}_{\mathbf{X}}^{j+1}), D_{\mathbf{x}} u(\hat{t}, \mathbf{Y}_{\mathbf{X}}^{j+1})\big\rangle_{\mathbb{R}^L} - \frac{1}{2}\mathrm{Tr}\big(\boldsymbol{\sigma}(\mathbf{Y}_{\mathbf{X}}^{j+1})\boldsymbol{\sigma}^\top(\mathbf{Y}_{\mathbf{X}}^{j+1}) D_{\mathbf{x}}^2 u(\hat{t}, \mathbf{Y}_{\mathbf{X}}^{j+1})\big) \\ &- g(\hat{t}, \mathbf{Y}_{\mathbf{X}}^{j+1})\Big] \cdot \tau^{j+1} + \big\langle D_{\mathbf{x}} u(t_j, \mathbf{Y}_{\mathbf{X}}^{j}), \mathbf{b}(\mathbf{Y}_{\mathbf{X}}^{j})\big\rangle_{\mathbb{R}^L} \cdot \tau^{j+1} + \big\langle D_{\mathbf{x}} u(t_j, \mathbf{Y}_{\mathbf{X}}^{j}), \boldsymbol{\sigma}(\mathbf{Y}_{\mathbf{X}}^{j})\boldsymbol{\xi}_{j+1}\big\rangle_{\mathbb{R}^L} \cdot \sqrt{\tau^{j+1}}\end{aligned}$$

$$+\frac{1}{2}\mathrm{Tr}\Big(D^2_{\mathbf{x}}u(t_j,\hat{\mathbf{Y}}^j_{\mathbf{X}})\big(\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\tau^{j+1}+\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\sqrt{\tau^{j+1}}\big)\big(\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\tau^{j+1}+\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\sqrt{\tau^{j+1}}\big)^\top\Big)$$
$$+g(t_j,\mathbf{Y}^j_{\mathbf{X}})\cdot\tau^{j+1}\Big\}\Big]\Big|.$$

Standard calculations then lead to

$$\tilde{\boldsymbol{I}}_{\boldsymbol{j}}\leq\boldsymbol{K_1}+\boldsymbol{K_2}+\boldsymbol{K_3}+\boldsymbol{K_4}+\boldsymbol{K_5}+\boldsymbol{K_6}\,, \tag{9.35}$$

where

$$\boldsymbol{K_1}:=\Big|\mathbb{E}\Big[\mathbf{1}_{A_j}\cdot\Big\{\Big\langle D_{\mathbf{x}}u(t_j,\mathbf{Y}^j_{\mathbf{X}}),\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\Big\rangle_{\mathbb{R}^L}-\Big\langle D_{\mathbf{x}}u(\hat{t},\mathbf{Y}^{j+1}_{\mathbf{X}}),\mathbf{b}(\mathbf{Y}^{j+1}_{\mathbf{X}})\Big\rangle_{\mathbb{R}^L}\Big\}\Big]\Big|\cdot\tau^{j+1}\,,$$

$$\boldsymbol{K_2}:=\frac{1}{2}\Big|\mathbb{E}\Big[\mathbf{1}_{A_j}\cdot\Big\{\mathrm{Tr}\Big(D^2_{\mathbf{x}}u(t_j,\hat{\mathbf{Y}}^j_{\mathbf{X}})\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\Big(\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\Big)^\top\Big)$$
$$-\mathrm{Tr}\Big(D^2_{\mathbf{x}}u(\hat{t},\hat{\mathbf{Y}}^{j+1}_{\mathbf{X}})\boldsymbol{\sigma}(\mathbf{Y}^{j+1}_{\mathbf{X}})\boldsymbol{\sigma}^\top(\mathbf{Y}^{j+1}_{\mathbf{X}})\Big)\Big\}\Big]\Big|\cdot\tau^{j+1}\,,$$

$$\boldsymbol{K_3}:=\frac{1}{2}\Big|\mathbb{E}\Big[\mathbf{1}_{A_j}\cdot\mathrm{Tr}\Big(D^2_{\mathbf{x}}u(t_j,\hat{\mathbf{Y}}^j_{\mathbf{X}})\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\Big(\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\Big)^\top\Big)\Big]\Big|\cdot\Big(\tau^{j+1}\Big)^{\frac{3}{2}}\,,$$

$$\boldsymbol{K_4}:=\frac{1}{2}\Big|\mathbb{E}\Big[\mathbf{1}_{A_j}\cdot\mathrm{Tr}\Big(D^2_{\mathbf{x}}u(t_j,\hat{\mathbf{Y}}^j_{\mathbf{X}})\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})^\top\Big)\Big]\Big|\cdot\Big(\tau^{j+1}\Big)^{\frac{3}{2}}\,,$$

$$\boldsymbol{K_5}:=\frac{1}{2}\Big|\mathbb{E}\Big[\mathbf{1}_{A_j}\cdot\mathrm{Tr}\Big(D^2_{\mathbf{x}}u(t_j,\hat{\mathbf{Y}}^j_{\mathbf{X}})\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})^\top\Big)\Big]\Big|\cdot\Big(\tau^{j+1}\Big)^{2}\,,$$

$$\boldsymbol{K_6}:=\frac{1}{2}\Big|\mathbb{E}\Big[\mathbf{1}_{A_j}\cdot\Big\{g(\hat{t},\mathbf{Y}^{j+1}_{\mathbf{X}})-g(t_j,\mathbf{Y}^j_{\mathbf{X}})\Big\}\Big]\Big|\cdot\tau^{j+1}\,.$$

Step 2: (Estimation of $\boldsymbol{K_1},\boldsymbol{K_2},\boldsymbol{K_3},\boldsymbol{K_4},\boldsymbol{K_5},\boldsymbol{K_6}$) We estimate the terms in (9.35) independently.

a) (Estimation of $\boldsymbol{K_1}$) In a first step, we rewrite $\boldsymbol{K_1}$.

$$\boldsymbol{K_1}\leq\boldsymbol{K_{1,1}}+\boldsymbol{K_{1,2}}\,, \tag{9.36}$$

where

$$\boldsymbol{K_{1,1}}:=\Big|\mathbb{E}\Big[\mathbf{1}_{A_j}\cdot\Big\langle D_{\mathbf{x}}u(t_j,\mathbf{Y}^j_{\mathbf{X}})-D_{\mathbf{x}}u(\hat{t},\mathbf{Y}^{j+1}_{\mathbf{X}}),\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\Big\rangle_{\mathbb{R}^L}\Big]\Big|\cdot\tau^{j+1}\,,$$
$$\boldsymbol{K_{1,2}}:=\Big|\mathbb{E}\Big[\mathbf{1}_{A_j}\cdot\Big\langle D_{\mathbf{x}}u(\hat{t},\mathbf{Y}^{j+1}_{\mathbf{X}}),\mathbf{b}(\mathbf{Y}^{j+1}_{\mathbf{X}})-\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\Big\rangle_{\mathbb{R}^L}\Big]\Big|\cdot\tau^{j+1}\,.$$

In the following, we investigate $\boldsymbol{K_{1,1}}$. We use the *mean value theorem* and *Taylor's formula* to get

$$\boldsymbol{K_{1,1}}=\Big|\sum_{i=1}^{L}\mathbb{E}\Big[\mathbf{1}_{A_j}\cdot b_i(\mathbf{Y}^j_{\mathbf{X}})\{\partial_{x_i}u(\hat{t},\mathbf{Y}^{j+1}_{\mathbf{X}})-\partial_{x_i}u(t_j,\mathbf{Y}^{j+1}_{\mathbf{X}})+\partial_{x_i}u(t_j,\mathbf{Y}^{j+1}_{\mathbf{X}})-\partial_{x_i}u(t_j,\mathbf{Y}^j_{\mathbf{X}})\}\Big]\Big|\cdot\tau^{j+1}$$

$$= \Big| \sum_{i=1}^{L} \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot b_i(\mathbf{Y}_{\mathbf{X}}^j) \cdot \partial_t \partial_{x_i} u(\hat{\hat{t}}, \mathbf{Y}_{\mathbf{X}}^{j+1}) \Big] \cdot (\hat{t} - t_j) \cdot \tau^{j+1}$$
$$+ \sum_{i=1}^{L} \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot b_i(\mathbf{Y}_{\mathbf{X}}^j) \cdot \langle D_{\mathbf{x}} \partial_{x_i} u(t_j, \mathbf{Y}_{\mathbf{X}}^j), \mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^j \rangle_{\mathbb{R}^L} \Big] \cdot (\hat{t} - t_j) \cdot \tau^{j+1} \Big] \cdot \tau^{j+1}$$
$$+ \frac{1}{2} \sum_{i=1}^{L} \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot b_i(\mathbf{Y}_{\mathbf{X}}^j) \cdot \mathrm{Tr}\Big(D_{\mathbf{x}}^2 \partial_{x_i} u(t_j, \hat{\mathbf{Y}}_{\mathbf{X}}^j) \{\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^j\} \{\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^j\}^\top \Big) \Big] \cdot \tau^{j+1} \Big|,$$

for some $\hat{\hat{t}}$ between t_j and $\hat{t}$, and for some $\hat{\mathbf{Y}}_{\mathbf{X}}^j$ between $\mathbf{Y}_{\mathbf{X}}^{j+1}$ and $\mathbf{Y}_{\mathbf{X}}^j$. Standard calculations further lead to

$$\boldsymbol{K}_{\mathbf{1,1}} \le \Big| \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \Big\langle D_{\mathbf{x}} \partial_t u(\hat{\hat{t}}, \mathbf{Y}_{\mathbf{X}}^{j+1}), \mathbf{b}(\mathbf{Y}_{\mathbf{X}}^j) \Big\rangle_{\mathbb{R}^L} \Big] \Big| \cdot \Big(\tau^{j+1} \Big)^2$$
$$+ \Big| \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \Big\langle D_{\mathbf{x}}^2 u(t_j, \mathbf{Y}_{\mathbf{X}}^j) \{\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^j\}, \mathbf{b}(\mathbf{Y}_{\mathbf{X}}^j) \Big\rangle_{\mathbb{R}^L} \Big] \Big| \cdot \tau^{j+1}$$
$$+ \frac{1}{2} \Big| \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \Big\langle D_{\mathbf{x}}^3 u(t_j, \hat{\mathbf{Y}}_{\mathbf{X}}^j) \{\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^j\} \{\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^j\}, \mathbf{b}(\mathbf{Y}_{\mathbf{X}}^j) \Big\rangle_{\mathbb{R}^L} \Big] \Big| \cdot \tau^{j+1}.$$

By Lemma 8.2, we then conclude

$$\boldsymbol{K}_{\mathbf{1,1}} \le \mathfrak{C}(\phi, g) \cdot \mathbf{E_1}(\mathbf{Y}_{\mathbf{X}}^j) \cdot \Big(\tau^{j+1} \Big)^2 + \mathfrak{C}(\phi, g) \cdot \mathbf{E_2}(\mathbf{Y}_{\mathbf{X}}^j) \cdot \Big(\tau^{j+1} \Big)^2 + \frac{1}{2} \mathfrak{C}(\phi, g) \cdot \mathbf{E_3}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j) \cdot \tau^{j+1}. \tag{9.37}$$

We proceed with the investigation of $\boldsymbol{K}_{\mathbf{1,2}}$. A first calculation yields

$$\boldsymbol{K}_{\mathbf{1,2}} \le \Big| \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \Big\langle D_{\mathbf{x}} u(\hat{t}, \mathbf{Y}_{\mathbf{X}}^{j+1}) - D_{\mathbf{x}} u(\hat{t}, \mathbf{Y}_{\mathbf{X}}^j), \mathbf{b}(\mathbf{Y}_{\mathbf{X}}^{j+1}) - \mathbf{b}(\mathbf{Y}_{\mathbf{X}}^j) \Big\rangle_{\mathbb{R}^L} \Big] \Big| \cdot \tau^{j+1}$$
$$+ \Big| \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \Big\langle D_{\mathbf{x}} u(\hat{t}, \mathbf{Y}_{\mathbf{X}}^j), \mathbf{b}(\mathbf{Y}_{\mathbf{X}}^{j+1}) - \mathbf{b}(\mathbf{Y}_{\mathbf{X}}^j) \Big\rangle_{\mathbb{R}^L} \Big] \Big| \cdot \tau^{j+1}.$$

We use Lemma 8.2 and *Taylor's formula* to get

$$\boldsymbol{K}_{\mathbf{1,2}} \le \mathfrak{C}(\phi, g) \cdot \mathbf{E_4}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j) \cdot \tau^{j+1} + \Big| \sum_{i=1}^{L} \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \partial_{x_i} u(\hat{t}, \mathbf{Y}_{\mathbf{X}}^j) \{ b_i(\mathbf{Y}_{\mathbf{X}}^{j+1}) - b_i(\mathbf{Y}_{\mathbf{X}}^j) \} \Big] \Big| \cdot \tau^{j+1}$$
$$\le \mathfrak{C}(\phi, g) \cdot \mathbf{E_4}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j) \cdot \tau^{j+1} + \Big| \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \langle D_{\mathbf{x}} \mathbf{b}(\mathbf{Y}_{\mathbf{X}}^j) \cdot \mathbf{b}(\mathbf{Y}_{\mathbf{X}}^j), D_{\mathbf{x}} u(\hat{t}, \mathbf{Y}_{\mathbf{X}}^j) \rangle_{\mathbb{R}^L} \Big] \Big| \cdot (\tau^{j+1})^2$$
$$+ \frac{1}{2} \Big| \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \langle D_{\mathbf{x}}^2 \mathbf{b}(\hat{\mathbf{Y}}_{\mathbf{X}}^j) \cdot \{\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^j\} \{\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^j\}, D_{\mathbf{x}} u(\hat{t}, \mathbf{Y}_{\mathbf{X}}^j) \rangle_{\mathbb{R}^L} \Big] \Big| \cdot \tau^{j+1},$$

for some $\hat{\mathbf{Y}}_{\mathbf{X}}^j$ between $\mathbf{Y}_{\mathbf{X}}^{j+1}$ and $\mathbf{Y}_{\mathbf{X}}^j$, and where we estimate $\|D_{\mathbf{x}} u(\hat{t}, \mathbf{Y}_{\mathbf{X}}^{j+1}) - D_{\mathbf{x}} u(\hat{t}, \mathbf{Y}_{\mathbf{X}}^j)\|_{\mathbb{R}^L} \le \mathfrak{C}(\phi, g) \cdot \|\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^j\|_{\mathbb{R}^L}$. Using Lemma 8.2 again, we conclude

$$\boldsymbol{K}_{\mathbf{1,2}} \le \mathfrak{C}(\phi, g) \cdot \mathbf{E_4}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j) \cdot \tau^{j+1} + \mathfrak{C}(\phi, g) \cdot \mathbf{E_5}(\mathbf{Y}_{\mathbf{X}}^j) \cdot \Big(\tau^{j+1} \Big)^2$$
$$+ \frac{1}{2} \mathfrak{C}(\phi, g) C_{D^2 \mathbf{b}} \cdot \mathbf{E_6}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j) \cdot \tau^{j+1}. \tag{9.38}$$

