Foundations of Computational Mathematics (2019) 19:1387–1430
https://doi.org/10.1007/s10208-018-09412-w

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS
The Journal of the Society for the Foundations of Computational Mathematics

# On a Randomized Backward Euler Method for Nonlinear Evolution Equations with Time-Irregular Coefficients

**Monika Eisenmann**[1] · **Mihály Kovács**[2] · **Raphael Kruse**[1] · **Stig Larsson**[2]

Received: 7 September 2017 / Revised: 13 November 2018 / Accepted: 20 November 2018 /
Published online: 3 January 2019


## Abstract

In this paper, we introduce a randomized version of the backward Euler method that is applicable to stiff ordinary differential equations and nonlinear evolution equations with time-irregular coefficients. In the finite-dimensional case, we consider Carathéodory-type functions satisfying a one-sided Lipschitz condition. After investigating the well-posedness and the stability properties of the randomized scheme, we prove the convergence to the exact solution with a rate of 0.5 in the root-mean-square norm assuming only that the coefficient function is square integrable with respect to the temporal parameter. These results are then extended to the approximation of infinite-dimensional evolution equations under monotonicity and Lipschitz conditions. Here, we consider a combination of the randomized backward Euler scheme with a Galerkin finite element method. We obtain error estimates that correspond to the regularity of the exact solution. The practicability of the randomized scheme is also illustrated through several numerical experiments.

**Keywords** Monte Carlo method · Evolution equations · Ordinary differential equations · Backward Euler method · Galerkin finite element method

Communicated by Arieh Iserles.

✉ Raphael Kruse
kruse@math.tu-berlin.de

Monika Eisenmann
meisenma@math.tu-berlin.de

Mihály Kovács
mihaly@chalmers.se

Stig Larsson
stig@chalmers.se

1 Institut für Mathematik, Technische Universität Berlin, Secr. MA 5-3, Straße des 17. Juni 136, 10623 Berlin, Germany

2 Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, 412 96 Gothenburg, Sweden

**Mathematics Subject Classification** 65C05 · 65L05 · 65L20 · 65M12 · 65M60

## 1 Introduction

The aim of this paper is to introduce a new numerical scheme to approximate the solution of an ordinary differential equation (ODE) of Carathéodory type

$$\begin{cases} \dot{u}(t) = f(t, u(t)), & \text{for almost all } t \in (0, T], \\ u(0) = u_0, \end{cases} \tag{1.1}$$

for $T \in (0, \infty)$, and of a non-autonomous evolution equation

$$\begin{cases} \dot{u}(t) + \mathcal{A}(t)u(t) = f(t), & \text{for almost all } t \in (0, T], \\ u(0) = u_0, \end{cases} \tag{1.2}$$

where $\mathcal{A}\colon [0, T] \times V \to V^*$ is a strongly monotone and Lipschitz continuous operator with respect to the second argument that is defined on a Gelfand triple $V \hookrightarrow H \cong H^* \hookrightarrow V^*$ for real Hilbert spaces $V$ and $H$.

We focus on the particular difficulty that the mappings $f$ and $\mathcal{A}$ are irregular with respect to the temporal parameter. More precisely, we do not impose any continuity conditions but only certain integrability requirements with respect to $t$. For a concise description of the general settings, we refer to Sects. 3 and 6, respectively. In particular, a precise statement of all conditions is given in Assumption 3.1 for (1.1) and in Assumption 6.1 for (1.2). To develop the idea of our scheme, we mostly focus on the ODE problem (1.1) in this introduction. The derivation of the numerical scheme for the evolution equation (1.2) follows analogously and will be introduced in detail in Sect. 6.

When considering a right-hand side $f$ that is only integrable, every deterministic algorithm can be "fooled" if it only uses information provided by point evaluations on prescribed (deterministic) points. One can easily construct suitable fooling functions for general classes of deterministic algorithms, for instance, based on adaptive strategies. In Sects. 5 and 8, we give examples of such fooling functions and investigate the numerical behavior. Further, we refer to the vast literature on the information-based complexity theory (IBC), which applies similar techniques to derive lower bounds for the error of certain classes of deterministic and randomized numerical algorithms. For instance, see [34,40] for a general introduction into IBC and [24,27,28] for applications to the numerical solution of initial value problems.

One way to construct numerical methods for the solution of initial value problems with time-irregular coefficients consists of allowing the algorithm to use additional information of the right-hand side $f$ as, for example, integrals of the form

$$f^n(x) := \frac{1}{t_n - t_{n-1}} \int_{t_{n-1}}^{t_n} f(s, x)\, \mathrm{d}s, \quad \text{for } x \in \mathbf{R}^d. \tag{1.3}$$

This approach is often found in the existence theory of ODEs and PDEs when a numerical method is used to construct analytical solutions to the initial value problems (1.1) and (1.2) under minimal regularity assumptions. The complexity of such methods has also been studied in [27] (and the references therein) for the numerical solution of ODEs. It is also the state-of-the-art method in many recent papers for the numerical solution of evolution equations of the form (1.2). For example, we refer to [5,15,25,33].

However, it is rarely discussed how a quantity such as $f^n(x)$ in (1.3) is obtained in practice. Strictly speaking, since the computation of $f^n(x)$ often requires the application of further numerical methods such as quadrature rules, algorithms relying on integrals such as (1.3) are, in general, not fully discrete solvers yet. More importantly, classical quadrature rules for the approximation of $f^n(x)$ are again based on deterministic point evaluations of $f$ and may therefore be "fooled".

Instead of using linear functionals such as (1.3), we propose the following *randomized* version of the backward Euler method. For $N \in \mathbf{N}$, a step size $k = \frac{T}{N}$, and a temporal grid $0 = t_0 < t_1 < \cdots < t_N = T$ with $t_n = nk$ for $n \in \{0, \dots, N\}$, the randomized scheme for the numerical solution (1.1) is then given by

$$\begin{cases} U^n = U^{n-1} + kf(\xi_n, U^n), & \text{for } n \in \{1, \dots, N\}, \\ U^0 = u_0, \end{cases} \tag{1.4}$$

where $\xi_n$ is a uniformly distributed random variable with values in the interval $[t_{n-1}, t_n]$. Note that we evaluate the right-hand side at random points between the grid points. Since the evaluation points vary every time the algorithm is called, it is not possible to construct a fooling function as described above.

We will prove in Theorem 4.7 that the numerical solution $U^n$ from (1.4) converges with (strong) order $\frac{1}{2}$ to the exact solution $u$ of (1.1), even if $f$ is only square integrable with respect to time. Due to the results in [24], this convergence rate is *optimal* in the sense that there exists no deterministic or randomized algorithm based on finitely many point evaluations of $f$ with a higher convergence rate within the class of all initial value problems satisfying Assumption 3.1.

The error analysis is based on the observation that the randomized scheme (1.4) is a hybrid of an implicit Runge–Kutta method and a Monte Carlo quadrature rule. In fact, if the ODE (1.1) is actually autonomous, that is, $f$ does not depend on $t$, then we recover the classical backward Euler method. On the other hand, if $f$ is independent of the state variable $u$, then the ODE (1.1) reduces to an integration problem and the randomized scheme (1.4) is the randomized Riemann sum for the approximation of $u_0 + \int_0^{t_n} f(s)\,\mathrm{d}s$ given by

$$U^n = u_0 + k \sum_{j=1}^{n} f(\xi_j), \quad \text{for } n \in \{1, \dots, N\}.$$

Observe that a randomized Riemann sum is a particular case of stratified sampling from Monte Carlo integration. It was introduced in [19,20] together with further, higher-order, quadrature rules. Our error analysis of the randomized scheme (1.4) combines techniques for the analysis of both time-stepping schemes and Monte Carlo integration. In particular, since we are interested in the discretization of evolution equations in later sections, we apply techniques for the numerical analysis of stiff ODEs developed in [22] and for stochastic ODEs in [3].

Before we give a more detailed account of the remainder of this paper, let us emphasize a few practical advantages of the randomized scheme (1.4):

1. The implementation of the randomized scheme (1.4) is as difficult as for the classical backward Euler method in terms of the requirements of solving a nonlinear system of equations. On the other hand, the scheme (1.4) does not require integrals such as $f^n(x)$ if the right-hand side is time-irregular.
2. The same is true for the computational effort. Compared to the classical backward Euler method, the randomized scheme (1.4) only requires in each step the additional simulation of a single scalar-valued random variable. In general, the resulting additional computational effort is negligible compared to the solution of a potentially high-dimensional nonlinear system of equations. More importantly, due to the randomization we avoid the potentially costly computation of the integrals $f^n(x)$.
3. In contrast to every deterministic method based on point evaluations of $f$, the randomized scheme (1.4) is independent of the particular representation of an integrable function. To be more precise, let $g_1$ and $g_2$ be two representations of the same equivalence class $g \in L^2(0, T)$. Then, it follows that $g_1(\xi_n) = g_2(\xi_n)$ with probability one, since $g_1 = g_2$ almost everywhere.

We remark that the last item is only valid as long as the random variable $\xi_n$ is indeed uniformly distributed in $[t_{n-1}, t_n]$. In practice, however, one usually applies a pseudo-random number generator which only draws values from the set of floating point numbers. Since this is a null set with respect to the Lebesgue measure, the argument given above is no longer valid. Of course, this problem affects any algorithm that uses the floating point arithmetic. Nevertheless, a randomized algorithm is often more robust regarding the particular choice of the representation of an equivalence class in $L^2(0, T)$ and, hence, more user-friendly. For instance, the mapping $(0, T) \ni t \mapsto (T - t)^{-\frac{1}{3}}$ causes problems for the classical backward Euler method as it will evaluate the mapping in the singularity at $t = T$. This problem does not occur for the randomized backward Euler method with probability one.

Let us also mention that randomized algorithms for the numerical solution of initial value problems have already been studied in the literature. In the ODE case, the complexity and optimality of such algorithms is considered in [12,24,28] under various degrees of smoothness of $f$. The time-irregular case studied in the present paper was first investigated in [38,39]. See also [26,30] for a more recent exposition of explicit randomized schemes.

The present paper extends the earlier results in several directions. In order to deal with possibly stiff ODEs, we consider a randomized version of the backward Euler method and prove its well-posedness and stability under a one-sided Lipschitz condition. In addition, we require only local Lipschitz conditions with respect to the state variable in order to obtain estimates on the local truncation error, thereby extending results from [30]. We also avoid any (local) boundedness condition on $f$ as, for example, in [12,26].

The stability properties also qualify this new randomized backward Euler method as a suitable temporal integrator for non-autonomous evolution equations with time-irregular coefficients. To the best of our knowledge, there is no work found in the literature that applies a randomized algorithm to the numerical solution of evolution equations of the form (1.2). Instead, the standard approach in the time-irregular case relies on the availability of suitable integrals of the right-hand side as in (1.3). In particular, we mention [15,25]. Further results on optimal rates under minimal regularity assumptions for linear parabolic PDEs can be found, e.g., in [5,9,21]. For semilinear parabolic problems, optimal error estimates are also found in [33], where a discontinuous Galerkin method in time and space is considered.

This paper is organized as follows. In Sect. 2, we shortly introduce the notation and recapture some important concepts of stochastic analysis that are relevant for this paper. In the following Sect. 3, we state the assumptions imposed on the ODE (1.1). We also discuss existence and regularity of the solution. In Sect. 4, we then prove the well-posedness and convergence of the randomized backward Euler method in the root-mean-square sense. The ODE part of this paper is completed in Sect. 5 by examining a numerical example.