Consequently, plugging (9.37) and (9.38) into (9.36) yields

$$\boldsymbol{K}_{\mathbf{1}} \le \mathfrak{C}(\phi, g) \cdot \Big\{ \mathbf{E_1}(\mathbf{Y}_{\mathbf{X}}^j) \cdot \tau^{j+1} + \mathbf{E_2}(\mathbf{Y}_{\mathbf{X}}^j) \cdot \tau^{j+1} + \frac{1}{2} \cdot \mathbf{E_3}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j) + \mathbf{E_4}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^j)$$

$$+\,\mathbf{E_5}(\mathbf{Y}^j_{\mathbf{X}})\cdot\tau^{j+1}+\frac{1}{2}C_{D^2\mathbf{b}}\cdot\mathbf{E_6}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}})\Big\}\cdot\tau^{j+1}\,. \tag{9.39}$$

b) (Estimation of $\boldsymbol{K_2}$) By the independence of $\xi^{(i)}_{j+1}$ and $\xi^{(k)}_{j+1}$ (note that $\boldsymbol{\xi}_{j+1}=(\xi^{(1)}_{j+1},...,\xi^{(L)}_{j+1})^\top$), we estimate

$$\begin{aligned}\boldsymbol{K_2}&=\frac{1}{2}\Big|\mathbb{E}\Big[\mathbf{1}_{A_j}\cdot\Big\{\Big\langle\Big\{D^2_{\mathbf{x}}u(t_j,\hat{\mathbf{Y}}^j_{\mathbf{X}})-D^2_{\mathbf{x}}u(t_j,\mathbf{Y}^j_{\mathbf{X}})\Big\}\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1},\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\Big\rangle_{\mathbb{R}^L}\\&\qquad+\Big\langle\Big\{D^2_{\mathbf{x}}u(t_j,\mathbf{Y}^j_{\mathbf{X}})\Big\}\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1},\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\Big\rangle_{\mathbb{R}^L}\\&\qquad-\mathrm{Tr}\Big(D^2_{\mathbf{x}}u(\hat{t},\mathbf{Y}^{j+1}_{\mathbf{X}})\Big\}\boldsymbol{\sigma}(\mathbf{Y}^{j+1}_{\mathbf{X}})\boldsymbol{\sigma}^\top(\mathbf{Y}^{j+1}_{\mathbf{X}})\Big)\Big\}\Big]\Big|\cdot\tau^{j+1}\\&\le\boldsymbol{K_{2,1}}+\boldsymbol{K_{2,1}}\,,\end{aligned} \tag{9.40}$$

where

$$\boldsymbol{K_{2,1}}:=\frac{1}{2}\Big|\mathbb{E}\Big[\mathbf{1}_{A_j}\cdot\Big\langle\Big\{D^2_{\mathbf{x}}u(t_j,\hat{\mathbf{Y}}^j_{\mathbf{X}})-D^2_{\mathbf{x}}u(t_j,\mathbf{Y}^j_{\mathbf{X}})\Big\}\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1},\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\xi}_{j+1}\Big\rangle_{\mathbb{R}^L}\Big]\Big|\cdot\tau^{j+1}\,,$$

$$\boldsymbol{K_{2,2}}:=\frac{1}{2}\Big|\mathbb{E}\Big[\mathbf{1}_{A_j}\cdot\Big\{\mathrm{Tr}\big(D^2_{\mathbf{x}}u(t_j,\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\sigma}^\top(\mathbf{Y}^j_{\mathbf{X}})\big)-\mathrm{Tr}\big(D^2_{\mathbf{x}}u(\hat{t},\mathbf{Y}^{j+1}_{\mathbf{X}})\boldsymbol{\sigma}(\mathbf{Y}^{j+1}_{\mathbf{X}})\boldsymbol{\sigma}^\top(\mathbf{Y}^{j+1}_{\mathbf{X}})\big)\Big\}\Big]\Big|\cdot\tau^{j+1}\,.$$

The investigation of $\boldsymbol{K_{2,1}}$ is similar to $\boldsymbol{K_{1,2}}$, and besides standard calculations, uses *Taylor's formula*, Lemma 8.2 and the fact that $\theta<1$ (see the representation of $\hat{\mathbf{Y}}^j_{\mathbf{X}}$ in (9.34)). We get

$$\boldsymbol{K_{2,1}}\le\frac{1}{2}L\mathfrak{C}(\phi,g)\cdot\mathbf{E_7}(\mathbf{Y}^j_{\mathbf{X}})\cdot\Big(\tau^{j+1}\Big)^2+\frac{1}{4}L\mathfrak{C}(\phi,g)\cdot\mathbf{E_8}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}})\cdot\tau^{j+1}\,. \tag{9.41}$$

We proceed with the expression of $\boldsymbol{K_{2,2}}$. A first calculation leads to

$$\boldsymbol{K_{2,2}}\le\boldsymbol{K_{2,2,1}}+\boldsymbol{K_{2,2,1}}\,, \tag{9.42}$$

where

$$\boldsymbol{K_{2,2,1}}:=\frac{1}{2}\Big|\mathbb{E}\Big[\mathbf{1}_{A_j}\cdot\mathrm{Tr}\Big(\Big\{D^2_{\mathbf{x}}u(t_j,\mathbf{Y}^j_{\mathbf{X}})-D^2_{\mathbf{x}}u(\hat{t},\mathbf{Y}^{j+1}_{\mathbf{X}})\Big\}\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\sigma}^\top(\mathbf{Y}^j_{\mathbf{X}})\Big)\Big]\Big|\cdot\tau^{j+1}$$

$$\boldsymbol{K_{2,2,2}}:=\frac{1}{2}\Big|\mathbb{E}\Big[\mathbf{1}_{A_j}\cdot\mathrm{Tr}\Big(D^2_{\mathbf{x}}u(\hat{t},\mathbf{Y}^{j+1}_{\mathbf{X}})\Big\{\boldsymbol{\sigma}(\mathbf{Y}^{j+1}_{\mathbf{X}})\boldsymbol{\sigma}^\top(\mathbf{Y}^{j+1}_{\mathbf{X}})-\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}})\boldsymbol{\sigma}^\top(\mathbf{Y}^j_{\mathbf{X}})\Big\}\Big)\Big]\Big|\cdot\tau^{j+1}\,.$$

The investigation of $\boldsymbol{K_{2,2,1}}$ is similar to $\boldsymbol{K_{1,1}}$, and uses the *mean value theorem* and *Taylor's formula*, as well as Lemma 8.2. Hence, we get

$$\boldsymbol{K_{2,2,1}}\le\mathfrak{C}(\phi,g)\cdot\Big\{\frac{1}{2}L\cdot\mathbf{E_9}(\mathbf{Y}^j_{\mathbf{X}})\cdot\big(\tau^{j+1}\big)^2+\frac{1}{2}L\cdot\mathbf{E_7}(\mathbf{Y}^j_{\mathbf{X}})\cdot\big(\tau^{j+1}\big)^2+\frac{1}{4}L\cdot\mathbf{E_8}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}})\cdot\tau^{j+1}\Big\}\,. \tag{9.43}$$

Furthermore, the treatment of $\boldsymbol{K_{2,2,2}}$ is very similar to $\boldsymbol{K_{1,2}}$, and we get

$$\boldsymbol{K_{2,2,2}}\le\mathfrak{C}(\phi,g)\cdot\Big\{\frac{1}{2}L\cdot\mathbf{E_{10}}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}})\cdot\tau^{j+1}+\frac{1}{2}L\cdot\mathbf{E_{11}}(\mathbf{Y}^j_{\mathbf{X}})\cdot\big(\tau^{j+1}\big)^2+\frac{1}{4}LC_{D^2\boldsymbol{\sigma}\boldsymbol{\sigma}^\top}\cdot\mathbf{E_6}(\mathbf{Y}^{j+1}_{\mathbf{X}},\mathbf{Y}^j_{\mathbf{X}})\cdot\tau^{j+1}\Big\}\,. \tag{9.44}$$

We combine (9.43) and (9.44) with (9.42), and plug the resulting expression, as well as (9.41) into (9.40) to finally get

$$\begin{aligned}\boldsymbol{K_2} \leq \mathfrak{C}(\phi, g) \cdot \Big\{ &\frac{1}{4} L C_{D^2 \boldsymbol{\sigma}\boldsymbol{\sigma}^\top} \cdot \mathbf{E_6}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^{j}_{\mathbf{X}}) + L \cdot \mathbf{E_7}(\mathbf{Y}^{j}_{\mathbf{X}}) \cdot \tau^{j+1} + \frac{1}{2} L \cdot \mathbf{E_8}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^{j}_{\mathbf{X}}) \\ &+ \frac{1}{2} L \cdot \mathbf{E_9}(\mathbf{Y}^{j}_{\mathbf{X}}) \cdot \tau^{j+1} + \frac{1}{2} L \cdot \mathbf{E_9}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^{j}_{\mathbf{X}}) + \frac{1}{2} L \cdot \mathbf{E_{11}}(\mathbf{Y}^{j}_{\mathbf{X}}) \cdot \tau^{j+1} \Big\} \cdot \tau^{j+1}. \qquad (9.45)\end{aligned}$$

c) (Estimation of $\boldsymbol{K_3}$) We add and substract $D^2_{\mathbf{x}} u(t_j, \mathbf{Y}^j_{\mathbf{X}})$, use independency and Lemma 8.2 to obtain

$$\begin{aligned}\boldsymbol{K_3} &= \frac{1}{2} \Big| \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \mathrm{Tr}\Big(\big\{ D^2_{\mathbf{x}} u(t_j, \hat{\mathbf{Y}}^j_{\mathbf{X}}) - D^2_{\mathbf{x}} u(t_j, \mathbf{Y}^j_{\mathbf{X}}) \big\} \mathbf{b}(\mathbf{Y}^j_{\mathbf{X}}) \big(\boldsymbol{\sigma}(\mathbf{Y}^j_{\mathbf{X}}) \boldsymbol{\xi}_{j+1} \big)^\top \Big) \Big] \Big| \cdot \big(\tau^{j+1} \big)^{\frac{3}{2}} \\ &\leq \frac{1}{2} \sqrt{L} \mathfrak{C}(\phi, g) \cdot \mathbf{E_{12}}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^{j}_{\mathbf{X}}) \cdot \big(\tau^{j+1} \big)^{\frac{3}{2}}, \qquad (9.46)\end{aligned}$$

where we estimate $\|D^2_{\mathbf{x}} u(t_j, \hat{\mathbf{Y}}^j_{\mathbf{X}}) - D^2_{\mathbf{x}} u(t_j, \mathbf{Y}^j_{\mathbf{X}})\|_{\mathbb{R}^{L\times L}} \leq \mathfrak{C}(\phi, g) \cdot \|\hat{\mathbf{Y}}^j_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}}\|_{\mathbb{R}^L}$, and $\|\hat{\mathbf{Y}}^j_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}}\|_{\mathbb{R}^L} \leq \|\mathbf{Y}^{j+1}_{\mathbf{X}} - \mathbf{Y}^j_{\mathbf{X}}\|_{\mathbb{R}^L}$ due to the representation of $\hat{\mathbf{Y}}^j_{\mathbf{X}}$ in (9.34).
d) (Estimation of $\boldsymbol{K_4}$) The treatment of $\boldsymbol{K_4}$ is similar to $\boldsymbol{K_3}$. We have

$$\boldsymbol{K_4} \leq \frac{1}{2} \sqrt{L} \mathfrak{C}(\phi, g) \cdot \mathbf{E_{12}}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^{j}_{\mathbf{X}}) \cdot \big(\tau^{j+1} \big)^{\frac{3}{2}}. \qquad (9.47)$$

e) (Estimation of $\boldsymbol{K_5}$) Lemma 8.2 and standard arguments immediately lead to

$$\boldsymbol{K_5} \leq \frac{1}{2} \mathfrak{C}(\phi, g) \cdot \mathbf{E_2}(\mathbf{Y}^{j}_{\mathbf{X}}) \cdot \big(\tau^{j+1} \big)^{2}. \qquad (9.48)$$

f) (Estimation of $\boldsymbol{K_6}$) Standard arguments and the *mean value theorem* yield

$$\begin{aligned}\boldsymbol{K_6} &\leq |\mathbf{E_{13}}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^{j}_{\mathbf{X}}, t_j)| \cdot \tau^{j+1} + \Big| \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \big\{ g(t_j, \mathbf{Y}^{j+1}_{\mathbf{X}}) - g(\hat{t}, \mathbf{Y}^{j+1}_{\mathbf{X}}) \big\} \Big] \Big| \cdot \tau^{j+1} \\ &= |\mathbf{E_{13}}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^{j}_{\mathbf{X}}, t_j)| \cdot \tau^{j+1} + \Big| \mathbb{E}\Big[\mathbf{1}_{A_j} \cdot \partial_t g(\hat{\hat{t}}, \mathbf{Y}^{j+1}_{\mathbf{X}}) \big\} \Big] \Big| \cdot (\hat{t} - t_j) \cdot \tau^{j+1} \\ &\leq |\mathbf{E_{13}}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^{j}_{\mathbf{X}}, t_j)| \cdot \tau^{j+1} + \mathfrak{C}(\phi, g) \cdot \mathbf{E_{14}}(\mathbf{Y}^{j}_{\mathbf{X}}) \cdot \big(\tau^{j+1} \big)^{2}.\end{aligned}$$

Since we can assume $\mathfrak{C}(\phi, g) \geq 1$ without restrictions, we have

$$\boldsymbol{K_6} \leq \mathfrak{C}(\phi, g) \cdot \Big\{ |\mathbf{E_{13}}(\mathbf{Y}^{j+1}_{\mathbf{X}}, \mathbf{Y}^{j}_{\mathbf{X}}, t_j)| + \mathbf{E_{14}}(\mathbf{Y}^{j}_{\mathbf{X}}) \cdot \tau^{j+1} \Big\} \cdot \tau^{j+1}. \qquad (9.49)$$

Step 3: Finally, combining (9.39), (9.45), (9.46), (9.47), (9.48) and (9.49) with (9.35) proves the assertion.

□

The following lemma estimates $\tilde{\boldsymbol{II_j}}$ from (9.30). Its proof is very similar to the proof of Lemma 9.9 and is thus omitted. In fact, by replacing A_j resp. $\mathbf{Y}^j_{\mathbf{X}}$ in the proof of Lemma 9.9 by $\overline{A}_j := \mathbf{1}_{\left\{\mathbf{Y}^j_{\mathbf{X}} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}} \mathbf{1}_{\left\{\overline{\mathbf{Y}}^j_{\mathbf{X}} = \mathbf{Y}^j_{\mathbf{X}} + \lambda_j \sqrt{\tau^{j+1}} \boldsymbol{n}\left(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\right)\right\}}$ resp. $\overline{\mathbf{Y}}^j_{\mathbf{X}}$, and observing that $\mathbf{Y}^{j+1}_{\mathbf{X}}$ is generated starting from $\overline{\mathbf{Y}}^j_{\mathbf{X}}$, yield the proof of Lemma 9.10.

Lemma 9.10. Assume **(B1)** − **(B3)**. Fix $(t, \mathbf{x}) \in [0,T) \times \boldsymbol{\mathcal{D}}$. Let $J \equiv J(t,\mathbf{x}) \in \mathbb{N}$ and $\{t_j\}_{j\geq J} \subset [t,T]$ be a mesh with local step sizes $\{\tau^{j+1}\}_{j=0}^{J-1}$. Let $\{(\mathbf{Y}_{\mathbf{X}}^j, Y_Z^j)\}_{j=0}^J$ solve Scheme 3. Then, for every $j \geq 0$, we have

$$\tilde{\boldsymbol{II}}_{\boldsymbol{j}} \leq \mathfrak{C}(\phi, g) \cdot \mathfrak{H}_2^{(j)} \cdot \tau^{j+1} \,,$$

where $\tilde{\boldsymbol{II}}_{\boldsymbol{j}}$ is given in (9.30), and $\mathfrak{C}(\phi, g) > 0$ is from Lemma 8.2.

The next lemma estimates $\tilde{\boldsymbol{III}}_{\boldsymbol{j}}$ from (9.31).

Lemma 9.11. Assume **(B1)** − **(B3)**. Fix $(t, \mathbf{x}) \in [0,T) \times \boldsymbol{\mathcal{D}}$. Let $J \equiv J(t,\mathbf{x}) \in \mathbb{N}$ and $\{t_j\}_{j\geq J} \subset [t,T]$ be a mesh with local step sizes $\{\tau^{j+1}\}_{j=0}^{J-1}$. Let $\{(\mathbf{Y}_{\mathbf{X}}^j, Y_Z^j)\}_{j=0}^J$ solve Scheme 3. Then, for every $j \geq 0$, we have

$$\tilde{\boldsymbol{III}}_{\boldsymbol{j}} \leq \mathfrak{C}(\phi, g) \cdot \mathfrak{H}_3^{(j)} \cdot \tau^{j+1} \,,$$

where $\tilde{\boldsymbol{III}}_{\boldsymbol{j}}$ is given in (9.31), and $\mathfrak{C}(\phi, g) > 0$ is from Lemma 8.2.

Proof. We take the conditional expectation *w.r.t.* $\mathbf{Y}_{\mathbf{X}}^j$ and use measureability arguments to obtain in a first calculation

$$\begin{aligned}
\tilde{\boldsymbol{III}}_{\boldsymbol{j}} = \Bigg| \mathbb{E}\Big[\mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}} \Big\{ & u(t_j, \mathbf{Y}_{\mathbf{X}}^j) - \mathbf{1}_{\left\{\overline{\mathbf{Y}}_{\mathbf{X}}^j = \Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\right\}} u\Big(t_j, \Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\Big) \\
& - \mathbf{1}_{\left\{\overline{\mathbf{Y}}_{\mathbf{X}}^j = \mathbf{Y}_{\mathbf{X}}^j + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\left(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\right)\right\}} u\Big(t_j, \mathbf{Y}_{\mathbf{X}}^j + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\Big)\Big)\Big\} \,\Big|\, \mathbf{Y}_{\mathbf{X}}^j \Big]\Big] \Bigg| \\
= \Bigg| \mathbb{E}\Big[\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}} \Big\{ & u(t_j, \mathbf{Y}_{\mathbf{X}}^j) - p_j \cdot u\Big(t_j, \Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\Big) \\
& - (1-p_j) \cdot u\Big(t_j, \mathbf{Y}_{\mathbf{X}}^j + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\Big)\Big)\Big\}\Big] \Bigg| \,,
\end{aligned}$$

where p_j is given in (8.5). We apply the *mean value theorem* twice to get

$$\begin{aligned}
\tilde{\boldsymbol{III}}_{\boldsymbol{j}} &= \Bigg| \mathbb{E}\Bigg[\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}} \Bigg\{ u(t_j, \mathbf{Y}_{\mathbf{X}}^j) - u(t_j, \mathbf{Y}_{\mathbf{X}}^j + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j))) \\
&\qquad + \frac{u(t_j, \mathbf{Y}_{\mathbf{X}}^j + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)) - u(t_j, \Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)))}{\|\mathbf{Y}_{\mathbf{X}}^j + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)) - \Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\|_{\mathbb{R}^L}} \cdot \lambda_j\sqrt{\tau^{j+1}} \Bigg\}\Bigg] \Bigg| \\
&\leq \mathbb{E}\Bigg[\mathbf{1}_{\left\{\mathbf{Y}_{\mathbf{X}}^j \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\right\}} \Bigg| - \Big\langle D_{\mathbf{x}} u(t_j, \hat{\mathbf{Y}}_{\mathbf{X}}^j), \boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)) \Big\rangle_{\mathbb{R}^L} \cdot \lambda_j\sqrt{\tau^{j+1}} \\
&\qquad + \frac{\Big\langle D_{\mathbf{x}} u(t_j, \hat{\hat{\mathbf{Y}}}_{\mathbf{X}}^j), \mathbf{Y}_{\mathbf{X}}^j + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)) - \Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j) \Big\rangle_{\mathbb{R}^L}}{\|\mathbf{Y}_{\mathbf{X}}^j + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)) - \Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^j)\|_{\mathbb{R}^L}} \cdot \lambda_j\sqrt{\tau^{j+1}} \Bigg|\Bigg] \,,
\end{aligned} \tag{9.50}$$

for some points $\hat{\mathbf{Y}}^j_{\mathbf{X}}$, $\hat{\hat{\mathbf{Y}}}^j_{\mathbf{X}}$ between $\mathbf{Y}^j_{\mathbf{X}}$ and $\mathbf{Y}^j_{\mathbf{X}} + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big)$, and between $\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})$ and $\mathbf{Y}^j_{\mathbf{X}} + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big)$, respectively.
Next, we define $\varphi\Big(\mathbf{z}, \boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big)\Big) := \Big\langle D_{\mathbf{x}}u(t_j,\mathbf{z}), \boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big)\Big\rangle_{\mathbb{R}^L}$, $\mathbf{z}\in\boldsymbol{\mathcal{D}}$. Since

$$\boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big) = \frac{\mathbf{Y}^j_{\mathbf{X}} + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big) - \Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})}{\Big\|\mathbf{Y}^j_{\mathbf{X}} + \lambda_j\sqrt{\tau^{j+1}}\boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big) - \Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big\|_{\mathbb{R}^L}},$$

we can rewrite (9.50) as follows,

$$\begin{aligned}\boldsymbol{I\tilde{I}I_j} &\leq \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\lambda_j\Big|\varphi\Big(\hat{\mathbf{Y}}^j_{\mathbf{X}}, \boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big)\Big) - \varphi\Big(\hat{\hat{\mathbf{Y}}}^j_{\mathbf{X}}, \boldsymbol{n}\Big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big)\Big)\Big|\Big]\cdot\sqrt{\tau^{j+1}} \\ &\leq \sup_{\mathbf{z}\in\boldsymbol{\mathcal{D}}}\sup_{\|\mathbf{v}\|_{\mathbb{R}^L}=1}\|D_{\mathbf{z}}\varphi(\mathbf{z},\mathbf{v})\|_{\mathbb{R}^L}\cdot\mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\lambda_j\|\hat{\mathbf{Y}}^j_{\mathbf{X}} - \hat{\hat{\mathbf{Y}}}^j_{\mathbf{X}}\|_{\mathbb{R}^L}\Big]\cdot\sqrt{\tau^{j+1}}\,. \qquad (9.51)\end{aligned}$$

Since $\|\hat{\mathbf{Y}}^j_{\mathbf{X}} - \hat{\hat{\mathbf{Y}}}^j_{\mathbf{X}}\|_{\mathbb{R}^L} \leq 2\lambda_j\sqrt{\tau^{j+1}}$ $\mathbb{P}-a.s.$, and for all $\mathbf{v}\in\mathbb{R}^L$ with $\|\mathbf{v}\|_{\mathbb{R}^L} = 1$

$$\|D_{\mathbf{z}}\varphi(\mathbf{z},\mathbf{v})\|_{\mathbb{R}^L} = \sqrt{\sum_{i=1}^{L}\Big|\langle D_{\mathbf{z}}\partial_{z_i}u(t_j,\mathbf{z}),\mathbf{v}\rangle_{\mathbb{R}^L}\Big|^2} \leq \mathfrak{C}(\phi,g)\,,$$

the assertion then follows from (9.51).

□

Next, we show convergence with optimal (weak) order 1 of the *a posteriori* error estimate (9.28) on a mesh with maximum mesh size $\tau^{max} > 0$.

Theorem 9.12. Assume **(B1)** − **(B3)** in Section 8.2. Fix $(t,\mathbf{x})\in[0,T)\times\boldsymbol{\mathcal{D}}$. Let $J\equiv J(t,\mathbf{x})\in\mathbb{N}$ and let $\{(\mathbf{Y}^j_{\mathbf{X}}, Y^j_Z)\}_{j=0}^J$ solve Scheme 3 on a mesh $\{t_j\}_{j=0}^J\subset[t,T]$ with local mesh sizes $\{\tau^{j+1}\}_{j=0}^{J-1}$ and maximum mesh size $\tau^{max} = \max_j\tau^{j+1}$. Then, there exists $\boldsymbol{C}\equiv\boldsymbol{C}(\phi,g,T) > 0$, such that

$$\mathfrak{C}(\phi,g)\cdot\sum_{j=0}^{J-1}\tau^{j+1}\Big\{\mathfrak{H}_1^{(j)} + \mathfrak{H}_2^{(j)} + \mathfrak{H}_3^{(j)}\Big\} \leq \boldsymbol{C}\cdot\tau^{max}\,.$$

Proof. In the following, $\boldsymbol{C} > 0$ is a constant, which might differ from line to line, but is always independent of τ^{max}.
Step 1: We independently bound $\{\mathbf{E_k}(\cdot)\}_{k=1,\dots,28}$ appearing in the error estimators $\{\mathfrak{H}_\ell^{(j)}\}_{j\geq0}$, $\ell = 1,2$, in (9.28).
a) Bounds for $\{\mathbf{E_k}(\cdot)\}_{k=1,\dots,14}$ in $\mathfrak{H}_1^{(j)}$: Due to **(B1)** − **(B3)**, the functions $\mathbf{b}$ and $\boldsymbol{\sigma}$, as well as their derivatives are bounded in $\overline{\boldsymbol{\mathcal{D}}}$ by some constant $\boldsymbol{C} > 0$. Let $j\geq0$. We immediately obtain

$$\mathbf{E_k}(\mathbf{Y}^j_{\mathbf{X}}) \leq \boldsymbol{C}\cdot\mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{D}}\setminus\boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big] \qquad \mathbf{k} = 1,2,5,7,9,11,14\,.$$

Moreover, by means of the representation of the increment $\|\mathbf{Y}_{\mathbf{X}}^{j+1} - \mathbf{Y}_{\mathbf{X}}^{j}\|_{\mathbb{R}^L}$ (see Scheme 3) and standard arguments used before, we further get

$$\mathbf{E_k}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^{j}) \leq \boldsymbol{C} \cdot \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^{j} \in \boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big] \cdot \tau^{j+1} \qquad \mathbf{k} = 3, 4, 6, 8, 10\,,$$

and

$$\mathbf{E_{12}}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^{j}) \leq \boldsymbol{C} \cdot \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^{j} \in \boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big] \cdot \Big(\tau^{j+1}\Big)^{\frac{1}{2}}\,.$$

Moreover, by *Taylor's formula*, we obtain

$$|\mathbf{E_{13}}(\mathbf{Y}_{\mathbf{X}}^{j+1}, \mathbf{Y}_{\mathbf{X}}^{j}, t_j)| \leq \boldsymbol{C} \cdot \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^{j} \in \boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big] \cdot \tau^{j+1}\,.$$

Consequently, by considering the representation of $\mathfrak{H}_1^{(j)}$ in Theorem 9.7, we obtain

$$\mathfrak{H}_1^{(j)} \leq \boldsymbol{C} \cdot \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^{j} \in \boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big] \cdot \tau^{j+1}\,. \tag{9.52}$$

b) Bounds for $\{\mathbf{E_k}(\cdot)\}_{k=15,\dots,28}$ in $\mathfrak{H}_2^{(j)}$: Similar to **a)**, we obtain

$$\mathfrak{H}_2^{(j)} \leq \boldsymbol{C} \cdot \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^{j} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}} \mathbf{1}_{\big\{\overline{\mathbf{Y}}_{\mathbf{X}}^{j} = \mathbf{Y}_{\mathbf{X}}^{j} + \lambda_j \sqrt{\tau^{j+1}} \boldsymbol{n}\big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^{j})\big)\big\}}\Big] \cdot \tau^{j+1}\,. \tag{9.53}$$

Step 2: Let $j \geq 0$. By means of the representation of λ_j in Scheme 3 (see (8.4), in particular), we get

$$\mathfrak{H}_3^{(j)} \leq \boldsymbol{C} \cdot \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^{j} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big]\,. \tag{9.54}$$

Step 3: We plug (9.52), (9.53) and (9.54) into (9.28), and use $\tau^{j+1} \leq \tau^{max}$, to get

$$\begin{aligned}
&\sum_{j=0}^{J-1} \tau^{j+1}\Big\{\mathfrak{H}_1^{(j)} + \mathfrak{H}_2^{(j)} + \mathfrak{H}_3^{(j)}\Big\} \\
&\quad \leq \boldsymbol{C} \cdot \tau^{max} \cdot \sum_{j=0}^{J-1} \tau^{j+1} \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^{j} \in \boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}} + \mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^{j} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}} \mathbf{1}_{\big\{\overline{\mathbf{Y}}_{\mathbf{X}}^{j} = \mathbf{Y}_{\mathbf{X}}^{j} + \lambda_j \sqrt{\tau^{j+1}} \boldsymbol{n}\big(\Pi_{\boldsymbol{\partial\mathcal{D}}}(\mathbf{Y}_{\mathbf{X}}^{j})\big)\big\}}\Big] \\
&\qquad + \boldsymbol{C} \cdot \tau^{max} \cdot \sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^{j} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big] \\
&\quad \leq \boldsymbol{C} \cdot \tau^{max} \cdot \Bigg\{T + \sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^{j} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big]\Bigg\}\,.
\end{aligned} \tag{9.55}$$

Due to Lemma 8.5, we have

$$\sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^{j} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1}}\big\}}\Big] < 2\,. \tag{9.56}$$

Step 4: We plug (9.56) into (9.55), which proves assertion.

□

9.3. A posteriori error analysis for the Euler method

Consider the Euler method

$$\mathbf{Y}_{\mathbf{X}}^{j+1} = \mathbf{Y}_{\mathbf{X}}^{j} + \mathbf{b}(\mathbf{Y}_{\mathbf{X}}^{j})\tau^{j+1} + \boldsymbol{\sigma}(\mathbf{Y}_{\mathbf{X}}^{j})\Big(\mathbf{W}_{t_{j+1}} - \mathbf{W}_{t_j}\Big) \qquad (j \geq 0)\,, \qquad \mathbf{Y}_{\mathbf{X}}^{0} = \mathbf{x}\,, \tag{9.57}$$

to approximate (6.1). The interpolating continuified Euler process $\boldsymbol{\mathcal{Y}}^{\mathbf{X}} \equiv \Big\{\boldsymbol{\mathcal{Y}}_t^{\mathbf{X}}\,; t \geq 0\Big\}$ of the $\{\mathbf{Y}_{\mathbf{X}}^{j}\}_{j\geq 0}$ is given by

$$\boldsymbol{\mathcal{Y}}_t^{\mathbf{X}} = \mathbf{Y}_{\mathbf{X}}^{j} + \mathbf{b}(\mathbf{Y}_{\mathbf{X}}^{j})(t-t_j) + \boldsymbol{\sigma}(\mathbf{Y}_{\mathbf{X}}^{j})\Big(\mathbf{W}_t - \mathbf{W}_{t_j}\Big) \qquad t \in [t_j, t_{j+1}] \qquad (j \geq 0)\,. \tag{9.58}$$

We consider the parabolic PDE (6.1) with $\boldsymbol{\sigma}(\mathbf{x}) \equiv \mathbb{I}$, $g \equiv 0$ for simplicity; *cf.* also Example 6.3. For the following, we fix $(t,\mathbf{x}) \in [0,T) \times \boldsymbol{\mathcal{D}}$ in (6.2) and let $\{t_j\}_{j=0}^{J} \subset [t,T]$ be a mesh with local step sizes $\{\tau^{j+1}\}_{j=0}^{J-1}$, where $J \equiv J(t,\mathbf{x}) \in \mathbb{N}$. We denote by

$$\tilde{\boldsymbol{\tau}} := \inf\Big\{s > t \,:\, \boldsymbol{\mathcal{Y}}_s^{\mathbf{X},t,\mathbf{x}} \in \boldsymbol{\partial\mathcal{D}} \text{ or } s \notin (t,T)\Big\} \tag{9.59}$$

the first exit time of $\boldsymbol{\mathcal{Y}}^{\mathbf{X},t,\mathbf{x}} \equiv \Big\{\boldsymbol{\mathcal{Y}}_s^{\mathbf{X},t,\mathbf{x}}\,;\ s \in [t,T]\Big\}$, which starts at time $t \in [0,T]$ in $\mathbf{x} \in \boldsymbol{\mathcal{D}}$. Motivated by [37, p. 181], the expression

$$\sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbf{1}_{\{\tilde{\boldsymbol{\tau}} \in (t_j, t_{j+1}]\}}\phi(t_{j+1}, \mathbf{Y}_{\mathbf{X}}^{j+1})\Big] \tag{9.60}$$

is now used to approximate (6.2). In this respect, (9.60) first localizes $\tilde{\boldsymbol{\tau}}$ in $(t_j, t_{j+1}]$, and if it is assured, proceeds via the approximation $\tilde{\boldsymbol{\tau}} \approx t_{j+1}$. The following theorem presents a related *a posteriori* error estimate for (9.57), (9.59). From its representation it is not difficult to obtain first order of convergence for (9.61) on families of (time-)meshes with maximum mesh size $\tau^{max} > 0$.