In Sect. 6, we introduce the setting for the irregular non-autonomous evolution equation (1.2) that we consider in the second part of this paper. Under some additional regularity assumptions on the exact solution, we prove the convergence of a fully discrete method that combines the randomized backward Euler scheme with a Galerkin finite element method. The additional regularity assumption is then discussed in more detail in Sect. 7. In particular, it is shown that the regularity condition is fulfilled for rather general classes of linear and semilinear parabolic PDEs. Finally, in Sect. 8, we demonstrate that this new randomized method can be applied to evolution equations. To this end, we present a numerical example which is based on the finite element software package FEniCS [31].

## 2 Preliminaries

In this section, we explain the necessary tools from probability theory and recall some important inequalities that are needed. First, we start by fixing the notation used in this paper.

We denote the set of all positive integers by $\mathbf{N}$ and the set of all real numbers by $\mathbf{R}$. In $\mathbf{R}^d$, $d \geq 1$, we denote the Euclidean norm by $|\cdot|$ which coincides with the absolute value of a real number for $d = 1$. The standard inner product in $\mathbf{R}^d$ is denoted by $(\cdot, \cdot)$. For a ball of radius $r$ with center $x \in \mathbf{R}^d$, we write $B_r(x) \subseteq \mathbf{R}^d$.

In the following, we will consider different spaces of functions with values in general Hilbert spaces. To this end, let $(H, (\cdot,\cdot)_H, \|\cdot\|_H)$ be a real Hilbert space and $T > 0$. We will denote the space of continuous functions on $[0, T]$ with values in $H$ by $C([0, T]; H)$ where the norm is given by

$$\|f\|_{C([0,T];H)} = \sup_{t\in[0,T]} \|f(t)\|_H.$$

It will also be important to consider functions which are a little more regular. For $0 < \gamma < 1$, we denote the space of Hölder continuous functions by $C^{\gamma}([0,T];H)$ with norm given by

$$\|f\|_{C^{\gamma}([0,T];H)} = \sup_{t\in[0,T]} \|f(t)\|_H + \sup_{\substack{s,t\in[0,T]\\ s\neq t}} \frac{\|f(s)-f(t)\|_H}{|s-t|^{\gamma}}.$$

For $p \in [1,\infty)$, we introduce the Bochner–Lebesgue space

$$\begin{aligned} &L^p(0,T;H)\\ &\quad = \big\{u : [0,T] \to H : u \text{ is strongly measurable and } \|u\|_{L^p(0,T;H)} < \infty\big\} \end{aligned}$$

where the norm is given by

$$\|u\|_{L^p(0,T;H)}^p = \int_0^T \|u(t)\|_H^p \,\mathrm{d}t.$$

In the case $H = \mathbf{R}$, we write $L^p(0,T)$.

The space of linear bounded operators from $H$ to a Banach space $(U, \|\cdot\|_U)$ is denoted by $\mathcal{L}(H,U)$ and in the case of $U = H$ we write $\mathcal{L}(H)$. The norm of this space is the usual operator norm given by

$$\|A\|_{\mathcal{L}(H,U)} = \sup_{v\in H, \|v\|_H=1} \|Av\|_U.$$

Since we are interested in a randomized scheme, we will briefly recall the most important probabilistic concepts needed in this paper. To this end, we consider a probability space $(\Omega,\mathcal{F},\mathbf{P})$ which consists of a measurable space $(\Omega,\mathcal{F})$ together with a finite measure $\mathbf{P}$ such that $\mathbf{P}(A) \in [0,1]$ for every $A \in \mathcal{F}$ and $\mathbf{P}(\Omega) = 1$. A mapping $X\colon \Omega \to H$ is called a random variable if it is measurable with respect to the $\sigma$-algebra $\mathcal{F}$ and the Borel $\sigma$-algebra $\mathcal{B}(H)$ in $H$, i.e., for every $B \in \mathcal{B}(H)$

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$$

is an element of $\mathcal{F}$. The integral of a random variable $X$ with respect to the measure $\mathbf{P}$ is often denoted by

$$\mathbf{E}[X] = \int_{\Omega} X(\omega)\,\mathrm{d}\mathbf{P}(\omega).$$

The space of $\mathcal{F}$-measurable random variables $X$ such that $\mathbf{E}[\|X\|_H]$ is finite is denoted by $L^1(\Omega,\mathcal{F},\mathbf{P};H)$.

For our purposes, it is important to consider the space $L^2(\Omega,\mathcal{F},\mathbf{P};H)$ of square-integrable $\mathcal{F}$-measurable random variables. This space is often abbreviated by $L^2(\Omega; H)$ if it is clear from the context which $\sigma$-algebra $\mathcal{F}$ and measure $\mathbf{P}$ is used. The space is endowed with the norm

$$\|X\|^2_{L^2(\Omega;H)} = \int_\Omega \|X(\omega)\|^2_H \,\mathrm{d}\mathbf{P}(\omega) = \mathbf{E}[\|X\|^2_H], \quad X \in L^2(\Omega; H).$$

Equipped with this norm and inner product

$$(X_1, X_2)_{L^2(\Omega;H)} = \int_\Omega (X_1(\omega), X_2(\omega))_H \,\mathrm{d}\mathbf{P}(\omega), \quad X_1, X_2 \in L^2(\Omega; H),$$

the space $L^2(\Omega; H)$ is a Hilbert space.

A further important concept is the independence of events $(A_n)_{n\in\mathbf{N}} \subset \mathcal{F}$. We call the events $(A_n)_{n\in\mathbf{N}}$ independent if for every finite subset $I \subset \mathbf{N}$

$$\mathbf{P}\Big(\bigcap_{n\in I} A_n\Big) = \prod_{n\in I} \mathbf{P}(A_n)$$

holds. This concept can be transferred to families $(\mathcal{F}_n)_{n\in\mathbf{N}}$ of $\sigma$-algebras. Such a family is called independent if for every finite subset $I \subset \mathbf{N}$ it follows that every choice of events $(A_n)_{n\in I}$ with $A_n \in \mathcal{F}_n$ are independent. Similarly, a family of $H$-valued random variables $(X_n)_{n\in\mathbf{N}}$ is called independent if the generated $\sigma$-algebras

$$\sigma(X_n) = \big\{X_n^{-1}(B) : B \in \mathcal{B}(H)\big\}$$

are independent.

A family $(\mathcal{F}_n)_{n\in\mathbf{N}}$ of $\sigma$-algebras is called a filtration if for every $n \in \mathbf{N}$ the $\sigma$-algebra $\mathcal{F}_n$ is a subset of $\mathcal{F}$ and $\mathcal{F}_n \subset \mathcal{F}_m$ holds for $n \le m$. Thus a random variable $X$ can be measurable with respect to $\mathcal{F}_m$ but not with respect to $\mathcal{F}_n$ for $n < m$. In some of the arguments in this paper it will be important to project an $\mathcal{F}_m$-measurable random variable to a smaller $\sigma$-algebra $\mathcal{F}_n$. To this end, we introduce the conditional expectation of $X$ with respect to $\mathcal{F}_n$: For a random variable $X \in L^1(\Omega, \mathcal{F}_m, \mathbf{P}; H)$, we introduce the $\mathcal{F}_n$-measurable random variable $\mathbf{E}[X|\mathcal{F}_n] \colon \Omega \to H$ which fulfills

$$\mathbf{E}[X\mathbb{1}_A] = \mathbf{E}[\mathbf{E}[X|\mathcal{F}_n]\mathbb{1}_A]$$

for every $A \in \mathcal{F}_m$ where $\mathbb{1}_A$ is the characteristic function with respect to $A$. The random variable $\mathbf{E}[X|\mathcal{F}_n]$ is uniquely determined by these postulations. An important property of the conditional expectation of $X \in L^1(\Omega, \mathcal{F}, \mathbf{P}; H)$ is the tower property which states that for two $\sigma$-algebras $\mathcal{F}_n$ and $\mathcal{F}_m$ of the filtration $(\mathcal{F}_n)_{n\in\mathbf{N}}$ with $\mathcal{F}_n \subseteq \mathcal{F}_m$ we obtain that

$$\mathbf{E}[\mathbf{E}[X|\mathcal{F}_n]|\mathcal{F}_m] = \mathbf{E}[\mathbf{E}[X|\mathcal{F}_m]|\mathcal{F}_n] = \mathbf{E}[X|\mathcal{F}_n].$$

In particular, if $X$ is already measurable with respect to $\mathcal{F}_n$ then $\mathbf{E}[X|\mathcal{F}_n] = X$ holds. If $\sigma(X)$ is independent of $\mathcal{F}_n$, we obtain that $\mathbf{E}[X|\mathcal{F}_n] = \mathbf{E}[X]$.

In the course of this paper, we will often use random variables which are uniformly distributed on a given temporal interval $(a, b)$. To denote such a random variable $\tau \colon \Omega \to \mathbf{R}$, we write $\tau \sim \mathcal{U}(a, b)$.

For a deeper insight of the probabilistic background, we refer the reader to [29].

The following inequalities will be helpful in order to give suitable a priori bounds for the solution of a differential equation and the solution of a numerical scheme.

**Lemma 2.1** (Discrete Gronwall lemma) *Let $(u_n)_{n\in\mathbf{N}}$ and $(b_n)_{n\in\mathbf{N}}$ be two nonnegative sequences which satisfy, for given $a \in [0, \infty)$ and $N \in \mathbf{N}$, that*

$$u_n \le a + \sum_{j=1}^{n-1} b_j u_j, \quad \textit{for all } n \in \{1, \dots, N\}.$$

*Then, it follows that*

$$u_n \le a \exp\Big(\sum_{j=1}^{n-1} b_j\Big), \quad \textit{for all } n \in \{1, \dots, N\},$$

*where we use the convention $\sum_{j=1}^{0} b_j = 0$.*

**Lemma 2.2** (Gronwall lemma) *If $u, a \in C([0, T])$ are nonnegative functions which satisfy, for given $b \in [0, \infty)$, that*

$$u(t) \le a(t) + b \int_0^t u(s)\, \mathrm{d}s, \quad \textit{for every } t \in [0, T],$$

*then*

$$u(t) \le \mathrm{e}^{bt} \max_{s\in[0,t]} a(s), \quad \textit{for every } t \in [0, T].$$

For a proof of the discrete Gronwall lemma, we refer the reader to [10]. A proof of Lemma 2.2 can be found in [23].

## 3 A Carathéodory-Type ODE Under a One-Sided Lipschitz Condition

In this section, we introduce an initial value problem involving an ordinary differential equation with a non-autonomous vector field of Carathéodory type that satisfies a one-sided Lipschitz condition. We give a precise statement of all conditions on the coefficient function in Assumption 3.1, which are sufficient to ensure the existence of a unique global solution. The same conditions will also be used for the error analysis of the randomized backward Euler method in Sect. 4. Further, we briefly investigate the temporal regularity of the solution $u$.

Let $T \in (0, \infty)$. We are interested in finding an absolutely continuous mapping $u\colon [0, T] \to \mathbf{R}^d$ that is a solution to the initial value problem

$$\begin{cases} \dot{u}(t) = f(t, u(t)), & \text{for almost all } t \in (0, T], \\ u(0) = u_0, \end{cases} \tag{3.1}$$

where $u_0 \in \mathbf{R}^d$ denotes the initial value. The following conditions on the right-hand side $f\colon [0, T] \times \mathbf{R}^d \to \mathbf{R}^d$ will ensure the existence of a unique global solution.

**Assumption 3.1** The mapping $f\colon [0, T] \times \mathbf{R}^d \to \mathbf{R}^d$ is measurable. Moreover, there exists a null set $\mathcal{N}_f \in \mathcal{B}([0, T])$ such that:

(i) There exists $\nu \in [0, \infty)$ such that

$$\big(f(t, x) - f(t, y), x - y\big) \le \nu |x - y|^2,$$

for all $x, y \in \mathbf{R}^d$ and $t \in [0, T] \setminus \mathcal{N}_f$.

(ii) There exists a mapping $g\colon [0, T] \to [0, \infty)$ with $g \in L^2(0, T; \mathbf{R})$ such that

$$|f(t, 0)| \le g(t), \quad \text{for all } t \in [0, T] \setminus \mathcal{N}_f.$$

(iii) For every compact set $K \subset \mathbf{R}^d$, there exists a mapping $L_K\colon [0, T] \to [0, \infty)$ with $L_K \in L^2(0, T; \mathbf{R})$ such that

$$|f(t, x) - f(t, y)| \le L_K(t) |x - y|$$

for all $x, y \in K$ and $t \in [0, T] \setminus \mathcal{N}_f$.

First, we note that from Assumption 3.1 (i) and (ii) we immediately get

$$\begin{aligned} \big(f(t, x), x\big) &= \big(f(t, x) - f(t, 0), x - 0\big) + \big(f(t, 0), x\big) \\ &\le \nu |x|^2 + g(t) |x| \end{aligned} \tag{3.2}$$

for all $x \in \mathbf{R}^d$ and $t \in [0, T] \setminus \mathcal{N}_f$.

Moreover, it is well known that Assumption 3.1 (ii) and (iii) are sufficient to ensure the existence of a unique local solution $u\colon [0, T_0) \to \mathbf{R}^d$ to the initial value problem (3.1) with a local existence time $T_0 \le T$ (see, for instance, [23, Chap. I, Thm 5.3]). Here, we recall that a mapping $u\colon [0, T_0) \to \mathbf{R}^d$ is a (local) *solution in the sense of Carathéodory* to (3.1) if $u$ is absolutely continuous and satisfies

$$u(t) = u_0 + \int_0^t f(s, u(s)) \, \mathrm{d}s \tag{3.3}$$

for all $t \in [0, T_0)$. Moreover, for almost all $t \in [0, T_0)$ with $|u(t)| > 0$ we have

$$|u(t)| \frac{\mathrm{d}}{\mathrm{d}t} |u(t)| = \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} |u(t)|^2 = \big(f(t, u(t)), u(t)\big) \le \nu |u(t)|^2 + g(t) |u(t)|,$$

due to (3.2). Hence, by canceling $|u(t)| > 0$ from both sides of the inequality we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}|u(t)| \le \nu|u(t)| + g(t)$$

for almost all $t \in [0, T_0)$ with $|u(t)| > 0$. After integrating this inequality from 0 to $t$, it follows

$$|u(t)| \le |u_0| + \int_0^t g(s)\,\mathrm{d}s + \int_0^t \nu|u(s)|\,\mathrm{d}s,$$

which holds for all $t \in [0, T_0)$. An application of the Gronwall lemma (see Lemma 2.2) yields

$$|u(t)| \le \mathrm{e}^{\nu t}\Big(|u_0| + \int_0^t g(s)\,\mathrm{d}s\Big) \tag{3.4}$$

for all $t \in [0, T_0)$. In particular, since $g \in L^2(0, T; \mathbf{R})$, we deduce from (3.4) that $u$ is in fact the unique global solution with $T_0 = T$.

Finally, let us investigate the regularity of the solution $u$. To this end, we define

$$K_u := \Big\{x \in \mathbf{R}^d : |x| \le \mathrm{e}^{\nu T}\Big(|u_0| + \int_0^T g(s)\,\mathrm{d}s\Big)\Big\}. \tag{3.5}$$

Clearly, $K_u \subset \mathbf{R}^d$ is a compact set that contains the origin and the complete curve $[0, T] \ni t \mapsto u(t) \in \mathbf{R}^d$ due to (3.4). Then, an application of Assumption 3.1 (iii) with $K = K_u$ yields

$$|f(t, u(t))| \le L_{K_u}(t)|u(t)| + |f(t, 0)| \le L_{K_u}(t)|u(t)| + g(t) \tag{3.6}$$

for all $t \in [0, T]\backslash\mathcal{N}_f$.

For arbitrary $s, t \in [0, T]$ with $s < t$ it follows from (3.3) that

$$|u(s) - u(t)| \le \int_s^t |f(z, u(z))|\,\mathrm{d}z.$$

Furthermore, after inserting (3.6), we have

$$\begin{aligned}|u(s) - u(t)| &\le \int_s^t L_{K_u}(z)|u(z)| + g(z)\,\mathrm{d}z\\ &\le \Big(1 + \sup_{z\in[0,T]}|u(z)|\Big)\int_s^t \big(L_{K_u}(z) + g(z)\big)\,\mathrm{d}z.\end{aligned}$$

Then, an application of the Cauchy–Schwarz inequality yields

$$|u(s) - u(t)| \le \big(1 + \|u\|_{C([0,T];\mathbf{R}^d)}\big)\big\|L_{K_u} + g\big\|_{L^2(0,T;\mathbf{R})}|s - t|^{\frac{1}{2}} \tag{3.7}$$

for all $s, t \in [0, T]$. This proves that $u$ is Hölder continuous with exponent $\frac{1}{2}$.

## 4 Error Analysis of the Randomized Backward Euler Method

This section is devoted to the error analysis of the randomized backward Euler method. Our error analysis partly relies on variational methods developed in [15] that have recently been adapted to stochastic problems in [3].

In this section, we consider the following randomized version of the backward Euler method: Let $N \in \mathbf{N}$ denote the number of temporal steps and set $k = \frac{T}{N}$ as the temporal step size. For given $N$ and $k$, we obtain an equidistant partition of the interval $[0, T]$ given by $t_n := kn$, $n \in \{0, \dots, N\}$. Further, let $\tau = (\tau_n)_{n \in \mathbf{N}}$ be a family of independent and $\mathcal{U}(0, 1)$-distributed random variables on a complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and let $\xi = (\xi_n)_{n \in \mathbf{N}}$ be the family of random variables given by $\xi_n = t_n + k\tau_n$ for $n \in \mathbf{N}$. Then, the numerical approximation $(U^n)_{n \in \{0,\dots,N\}}$ of the solution $u$ is determined by the recursion

$$\begin{cases} U^n = U^{n-1} + kf(\xi_n, U^n), & \text{for } n \in \{1, \dots, N\}, \\ U^0 = u_0. \end{cases} \tag{4.1}$$

When investigating the solvability of this implicit equation, the mild step size restriction $k\nu < 1$ becomes necessary due to the implicit structure of the scheme. When considering a dissipative equation which is the case when $\nu \le 0$, the restriction disappears. This case corresponds to the setting of the monotone operators in Sect. 6.

Note that (4.1) is an implicit Runge–Kutta method with one stage and a randomized node. More precisely, in each step we apply one member of the following family of implicit Runge–Kutta methods determined by the Butcher tableau

$$\begin{array}{c|c} \theta & 1 \\ \hline & 1 \end{array} \tag{4.2}$$

where the value of the parameter $\theta \in [0, 1]$ is determined by the random variable $\tau_j$ in the $j$-th step.

Further, the resulting sequence $(U^n)_{n \in \{0,\dots,N\}}$ consists of random variables, since we artificially inserted randomness into the numerical method. From a probabilistic point of view, $(U^n)_{n \in \{0,\dots,N\}}$ is in fact a discrete-time stochastic process that takes values in $\mathbf{R}^d$ and is adapted to the complete filtration $(\mathcal{F}_n)_{n \in \mathbf{N}}$. Here, $\mathcal{F}_n \subset \mathcal{F}$ is the smallest complete $\sigma$-algebra such that the subfamily $(\tau_j)_{j \in \{1,\dots,n\}}$ is measurable. Note that $\mathcal{F}_n \subset \mathcal{F}_m$, whenever $n \le m$. More precisely,

$$\begin{aligned} \mathcal{F}_0 &:= \sigma\big(\mathcal{N} \in \mathcal{F} \ : \ \mathbf{P}(\mathcal{N}) = 0\big), \\ \mathcal{F}_n &:= \sigma\big(\sigma(\tau_j \ : \ j \in \{1, \dots, n\}) \cup \mathcal{F}_0\big), \quad n \in \mathbf{N}. \end{aligned} \tag{4.3}$$

In particular, each $\mathbf{P}$-null set (and each subset of a $\mathbf{P}$-null set) is contained in every $\sigma$-algebra $\mathcal{F}_n$, $n \in \mathbf{N}_0$.

Next, let us introduce the following set $\mathcal{G}_N^2$ of *square-integrable and adapted grid functions*. For each $N \in \mathbf{N}$ this set is defined by

$$\mathcal{G}_N^2 := \big\{ Z\colon \{0,\dots,N\} \times \Omega \to \mathbf{R}^d : Z^0 = z_0 \in \mathbf{R}^d, \\ Z^n, f(\xi_n, Z^n) \in L^2(\Omega, \mathcal{F}_n, \mathbf{P}; \mathbf{R}^d) \text{ for } n \in \{1,\dots,N\}\big\}.$$

Take note that $z_0 \in \mathbf{R}^d$ is an arbitrary deterministic initial value and that the condition $Z^n \in L^2(\Omega, \mathcal{F}_n, \mathbf{P}; \mathbf{R}^d)$ ensures that $Z^n$ is square-integrable as well as measurable with respect to the $\sigma$-algebra $\mathcal{F}_n$. First, we will show that the randomized backward Euler method (4.1) with a sufficiently large number $N \in \mathbf{N}$ of steps uniquely determines an element in $\mathcal{G}_N^2$.

We begin by proving the existence of a solution to the implicit scheme. First, we state two technical lemmata to prove the existence and measurability of a solution.

**Lemma 4.1** *For $R \in (0,\infty)$, let $h\colon \overline{B_R(0)} \subseteq \mathbf{R}^d \to \mathbf{R}^d$ be continuous and fulfill the condition*

$$(h(x), x) \geq 0, \quad \textit{for every } x \in \partial B_R(0).$$

*Then, there exists at least one $x_0 \in \overline{B_R(0)}$ such that $h(x_0) = 0$.*

A proof of Lemma 4.1 is found, for instance, in [16, Sec. 9.1].

***Remark 4.2*** For a symmetric, positive definite $Q \in \mathbf{R}^{d,d}$ Lemma 4.1 can be extended as follows. If a function $h\colon B_{Q,R} \subseteq \mathbf{R}^d \to \mathbf{R}^d$ where $B_{Q,R}$ is given by

$$B_{Q,R} = \{x \in \mathbf{R}^d : (Qx, x) \leq R^2\}$$

is continuous and fulfills

$$(Qh(x), x) \geq 0, \quad \text{for every } x \in \partial B_{Q,R},$$

then there exists $x_0 \in B_{Q,R}$ such that $h(x_0) = 0$. This extension of Lemma 4.1 can be proved by exploiting that

$$(Qh(x), x) \geq 0, \quad \text{for every } x \in \mathbf{R}^d \text{ with } (Qx, x) = R^2,$$

can be rewritten as

$$(Q^{\frac{1}{2}} h(Q^{-\frac{1}{2}} y), y) \geq 0, \quad \text{for every } y \in \mathbf{R}^d \text{ with } (y, y) = R^2,$$

using the transformation $y = Q^{\frac{1}{2}} x$.

The next result is needed in order to prove the measurability of the sequence generated by the implicit numerical method (4.1). For a closely related result, we refer to [17, Lem. 3.8]. The proof presented here follows an approach from [13, Prop. 1] that can easily be extended to more general situations.