Theorem 9.13. Assume **(B1)** $-$ **(B3)** in Section 8.2. Fix $(t,\mathbf{x}) \in [0,T) \times \boldsymbol{\mathcal{D}}$. Let $J \equiv J(t,\mathbf{x}) \in \mathbb{N}$, and $\{t_j\}_{j=0}^{J} \subset [t,T]$ be a mesh with local step sizes $\{\tau^{j+1}\}_{j=0}^{J-1}$. Let $\{\boldsymbol{\mathcal{Y}}_{t_j}^{\mathbf{X},t,\mathbf{x}}\}_{j=0}^{J}$ solve (9.58). Then, for $\mathfrak{C}(\phi, g) \geq 1$ from Lemma 8.2,

$$\left|u(t,\mathbf{x}) - \sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbf{1}_{\{\tilde{\boldsymbol{\tau}} \in (t_j, t_{j+1}]\}}\phi(t_{j+1}, \mathbf{Y}_{\mathbf{X}}^{j+1})\Big]\right| \leq \mathfrak{C}(\phi, g) \cdot \sum_{j=0}^{J-1} \tau^{j+1}\Big\{\tilde{\mathfrak{H}}_1^{(j)} + \tilde{\mathfrak{H}}_2^{(j)}\Big\}\,, \tag{9.61}$$

where the *a posteriori* error estimators $\{\tilde{\mathfrak{H}}_\ell^{(j)}\}_{\ell=1}^{2}$ are given by

$$\begin{aligned}
\tilde{\mathfrak{H}}_1^{(j)} &:= \frac{1}{2} \cdot \tilde{\mathbf{E}}_{\mathbf{1}}(\mathbf{Y}_{\mathbf{X}}^{j}) \cdot \tau^{j+1} + \frac{1}{4}L \cdot \tilde{\mathbf{E}}_{\mathbf{2}}(\mathbf{Y}_{\mathbf{X}}^{j}) \cdot \tau^{j+1} + \frac{1}{2}L \cdot \tilde{\mathbf{E}}_{\mathbf{3}}(\mathbf{Y}_{\mathbf{X}}^{j}) \cdot \tau^{j+1} \\
&\quad + \Big\{C_{D^2\mathbf{b}} \cdot \sqrt{\tilde{\mathbf{E}}_{\mathbf{4}}(\mathbf{Y}_{\mathbf{X}}^{j})} + \frac{1}{2}LC_{D^3\mathbf{b}} + LC_{D^2\mathbf{b}}\Big\} \\
&\quad \cdot \sqrt{\frac{1}{15}\tilde{\mathbf{E}}_{\mathbf{4}}(\mathbf{Y}_{\mathbf{X}}^{j}) \cdot \tau^{j+1} + \frac{L}{8} \cdot \tilde{\mathbf{E}}_{\mathbf{5}}(\mathbf{Y}_{\mathbf{X}}^{j})} \cdot \Big(\tau^{j+1}\Big)^{\frac{3}{2}}\,, \\
\tilde{\mathfrak{H}}_2^{(j)} &:= \tilde{\mathbf{E}}_{\mathbf{6}}(\mathbf{Y}_{\mathbf{X}}^{j})\,,
\end{aligned}$$

where $C_{D^2\mathbf{b}} := \sup_{\mathbf{y}\in\overline{\boldsymbol{\mathcal{D}}}} \|D^2\mathbf{b}(\mathbf{y})\|_{\mathcal{L}^2}$, $C_{D^3\mathbf{b}} := \sup_{\mathbf{y}\in\overline{\boldsymbol{\mathcal{D}}}} \|D^3\mathbf{b}(\mathbf{y})\|_{\mathcal{L}^3}$, and with computable terms

1. $\tilde{\mathbf{E}}_1(\mathbf{Y}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\{t_j<\tilde{\boldsymbol{\tau}}\}}\|D\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\cdot\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\|_{\mathbb{R}^L}\Big]$,
2. $\tilde{\mathbf{E}}_2(\mathbf{Y}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\{t_j<\tilde{\boldsymbol{\tau}}\}}\|D^2\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\|_{\mathcal{L}^2}\Big]$,
3. $\tilde{\mathbf{E}}_3(\mathbf{Y}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\{t_j<\tilde{\boldsymbol{\tau}}\}}\|D\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\|_{\mathcal{L}^1}\Big]$,
4. $\tilde{\mathbf{E}}_4(\mathbf{Y}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\{t_j<\tilde{\boldsymbol{\tau}}\}}\|\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\|^2_{\mathbb{R}^L}\Big]$,
5. $\tilde{\mathbf{E}}_5(\mathbf{Y}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\{t_j<\tilde{\boldsymbol{\tau}}\}}\Big]$,
6. $\tilde{\mathbf{E}}_6(\mathbf{Y}^j_{\mathbf{X}}) := \mathbb{E}\Big[\mathbf{1}_{\{\tilde{\boldsymbol{\tau}}\in(t_j,t_{j+1}]\}}\Big(C_{\partial\phi} + C_{D\phi}\|\mathbf{b}(\mathbf{Y}^j_{\mathbf{X}})\|_{\mathbb{R}^L} + \tfrac{1}{2}LC_{D^2\phi}\Big)\Big]$,

where

$$C_{\partial\phi} := \sup_{(s,\mathbf{y})\in[0,T]\times\overline{\boldsymbol{\mathcal{D}}}} |\partial\phi(s,\mathbf{y})|\,, \qquad C_{D\phi} := \sup_{(s,\mathbf{y})\in[0,T]\times\overline{\boldsymbol{\mathcal{D}}}} \|D\phi(s,\mathbf{y})\|_{\mathbb{R}^L}\,,$$
$$C_{D^2\phi} := \sup_{(s,\mathbf{y})\in[0,T]\times\overline{\boldsymbol{\mathcal{D}}}} \|D^2\phi(s,\mathbf{y})\|_{\mathbb{R}^{L\times L}}\,.$$

Proof. (*Sketch of the proof of Theorem 9.13*) We add and substract the term $\mathbb{E}\Big[\phi(\tilde{\boldsymbol{\tau}},\boldsymbol{\mathcal{Y}}^{\mathbf{X},t,\mathbf{x}}_{\tilde{\boldsymbol{\tau}}})\Big]$ in the expression on the left-hand side of (9.61) to get

$$\begin{aligned}&\Big|u(t,\mathbf{x}) - \sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbf{1}_{\{\tilde{\boldsymbol{\tau}}\in(t_j,t_{j+1}]\}}\phi(t_{j+1},\mathbf{Y}^{j+1}_{\mathbf{X}})\Big]\Big|\\ &\quad\le \underbrace{\Big|u(t,\mathbf{x}) - \mathbb{E}\Big[\phi(\tilde{\boldsymbol{\tau}},\boldsymbol{\mathcal{Y}}^{\mathbf{X},t,\mathbf{x}}_{\tilde{\boldsymbol{\tau}}})\Big]\Big|}_{=:\tilde{\boldsymbol{I}}} + \underbrace{\Big|\mathbb{E}\Big[\phi(\tilde{\boldsymbol{\tau}},\boldsymbol{\mathcal{Y}}^{\mathbf{X},t,\mathbf{x}}_{\tilde{\boldsymbol{\tau}}})\Big] - \sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbf{1}_{\{\tilde{\boldsymbol{\tau}}\in(t_j,t_{j+1}]\}}\phi(t_{j+1},\mathbf{Y}^{j+1}_{\mathbf{X}})\Big]\Big|}_{=:\tilde{\boldsymbol{II}}}\,.\end{aligned}$$

a) Estimation of $\tilde{\boldsymbol{I}}$: The estimation of $\tilde{\boldsymbol{I}}$ is similar to Theorem 3.1 in the first part, where an *a posteriori* error estimate on $[0,T]\times\mathbb{R}^L$ is presented, and conceptually uses the same tools due to the Wiener process involved in (9.57) resp. (9.58). By means of the fact that $\mathbb{E}\Big[\phi(\tilde{\boldsymbol{\tau}},\boldsymbol{\mathcal{Y}}^{\mathbf{X},t,\mathbf{x}}_{\tilde{\boldsymbol{\tau}}})\Big] = \mathbb{E}\Big[u(\tilde{\boldsymbol{\tau}},\boldsymbol{\mathcal{Y}}^{\mathbf{X},t,\mathbf{x}}_{\tilde{\boldsymbol{\tau}}})\Big]$, *Itô's formula*, the identity in the first line of (6.1), Lemma 8.2, as well as Malliavin calculus techniques, we obtain

$$\tilde{\boldsymbol{I}} \le \mathfrak{C}(\phi,g)\cdot\sum_{j=0}^{J-1}\tau^{j+1}\tilde{\mathfrak{H}}^{(j)}_1\,.$$

b) Estimation of $\tilde{\boldsymbol{II}}$: Using *Itô's formula* immediately leads to

$$\tilde{\boldsymbol{II}} = \Big|\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbf{1}_{\{\tilde{\boldsymbol{\tau}}\in(t_j,t_{j+1}]\}}\Big(\phi(t_{j+1},\mathbf{Y}^{j+1}_{\mathbf{X}}) - \phi(\tilde{\boldsymbol{\tau}},\boldsymbol{\mathcal{Y}}^{\mathbf{X},t,\mathbf{x}}_{\tilde{\boldsymbol{\tau}}})\Big)\Big]\Big|$$

$$\leq \sum_{j=0}^{J-1} \Bigg| \mathbb{E}\Big[\mathbf{1}_{\{\tilde{\boldsymbol{\tau}} \in (t_j, t_{j+1}]\}} \int_{\tilde{\tau}}^{t_{j+1}} \partial_s \phi(s, \boldsymbol{\mathcal{Y}}_s^{\mathbf{X},t,\mathbf{x}}) + \Big\langle \mathbf{b}(\mathbf{Y}_{\mathbf{X}}^j), D_{\mathbf{x}} \phi(s, \boldsymbol{\mathcal{Y}}_s^{\mathbf{X},t,\mathbf{x}}) \Big\rangle_{\mathbb{R}^L} + \frac{1}{2} \mathrm{Tr}\Big(D_{\mathbf{x}}^2 \phi(s, \boldsymbol{\mathcal{Y}}_s^{\mathbf{X},t,\mathbf{x}}) \Big) \mathrm{d}s \Big] + \mathbb{E}\Big[\mathbf{1}_{\{\tilde{\boldsymbol{\tau}} \in (t_j, t_{j+1}]\}} \int_{\tilde{\tau}}^{t_{j+1}} \Big\langle D_{\mathbf{x}} \phi(s, \boldsymbol{\mathcal{Y}}_s^{\mathbf{X},t,\mathbf{x}}), \mathrm{d}\mathbf{W}_s \Big\rangle_{\mathbb{R}^L} \Big] \Bigg| .$$

Since $\mathbf{1}_{\{\tilde{\boldsymbol{\tau}} \in (t_j, t_{j+1}]\}} = 1 - \mathbf{1}_{\{\tilde{\boldsymbol{\tau}} \leq t_j\}} - \mathbf{1}_{\{\tilde{\boldsymbol{\tau}} > t_{j+1}\}}$, and due to arguments from probability theory, we get

$$\mathbb{E}\Big[\mathbf{1}_{\{\tilde{\boldsymbol{\tau}} \in (t_j, t_{j+1}]\}} \int_{\tilde{\tau}}^{t_{j+1}} \Big\langle D_{\mathbf{x}} \phi(s, \boldsymbol{\mathcal{Y}}_s^{\mathbf{X},t,\mathbf{x}}), \mathrm{d}\mathbf{W}_s \Big\rangle_{\mathbb{R}^L} \Big] = 0 .$$

Thus, we conclude

$$\tilde{\boldsymbol{II}} \leq \mathfrak{C}(\phi, g) \cdot \sum_{j=0}^{J-1} \tau^{j+1} \tilde{\mathfrak{H}}_2^{(j)} ,$$

where we assume $\mathfrak{C}(\phi, g) \geq 1$ without restrictions. □

Remark 9.3. The terms in $\tilde{\mathfrak{H}}_1^{(\cdot)}$ address effects inside $\boldsymbol{\mathcal{D}}$, and due to their derivation, are very similar to Theorem 3.1. In contrast, $\tilde{\mathfrak{H}}_2^{(\cdot)}$ captures effects concerning 'stopping'. In the framework of Example 6.3 for instance, we have

$$\tilde{\mathfrak{H}}_1^{(j)} \equiv 0 , \quad \tilde{\mathfrak{H}}_2^{(j)} = \Big(C_{\partial\phi} + \tfrac{1}{2} L C_{D^2\phi} \Big) \cdot \mathbb{E}\Big[\mathbf{1}_{\{\tilde{\boldsymbol{\tau}} \in (t_j, t_{j+1}]\}} \Big] .$$

The practical application of the *a posteriori* error estimate (9.61) is restricted by terms $\tilde{\mathbf{E}}_{\mathbf{1}}(\cdot), ..., \tilde{\mathbf{E}}_{\mathbf{6}}(\cdot)$ in $\tilde{\mathfrak{H}}_1^{(\cdot)}$ and $\tilde{\mathfrak{H}}_2^{(\cdot)}$, and even of (9.60) itself, which involves the abstract stopping time $\tilde{\boldsymbol{\tau}}$. A possibility to tackle this problem might be the use of concepts from [39] (*cf.* Section 7.1), which, however, requires (additional) approximations.

10. Adaptive weak Euler methods: algorithm and convergence

The Theorems 9.1 resp. 9.7 provide *a posteriori* error estimates for the approximation of (0.7) resp. (6.2). In this chapter, we use these results to set up adaptive methods that automatically steer successive local mesh size selection to meet a pre-assigned tolerance $\texttt{Tol} > 0$ of the overall errors.

10.1. Adaptive weak Euler method for the elliptic PDE (0.6)

The algorithm below combines Scheme 2 with an automatic step size selection procedure, which is based on the *a posteriori* error estimate in Section 9.1. In every step indexed by $j \in \mathbb{N}_0$, we first check the 'safeguard' criterion **I** (see Algorithm 10.1 below), which holds if the computation of $\mathbf{Y}_{\mathbf{X}}^j$ through **II (2) b)** and **II (3)** in the previous step took place with a step size being 'too large'. Next, we check via **II** if $\mathbf{Y}_{\mathbf{X}}^j$ is 'close' to the boundary $\boldsymbol{\partial\mathcal{D}}$ and then may possibly be projected onto $\boldsymbol{\partial\mathcal{D}}$; and if not, we compute the subsequent iterates $\mathbf{Y}_{\mathbf{X}}^{j+1}, Y_V^{j+1}, Y_Z^{j+1}$ with the help of step size τ^{j+1}. With this information at hand, we compute the *a posteriori* error estimators $\{\mathfrak{G}_\ell^{(j)}\}_{\ell=1}^3$ and check for a given tolerance $\texttt{Tol} > 0$, if the criterion in **III** which also incorporates the already computed $\{\mathfrak{G}_\ell^{(k)}\}_{k=0}^{j-1}$, $(\ell = 1, 2, 3)$, and which is

$$\sum_{k=0}^{j} \tau^{k+1}\Big\{\mathfrak{G}_1^{(k)} + \mathfrak{G}_2^{(k)} + \mathfrak{G}_3^{(k)}\Big\} \overset{!}{\leq} \texttt{Tol} \cdot \left(1 + a_j \cdot \sum_{k=0}^{j} \left\{\tau^{k+1}\mathbb{E}\Big[\mathbf{1}_{\{J^* > k\}}\Big] + \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}_{\mathbf{X}}^k \in \boldsymbol{\mathcal{S}}_{\tau^{k+1}}\big\}}\Big]\right\}\right), \tag{10.1}$$

is met or not to decide whether to coarsen or refine the current mesh size τ^{j+1}. Different to Algorithm 4.1 in the first part, criterion (10.1) involves the error estimators $\{\mathfrak{G}_\ell^{(k)}\}_{\ell=1}^3$ with $k \leq j$ in summarized form, and aims for $\texttt{Tol}$ times a 'temporal weight' as corresponding upper bound to adress stopping. Hence, a summation also appears on the right-hand side of (10.1) to address accumulated errors; see Theorem 10.2 below. Furthermore, $\{a_j\}_{j\geq 0}$ is a sequence of *additional* 'weights' with $1 \leq a_{j-1} \leq a_j \leq \mathsf{C}$ for some $\mathsf{C} > 0$, which may be given by the user; see Chapter 11 for a detailed explanation. A suitable choice may be $a_j := \min\Big(1 + L \cdot t_j, 1 + L \cdot (C + k)\Big)$, for some $k \in \mathbb{N}$, where $C > 0$ is the constant from Lemma 8.4.

Algorithm 10.1. Fix $\texttt{Tol} > 0$ and $\tau^1 \geq \texttt{Tol}$. Let $(\tau^j, \mathbf{Y}^j_{\mathbf{X}}, Y^j_V, Y^j_Z)$ and a_j be given for some $j \geq 0$. Define $\tau^{j+1,0} := \tau^j$.

I ('`Safeguard`') If $\mathbf{Y}^j_{\mathbf{X}} \notin \boldsymbol{\mathcal{D}}$, set $J^* := j$, $\mathbf{Y}^{J^*}_{\mathbf{X}} := \Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})$, $Y^{J^*}_V := Y^j_V$, $Y^{J^*}_Z := Y^j_Z$, and **STOP**.

— For $\ell = 0, 1, 2, ...$ do:

II **(1)** ('`Localization`') If $\mathbf{Y}^j_{\mathbf{X}} \in \boldsymbol{\mathcal{D}} \setminus \boldsymbol{\mathcal{S}}_{\tau^{j+1,\ell}}$, set $\overline{\mathbf{Y}}^j_{\mathbf{X}} := \mathbf{Y}^j_{\mathbf{X}}$ and go to **II (3)**.

(2) ('`Localization`') If $\mathbf{Y}^j_{\mathbf{X}} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1,\ell}}$, draw a *Bernoulli* distributed random variable $\mathtt{U}$ with parameter p_j given in (8.5).

a) ('`Projection`') If $\mathtt{U} = 1$, set $J^* := j$, $\mathbf{Y}^{J^*}_{\mathbf{X}} := \Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})$, $Y^{J^*}_V := Y^j_V$, $Y^{J^*}_Z := Y^j_Z$, and go to **III**.

b) ('`Bouncing back`') If $\mathtt{U} = 0$, set $\overline{\mathbf{Y}}^j_{\mathbf{X}} := \mathbf{Y}^j_{\mathbf{X}} + \lambda_j \sqrt{\tau^{j+1,\ell}}\boldsymbol{n}\Big(\Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big)$ and go to **II (3)**.

(3) ('`Solve`') Set $\mathbf{Y}^j_{\mathbf{X}} := \overline{\mathbf{Y}}^j_{\mathbf{X}}$. Compute $\mathbf{Y}^{j+1}_{\mathbf{X}}$, Y^{j+1}_V and Y^{j+1}_Z via Scheme 1 with step size $\tau^{j+1,\ell}$.

III ('`Computation`') Compute $\{\mathfrak{G}^{(j,\ell)}_i\}_{i=1}^3$, *i.e.*, $\{\mathfrak{G}^{(j)}_i\}_{i=1}^3$ with step size $\tau^{j+1,\ell}$, set

$$\boldsymbol{\mathfrak{G}}^{(j,\ell)} := \sum_{k=0}^{j-1} \tau^{k+1}\Big\{\mathfrak{G}^{(k)}_1 + \mathfrak{G}^{(k)}_2 + \mathfrak{G}^{(k)}_3\Big\} + \tau^{j+1,\ell}\Big\{\mathfrak{G}^{(j,\ell)}_1 + \mathfrak{G}^{(j,\ell)}_2 + \mathfrak{G}^{(j,\ell)}_3\Big\},$$

$$\mathtt{I}_{j,\ell} := \sum_{k=0}^{j-1}\left\{\tau^{k+1}\mathbb{E}\Big[\mathbf{1}_{\{J^*>k\}}\Big] + \mathbb{E}\Big[\mathbf{1}_{\Big\{\mathbf{Y}^k_{\mathbf{X}} \in \boldsymbol{\mathcal{S}}_{\tau^{k+1}}\Big\}}\Big]\right\} + \tau^{j+1,\ell}\mathbb{E}\Big[\mathbf{1}_{\{J^*>k\}}\Big] + \mathbb{E}\Big[\mathbf{1}_{\Big\{\mathbf{Y}^j_{\mathbf{X}} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1,\ell}}\Big\}}\Big],$$

and decide:

(1) If $\boldsymbol{\mathfrak{G}}^{(j,\ell)} > \texttt{Tol} \cdot \Big(1 + a_j \cdot \mathtt{I}_{j,\ell}\Big)$, set $\tau^{j+1,\ell+1} := \frac{\tau^{j+1,\ell}}{2}$, and $\ell \hookrightarrow \ell + 1$.

(2) If $\frac{\texttt{Tol}}{2} \cdot \Big(1 + a_j \cdot \mathtt{I}_{j,\ell}\Big) \leq \boldsymbol{\mathfrak{G}}^{(j,\ell)} \leq \texttt{Tol} \cdot \Big(1 + a_j \cdot \mathtt{I}_{j,\ell}\Big)$, and

a) if $\mathtt{U} = 1$, **STOP**.

b) if $\mathtt{U} = 0$, set $\tau^{j+1} := \tau^{j+1,\ell}$, $t_{j+1} := t_j + \tau^{j+1}$, and $j \hookrightarrow j + 1$.

(3) If $\boldsymbol{\mathfrak{G}}^{(j,\ell)} < \frac{\texttt{Tol}}{2} \cdot \Big(1 + a_j \cdot \mathtt{I}_{j,\ell}\Big)$, and

a) if $\mathtt{U} = 1$, **STOP**.

b) if $\mathtt{U} = 0$, set $\tau^{j+1} := \tau^{j+1,\ell}$, $t_{j+1} := t_j + \tau^{j+1}$. Then set $\tau^{j+1} := 2\tau^{j+1}$ and $j \hookrightarrow j + 1$.

After the admissible step size τ^{j+1} has been generated through the (finite) sequence $\{\tau^{j+1,\ell}\}_{\ell\geq 0}$ of refinements **III (1)** which precedes a single potential coarsening step **III (3)**, the new local error (up to time t_{j+1}) is bounded by (10.1).

The following theorem validates termination and convergence (*w.r.t.* '`Tol`') of Algorithm 10.1; its proof is similar to Theorem 4.2 in the first part, and exploits the stability results in Lemmas 8.3 and 8.4.

Theorem 10.2. Let $\texttt{Tol} > 0$. Suppose **(A1)** – **(A3)** in Section 8.1. Further suppose that the step sizes $\{\tau^{j+1}\}_{j\geq 0}$ are sufficiently small according to Lemma 8.4, or either **(A1*)** or **(A1**)** are valid. Fix $\mathbf{x} \in \boldsymbol{\mathcal{D}}$. Then, Algorithm 10.1 generates each local step size τ^{j+1} after $\mathcal{O}\Big(\log(\texttt{Tol}^{-1})\Big)$ many iterations and terminates after $\mathbb{E}\Big[J^*\Big] = \mathcal{O}\Big(\texttt{Tol}^{-1}\Big)$ time steps. Furthermore, the 4−tuple $\Big\{(\tau^{j+1}, \mathbf{Y}^j_{\mathbf{X}}, Y^j_V, Y^j_Z)\Big\}_{j\geq 0}$ satisfies

$$\left| u(\mathbf{x}) - \mathbb{E}\Big[\phi(\mathbf{Y}^{J^*}_{\mathbf{X}})Y^{J^*}_V + Y^{J^*}_Z\Big]\right| \leq \mathbf{C}\cdot\texttt{Tol}\,, \tag{10.2}$$

where $\mathbf{C} > 0$ depends on $\boldsymbol{C}(\phi, g) > 0$ from Lemma 8.1, $C > 0$ from Lemma 8.4, and the upper bound of $\max\limits_{j\geq 0} a_j$.

In the following proof, let $\boldsymbol{C} \geq 1$ be a constant which might differ from line to line, but is always independent of τ^{j+1} and $j \geq 0$.

Proof. **a) Termination for each** $j \geq 0$**:** Fix $j \geq 0$, $\ell \geq 0$, and use (9.22), (9.23) and (9.24) to deduce

$$\tau^{j+1,\ell}\Big\{\mathfrak{G}^{(j,\ell)}_1 + \mathfrak{G}^{(j,\ell)}_2 + \mathfrak{G}^{(j,\ell)}_3\Big\} \leq \boldsymbol{C}\tau^{j+1,\ell}\left\{\tau^{j+1,\ell}\mathbb{E}\Big[\mathbf{1}_{\{J^*>j\}}\Big] + \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1,\ell}}\big\}}\Big]\right\}. \tag{10.3}$$

Hence, we generate a finite sequence $\{\tau^{j+1,\ell}\}_{\ell=1}^{\ell^*_{j+1}}$ with $\tau^{j+1,\ell} = \frac{\tau^{j+1,0}}{2^\ell}$, $\ell = 0, ..., \ell^*_{j+1}$, according to the refinement mechanism **III (1)** in Algorithm 4.1 until either **III (2)**, or **III (3)** is met. In view of (10.3), as well as

$$\begin{aligned} \mathfrak{G}^{(j,\ell)} &\leq \texttt{Tol}\cdot\left\{1 + a_{j-1}\cdot\sum_{k=0}^{j-1}\tau^{k+1}\mathbb{E}\Big[\mathbf{1}_{\{J^*>k\}}\Big] + \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}^k_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{k+1}}\big\}}\Big]\right\} \\ &\quad + \boldsymbol{C}\tau^{j+1,\ell}\left\{\tau^{j+1,\ell}\mathbb{E}\Big[\mathbf{1}_{\{J^*>j\}}\Big] + \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}^j_{\mathbf{X}}\in\boldsymbol{\mathcal{S}}_{\tau^{j+1,\ell}}\big\}}\Big]\right\} \\ &\overset{!}{\leq} \texttt{Tol}\cdot\Big(1 + a_j\cdot\texttt{I}_{j,\ell}\Big)\,, \end{aligned} \tag{10.4}$$

and the fact that $a_j \geq 1$ and $a_j - a_{j-1} \geq 0$, we find out that $\ell = \left\lceil \log\Big(\frac{\tau^{j+1,0}\boldsymbol{C}}{\texttt{Tol}}\Big)/\log(2)\right\rceil$ is the smallest number such that

$$\boldsymbol{C}\cdot\tau^{j+1,\ell} = \boldsymbol{C}\cdot\frac{\tau^{j+1,0}}{2^\ell} \overset{!}{\leq} \texttt{Tol}\,,$$

which, however, implies (10.4). Consequently,

$$0 \leq \ell^*_{j+1} \leq \left\lceil\frac{\log\Big(\frac{\tau^{j+1,0}\boldsymbol{C}}{\texttt{Tol}}\Big)}{\log(2)}\right\rceil\,, \tag{10.5}$$

which yields a maximum of $\mathcal{O}\Big(\log(\mathtt{Tol}^{-1})\Big)$ (refinement) steps to accept the local step size $\tau^{j+1} := \tau^{j+1,\ell^*_{j+1}} = \frac{\tau^{j+1,0}}{2^{\ell^*_{j+1}}}$.

b) Global termination: We show by induction that

$$\tau^{j+1} \geq \frac{\mathtt{Tol}}{2\boldsymbol{C}} \qquad (j \geq 0)\,. \tag{10.6}$$

This means that all step sizes are bounded from below by $\frac{\mathtt{Tol}}{2\boldsymbol{C}}$, which, in particular, can be considered as the smallest step size that Algorithm 10.1 is able to generate.
The base case follows by the choice of the initial mesh size $\tau^1 \geq \mathtt{Tol}$. Now suppose that we have generated $\mathbf{Y}^j_{\mathbf{X}}$, Y^j_V and Y^j_Z with step size $\tau^j \geq \frac{\mathtt{Tol}}{2\boldsymbol{C}}$ and no 'stopping' has been detected before. In order to successfully compute $\mathbf{Y}^{j+1}_{\mathbf{X}}$ (and Y^j_V, Y^{j+1}_Z), we set $\tau^{j+1,0} := \tau^j$ (if **III (2)** occurred in the generation of τ^j), or $\tau^{j+1,0} := 2\tau^j$ (if **III (3)** occurred in the generation of τ^j). In all cases, $\tau^{j+1,0} \geq \frac{\mathtt{Tol}}{2\boldsymbol{C}}$. Via **a)**, we generate a finite sequence $\{\tau^{j+1,\ell}\}_{\ell=0}^{\ell^*_{j+1}}$ until either **III (2)**, or **III (3)** is met, and then generate $\mathbf{Y}^{j+1}_{\mathbf{X}}$ (and Y^j_V, Y^{j+1}_Z) with step size $\tau^{j+1} := \tau^{j+1,\ell^*_{j+1}} = \frac{\tau^{j+1,0}}{2^{\ell^*_{j+1}}}$. Since $\lceil x \rceil \leq 1 + x$, $x \in \mathbb{R}$, we conclude by means of (10.5) that

$$\tau^{j+1} := \tau^{j+1,\ell^*_{j+1}} = \frac{\tau^{j+1,0}}{2^{\ell^*_{j+1}}} \geq \frac{\mathtt{Tol}}{2\boldsymbol{C}}\,.$$

Having (10.6) at hand, we conclude that

$$\mathbb{E}[t_{J^*}] = \sum_{j=0}^{\infty} \tau^{j+1}\mathbb{E}\Big[\mathbf{1}_{\{J^*>j\}}\Big] \geq \mathbb{E}\Big[J^*\Big] \cdot \frac{\mathtt{Tol}}{2\boldsymbol{C}}\,.$$

By Lemma 8.4, we may infer $\mathbb{E}\Big[J^*\Big] = \mathcal{O}\Big(\mathtt{Tol}^{-1}\Big)$.

c) Convergence rate: Let $j \geq 0$. We (upper) bound the right-hand side of (10.1) (independent of j). By means of the boundedness of $a_j \leq \mathsf{C}$, and Lemmas 8.3 and 8.4, we immediately obtain

$$\begin{aligned}
\mathtt{Tol} \cdot \Bigg(1{+}a_j \cdot \sum_{k=0}^{j} \Big\{\tau^{k+1}\mathbb{E}\Big[\mathbf{1}_{\{J^*>k\}}\Big] + \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}^k_{\mathbf{X}} \in \boldsymbol{\mathcal{S}}_{\tau^{k+1}}\big\}}\Big]\Big\}\Bigg)& \\
\leq \mathtt{Tol} \cdot \Bigg(1 + \mathsf{C} \cdot \sum_{k=0}^{\infty} \Big\{\tau^{k+1}\mathbb{E}\Big[\mathbf{1}_{\{J^*>k\}}\Big] + \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}^k_{\mathbf{X}} \in \boldsymbol{\mathcal{S}}_{\tau^{k+1}}\big\}}\Big]\Big\}\Bigg)& \\
\leq \mathtt{Tol} \cdot \Bigg(1 + \mathsf{C} \cdot \Bigg(\mathbb{E}[t_{J^*}] + \sum_{k=0}^{\infty} \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}^k_{\mathbf{X}} \in \boldsymbol{\mathcal{S}}_{\tau^{k+1}}\big\}}\Big]\Bigg)\Bigg)& \\
\leq \mathtt{Tol} \cdot \Bigg(1 + \mathsf{C} \cdot \Big(C + 2\Big)\Bigg)\,,&
\end{aligned} \tag{10.7}$$

where $C > 0$ is from Lemma 8.4. Now, (10.2) immediately follows from the *a posteriori* error estimate (9.1), the tolerance criterion in Algorithm 10.1 (*cf.* (10.1)), and (10.7).

□

10.2. Adaptive weak Euler method for the parabolic PDE (6.1)

By following a similar guideline in the setup of Algorithm 10.1, we present an adaptive algorithm (see Algorithm 10.3 below) for the (pointwise, *i.e.*, for $(t, \mathbf{x}) \in [0,T) \times \boldsymbol{\mathcal{D}}$ fixed) approximation of PDE (6.1). Similar to Theorem 10.2 in Section 10.1, Theorem 10.4 validates termination and convergence of the adaptive method.

Algorithm 10.3. Fix $\mathtt{Tol} > 0$ and $\tau^1 \geq \mathtt{Tol}$. Let $(\tau^j, \mathbf{Y}^j_{\mathbf{X}}, Y^j_Z)$ and $a_j = 1 + L \cdot t_j$ be given for some $j \geq 0$. Define $\tau^{j+1,0} := \tau^j$.

I (‘`Safeguard`’) If $\mathbf{Y}^j_{\mathbf{X}} \notin \boldsymbol{\mathcal{D}}$, set $J^* := j$, $\mathbf{Y}^{J^*}_{\mathbf{X}} := \Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})$, $Y^{J^*}_Z := Y^j_Z$, and **STOP**.