**Lemma 4.3** *Let $\tilde{\mathcal{F}}$ be a complete sub $\sigma$-algebra of the $\sigma$-algebra $\mathcal{F}$, $\mathcal{M} \in \tilde{\mathcal{F}}$ with $\mathbf{P}(\mathcal{M}) = 1$ and $h \colon \Omega \times \mathbf{R}^d \to \mathbf{R}^d$ such that the following conditions are fulfilled.*

(i) *The mapping $x \mapsto h(\omega, x)$ is continuous for every $\omega \in \mathcal{M}$.*
(ii) *The mapping $\omega \mapsto h(\omega, x)$ is $\tilde{\mathcal{F}}$-measurable for every $x \in \mathbf{R}^d$.*
(iii) *For every $\omega \in \mathcal{M}$, there exists a unique root of the function $h(\omega, \cdot)$.*

*Define the mapping*

$$U \colon \Omega \to \mathbf{R}^d, \quad \omega \mapsto U(\omega),$$

*where $U(\omega)$ is the unique root of $h(\omega, \cdot)$ for $\omega \in \mathcal{M}$ and $U(\omega)$ is arbitrary for $\omega \in \Omega \backslash \mathcal{M}$. Then, $U$ is $\tilde{\mathcal{F}}$-measurable.*

***Proof*** Define the (multivalued) mapping

$$U_\varepsilon \colon \Omega \to \mathcal{P}(\mathbf{R}^d), \quad U_\varepsilon(\omega) := \{x \in \mathbf{R}^d : h(\omega, x) \in B_\varepsilon(0)\}$$

for $\varepsilon > 0$. We first show for an arbitrary open set $A \in \mathcal{B}(\mathbf{R}^d)$ that the set

$$\begin{aligned} U_\varepsilon^{-1}(A) &:= \{\omega \in \Omega : \text{ there exists } x \in A \text{ such that } h(\omega, x) \in B_\varepsilon(0)\} \\ &= \bigcup_{x \in A} \{\omega \in \Omega : h(\omega, x) \in B_\varepsilon(0)\} \end{aligned}$$

is an element of $\tilde{\mathcal{F}}$. To this end, first note that $h(\cdot, x)^{-1}(B_\varepsilon(0)) \in \tilde{\mathcal{F}}$ since $\omega \mapsto h(\omega, x)$ is $\tilde{\mathcal{F}}$-measurable. Then, it follows that

$$\begin{aligned} U_\varepsilon^{-1}(A \cap \mathbf{Q}^d) &= \bigcup_{x \in A \cap \mathbf{Q}^d} \{\omega \in \Omega : h(\omega, x) \in B_\varepsilon(0)\} \\ &= \bigcup_{x \in A \cap \mathbf{Q}^d} h(\cdot, x)^{-1}(B_\varepsilon(0)) \in \tilde{\mathcal{F}}. \end{aligned}$$

It remains to verify the equality

$$U_\varepsilon^{-1}(A) = U_\varepsilon^{-1}(A \cap \mathbf{Q}^d). \tag{4.4}$$

It is clear that $U_\varepsilon^{-1}(A \cap \mathbf{Q}^d)$ is a subset of $U_\varepsilon^{-1}(A)$.

To prove $U_\varepsilon^{-1}(A) \subseteq U_\varepsilon^{-1}(A \cap \mathbf{Q}^d)$, we consider two cases. If $U_\varepsilon^{-1}(A)$ is a subset of $\Omega \backslash \mathcal{M}$, then it is a null set and lies in $\tilde{\mathcal{F}}$ due to the completeness of the $\sigma$-algebra. Else, we can assume that there exist $\omega \in U_\varepsilon^{-1}(A) \cap \mathcal{M}$ and $x_0 \in A$ with $h(\omega, x_0) \in B_\varepsilon(0)$. In particular, we note that the function $x \mapsto h(\omega, x)$ is continuous, since $\omega \in \mathcal{M}$. Further, observe that $A$ is an open neighborhood of $x_0$ and $B_\varepsilon(0)$ is an open neighborhood of $h(\omega, x_0)$. Since $B_\varepsilon(0)$ is open, the continuity of $h$ implies that the set

$$C := h(\omega, \cdot)^{-1}(B_\varepsilon(0))$$

is an open set in $\mathbf{R}^d$ with $x_0 \in C$. Thus, $C \cap A$ is non-empty and open. Therefore, there exists $\overline{x} \in (C \cap A) \cap \mathbf{Q}^d$ such that $h(\omega, \overline{x}) \in B_\varepsilon(0)$. This implies $\omega \in U_\varepsilon^{-1}(A \cap \mathbf{Q}^d)$ and completes the proof of (4.4). Consequently, $U_\varepsilon^{-1}(A) \in \tilde{\mathcal{F}}$ for each open set $A \in \mathcal{B}(\mathbf{R}^d)$.

Next, recall that for each $\omega \in \mathcal{M}$ the image of $U$ is defined as the unique element of $h(\omega, \cdot)^{-1}(\{0\})$. Thus, the set

$$U_0(\omega) := \bigcap_{j \in \mathbf{N}} U_{\frac{1}{j}}(\omega)$$

consists of a single element which coincides with $U(\omega)$. Therefore, we obtain

$$\begin{aligned}
&\mathcal{M} \cap U^{-1}(A)\\
&\quad= \mathcal{M} \cap \{\omega \in \Omega : \text{ there exists } x \in A \text{ such that } h(\omega, x) = 0\}\\
&\quad= \mathcal{M} \cap \bigcap_{j \in \mathbf{N}} \{\omega \in \Omega \ : \ \text{ there exists } x \in A \text{ such that } h(\omega, x) \in B_{\frac{1}{j}}(0)\}\\
&\quad= \mathcal{M} \cap \bigcap_{j \in \mathbf{N}} U_{\frac{1}{j}}^{-1}(A) = \mathcal{M} \cap \bigcap_{j \in \mathbf{N}} U_{\frac{1}{j}}^{-1}(A \cap \mathbf{Q}^d) \in \tilde{\mathcal{F}},
\end{aligned}$$

which also implies $U^{-1}(A) \in \tilde{\mathcal{F}}$ for every open set $A \in \mathcal{B}(\mathbf{R}^d)$ due to the completeness of $\tilde{\mathcal{F}}$. From this, the measurability of the mapping $\omega \mapsto U(\omega)$ follows. □

**Lemma 4.4** *Let Assumption 3.1 be satisfied. Then, for each $N \in \mathbf{N}$ with $\frac{T}{N}\nu = k\nu < 1$ there exists a unique solution $U = (U^n)_{n \in \{0,\dots,N\}} \in \mathcal{G}_N^2$ to the implicit scheme (4.1).*

***Proof*** The assertion $U^n \in L^2(\Omega, \mathcal{F}_n, \mathbf{P}; \mathbf{R}^d)$ is proved using an inductive argument for $n \in \{0, \dots, N\}$. Since $U^0 \equiv u_0 \in L^2(\Omega, \mathcal{F}_0, \mathbf{P}; \mathbf{R}^d)$, the case $n = 0$ is evident. Next, assuming $U^{n-1} \in L^2(\Omega, \mathcal{F}_{n-1}, \mathbf{P}; \mathbf{R}^d)$ exists, we define the set

$$\mathcal{M} = \{\omega \in \Omega : g(\xi_n(\omega)) < \infty,\ |U^{n-1}(\omega)| < \infty \text{ and } \xi_n(\omega) \in [0, T]\backslash \mathcal{N}_f\},$$

where $\mathcal{N}_f \in \mathcal{B}([0, T])$ is the null set from Assumption 3.1. Since $\|g\|_{L^2(0,T;\mathbf{R})} < \infty$ and $\|U^{n-1}\|_{L^2(\Omega;\mathbf{R}^d)} < \infty$ the set fulfills $\mathbf{P}(\mathcal{M}) = 1$. We define the function $h_n$ by

$$h_n \colon \Omega \times \mathbf{R}^d \to \mathbf{R}^d, \quad h_n(\omega, x) = x - U^{n-1}(\omega) - kf(\xi_n(\omega), x). \tag{4.5}$$

In the following, we consider a fixed $\omega \in \mathcal{M}$. Then, the mapping $h_n(\omega, \cdot)$ is continuous by Assumption 3.1 (iii). Further we write

$$R = R(\omega) = \frac{1}{1 - \nu k}\big(|U^{n-1}(\omega)| + kg(\xi_n(\omega))\big).$$

Thus, for each $x \in \mathbf{R}^d$ with $|x| = R$ this implies

$$
\begin{aligned}
(h_n(\omega, x), x) &= |x|^2 - (U^{n-1}(\omega), x) - k(f(\xi_n(\omega), x), x) \\
&\geq R^2 - |U^{n-1}(\omega)|R - k\nu R^2 - kg(\xi_n(\omega))R \\
&= R^2 - k\nu R^2 - (|U^{n-1}(\omega)| + kg(\xi_n(\omega)))R \\
&\geq (1-\nu k)R^2 - (1-\nu k)R^2 = 0.
\end{aligned}
$$

Hence, by Lemma 4.1, for every $\omega \in \mathcal{M}$ there exists $x = x(\omega) \in \mathbf{R}^d$ such that $h_n(\omega, x) = 0$ holds. This $x$ is always unique: Assume there exists $\omega \in \mathcal{M}$ and $x, y \in \mathbf{R}^d$ such that

$$
x = U^{n-1}(\omega) + kf(\xi_n(\omega), x) \quad \text{and} \quad y = U^{n-1}(\omega) + kf(\xi_n(\omega), y)
$$

hold. Then, we can write for the difference

$$
\begin{aligned}
|x-y|^2 &= k\big(f(\xi_n(\omega), x) - f(\xi_n(\omega), y), x-y\big) \\
&\leq k\nu|x-y|^2 < |x-y|^2
\end{aligned}
$$

which implies $x = y$. Thus, the function $h_n \colon \Omega \times \mathbf{R}^d \to \mathbf{R}^d$ is $\mathcal{F}_n$-measurable in the first entry, continuous in the second and has a unique root $x$ for every $\omega \in \mathcal{M}$. Then, Lemma 4.3 implies that the function

$$
U^n \colon \Omega \to \mathbf{R}^d, \quad \omega \mapsto U^n(\omega),
$$

where $U^n(\omega)$ is the unique root of $h_n(\omega, \cdot)$ for $\omega \in \mathcal{M}$ and $U^{n-1}(\omega)$ for $\omega \in \Omega \backslash \mathcal{M}$ is $\mathcal{F}_n$-measurable.

It remains to prove that $U^n$ is finite with respect to the $L^2(\Omega; \mathbf{R}^d)$-norm. Using (3.2), it follows

$$
\begin{aligned}
\|U^n\|^2_{L^2(\Omega;\mathbf{R}^d)} &= \mathbf{E}\big[(U^{n-1}, U^n) + k(f(\xi_n, U^n), U^n)\big] \\
&\leq \|U^{n-1}\|_{L^2(\Omega;\mathbf{R}^d)}\|U^n\|_{L^2(\Omega;\mathbf{R}^d)} + k\mathbf{E}\big[\nu|U^n|^2 + g(\xi_n)|U^n|\big] \\
&\leq \|U^{n-1}\|_{L^2(\Omega;\mathbf{R}^d)}\|U^n\|_{L^2(\Omega;\mathbf{R}^d)} + k\nu\|U^n\|^2_{L^2(\Omega;\mathbf{R}^d)} \\
&\quad + \sqrt{k}\|g\|_{L^2(0,T;\mathbf{R})}\|U^n\|_{L^2(\Omega;\mathbf{R}^d)}
\end{aligned}
$$

and therefore

$$
\|U^n\|_{L^2(\Omega;\mathbf{R}^d)} \leq \frac{1}{1-k\nu}\big(\|U^{n-1}\|_{L^2(\Omega;\mathbf{R}^d)} + \sqrt{k}\|g\|_{L^2(0,T;\mathbf{R})}\big).
$$

The last step is to prove that the function $f(\xi_n, U^n)$ is also an element of $L^2(\Omega, \mathcal{F}_n, \mathbf{P}; \mathbf{R}^d)$. The mapping $\omega \mapsto f(\xi_n(\omega), U^n(\omega))$ is $\mathcal{F}_n$-measurable since $f$ is measurable and both $\xi_n$ and $U^n$ are $\mathcal{F}_n$-measurable. Since both $U^n$ and $U^{n-1}$ are elements of $L^2(\Omega; \mathbf{R}^d)$, we can write

$$\|f(\xi_n, U^n)\|_{L^2(\Omega;\mathbf{R}^d)} = \|\tfrac{1}{k}(U^n - U^{n-1})\|_{L^2(\Omega;\mathbf{R}^d)}.$$

Thus, $f(\xi_n, U^n)$ is finite in the $L^2(\Omega; \mathbf{R}^d)$-norm. □

The following stability lemma will play an important role in the error analysis of the randomized backward Euler method. Its proof is based on techniques developed in [3]. For its formulation, we introduce the *local residual* $(\rho_N^n(V))_{n\in\{0,\dots,N\}}$, $N \in \mathbf{N}$, of an arbitrary grid function $V = (V^n)_{n\in\{0,\dots,N\}} \in \mathcal{G}_N^2$. More precisely, for every $n \in \{1,\dots,N\}$ we define $\rho_N^n(V)$ by

$$\rho_N^n(V) = kf(\xi_n, V^n) - V^n + V^{n-1}. \tag{4.6}$$

Since $(V^n)_{n\in\{0,\dots,N\}} \in \mathcal{G}_N^2$ it directly follows that $\rho_N^n(V) \in L^2(\Omega, \mathcal{F}_n, \mathbf{P}; \mathbf{R}^d)$ for every $n \in \{0,\dots,N\}$.

**Lemma 4.5** *Let Assumption 3.1 be satisfied. For $N \in \mathbf{N}$, let $(U^n)_{n\in\{0,\dots,N\}} \in \mathcal{G}_N^2$ be the grid function generated by (4.1) with step size $k = \frac{T}{N}$. If $\nu k < \frac{1}{4}$, then for every $V \in \mathcal{G}_N^2$ it holds true that*

$$\|U^n - V^n\|_{L^2(\Omega;\mathbf{R}^d)} \le \mathrm{e}^{(2\nu+1)t_n}\Big(|U^0 - V^0|^2 + \sum_{j=1}^{n}\Big(2\|\rho_N^j(V)\|^2_{L^2(\Omega;\mathbf{R}^d)} + \frac{2}{k}\|\mathbf{E}[\rho_N^j(V)|\mathcal{F}_{j-1}]\|^2_{L^2(\Omega;\mathbf{R}^d)}\Big)\Big)^{\frac{1}{2}}$$

*for every $n \in \{1,\dots,N\}$.*

***Proof*** Let $N \in \mathbf{N}$ and $V = (V^j)_{j\in\{0,\dots,N\}} \in \mathcal{G}_N^2$ be arbitrary. Set $E^j := U^j - V^j$ for each $j \in \{0,\dots,N\}$. Since $(a - b, a) = \frac{1}{2}\big(|a|^2 - |b|^2 + |a - b|^2\big)$ for all $a, b \in \mathbf{R}^d$ we get for every $j \in \{1,\dots,N\}$

$$\begin{aligned}|E^j|^2 - |E^{j-1}|^2 + |E^j - E^{j-1}|^2 &= 2\big(E^j - E^{j-1}, E^j\big)\\ &= 2k\big(f(\xi_j, U^j) - f(\xi_j, V^j), E^j\big)\\ &\quad + 2\big(kf(\xi_j, V^j) - V^j + V^{j-1}, E^j\big).\end{aligned}$$

Next, note that $\mathbf{P}(\xi_j \in \mathcal{N}_f) = 0$, where $\mathcal{N}_f$ denotes the null set from Assumption 3.1. Hence, we can apply Assumption 3.1 (i) to the first term on a set with probability one. In addition, we insert (4.6) into the second term and obtain the inequality

$$\begin{aligned}&|E^j|^2 - |E^{j-1}|^2 + |E^j - E^{j-1}|^2\\ &\quad\le 2\nu k|E^j|^2 + 2\big(\rho_N^j(V), E^j - E^{j-1}\big) + 2\big(\rho_N^j(V), E^{j-1}\big) \quad \text{almost surely.}\end{aligned}$$

After taking the expected value, we further observe that

$$\begin{aligned}\mathbf{E}\big[\big(\rho_N^j(V), E^{j-1}\big)\big] &= \big\langle\rho_N^j(V), E^{j-1}\big\rangle_{L^2(\Omega;\mathbf{R}^d)}\\ &= \big\langle\mathbf{E}[\rho_N^j(V)|\mathcal{F}_{j-1}], E^{j-1}\big\rangle_{L^2(\Omega;\mathbf{R}^d)},\end{aligned}$$

since $E^{j-1}$ is $\mathcal{F}_{j-1}$-measurable. Then, applications of the Cauchy–Schwarz inequality and the weighted Young inequality yield

$$\begin{aligned} &2\big\langle \mathbf{E}[\rho_N^j(V)|\mathcal{F}_{j-1}], E^{j-1}\big\rangle_{L^2(\Omega;\mathbf{R}^d)} \\ &\quad \le \frac{1}{k}\big\|\mathbf{E}[\rho_N^j(V)|\mathcal{F}_j]\big\|^2_{L^2(\Omega;\mathbf{R}^d)} + k\big\|E^{j-1}\big\|^2_{L^2(\Omega;\mathbf{R}^d)}. \end{aligned}$$

In the same way, the Cauchy–Schwarz and Young inequalities also yield

$$\mathbf{E}\big[2\big(\rho_N^j(V), E^j - E^{j-1}\big)\big] \le \big\|\rho_N^j(V)\big\|^2_{L^2(\Omega;\mathbf{R}^d)} + \big\|E^j - E^{j-1}\big\|^2_{L^2(\Omega;\mathbf{R}^d)}.$$

Altogether, we have shown that

$$\begin{aligned} &\|E^j\|^2_{L^2(\Omega;\mathbf{R}^d)} - \|E^{j-1}\|^2_{L^2(\Omega;\mathbf{R}^d)} \\ &\quad \le 2\nu k\|E^j\|^2_{L^2(\Omega;\mathbf{R}^d)} + \big\|\rho_N^j(V)\big\|^2_{L^2(\Omega;\mathbf{R}^d)} \\ &\qquad + \frac{1}{k}\big\|\mathbf{E}[\rho_N^j(V)|\mathcal{F}_{j-1}]\big\|^2_{L^2(\Omega;\mathbf{R}^d)} + k\big\|E^{j-1}\big\|^2_{L^2(\Omega;\mathbf{R}^d)}, \end{aligned}$$

where we canceled the term $\|E^j - E^{j-1}\|^2_{L^2(\Omega;\mathbf{R}^d)}$ on both sides of the inequality. Then, after some rearrangements and summing this inequality for all $j \in \{1, \dots, n\}$ with arbitrary $n \in \{1, \dots, N\}$ we obtain

$$\begin{aligned} &(1 - 2\nu k)\|E^n\|^2_{L^2(\Omega;\mathbf{R}^d)} \\ &\quad \le (1 - 2\nu k)\|E^0\|^2_{L^2(\Omega;\mathbf{R}^d)} + (2\nu + 1)k\sum_{j=1}^n \big\|E^{j-1}\big\|^2_{L^2(\Omega;\mathbf{R}^d)} \\ &\qquad + \sum_{j=1}^n \Big(\big\|\rho_N^j(V)\big\|^2_{L^2(\Omega;\mathbf{R}^d)} + \frac{1}{k}\big\|\mathbf{E}[\rho_N^j(V)|\mathcal{F}_{j-1}]\big\|^2_{L^2(\Omega;\mathbf{R}^d)}\Big). \end{aligned}$$

Next, note that from the assumption $\nu k < \frac{1}{4}$ it follows $(1 - 2\nu k)^{-1} \le 2$. Therefore,

$$\begin{aligned} \|E^n\|^2_{L^2(\Omega;\mathbf{R}^d)} &\le \|E^0\|^2_{L^2(\Omega;\mathbf{R}^d)} + 2(2\nu + 1)k\sum_{j=1}^n \big\|E^{j-1}\big\|^2_{L^2(\Omega;\mathbf{R}^d)} \\ &\quad + \sum_{j=1}^n \Big(2\big\|\rho_N^j(V)\big\|^2_{L^2(\Omega;\mathbf{R}^d)} + \frac{2}{k}\big\|\mathbf{E}[\rho_N^j(V)|\mathcal{F}_{j-1}]\big\|^2_{L^2(\Omega;\mathbf{R}^d)}\Big). \end{aligned}$$

Finally, applying a discrete Gronwall lemma (Lemma 2.1) completes the proof.

The second ingredient in the error analysis is an estimate of the local residual of the exact solution. For its formulation, we need to represent the exact solution by a grid function. This is easily achieved by restricting $u$ to the temporal points $t_n = nk$, $n \in \{0, \dots, N\}$, with $k = \frac{T}{N}$ and $N \in \mathbf{N}$. More precisely, we define the *restriction* $u|_N$ of $u$ to the grid points $(t_n)_{n \in \{0,\dots,N\}}$ by

$$[u|_N]^n := u(t_n) \tag{4.7}$$

for all $n \in \{0, \dots, N\}$. Since $u|_N$ is deterministic, we clearly have $[u|_N]^n = u(t_n) \in L^2(\Omega, \mathcal{F}_n, \mathbf{P}; \mathbf{R}^d)$ for all $n \in \{0, \dots, N\}$. In addition, as in (3.6) we have

$$\begin{aligned}\|f(\xi_n, u(t_n))\|_{L^2(\Omega;\mathbf{R}^d)} &\le \big\| L_{K_u}(\xi_n)|u(t_n)| + g(\xi_n)\big\|_{L^2(\Omega;\mathbf{R})}\\ &\le \frac{1}{\sqrt{k}}\big(1 + \|u\|_{C([0,T];\mathbf{R}^d)}\big)\big\|L_{K_u} + g\big\|_{L^2(t_{n-1},t_n;\mathbf{R})} < \infty.\end{aligned}$$

This shows that $u|_N \in \mathcal{G}^2_N$ for every $N \in \mathbf{N}$.

**Lemma 4.6** *Let Assumption 3.1 be satisfied. Then, for all $N \in \mathbf{N}$ and $n \in \{1, \dots, N\}$ the local residual (4.6) of the exact solution $u$ to the initial value problem (3.1) is bounded by*

$$\begin{aligned}\big\|\rho^n_N(u|_N)\big\|_{L^2(\Omega;\mathbf{R}^d)} &\le \big(1 + \|u\|_{C([0,T];\mathbf{R}^d)}\big)\Big(1 + T^{\frac{1}{2}}\big\|L_{K_u} + g\big\|_{L^2(0,T;\mathbf{R})}\Big)\\ &\quad \times \Big(\|g\|_{L^2(t_{n-1},t_n;\mathbf{R})} + \big\|L_{K_u}\big\|_{L^2(t_{n-1},t_n;\mathbf{R})}\Big)k^{\frac{1}{2}}\end{aligned} \tag{4.8}$$

*and*

$$\begin{aligned}&\big\|\mathbf{E}\big[\rho^n_N(u|_N)|\mathcal{F}_{n-1}\big]\big\|_{L^2(\Omega;\mathbf{R}^d)}\\ &\quad \le \big(1 + \|u\|_{C([0,T];\mathbf{R}^d)}\big)\big\|L_{K_u} + g\big\|_{L^2(0,T;\mathbf{R})}\big\|L_{K_u}\big\|_{L^2(t_{n-1},t_n;\mathbf{R})}k.\end{aligned} \tag{4.9}$$