— For $\ell = 0, 1, 2, \ldots$ do:

II **(1)** (‘`Localization`’) If $\mathbf{Y}^j_{\mathbf{X}} \notin \boldsymbol{\mathcal{S}}_{\tau^{j+1,\ell}}$, set $\overline{\mathbf{Y}}^j_{\mathbf{X}} := \mathbf{Y}^j_{\mathbf{X}}$ and go to **II (3)**.

(2) (‘`Localization`’) If $\mathbf{Y}^j_{\mathbf{X}} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1,\ell}}$, draw a *Bernoulli* distributed random variable $\mathtt{U}$ with parameter p_j given in (8.5).

a) (‘`Projection`’) If $\mathtt{U} = 1$, set $J^* := j$, $\mathbf{Y}^{J^*}_{\mathbf{X}} := \Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})$, $Y^{J^*}_Z := Y^j_Z$, and go to **III**.

b) (‘`Bouncing back`’) If $\mathtt{U} = 0$, set $\overline{\mathbf{Y}}^j_{\mathbf{X}} := \mathbf{Y}^j_{\mathbf{X}} + \lambda_j \sqrt{\tau^{j+1,\ell}} \boldsymbol{n}\Big(\Pi_{\partial\boldsymbol{\mathcal{D}}}(\mathbf{Y}^j_{\mathbf{X}})\Big)$ and go to **II (3)**.

(3) (‘`Solve`’) Set $\mathbf{Y}^j_{\mathbf{X}} := \overline{\mathbf{Y}}^j_{\mathbf{X}}$. Compute $\mathbf{Y}^{j+1}_{\mathbf{X}}$ and Y^{j+1}_Z via Scheme 3 with step size $\tau^{j+1,\ell}$.

III (‘`Computation`’) Compute $\{\mathfrak{H}^{(j,\ell)}_i\}_{i=1}^3$, *i.e.*, $\{\mathfrak{H}^{(j)}_i\}_{i=1}^3$ with step size $\tau^{j+1,\ell}$, set

$$\boldsymbol{\mathfrak{H}}^{(j,\ell)} := \sum_{k=0}^{j-1} \tau^{k+1}\Big\{\mathfrak{H}^{(k)}_1 + \mathfrak{H}^{(k)}_2 + \mathfrak{H}^{(k)}_3\Big\} + \tau^{j+1,\ell}\Big\{\mathfrak{H}^{(j,\ell)}_1 + \mathfrak{H}^{(j,\ell)}_2 + \mathfrak{H}^{(j,\ell)}_3\Big\},$$

$$\mathtt{I}_{j,\ell} := \sum_{k=0}^{j-1}\left\{\tau^{k+1}\mathbb{E}\Big[\mathbf{1}_{\{J^*>k\}}\Big] + \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}^k_{\mathbf{X}} \in \boldsymbol{\mathcal{S}}_{\tau^{k+1}}\big\}}\Big]\right\} + \tau^{j+1,\ell}\mathbb{E}\Big[\mathbf{1}_{\{J^*>k\}}\Big] + \mathbb{E}\Big[\mathbf{1}_{\big\{\mathbf{Y}^j_{\mathbf{X}} \in \boldsymbol{\mathcal{S}}_{\tau^{j+1,\ell}}\big\}}\Big],$$

and decide:

(1) If $\boldsymbol{\mathfrak{H}}^{(j,\ell)} > \mathtt{Tol} \cdot \Big(1 + a_j \cdot \mathtt{I}_{j,\ell}\Big)$, set $\tau^{j+1,\ell+1} := \frac{\tau^{j+1,\ell}}{2}$, and $\ell \hookrightarrow \ell + 1$.

(2) If $\frac{\mathtt{Tol}}{2} \cdot \Big(1 + a_j \cdot \mathtt{I}_{j,\ell}\Big) \leq \boldsymbol{\mathfrak{H}}^{(j,\ell)} \leq \mathtt{Tol} \cdot \Big(1 + a_j \cdot \mathtt{I}_{j,\ell}\Big)$, and

a) if $\mathtt{U} = 1$, **STOP**.

b) if $\mathtt{U} = 0$, set $\tau^{j+1} := \tau^{j+1,\ell}$, $t_{j+1} := t_j + \tau^{j+1}$, and $j \hookrightarrow j + 1$.

(3) If $\boldsymbol{\mathfrak{H}}^{(j,\ell)} < \frac{\mathtt{Tol}}{2} \cdot \Big(1 + a_j \cdot \mathtt{I}_{j,\ell}\Big)$, and

a) if $\mathtt{U} = 1$, **STOP**.

b) if $\mathtt{U} = 0$, set $\tau^{j+1} := \tau^{j+1,\ell}$, $t_{j+1} := t_j + \tau^{j+1}$. Then set $\tau^{j+1} := 2\tau^{j+1}$ and $j \hookrightarrow j+1$.

STOP, if $t_j \geq T$ for some j and set $J := j$, $J^* := J$, $\mathbf{Y}_{\mathbf{X}}^{J^*} := \mathbf{Y}_{\mathbf{X}}^{J}$, $Y_Z^{J^*} := Y_Z^J$.

Very similar to Theorem 10.2, the following theorem states termination and convergence (*w.r.t.* '$\mathtt{Tol}$') of Algorithm 10.3. Its proof simplifies due to the *a priori* knowledge of the terminal time $T > 0$, and is thus omitted.

Theorem 10.4. Let $\mathtt{Tol} > 0$. Suppose **(B1)** – **(B3)** in Section 8.2. Fix $(t, \mathbf{x}) \in [0, T) \times \boldsymbol{\mathcal{D}}$. Then, Algorithm 10.3 generates each local step size τ^{j+1} after $\mathcal{O}\Big(\log(\mathtt{Tol}^{-1})\Big)$ many iterations and terminates after $\mathcal{O}\Big(\mathtt{Tol}^{-1}\Big)$ time steps. Furthermore, the 3−tuple $\Big\{(\tau^{j+1}, \mathbf{Y}_{\mathbf{X}}^j, Y_Z^j)\Big\}_{j\geq 0}$ satisfies

$$\Big|u(t, \mathbf{x}) - \mathbb{E}\Big[\phi(t_{J^*}, \mathbf{Y}_{\mathbf{X}}^{J^*}) + Y_Z^{J^*}\Big]\Big| \leq \mathbf{C} \cdot \mathtt{Tol}\,,$$

where $\mathbf{C} > 0$ depends on $\mathfrak{C}(\phi, g) \geq 1$ from Lemma 8.2, and $T > 0$.

11. Monte-Carlo simulations for adaptive weak Euler methods

All simulations are conducted via `MATLAB` (version: 2017a and 2019a).

In the practical implementation of Algorithm 10.1 (and also for Algorithm 10.3), we replace all expectations $\mathbf{E}_{\boldsymbol{\ell}}(\cdot)$, $\boldsymbol{\ell} = 1, ..., 15$ contained in $\left\{\mathfrak{G}_{\ell}^{(\cdot)}\right\}_{\ell=1}^{3}$ (in Algorithm 10.1) by their empirical means $\mathbf{E}_{\boldsymbol{\ell}}^{(\mathtt{M})}(\cdot)$, $\boldsymbol{\ell} = 1, ..., 15$, and write $\left\{\mathfrak{G}_{\ell}^{(\cdot),(\mathtt{M})}\right\}_{\ell=1}^{3}$ for the related (empirical) error estimators. Moreover, in every simulation concerning (0.6), we chose $a_j := \min\Big(1 + L \cdot t_j, 1 + L \cdot C\Big)$ (with $C > 0$ from Lemma 8.4) as *additional* 'weight' in Algorithm 10.1.

As seen in Examples 6.2 and 6.4 (and also in (most of) the examples below), the plots of the step sizes generated via Algorithms 10.1 and 10.3 conceptually have the same structure ('U'-profile). This related dynamics can be classified into three phases (*cf.* Figure 6.2):

- *Initial phase:* In this stage, no stopping dynamics takes place, and step sizes are refined until first samples enter the boundary strips.
- *Bulk phase:* Here, the step sizes attain smallest values and remain constant over a certain period of time. In this period, the majority of samples are projected onto the boundary $\partial\mathcal{D}$, where the smallness of the step sizes guarantee accuracy of those projections.
- *End phase:* In this phase, the leftover samples, *i.e.*, the 'outlier-samples', are gradually forced to be projected onto $\partial\mathcal{D}$ by a gradual coarsening of the step sizes. Besides the guidance of refinement/coarsening through the error estimators, the 'temporal weight' in (10.1) is mainly responsible for this additional dynamics towards the end.

So far, the data requirements of Examples 6.2 and 6.4 coincide with the assumptions in Chapter 8. In the following two sections, we (mostly) consider data setups for simulations of (0.6) and (6.1), which go beyond the scope of our theoretical backup. Even in these cases, where *e.g.* $\mathcal{D}$ has a reentrant corner (see Example 11.2), or boundary and terminal data are incompatible (see Examples 11.4 and 11.5), Algorithms 10.1 and 10.3 still yield promising results.

11.1. Simulations for the elliptic PDE (0.6)

The following example with a quadratic inhomogeneity g and a quartic boundary condition ϕ is taken from [12, Problem 4.2.2b]. The discussion in [9, Subsec. 5.5, Fig. 9] states sub-optimal practical performance of Scheme 2, if run on a fixed mesh; see Figure 11.1 (C) below. In contrast to this, the results in Example 11.1 below show that complementary 'time adaptivity' (see Algorithm 10.1) recovers optimal convergence.

Example 11.1 (see [12, Problem 4.2.2b]). Let $L = 32$ and $\boldsymbol{\mathcal{D}} := \left\{\mathbf{x} \in \mathbb{R}^L \,:\, \|\mathbf{x}\|_{\mathbb{R}^L} < 1\right\}$. Consider (0.6) with $\boldsymbol{\sigma}(\mathbf{x}) = \mathbb{I}$, $\mathbf{b}(\mathbf{x}) \equiv \mathbf{0}$, $c(\mathbf{x}) \equiv 0$ and

$$g(\mathbf{x}) = -\sum_{i=1}^{L} i \cdot x_i^2\,, \qquad \phi(\mathbf{x}) = \frac{1}{6}\sum_{i=1}^{L} i \cdot x_i^4 \qquad (\mathbf{x} \in \boldsymbol{\mathcal{D}})\,.$$

For $\mathtt{M} = 10^5$ fixed, we investigate the convergence rate for the approximation $\mathtt{u}^{(\mathtt{M})}(\mathbf{x})$ of the solution $u(\mathbf{x}) = \phi(\mathbf{x})$ at $\mathbf{x} = ({}^5\!/_{100}, \dots, {}^5\!/_{100})^\top$. Figure 11.1 (B) illustrates optimal order of convergence 1 with respect to `Tol` — as opposed to sub-optimal order $\frac{1}{2}$ (as also found in [9, Subsec. 5.5, Fig. 9]) on uniform-meshes; see Figure 11.1 (C). It seems that time adaptivity in this respect preserves the theoretically stated first order convergence even for 'complicated' functions g and ϕ involved in (0.6).

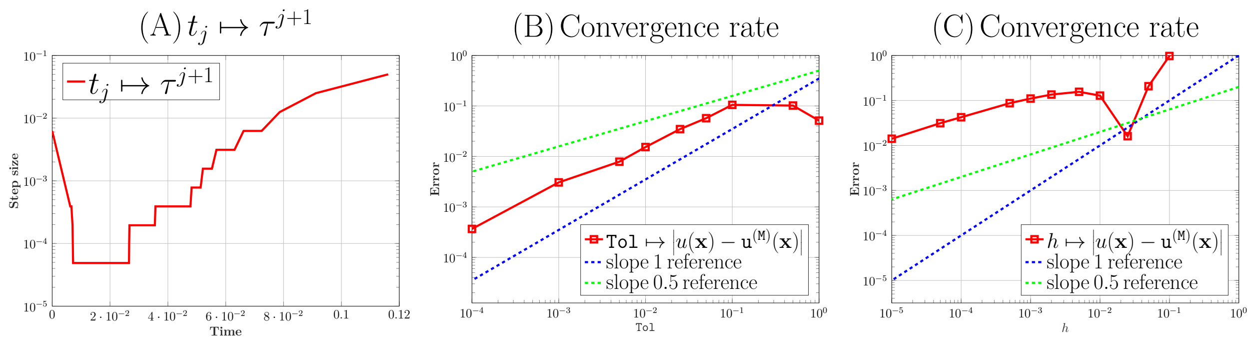


Figure 11.1.: (A) Semi-Log plot of the (adaptive) step sizes generated via Algorithm 10.1 (with `Tol` $= 0.01$). (B) Convergence rate (error) Log-log plot via Algorithm 10.1. (C) Convergence rate (error) Log-log plot via Scheme 2 on uniform meshes with step size h.

The next example from [9, Example 1] considers a data setup for (0.6) in a 3−dimensional, non-convex domain $\boldsymbol{\mathcal{D}}$ with a reentrant corner. Computational studies in [9, Example 1] with the original method [61] on uniform meshes (again) state a suboptimal performance when approximating the solution at a point close to the corner. Although this experimental framework violates our assumptions, we see that Algorithm 10.1 even in this case performs very well, and achieves the desired order of convergence; see Figure 11.2 (D) below.

Example 11.2 (see [9, Example 1]). Let $\boldsymbol{\mathcal{D}} := \Big\{\mathbf{x} = (x_1, x_2, x_3)^\top \in \mathbb{R}^3 : \|\mathbf{x}\|_{\mathbb{R}^3} < 1\Big\} \setminus \Big\{\mathbf{x} = (x_1, x_2, x_3)^\top : x_1 \geq 0.67\,,\ x_2 \geq 0.67\,,\ x_3 \geq 0.67\Big\}$. Consider (0.6) with

$$\boldsymbol{\sigma}(\mathbf{x}) = \begin{bmatrix} \sqrt{1+|x_3|} & 0 & 0 \\ \frac{1}{2}\sqrt{1+|x_1|} & \sqrt{\frac{3}{4}}\sqrt{1+|x_1|} & 0 \\ 0 & \frac{1}{2}\sqrt{1+|x_2|} & \sqrt{\frac{3}{4}}\sqrt{1+|x_2|} \end{bmatrix}, \quad \mathbf{b}(\mathbf{x}) = \begin{bmatrix} x_2 \\ x_3 \\ x_1 \end{bmatrix}, \quad c(\mathbf{x}) \equiv 0\,,$$

$$g(\mathbf{x}) = -\Big(x_2^2 x_3 + x_3^2 x_1 + x_1^2 x_2 + \frac{1}{2}\Big(x_3\sqrt{1+|x_1|}\sqrt{1+|x_3|} + x_1\sqrt{\frac{3}{4}}\sqrt{1+|x_1|}\sqrt{1+|x_2|}\Big)\Big)\,,$$

$$\phi(\mathbf{x}) = x_1 x_2 x_3\,.$$

The corresponding solution is given by $u(\mathbf{x}) = x_1 x_2 x_3$. We fix $\mathbf{x} = (0.57, 0.57, 0.57)^\top$, and use Algorithm 10.1 (with `Tol` $= 0.01$, `M` $= 10^5$) to approximate $u(\mathbf{x})$ (see Figures 11.2 (B) and (C) for resulting features). As stated in [9, Example 1], the suboptimal performance there is due the frequent overshoots of the inner corner of $\boldsymbol{\mathcal{D}}$, where an '*ad-hoc*' value $\lambda_j \equiv 2$ in (8.4) is used within the related method (on uniform meshes) to characterize the boundary strips. In contrast, the (automatic) choice of adaptive step sizes in combination with the flexible choice of λ_j according to (8.4) accurately identifies particular boundary strips, *i.e.*, leads to proper projections onto $\boldsymbol{\partial\mathcal{D}}$, and yields a first order convergent approximation.

Within the framework of the last example in this section, we discuss aspects of Algorithm 10.1 concerning the choice of `M` and `Tol` with respect to the absolute error, when used for approximations at different points at whose neighborhoods the (true) solution does not resp. drastically change. For this reason, we consider a convection dominated example taken from [44, Example 2] (slightly modified here) which has an internal boundary layer; see Figure 11.3.