***Proof*** Fix $N \in \mathbf{N}$ and $n \in \{1, \dots, N\}$ arbitrarily. First recall that

$$\rho^n_N(u|_N) = kf(\xi_n, u(t_n)) - u(t_n) + u(t_{n-1}).$$

Inserting (3.3) yields

$$\begin{aligned}\rho^n_N(u|_N) &= k\big(f(\xi_n, u(t_n)) - f(\xi_n, u(\xi_n))\big)\\ &\quad + kf(\xi_n, u(\xi_n)) - \int_{t_{n-1}}^{t_n} f(s, u(s))\,\mathrm{d}s.\end{aligned} \tag{4.10}$$

Since $f$ and $u$ are deterministic, the only source of randomness in this expression is the random variable $\xi_n$. Further, since $\xi_n$ is independent of $\mathcal{F}_{n-1}$ we obtain

$$\mathbf{E}\big[\rho^n_N(u|_N)|\mathcal{F}_{n-1}\big] = \mathbf{E}\big[\rho^n_N(u|_N)\big] = k\mathbf{E}\big[f(\xi_n, u(t_n)) - f(\xi_n, u(\xi_n))\big],$$

where we also used that

$$k\mathbf{E}\big[f(\xi_n, u(\xi_n))\big] = \int_{t_{n-1}}^{t_n} f(s, u(s))\,\mathrm{d}s. \tag{4.11}$$

Since $\mathbf{P}(\xi_n \in \mathcal{N}_f) = 0$, we can apply Assumption 3.1 (iii) with the compact set $K = K_u \subset \mathbf{R}^d$ defined in (3.5) inside the expectation. This yields

$$\begin{aligned}\big\|\mathbf{E}\big[\rho_N^n(u|_N)|\mathcal{F}_{n-1}\big]\big\|_{L^2(\Omega;\mathbf{R}^d)} &= \big|\mathbf{E}\big[\rho_N^n(u|_N)\big]\big|\\ &\le k\mathbf{E}\big[|f(\xi_n, u(t_n)) - f(\xi_n, u(\xi_n))|\big]\\ &\le k\mathbf{E}\big[L_{K_u}(\xi_n)\,|u(t_n) - u(\xi_n)|\big]\\ &\le k\big(\mathbf{E}\big[L_{K_u}(\xi_n)^2\big]\big)^{\frac12}\big\|u(t_n) - u(\xi_n)\big\|_{L^2(\Omega;\mathbf{R}^d)}.\end{aligned}$$

Then, we make use of the Hölder continuity (3.7) of $u$ and obtain

$$\big\|u(t_n) - u(\xi_n)\big\|_{L^2(\Omega;\mathbf{R}^d)} \le \big(1 + \|u\|_{C([0,T];\mathbf{R}^d)}\big)\big\|L_{K_u} + g\big\|_{L^2(0,T;\mathbf{R})}k^{\frac12}.$$

In addition, we note that

$$k^{\frac12}\big(\mathbf{E}\big[L_{K_u}(\xi_n)^2\big]\big)^{\frac12} = \Big(\int_{t_{n-1}}^{t_n} L_{K_u}(s)^2\,\mathrm{d}s\Big)^{\frac12}.$$

Hence,

$$\begin{aligned}&\big\|\mathbf{E}\big[\rho_N^n(u|_N)|\mathcal{F}_{n-1}\big]\big\|_{L^2(\Omega;\mathbf{R}^d)}\\ &\quad\le \big(1 + \|u\|_{C([0,T];\mathbf{R}^d)}\big)\big\|L_{K_u} + g\big\|_{L^2(0,T;\mathbf{R})}\Big(\int_{t_{n-1}}^{t_n} L_{K_u}(s)^2\,\mathrm{d}s\Big)^{\frac12}k,\end{aligned}$$

which proves assertion (4.9).

It remains to show (4.8). To this end, we directly apply the $L^2(\Omega;\mathbf{R}^d)$-norm to (4.10) and obtain

$$\begin{aligned}\big\|\rho_N^n(u|_N)\big\|_{L^2(\Omega;\mathbf{R}^d)} &\le k\big\|f(\xi_n, u(t_n)) - f(\xi_n, u(\xi_n))\big\|_{L^2(\Omega;\mathbf{R}^d)}\\ &\quad + \Big\|kf(\xi_n, u(\xi_n)) - \int_{t_{n-1}}^{t_n} f(s, u(s))\,\mathrm{d}s\Big\|_{L^2(\Omega;\mathbf{R}^d)}.\end{aligned} \tag{4.12}$$

By similar arguments as above, we derive the following estimate for the first term:

$$\begin{aligned}&k\big\|f(\xi_n, u(t_n)) - f(\xi_n, u(\xi_n))\big\|_{L^2(\Omega;\mathbf{R}^d)}\\ &\quad\le k\big(\mathbf{E}\big[L_{K_u}(\xi_n)^2|u(t_n) - u(\xi_n)|^2\big]\big)^{\frac12}\\ &\quad\le k^{\frac12}T^{\frac12}\big(1 + \|u\|_{C([0,T];\mathbf{R}^d)}\big)\big\|L_{K_u} + g\big\|_{L^2(0,T;\mathbf{R})}\Big(\int_{t_{n-1}}^{t_n} L_{K_u}(s)^2\,\mathrm{d}s\Big)^{\frac12},\end{aligned}$$

where we also made use of the estimate $k \le T$ in the last step.

Regarding the second summand in (4.12), we first observe that

$$\begin{aligned}
&\Big\| kf(\xi_n, u(\xi_n)) - \int_{t_{n-1}}^{t_n} f(s, u(s))\,\mathrm{d}s \Big\|^2_{L^2(\Omega;\mathbf{R}^d)} \\
&\quad = \big\| kf(\xi_n, u(\xi_n)) \big\|^2_{L^2(\Omega;\mathbf{R}^d)} - 2\mathbf{E}\Big[\Big(kf(\xi_n, u(\xi_n)), \int_{t_{n-1}}^{t_n} f(s, u(s))\,\mathrm{d}s\Big)\Big] \\
&\qquad + \Big| \int_{t_{n-1}}^{t_n} f(s, u(s))\,\mathrm{d}s \Big|^2 \\
&\quad = \big\| kf(\xi_n, u(\xi_n)) \big\|^2_{L^2(\Omega;\mathbf{R}^d)} - \Big| \int_{t_{n-1}}^{t_n} f(s, u(s))\,\mathrm{d}s \Big|^2 \\
&\quad \le \big\| kf(\xi_n, u(\xi_n)) \big\|^2_{L^2(\Omega;\mathbf{R}^d)},
\end{aligned}$$

due to (4.11). Moreover, since $0 \in K_u \subset \mathbf{R}^d$, we derive from Assumption 3.1 (ii) and (iii) that

$$\begin{aligned}
&\big\| kf(\xi_n, u(\xi_n)) \big\|_{L^2(\Omega;\mathbf{R}^d)} \\
&\quad \le k \big\| f(\xi_n, 0) \big\|_{L^2(\Omega;\mathbf{R}^d)} + k \big\| f(\xi_n, u(\xi_n)) - f(\xi_n, 0) \big\|_{L^2(\Omega;\mathbf{R}^d)} \\
&\quad \le k \big(\mathbf{E}\big[|g(\xi_n)|^2\big]\big)^{\frac{1}{2}} + k \big(\mathbf{E}\big[L_{K_u}(\xi_n)^2 |u(\xi_n)|^2\big]\big)^{\frac{1}{2}} \\
&\quad \le k^{\frac{1}{2}} \Big( \int_{t_{n-1}}^{t_n} g(s)^2\,\mathrm{d}s \Big)^{\frac{1}{2}} + k^{\frac{1}{2}} \|u\|_{C([0,T];\mathbf{R}^d)} \Big( \int_{t_{n-1}}^{t_n} L_{K_u}(s)^2\,\mathrm{d}s \Big)^{\frac{1}{2}}.
\end{aligned}$$

In summary, we have shown that

$$\begin{aligned}
\big\| \rho_N^n(u|_N) \big\|_{L^2(\Omega;\mathbf{R}^d)} &\le \big(1 + \|u\|_{C([0,T];\mathbf{R}^d)}\big)\Big(1 + T^{\frac{1}{2}} \big\| L_{K_u} + g \big\|_{L^2(0,T;\mathbf{R})}\Big) \\
&\qquad \times \Big( \big\| g \big\|_{L^2(t_{n-1},t_n;\mathbf{R})} + \big\| L_{K_u} \big\|_{L^2(t_{n-1},t_n;\mathbf{R})} \Big) k^{\frac{1}{2}}.
\end{aligned}$$

This completes the proof of (4.8). □

We are now well prepared to state and prove the main result of this section.

**Theorem 4.7** *Let Assumption 3.1 be satisfied. Let $(U^n)_{n\in\{0,\dots,N\}} \in \mathcal{G}_N^2$, $N \in \mathbf{N}$, be the grid function generated by the randomized backward Euler method (4.1) with step size $k = \frac{T}{N}$. If $\nu k < \frac{1}{4}$, then there exists a constant $C$ independent of $N$ and $k$ such that*

$$\max_{n\in\{0,\dots,N\}} \big\| U^n - u(t_n) \big\|_{L^2(\Omega;\mathbf{R}^d)} \le C k^{\frac{1}{2}}. \tag{4.13}$$

***Proof*** Let us fix an arbitrary $N \in \mathbf{N}$ such that $\nu k < \frac{1}{4}$. First note that the sequence $(U^n)_{n\in\{0,\dots,N\}} \in \mathcal{G}_N^2$ is well defined by Lemma 4.4. Furthermore, as we already discussed above, the restriction $u|_N$ defined in (4.7) is also an element of $\mathcal{G}_N^2$. Hence, Lemma 4.5 is applicable with $V = u|_N$. Using that $U^0 = u_0 = u(t_0)$, we therefore obtain

$$\begin{aligned}&\big\|U^n - u(t_n)\big\|^2_{L^2(\Omega;\mathbf{R}^d)}\\ &\quad\le \mathrm{e}^{2(2\nu+1)t_n} \sum_{j=1}^{n} \Big( 2\big\|\rho_N^j(u|_N)\big\|^2_{L^2(\Omega;\mathbf{R}^d)} + \frac{2}{k}\big\|\mathbf{E}\big[\rho_N^j(u|_N) \,|\, \mathcal{F}_{j-1}\big]\big\|^2_{L^2(\Omega;\mathbf{R}^d)}\Big)\end{aligned}$$

for every $n \in \mathbf{N}$. After taking the maximum over $n \in \{0, \ldots, N\}$ it remains to estimate the two sums over the local residuals of the exact solution. From Lemma 4.6, we get

$$\begin{aligned}2\sum_{j=1}^{N} \|\rho_N^j(u|_N)\|^2_{L^2(\Omega;\mathbf{R}^d)} &\le C_1 k \sum_{j=1}^{N} \big(\|g\|_{L^2(t_{j-1},t_j;\mathbf{R})} + \|L_{K_u}\|_{L^2(t_{j-1},t_j;\mathbf{R})}\big)^2\\ &\le 2C_1 k \sum_{j=1}^{N} \big(\|g\|^2_{L^2(t_{j-1},t_j;\mathbf{R})} + \|L_{K_u}\|^2_{L^2(t_{j-1},t_j;\mathbf{R})}\big)\\ &= 2C_1\big(\|g\|^2_{L^2(0,T;\mathbf{R})} + \|L_{K_u}\|^2_{L^2(0,T;\mathbf{R})}\big)k,\end{aligned}$$

where the constant $C_1$ is given by

$$C_1 = 2\big(1 + \|u\|_{C([0,T];\mathbf{R}^d)}\big)^2\Big(1 + T^{\frac{1}{2}}\big\|L_{K_u} + g\big\|_{L^2(0,T;\mathbf{R})}\Big)^2.$$

In addition, Lemma 4.6 also yields

$$\begin{aligned}&\frac{2}{k}\sum_{j=1}^{n} \big\|\mathbf{E}\big[\rho_N^j(u|_N) \,|\, \mathcal{F}_{j-1}\big]\big\|^2_{L^2(\Omega;\mathbf{R}^d)}\\ &\quad\le C_2 k \sum_{j=1}^{N} \|L_{K_u}\|^2_{L^2(t_{j-1},t_j;\mathbf{R})} = C_2\|L_{K_u}\|^2_{L^2(0,T;\mathbf{R})}k,\end{aligned}$$

with

$$C_2 = 2\big(1 + \|u\|_{C([0,T];\mathbf{R}^d)}\big)^2\big\|L_{K_u} + g\big\|^2_{L^2(0,T;\mathbf{R})}.$$

Altogether, this proves (4.13) with

$$C = \mathrm{e}^{(2\nu+1)T}\sqrt{\max(2C_1, C_2)}\Big(\|g\|_{L^2(0,T;\mathbf{R})} + \|L_{K_u}\|_{L^2(0,T;\mathbf{R})}\Big).$$

□

## 5 Numerical Experiments for ODEs

A simple, yet useful problem to demonstrate the usability of the randomized backward Euler method (4.1) is the Prothero–Robinson example from [36] (see also [22, Sec. IV.15]) which is given by

$$\begin{cases} \dot{u}(t) = \lambda(u(t) - g(t)) + \dot{g}(t), & \text{for almost all } t \in (0, T], \\ u(0) = g(0), \end{cases} \tag{5.1}$$

for $\lambda \in \mathbf{R}$ and $g \in H^1(0, T)$. Here, $H^1(0, T)$ denotes the standard Sobolev space of square-integrable and weakly differentiable functions. It is easy to verify that $u = g$ is a solution to (5.1) in the sense of Carathéodory. The right-hand side $f : [0, T] \times \mathbf{R} \to \mathbf{R}$ is given by

$$f(t, x) := \lambda(x - g(t)) + \dot{g}(t), \quad t \in [0, T], \; x \in \mathbf{R},$$

which fulfills Assumption 3.1, as can easily be shown.

For a numerical example, we choose $T = 1$ and a function $g$ which is oscillating with a period $2p$, for $p = 2^{-K}$, $K \in \mathbf{N}$. To this end, we use a continuous, piecewise linear function $g$. This function is chosen such that it fulfills

$$g(ip) = \begin{cases} p, & \text{for } i \in \{0, \dots, 2^K\} \text{ odd}, \\ 0, & \text{for } i \in \{0, \dots, 2^K\} \text{ even}, \end{cases}$$

and the affine linear interpolation of these values for all other $t \in [0, T]$. Further, the function $g$ has a weak derivative in $L^2(0, 1)$. For the implementation, we take the following representation for $\dot{g}$ given by

$$\dot{g}(t) = \begin{cases} -1, & \text{for } t \in [ip, (i+1)p), \; i \in \{0, \dots, 2^K - 1\} \text{ odd}, \\ 1, & \text{for } t \in [ip, (i+1)p), \; i \in \{0, \dots, 2^K - 1\} \text{ even}. \end{cases}$$

For every equidistant step size $k = 2^{-n} > p$, $n \in \mathbf{N}$ with $n < K$, the classical backward Euler method only evaluates the mapping $g$ in the grid points, where $g$ is equal to zero and where the chosen representation of $\dot{g}$ is equal to 1. Therefore, for all such step sizes, the classical backward Euler method cannot distinguish between the problem (5.1) and the initial value problem

$$\begin{cases} \dot{v}(t) = \lambda v(t) + 1, & \text{for all } t \in (0, T], \\ v(0) = g(0). \end{cases}$$

Since $u = g \neq v$, it is not surprising that the classical backward Euler method does not yield a good approximation of the correct solution. Only for $k < p$ it becomes visible that the classical backward Euler method converges to the exact solution $u = g$.

On the other hand, the randomized scheme (4.1) is not so easily "fooled" by the highly oscillating function $g$. It already yields more reliable results for step sizes $k > p$, since it evaluates $g$ and $\dot{g}$ not only in extremal points. In Fig. 1, we indeed see that the error of the randomized scheme (4.1) measured in the $L^2(\Omega, \mathbf{R})$-norm is significantly smaller than that of the classical backward Euler method.

Obviously, a simple way to correct the backward Euler method would be to choose a different temporal grid. For instance, one might use a non-equidistant partition of $[0, T]$ or an adaptive version of the backward Euler method. However, no matter what deterministic strategy is used, it is always possible to construct a similar "fooling" function $f : [0, T] \times \mathbf{R} \to \mathbf{R}$ that satisfies Assumption 3.1 and deceives the deterministic algorithm to approximate the wrong initial value problem for all computationally feasible numbers of function evaluations.

A further interesting aspect of problem (5.1) is the fact that for $\lambda < 0$ it has a dissipative structure, i.e., there exists $\nu \in [0, \infty)$ such that

$$\big(f(t, x) - f(t, y), x - y\big) \leq -\nu |x - y|^2,$$

holds for all $x, y \in \mathbf{R}$ and $t \in [0, 1]$. It is well known as discussed in [22] that this structure of the problem can be exploited more efficiently with an implicit scheme in comparison to explicit Runge–Kutta methods. Here, we will compare the randomized backward Euler method (4.1) with its explicit randomized counterpart

$$\begin{cases} U^n = U^{n-1} + k f(\xi_n, U^{n-1}), & \text{for } n \in \{1, \dots, N\}, \\ U^0 = u_0, \end{cases} \tag{5.2}$$

which has been studied in [12,24,26,30]. In this particular example, we obtain the scheme

$$\begin{cases} U^n = (1 + k\lambda) U^{n-1} - k\lambda g(\xi_n) + k\dot{g}(\xi_n), & \text{for } n \in \{1, \dots, N\}, \\ U^0 = u_0. \end{cases}$$

This will lead to an oscillating numerical solution with a high amplitude if $|1 + k\lambda| > 1$ holds true. For $\lambda < 0$, this is the case if $k < -\frac{2}{\lambda}$.

In the numerical examples that lead to Fig. 1, we considered the value $p = 2^{-12}$ and step sizes $k = 2^{-n}$ for $n \in \{5, \dots, 14\}$. To evaluate the $L^2(\Omega; \mathbf{R})$-norm, we considered 1000 Monte Carlo iterations. In the plot on the left hand side, we used the value $\lambda = 2$ and compared the classical backward Euler method with scheme (4.1). As we expected from the discussion above, two different phases of the example become well visible. For $n \in \{5, \dots, 11\}$, the classical backward Euler method does not offer an accurate numerical solution. The error of the randomized backward Euler method decreases with a rate of approximately 0.5. When $n$ changes from 11 to 12, both schemes improve drastically since they are now able to fully resolve the oscillations of the solution. In the last part, for $n \in \{12, 13, 14\}$ the errors of both schemes decrease with a larger rate. Also here, the randomized scheme appears to have a higher rate of

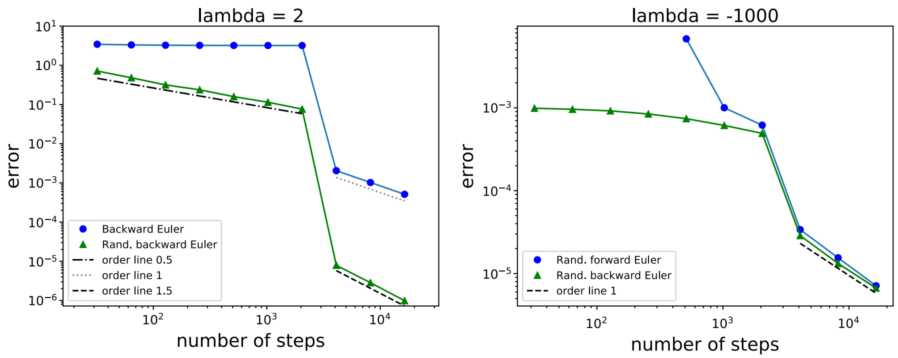


**Fig. 1** Left: $L^2$-convergence of the classical backward Euler method and scheme (4.1) to the IVP (5.1) with $\lambda = 2$. Right: $L^2$-convergence of schemes (4.1) and (5.2) to (5.1) with $\lambda = -1000$

convergence, 1.5, than the classical scheme which converges with rate 1. Note that the rate of 1.5 is in line with those of randomized quadrature rules (see [30]).

In the plot on the right-hand side in Fig. 1, we considered the case $\lambda = -1000$ and compared the randomized backward Euler method (4.1) with the randomized forward Euler method (5.2). Here, we only plotted errors smaller than 1, since the explicit scheme produces strongly oscillating numerical solutions with a very large amplitude for step sizes which are not small enough. The first occurring error of the scheme (5.2) in the plot appears for $2^9 = 512$ temporal steps. This was expected since the explicit scheme only leads to a non-exploding solution for step sizes $k$ with $|1 + k\lambda| < 1$.

To sum up, the numerical experiments in this section indicate that the randomized backward Euler method is especially advantageous compared to deterministic methods if the problem has very irregular coefficients. Compared to explicit randomized Runge–Kutta methods such as (5.2), we also obtain more reliable results for rather large step sizes when considering problems with a dissipative structure. Both points qualify the scheme (4.1) for the numerical treatment of monotone evolution equations with time-irregular coefficients. This will be studied in more detail in the following sections.

## 6 A Non-autonomous Nonlinear Evolution Equation with Time-Irregular Coefficients

In this section, we now turn our attention to the second class of initial value problems we consider in this paper. More precisely, we are interested in non-autonomous and possibly nonlinear evolution equations of the form

$$\begin{cases} \dot{u}(t) + \mathcal{A}(t)u(t) = f(t), & \text{for almost all } t \in (0, T], \\ u(0) = u_0. \end{cases} \tag{6.1}$$

In order to make this rather abstract setting more precise, we start by introducing the real, separable Hilbert spaces $(V, (\cdot, \cdot)_V, \|\cdot\|_V)$ and $(H, (\cdot, \cdot)_H, \|\cdot\|_H)$. Here, we assume that the space $V$ is densely embedded in the space $H$. Thus, we obtain the Gelfand triple

$$V \overset{d}{\hookrightarrow} H \cong H^* \overset{d}{\hookrightarrow} V^*,$$

where $H^*$ and $V^*$ are the dual spaces of $H$ and $V$, respectively. These spaces are equipped with the induced dual norms.

We impose the following conditions on $\mathcal{A}$. Note that, as it is customary, we usually write $\mathcal{A}(t)v$ instead of $\mathcal{A}(t, v)$.

**Assumption 6.1** The mapping $\mathcal{A}\colon [0, T] \times V \to V^*$ fulfills the conditions:

(i) For every $v_1, v_2 \in V$ the mapping $[0, T] \ni t \mapsto \langle \mathcal{A}(t)v_1, v_2\rangle_{V^*,V}$ is measurable.
(ii) There exists a constant $M \geq 0$ such that $\|\mathcal{A}(t)0\|_{V^*} \leq M$ for every $t \in [0, T]$.
(iii) There exists $L \in (0, \infty)$ such that for all $t \in [0, T]$ it holds true that

$$\|\mathcal{A}(t)v_1 - \mathcal{A}(t)v_2\|_{V^*} \leq L\|v_1 - v_2\|_V, \quad \text{for all } v_1, v_2 \in V.$$

(iv) There exists $\mu \in (0, \infty)$ such that for all $t \in [0, T]$ it holds true that

$$\langle \mathcal{A}(t)v_1 - \mathcal{A}(t)v_2, v_1 - v_2\rangle_{V^*,V} \geq \mu\|v_1 - v_2\|_V^2, \quad \text{for all } v_1, v_2 \in V.$$

***Remark 6.2*** Instead of Assumption 6.1 (iv), we can ask for the weaker condition

(iv′) There exist $\mu \in (0, \infty)$ and $\kappa \in [0, \infty)$ such that for all $t \in [0, T]$ it holds true that

$$\langle \mathcal{A}(t)v_1 - \mathcal{A}(t)v_2, v_1 - v_2\rangle_{V^*,V} \geq \mu\|v_1 - v_2\|_V^2 - \kappa\|v_1 - v_2\|_H^2,$$

for all $v_1, v_2 \in V$.