Example 11.3. Let $\boldsymbol{\mathcal{D}} := (0,1)^2$ and $\varepsilon = 10^{-5}$. Consider (0.6) with $\boldsymbol{\sigma}(\mathbf{x}) \equiv \sqrt{2\varepsilon} \cdot \mathbb{I}$, $\mathbf{b}(\mathbf{x}) = (1,0)^\top$, $c(\mathbf{x}) \equiv -1$. The functions g and ϕ are chosen such that the true solution is given by

$$u(\mathbf{x}) = \frac{1}{2}\Big(1 + \tanh\Big(\frac{x_1 - 0.5}{0.05}\Big)\Big) \qquad \Big(\mathbf{x} = (x_1, x_2)^\top \in \boldsymbol{\mathcal{D}}\Big)\,.$$

We fix $\boldsymbol{x_1} = (0.8, 0.5)^\top \in \boldsymbol{\mathcal{D}}$ and $\boldsymbol{x_2} = (0.8, 0.5)^\top \in \boldsymbol{\mathcal{D}}$. Tables 11.1 and 11.2 display errors in relation to `Tol` and `M` for the approximation of $u(\boldsymbol{x_1})$ resp. $u(\boldsymbol{x_2})$. As we can see in the tables below, the approximation of $u(\boldsymbol{x_2})$ via Algorithm 10.1, *i.e.*, at a point inside the convection dominated area needs very small values for `Tol`, more Monte-Carlo samples `M` and much more time steps to aim for a 'small' corresponding error, as opposed to the approximation of $u(\boldsymbol{x_1})$, where $\boldsymbol{x_1}$ is located outside the internal boundary layer; see Figure 11.3.

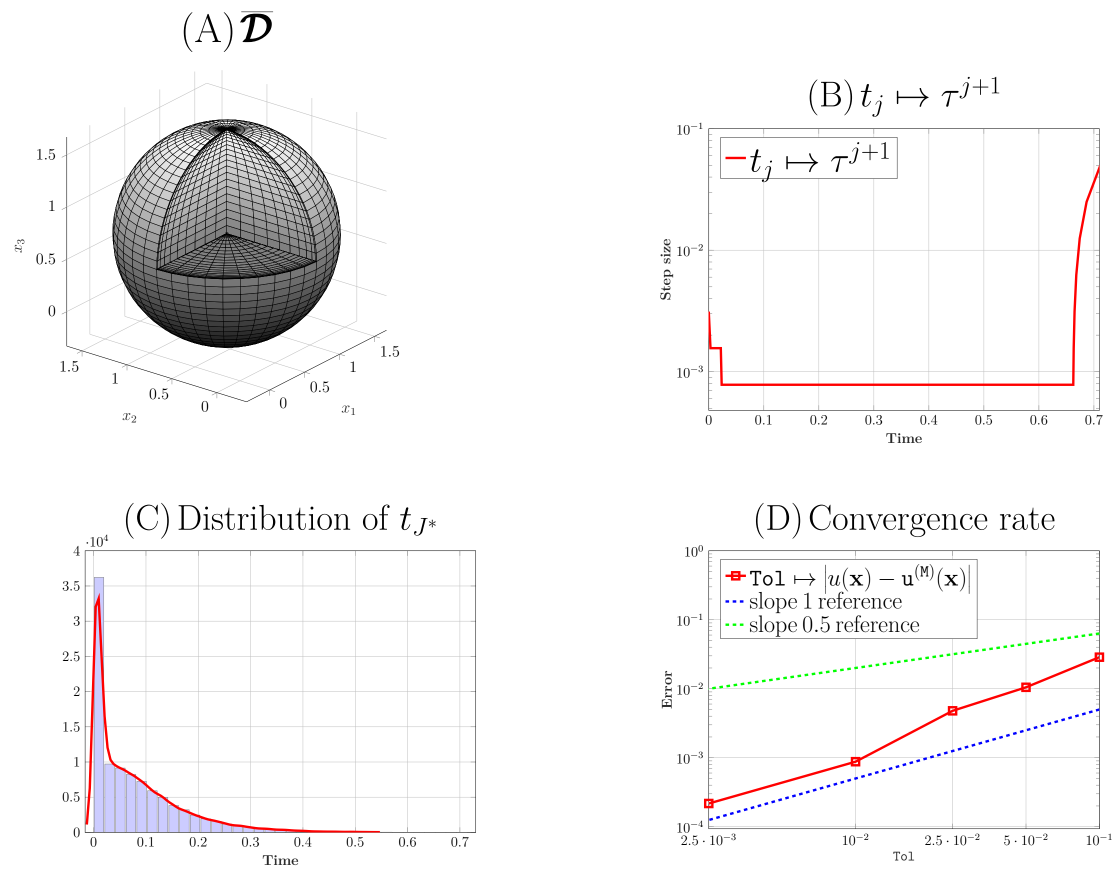


Figure 11.2.: (A) Representation of $\overline{\mathcal{D}}$. (B) Semi-Log plot of the (adaptive) step sizes generated via Algorithm 10.1. (C) Shape of the distribution of t_{J^*} illustrated via a histogram plot. (D) Convergence rate (error) Log-log plot via Algorithm 10.1 ($\mathtt{M} = 5 \cdot 10^5$, $\mathbf{x} = (0.57, 0.57, 0.57)^\top$).

Tol	M	time steps	**Error**
0.1	10^3	8	0.09
0.1	10^5	9	0.084
0.01	10^3	15	0.014
0.01	10^5	16	0.012
0.001	10^2	329	0.0012
0.001	10^4	331	0.001
0.0001	10^3	7439	$1.25 \cdot 10^{-4}$

Table 11.1.: $\boldsymbol{x_1} = (0.8, 0.5)^\top$

Tol	M	time steps	**Error**
0.1	10^3	10	0.89
0.1	10^5	11	0.088
0.01	10^3	54	0.223
0.01	10^5	55	0.219
0.001	10^3	1090	0.023
0.001	10^4	1091	0.023
0.0001	10^3	19598	$8.6 \cdot 10^{-4}$

Table 11.2.: $\boldsymbol{x_2} = (0.5, 0.5)^\top$

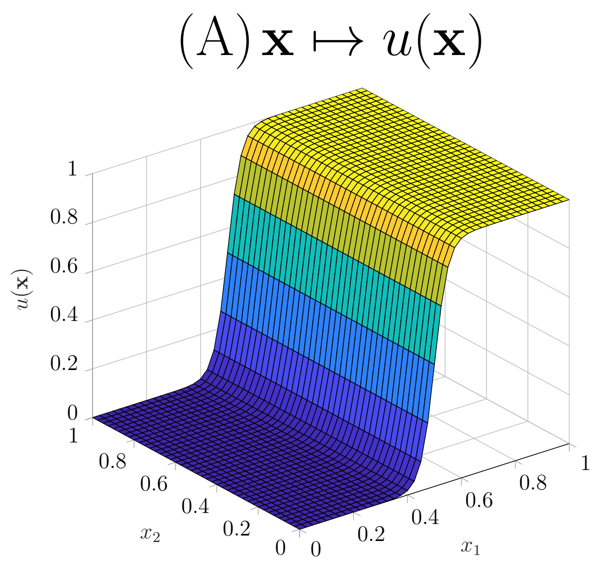


Figure 11.3.: (A) Solution in $\boldsymbol{\mathcal{D}}$.

11.2. Simulations for the parabolic PDE (6.1)

The following example from [21] ('time-reversed' here in our setting) considers incompatible terminal and boundary data in (6.1), and hence violates assumption **(B3)** in Section 8.2. We observe encouraging results with Algorithm 10.3 also in this case, where different structures of step size plots ('L'-profiles) occur, depending on the point in space *and* time at which the solution is approximated; see Figure 11.4 (D) below.

Example 11.4 (see [21, Subsec. 3.2])**.** Let $T = 1$, and $\boldsymbol{\mathcal{D}} := (0,1)^2$. Consider (6.1) with $\boldsymbol{\sigma}(\mathbf{x}) \equiv \sqrt{0.4} \cdot \mathbb{I}$, $\mathbf{b}(\mathbf{x}) \equiv \mathbf{0}$, $g(t,\mathbf{x}) \equiv 0$, and

$$\phi(t,\mathbf{x}) = \begin{cases} \sin\Big(\frac{5\pi}{4}x_1 + \frac{3\pi}{4}\Big)\sin\Big(\frac{5\pi}{4}x_2 + \frac{3\pi}{4}\Big)\,, & t = T \\ 0\,, & t < T\,. \end{cases}$$

For $\mathtt{M} = 10^4$ and $\mathtt{Tol} = 10^{-4}$ fixed, Figure 11.4 below displays step size plots for the approximation of $u(0.98,\mathbf{x})$ at two different values for $\mathbf{x}$: one in the spatio-temporal incompatibility region (*cf.* (B)), and one in the interior of $\boldsymbol{\mathcal{D}}$ (*cf.* (C)). In both step size plots, we observe a 'L'-profile structure — opposed to a 'U'-structure as before, which is due to the temporal dynamics involved here, since 'not enough' samples are projected onto $\boldsymbol{\partial\mathcal{D}}$ within the short time interval $[0.98, 1]$. Consequently, a gradual coarsening, *i.e.*, the '*end phase*' does not occur — opposed to Figures 11.5 (B) and (C), where for $\mathtt{Tol} = 0.01$ fixed, related step size plots are shown at a different time $t = 0$, at which the incompatibility has dissolved. Moreover, the *reversed* 'L'-profile structure of the step size plot in Figure 11.5 (B) results from the position of the (spatial) point (close to $\boldsymbol{\partial\mathcal{D}}$) at which u is approximated, and the choice of $\mathtt{Tol}$: here, samples are already located in the boundary strips at the beginning, which is why the '*initial phase*' does not take place. Furthermore, in order to make up for the incompatibility effect, more time steps, *i.e.*, $\max_{m=1,\dots,\mathtt{M}} J^*(\omega_m) = 154$, are needed in Figure 11.4 (B) as opposed to $\max_{m=1,\dots,\mathtt{M}} J^*(\omega_m) = 42$ in Figure 11.4 (C). On the contrary, at time $t = 0$, more time steps, *i.e.*, $\max_{m=1,\dots,\mathtt{M}} J^*(\omega_m) = 128$, are needed for an approximation at the point in the interior of $\boldsymbol{\mathcal{D}}$ (*cf.* Figure 11.5 (C)) as opposed to $\max_{m=1,\dots,\mathtt{M}} J^*(\omega_m) = 57$ for an approximation at the point close to $\boldsymbol{\partial\mathcal{D}}$ (*cf.* Figure 11.5 (B)).

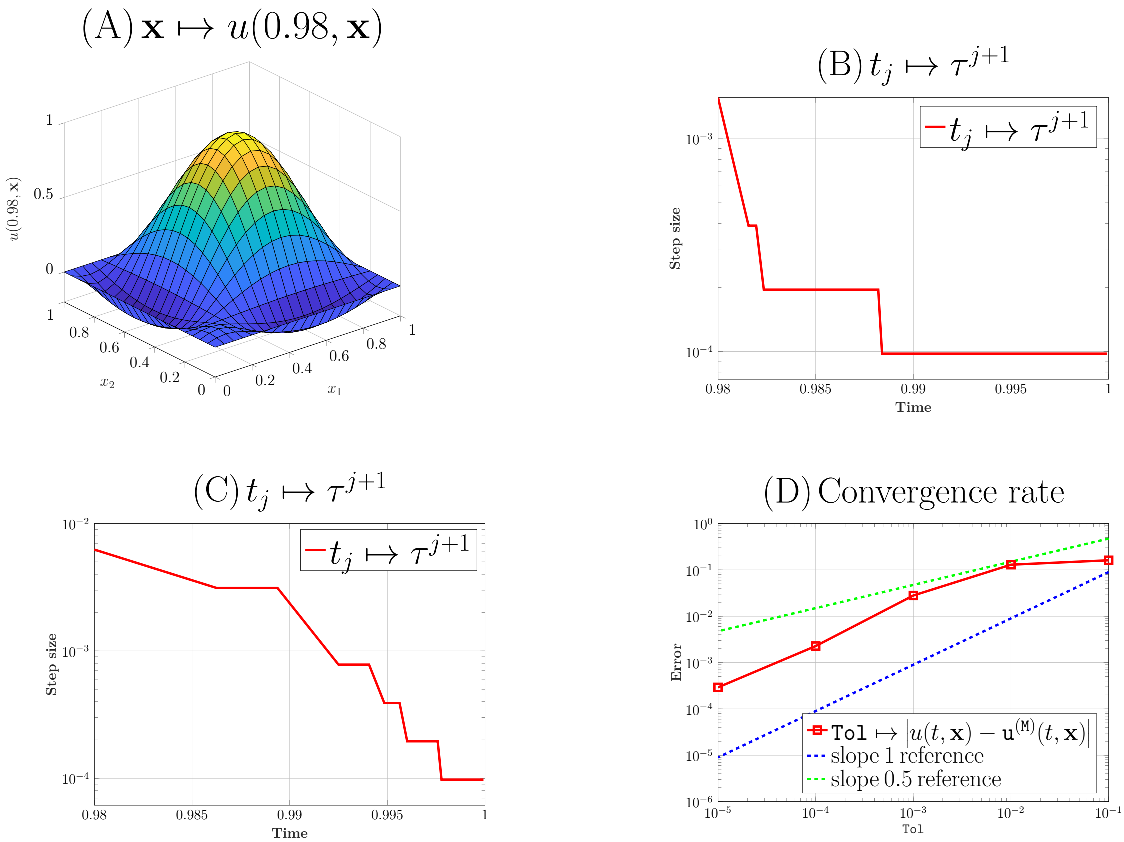


Figure 11.4.: (A) Solution in $\boldsymbol{\mathcal{D}}$ at $t = 0.98$. Semi-Log plots of (adaptive) step sizes for $t = 0.98$ and (B) $\mathbf{x} = (0.05, 0.6)^\top$, (C) $\mathbf{x} = (0.6, 0.6)^\top$. (D) Convergence rate (error) Log-log plot via Algorithm 10.3 ($\mathtt{M} = 5{\cdot}10^5$, $t = 0.98$, $\mathbf{x} = (0.05, 0.6)^\top$).

Example 11.5 below is from [70, Subsec. 6.2] (again, 'time-reversed' here) and similarly to Example 11.4, exhibits an incompatibility of boundary and terminal data at times close to $T > 0$, but in a high-dimensional domain $\boldsymbol{\mathcal{D}}$. Within the framework of this example — also complementing phenomena of Algorithm 10.3 in Example 11.4, we continue the investigation of different aspects of Algorithm 10.3 concerning the choice of M and Tol with respect to the absolute error.

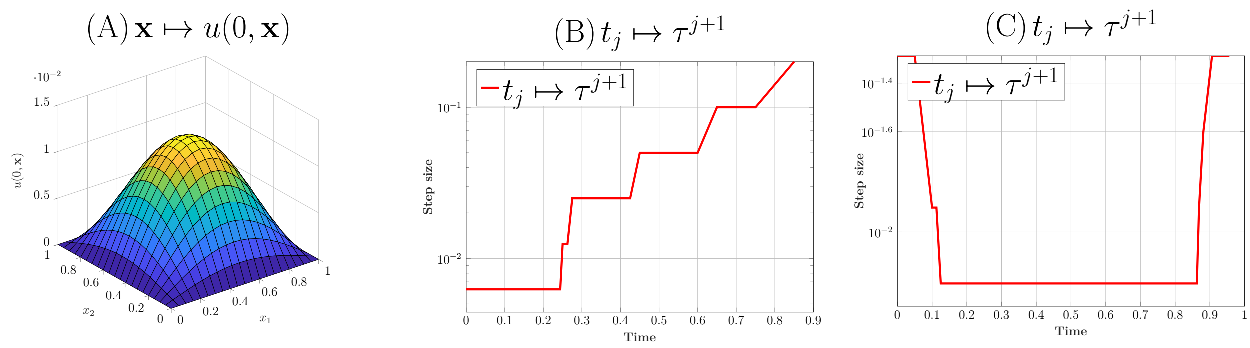


Figure 11.5.: (A) Solution in $\boldsymbol{\mathcal{D}}$ at $t = 0$. Semi-Log plots of (adaptive) step sizes for $t = 0$ and (B) $\mathbf{x} = (0.05, 0.6)^\top$, (C) $\mathbf{x} = (0.6, 0.6)^\top$.

Example 11.5 (see [70, Subsec. 6.2])**.** Let $T = 1$, and $\boldsymbol{\mathcal{D}} := (0,1)^{10}$. Consider (6.1) with $\boldsymbol{\sigma}(\mathbf{x}) \equiv \sqrt{2} \cdot \mathbb{I}$, $\mathbf{b}(\mathbf{x}) \equiv \mathbf{0}$, $g(t, \mathbf{x}) \equiv 0$, and

$$\phi(t, \mathbf{x}) = \begin{cases} 1\,, & t = T \\ 0\,, & t < T\,. \end{cases}$$

We fix $\mathbf{x} = (0.05, 0.5, ..., 0.5)^\top \in \boldsymbol{\mathcal{D}}$. Tables 11.3 – 11.6 display errors in relation to `Tol` and `M` for the approximation of $u(t, \mathbf{x})$ at different times t. As we can see in the tables below, the approximation via Algorithm 10.3 at times close to T, *i.e.*, where the incompatibility effect is present, needs much more Monte-Carlo samples `M` and very small values for `Tol` to aim for a 'small' corresponding error, as opposed to the approximation at times 'away' from T, where the incompatibility effect dissolves.

`Tol`	`M`	time steps	**Error**
0.01	10^2	8	0.16
0.01	10^4	8	0.14
0.001	10^2	53	0.054
0.001	10^4	53	0.053
0.0001	10^2	782	0.053
0.0001	10^4	772	0.014

Table 11.3.: $t = 0.99$

`Tol`	`M`	time steps	**Error**
0.01	10^2	25	0.012
0.01	10^4	33	0.01
0.001	10^2	184	0.01
0.001	10^4	257	0.002
0.0001	10^2	4067	0.002
0.0001	10^4	4061	$7 \cdot 10^{-4}$

Table 11.4.: $t = 0.95$

`Tol`	`M`	time steps	**Error**
0.01	10^2	26	$9 \cdot 10^{-5}$
0.01	10^4	50	$9 \cdot 10^{-5}$
0.001	10^2	217	$9 \cdot 10^{-5}$
0.001	10^5	513	$3 \cdot 10^{-5}$

Table 11.5.: $t = 0.9$

`Tol`	`M`	time steps	**Error**
0.01	10^2	23	$3 \cdot 10^{-13}$
0.01	10^4	50	$3 \cdot 10^{-13}$

Table 11.6.: $t = 0.7$

Bibliography

[1] O. Aboura, *Weak error expansion of the implicit euler scheme*, Preprint (2013).

[2] M. Ali, K. Urban, *HT-AWGM: a hierarchical Tucker-adaptive wavelet Galerkin method for high-dimensional elliptic problems*, Adv. Comput. Math. **46**, (2020).

[3] M. Bachmayr, W. Dahmen, *Adaptive near-optimal rank tensor approximation for high-dimensional operator equations*, Found. Comput. Math. **15**, pp. 839–898 (2015).

[4] M. Bachmayr, W. Dahmen, *Adaptive low-rank methods for problems on Sobolev spaces with error control in L_2*, ESAIM: M2AN **50**, pp. 1107–1136 (2016).

[5] M. Bachmayr, R. Schneider, A. Uschmajew, *Tensor networks and hierarchical tensors for the solution of high-dimensional partial differential equations*, Found. Comput. Math. **16**, pp. 1423–1472 (2016).

[6] S. Bartels, *Numerical approximation of partial differential equations*, Springer (2016).

[7] R. Becker, R. Rannacher, *An optimal control approach to a posteriori error estimation in finite element methods*, Acta Numerica, pp. 1–102 (2001).

[8] L. Belenki, L. Diening, C. Kreuzer, *Optimality of an adaptive finite element method for the p-Laplacian equation*, IMA J. Numer. Anal. **32**, pp. 484–510 (2012).

[9] F. Bernal, J.A. Acebron, *A comparison of higher-order weak numerical schemes for stopped stochastic differential equations*, Commun. Comput. Phys. **20**, pp. 703–732 (2016).

[10] P. Binev, W. Dahmen, R. DeVore, *Adaptive finite element methods with convergence rate*, Numer. Math. **97**, pp. 219–268 (2004).

[11] F.M. Buchmann, *Simulation of stopped diffusions*, J. Comp. Phys. **202**, pp. 446–462 (2005).

[12] F.M. Buchmann, W.P. Petersen, *An exit probability approach to solving high dimensional Dirichlet problems*, SIAM J. Sci. Comput. **28**, pp. 1153–1166 (2006).

[13] F.M. Buchmann, W.P. Petersen, *Solving Dirichlet problems numerically using the Feynman-Kac representation*, BIT **43**, no. 3, 519-540, (2003).

[14] F.M. Buchmann, W.P. Petersen, *Weak approximation of stopped diffusions*, Monte Carlo and quasi-Monte Carlo methods 2004, pp. 31–44, Springer, Berlin, (2006).

[15] H.J. Bungartz, M. Griebel, *Sparse grids*, Acta Numer. **13**, pp. 1–121 (2004).

[16] K. Böttcher, R. Rannacher, *Adaptive error control in solving ordinary differential equations by the discontinuous galerkin method*, Preprint (1996).

[17] C. Carstensen, M. Feischl, M. Page, D. Praetorius, *Axioms of adaptivity*, Comp. Math. Appl. **67**, pp. 1195–1253 (2014).

[18] J.M. Cascon, C. Kreuzer, R.H. Nochetto, K.G. Siebert, *Quasi-optimal convergence rate for an adaptive finite element method*, SIAM J. Numer. Anal. **46**, pp. 2524–2550 (2008).

[19] S. Cerrai, *Second Order PDE's in Finite and Infinite Dimension*, Lecture Notes in Mathematics **1762**, Springer (2001).

[20] Z. Chen, J. Feng, *An adaptive finite element algorithm with reliable and efficient error control for linear parabolic problems*, Math. Comp. **73**, pp. 1167-1193 (2004).

[21] Q. Chen, Z. Qin, R. Temam, *Treatment of incompatible initial and boundary data for parabolic equations in higher dimension*, Math. Comp. **80**, pp. 2071–2096 (2011).

[22] P.-L. Chow, *Stochastic partial differential equations*, Second edition, Advances in Applied Mathematics, CRC Press (2015).

[23] W. Dahmen, R. DeVore, L. Grasedyck, E. Süli, *Tensor-sparsity of solutions to high-dimensional elliptic partial differential equations*, Found. Comput. Math. **16**, pp. 813–874 (2016).

[24] A. Debussche, *Weak approximation of stochastic partial differential equations: the nonlinear case.*, Math. Comp. **80**, pp. 89–117 (2011).

[25] A. Debussche, J. Printems, *Weak order for the discretization of the stochastic heat equation*, Math. Comp. **78**, pp. 845–863 (2009).

[26] L. Diening, C. Kreuzer, *On the threshold condition for Dörfler marking*, Comput. Meth. Appl. Math. **21**, pp. 557–567 (2021).

[27] W. Dörfler, *A convergent adaptive algorithm for Poisson's equation*, SIAM J. Numer. Anal. **33**, pp. 1106–1124 (1996).

[28] A. Dzougoutov, K.-S. Moon, E. von Schwerin, A. Szepessy, R. Tempone, *Adaptive Monte Carlo algorithms for stopped diffusion*, Lect. Notes Comput. Sci. Eng. **44**, Springer (2005).

[29] K. Eriksson, D. Estep, P. Hansbo, C. Johnson, *Introduction to adaptive methods for differential equations*, Acta Numerica, pp. 105–158 (1995).

[30] D. Estep, *A posteriori error bounds and global error control for approximation of ordinary differential equations*, SIAM J. Numer. Anal. **32**, pp. 1–48 (1995).

[31] L.C. Evans, *Partial differential equations*, Graduate Studies in Mathematics **19**, American Mathematical Society (2010).

[32] W. Fang, M.B. Giles, *Adaptive Euler-Maruyama methods for SDEs with non-globally lipschitz drift.*, Ann. Appl. Probab. **30**, pp. 526–560 (2020).

[33] M. Feischl, T. Führer, D. Praetorius, *Adaptive FEM with optimal convergence rates for a certain class of nonsymmetric and possibly nonlinear problems*, SIAM J. Numer. Anal. **52**, pp. 601–625 (2014).

[34] G.D. Gaspoz, K. Siebert, C. Kreuzer, D.A. Ziegler, *A convergent time-space adaptive dG(s) finite element method for parabolic problems by equal error distribution*, IMA J. Numer. Anal. **39**, pp. 650–686 (2019).

[35] D. Gilbarg, N.S. Trudinger, *Elliptic partial differential equations of second order*, Classics in Mathematics, Springer (2001).

[36] M.B. Giles, E. Süli, *Adjoint methods for PDEs: a posteriori error analysis and postprocessing by duality*, Acta Numerica, pp. 145–236 (2002).

[37] E. Gobet. *Monte-Carlo methods and stochastic processes*, CRC Press, Boca Raton (2016).

[38] E. Gobet, *Weak approximation of killed diffusion using Euler schemes*, Stoch. Proc. Appl. **87**, pp. 167–197 (2000).

[39] E. Gobet, *Euler schemes and half-space approximation for the simulation of diffusion in a domain*, ESAIM: Probability and Statistics **5**, pp. 261–297 (2001).

[40] E. Gobet, S. Menozzi, *Stopped diffusion processes: boundary corrections and overshoot*, Stoch. Proc. Appl. **120**, pp. 130–162 (2010).

[41] M. Griebel, H. Harbrecht, *Singular vaue decomposition versus sparse grids: refined complexity estimates*, IMA J. Numer. Anal. **39**, pp. 1652–1671 (2019).

[42] D. Griffiths, D. Higham *Numerical methods for ordinary differential equations*, Springer (2010).

[43] I. Gyöngy, S. Sabanis, D. Siska, *Convergence of tamed Euler schemes for a class of stochastic evolution equations*, Stoch. PDE: Anal. Comp. **4**, pp. 225–245 (2016).

[44] Z. Jannesari, M. Tatari, *An adaptive strategy for solving convection dominated diffusion equation*, Comput. Appl. Math. **39**, (2020).

[45] M. Januszewski, M. Kosutur, *Accelerating numerical solution of stochastic differential equations with CUDA*, `OpenReview.net` (2019).

[46] A. Jentzen, P. Kloeden, *The numerical approximation of stochastic partial differential equations*, Milan J. Math. **77**, pp. 205–244 (2009).

[47] C. Johnson, *Numerical solutions of partial differential equations by the finite element method*, Camebridge Univ. Press (1987).

[48] C. Johnson, *Error estimates and adaptive time-step control for a class of one-step methods for stiff ordinary differential equations*, SIAM J. Numer. Anal. **25**, pp. 908–926 (1988).

[49] I. Karatzas, S. Shreve, *Brownian motion and stochastic calculus*, Graduate Texts in Mathematics, **113**, Springer (1991).

[50] C. Kelly, G.J. Lord, *Adaptive euler methods for stochastic systems with non-globally lipschitz coefficients*, Numer Algor (2021).

[51] C. Kelly, G.J. Lord, *Adaptive time-stepping strategies for nonlinear stochastic systems*, IMA J. Numer. Anal. **38**, pp. 1523–1549 (2018).

[52] P.E. Kloeden, E. Platen, *Numerical solution of stochastic differential equations*, Application of Mathematics **23**, Springer (1992).

[53] C. Kreuzer, C.A. Möller, A. Schmidt, K.G. Siebert, *Design and convergence analysis for an adaptive discretization of the heat equation*, IMA Journal Num. Anal. **32**, pp. 1375–1403 (2012).

[54] O.A. Ladyzenskaja, V.A. Solonnikov, N.N. Ural'ceva, *Linear and quasilinear equations of parabolic type*, American Mathematical Society, Providence, Transl. Math. Monogr. **23**, (1968).

[55] H. Lamba, J.C. Mattingly, A.M. Stuart, *An adaptive Euler-Maruyama scheme for SDEs: convergence and stability*, IMA J. Numer. Anal. **27**, pp. 479–506 (2007).

[56] B.J. Leimkuhler, A. Sharma, M.V. Tretyakov, *Simplest random walk for approximating Robin boundary value problems and ergodic limits of reflected diffusions*, Ann. Appl. Prob. (to appear).

[57] A.K. Majee, A. Prohl, *A posteriori error estimation and space-time adaptivity for a linear stochastic PDE with additive noise*, IMA J. Numer. Anal. (accepted).

[58] F. Merle, A. Prohl, *An adaptive time-stepping method based on a posteriori weak error analysis for large SDE systems*, Numer. Math. **149**, pp. 417-462 (2021).

[59] F. Merle, A. Prohl, *A posteriori error analysis and Adaptivity for high-dimensional elliptic and parabolic boundary value problems*, submitted (2022).

[60] G.N. Milstein, *Weak approximation of a diffusion process in a bounded domain*, Stochastics Stochastics Rep. **62**, pp. 147–200 (1997).

[61] G.N. Milstein, M.V. Tretyakov, *The simplest random walks for the Dirichlet problem*, Theory Probab. Appl. **47**, pp. 53–68 (2003).

[62] G.N. Milstein, M.V. Tretyakov, *Stochastic numerics for mathematical physics*, Scientific Computation, Springer (2004).

[63] M.L. Delle Monache, L. Thibault, B. Piccoli, R. Stern, D. Work, *Traffic reconstruction using autonomous vehicles*, SIAM J. Appl. Math. **79**, pp. 1748–1767 (2019).

[64] K.-S. Moon, A. Szepessy, R. Tempone, and G.E. Zouraris, *Convergence rates for adaptive weak approximation of stochastic differential equations*, Stoch. Anal. Appl. **23**, pp. 511–558 (2005).

[65] P. Morin, R.H. Nochetto, K.G. Siebert, *Convergence of adaptive finite element methods*, SIAM Review **44**, pp. 631–658 (2002).

[66] R.H. Nochetto, K.G. Siebert, A. Veeser, *Theory of adaptive finite element methods: an introduction.* in: *Multiscale, nonlinear, and adaptive approximation*, ed. by R. DeVore, A. Kunoth, pp. 409–542, Springer, Berlin (2009).

[67] R.H. Nochetto, G. Savare, C. Verdi, *A posteriori error estimates for variable time-step discretizations of nonlinear evolution equations*, Comm. Pure Appl. Math. **LIII**, pp. 525–589 (2000).

[68] D. Nualart, *The Malliavin calculus and related topics*, Probability and its Applications, Springer (2006).

[69] B. Oksendal, *Stochastic differential equations*, Sixth edition, Springer (2013).

[70] T. v. Petersdorff, C. Schwab, *Numerical solution of parabolic equations in high dimensions*, ESAIM: M2AN **38**, pp. 93–127 (2004).

[71] D. Pflüger, *Spatially adaptive sparse grids for high-dimensional problems*, Verlag Dr. Hut, München, Ph.D. thesis (2010).

[72] A. Prohl, C. Schellnegger *Adaptive concepts for stochastic partial differential equations*, J. Sci. Comput. **80**, pp. 444–474 (2019).

[73] A. Quarteroni, A. Valli, *Numerical approximation of partial differential equations*, Springer (1994).

[74] R. Schneider, A. Uschmajew, *Approximation rates for the hierarchical tensor format in periodic Sobolev spaces*, J. Complex. **30**, pp. 56–71 (2014).

[75] R. Stevenson, *Optimality of a standard adaptive finite element method*, Found. Comput. Math. **7**, pp. 245–269 (2007).

[76] A. Szepessy, R. Tempone, and G.E. Zouraris, *Adaptive weak approximation of stochastic differential equations*, Comm. Pure Appl. Math. **54**, pp. 1169–1214 (2001).

[77] D. Talay, L. Tubaro, *Expansion of the global error for numerical schemes solving stochastic differential equations*, Stochastic Anal. Appl. **8**, pp. 483–509 (1990).

[78] L.N. Trefethen, *Cubature, Approximation, and isotropy in the hyperspace*, SIAM Review **59**, pp. 469–491 (2017).

[79] R. Verfürth, *A posteriori error estimates for finite element discretizations of the heat equation*, Calcolo **40**, pp. 195–212 (2003).

[80] R. Verfürth, *A posteriori error estimation techniques for finite element methods*, Oxford University Press (2013).

[81] T. Witelski, M. Bowen, *Methods of mathematical modelling*, Springer (2015).

[82] Z. Zhang, G. Karniadakis, *Numerical methods for stochastic partial differential equations with white noise*, Springer (2017).