Using this Gårding-type inequality, the following proofs can be done in an analogous manner with a further application of Gronwall's inequality. This additional argument leads to a constant $C$ in Theorem 6.7 below that grows exponentially in time. For simplicity, we will only treat the case $\kappa = 0$ in the following.

Before we analyze the convergence of the numerical scheme (6.5) defined below, let us recall the existence of a unique solution to the abstract problem (6.1). We will consider the concept of weak solutions for abstract non-autonomous problems of the form (6.1), i.e., we call a function

$$u \in \mathcal{W}(0, T) = \left\{v \in L^2(0, T; V) : \dot{v} \in L^2(0, T; V^*)\right\}$$

a *weak solution* to (6.1) if $u(0) = u_0$ is fulfilled and if the integral equality

$$\int_0^T \langle \dot{u}(t) + \mathcal{A}(t)u(t), v(t)\rangle_{V^*,V}\, \mathrm{d}t = \int_0^T \langle f(t), v(t)\rangle_{V^*,V}\, \mathrm{d}t \tag{6.2}$$

is satisfied for every $v \in L^2(0, T; V)$. Note that evaluating the abstract function $u$ at the initial time is well defined since the space $\mathcal{W}(0, T)$ is embedded in the space $C([0, T]; H)$. An introduction to this concept of solutions can be found in, for example, [14,16] or [37].

**Proposition 6.3** *Let Assumption 6.1 be satisfied. Then, for every given* $f \in L^2(0,T;H)$ *and initial value* $u_0 \in H$ *there exists a unique weak solution* $u \in \mathcal{W}(0,T)$ *to the problem* (6.1).

Most proofs for this kind of statement that can be found in the literature are either for linear problems (see, for example, [42, Cor. 23.26] or [14, Satz 8.3.6]) or for nonlinear problems in a Browder–Minty setting (compare, for example, [43, Thm 30.A], [14, Satz 8.4.2] or [37, Theorem 8.9]). Our assertion is intermediate since we consider nonlinear operators that are still Lipschitz continuous. Therefore, the aforementioned references for nonlinear problems can be used, but we note that also small modifications of the proofs for linear problems would be sufficient.

***Remark 6.4*** Note that for mere existence results, it is sufficient to assume $f \in L^2(0,T;V^*) + L^1(0,T;H)$. The last proposition and some of the following statements would also hold under this more general condition. To obtain a rate of convergence for the numerical scheme, the additional assumption $f \in L^2(0,T;H)$ will be essential.

In the following, we will consider a full discretization of the problem (6.1), i.e., we will discretize the equation both in time and in space. For this purpose, let $N \in \mathbf{N}$ denote the number of temporal steps and set $k = \frac{T}{N}$ as the temporal step size. For this particular $N$ and $k$, we obtain an equidistant partition of the interval $[0,T]$ given by $t_n := kn$, $n \in \{0,\dots,N\}$. Further, we introduce the family of independent and $\mathcal{U}(0,1)$-distributed random variables $\tau = (\tau_n)_{n\in\mathbf{N}}$ on a complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and write $\xi_n = t_n + k\tau_n$ for $n \in \mathbf{N}$. Let $(\mathcal{F}_n)_{n\in\{0,\dots,N\}}$ be the complete filtration which is induced by $(\xi_n)_{n\in\{0,\dots,N\}}$ (compare with (4.3)).

For the space discretization, we consider an abstract Galerkin method. To this end, let $(V_h)_{h\in(0,1)}$ be a sequence of finite-dimensional subspaces of $V$ each endowed with the inner product $(\cdot,\cdot)_H$ and the norm $\|\cdot\|_H$ of $H$. Further, for each $h \in (0,1)$ we denote by $P_h\colon H \to V_h$ the orthogonal projection onto the Galerkin space $V_h$ with respect to $(\cdot,\cdot)_H$. More precisely, for each $v \in H$ we define $P_h v$ as the uniquely determined element in $V_h$ that satisfies

$$(P_h v, w_h)_H = (v, w_h)_H, \quad \text{for all } w_h \in V_h. \tag{6.3}$$

In order to formulate the Eq. (6.1) in a suitable discrete setting, we also introduce a discrete version $\mathcal{A}_h\colon [0,T] \times V_h \to V_h$ of the operator $\mathcal{A}$. This is accomplished in the same way as above by defining $\mathcal{A}_h(t)v_h$ for given $t \in [0,T]$ and $v_h \in V_h$ as the unique element in $V_h$ that fulfills

$$\big(\mathcal{A}_h(t)v_h, w_h\big)_H = \langle \mathcal{A}(t)v_h, w_h\rangle_{V^*,V} \tag{6.4}$$

for every $w_h \in V_h$. The existence of a unique $\mathcal{A}_h(t)v_h \in V_h$ follows directly from the Riesz representation theorem.

Our aim is to examine the numerical scheme

$$\begin{cases} U_h^n + k\mathcal{A}_h(\xi_n)U_h^n = kP_h f(\xi_n) + U_h^{n-1}, & \text{for } n \in \{1,\dots,N\}, \\ U_h^0 = P_h u_0. \end{cases} \tag{6.5}$$

Note that, as in the finite-dimensional case in Sect. 4, the numerical approximation $(U_h^n)_{n\in\{0,\dots,N\}}$ consists of a family of random variables taking values in $V_h$. Before we analyze the convergence of the scheme (6.5), the following lemma shows that $(U_h^n)_{n\in\{0,\dots,N\}}$ is indeed well defined for every value of the step size $k$.

**Lemma 6.5** *Let Assumption 6.1 be satisfied. Then for every inhomogeneity $f \in L^2(0,T;H)$, every initial value $u_0 \in H$, and every step size $k = \frac{T}{N}$, $N \in \mathbf{N}$, there exists a unique solution $(U_h^n)_{n\in\{0,\dots,N\}}$ to the implicit scheme (6.5) such that for every $n \in \{1,\dots,N\}$ the element $U_h^n$ is $\mathcal{F}_n$-measurable and $U_h^n(\omega) \in V_h$ for almost every $\omega \in \Omega$.*

***Proof*** Let $h \in (0,1)$ be fixed. To prove the existence of a suitable solution to (6.5), we introduce an equivalent problem in $\mathbf{R}^d$ with $d = \dim(V_h)$ such that we can apply arguments from Sect. 4 to prove the existence of a unique solution $(U_h^n)_{n\in\{0,\dots,N\}}$. To this end, we consider a basis $\{\psi_1,\dots,\psi_d\}$ of the finite-dimensional space $V_h$ and test (6.5) with a basis element $\psi_j$, $j \in \{1,\dots,d\}$. Then, (6.5) can equivalently be rewritten as the following system of scalar equations

$$\begin{cases} (U_h^n + k\mathcal{A}_h(\xi_n)U_h^n, \psi_j)_H = (kP_h f(\xi_n) + U_h^{n-1}, \psi_j)_H, & \text{for } n \in \{1,\dots,N\}, \\ (U_h^0, \psi_j)_H = (u_0, \psi_j)_H, \end{cases} \tag{6.6}$$

for all $j \in \{1,\dots,d\}$. Since the inhomogeneity $P_h f \in L^2(0,T;V_h)$ takes values in $V_h$, it can be represented by

$$P_h f = \sum_{i=1}^d f_{h,i}\psi_i,$$

where $f_{h,i} \in L^2(0,T;\mathbf{R})$ for each $i \in \{1,\dots,d\}$. In order to prove the existence of the $V_h$-valued random variable $U_h^n$, we will show that there exist measurable functions $\alpha_{h,i}^n \colon \Omega \to \mathbf{R}$, $i \in \{1,\dots,d\}$, $n \in \{0,\dots,N\}$, such that

$$U_h^n = \sum_{i=1}^d \alpha_{h,i}^n \psi_i$$

satisfies (6.5). For $n = 0$, this follows at once.

For the case $n > 0$, let us denote the vector of all coordinates $(\alpha_{h,i}^n)_{i\in\{1,\dots,d\}}$ and $(f_{h,i})_{i\in\{1,\dots,d\}}$ by

$$\mathbf{u_h^n}(\omega) := \big(\alpha_{h,i}^n(\omega)\big)_{i\in\{1,\dots,d\}} \quad \text{and} \quad \mathbf{f_h}(t) := \big(f_{h,i}(t)\big)_{i\in\{1,\dots,d\}}$$

for almost every $\omega \in \Omega$ and $t \in [0,T]$. Furthermore, we denote the mass matrix in $\mathbf{R}^{d,d}$ by

$$\mathbf{M_h} = \big((\psi_i,\psi_j)_H\big)_{i,j\in\{1,\dots,d\}}.$$

It is easily seen that $\mathbf{M_h} \in \mathbf{R}^{d,d}$ is symmetric and positive definite. In order to obtain a corresponding representation for $\mathcal{A}_h(t)\colon V_h \to V_h$, $t \in [0, T]$, we introduce $\mathbf{A_h}\colon [0, T] \times \mathbf{R}^d \to \mathbf{R}^d$ such that for $t \in [0, T]$ and $\mathbf{x} \in \mathbf{R}^d$ the vector $\mathbf{A_h}(t, \mathbf{x}) \in \mathbf{R}^d$ is determined by

$$\sum_{i=1}^{d} \big[\mathbf{A_h}(t, \mathbf{x})\big]_i \psi_i = \mathcal{A}_h(t) v_{\mathbf{x}} \in V_h,$$

where $v_{\mathbf{x}} = \sum_{i=1}^{d} \mathbf{x}_i \psi_i$. Then, (6.6) can equivalently be written as

$$\mathbf{M_h u_h^n} + k\mathbf{M_h A_h}(\xi_n, \mathbf{u_h^n}) = k\mathbf{M_h f_h}(\xi_n) + \mathbf{M_h u_h^{n-1}},$$

or simply

$$\mathbf{u_h^n} = \mathbf{u_h^{n-1}} + k\big(\mathbf{f_h}(\xi_n) - \mathbf{A_h}(\xi_n, \mathbf{u_h^n})\big).$$

In order to transfer the monotonicity and Lipschitz continuity of $\mathcal{A}_h$ to its counterpart, we introduce the following inner product and norm in $\mathbf{R}^d$:

$$(\mathbf{x}, \mathbf{y})_{\mathbf{M_h}} = \mathbf{x}^T \mathbf{M_h y} \quad \text{and} \quad \|\mathbf{x}\|_{\mathbf{M_h}} = \sqrt{(\mathbf{x}, \mathbf{x})_{\mathbf{M_h}}}$$

for $\mathbf{x}, \mathbf{y} \in \mathbf{R}^d$. This particular choice of inner product coincides with the inner product of $H$ of the elements $u_{\mathbf{x}}, v_{\mathbf{x}} \in H$ given by

$$u_{\mathbf{x}} = \sum_{i=1}^{d} \mathbf{x}_i \psi_i \quad \text{and} \quad v_{\mathbf{y}} = \sum_{i=1}^{d} \mathbf{y}_i \psi_i$$

for $\mathbf{x}, \mathbf{y} \in \mathbf{R}^d$, i.e., the following equalities hold:

$$(\mathbf{x}, \mathbf{y})_{\mathbf{M_h}} = (u_{\mathbf{x}}, v_{\mathbf{y}})_H \quad \text{and} \quad \|\mathbf{x}\|_{\mathbf{M_h}} = \|u_{\mathbf{x}}\|_H.$$

To prove the existence of an element $\mathbf{u_h^n}(\omega)$ for almost every $\omega \in \Omega$, we use Lemma 4.1. To this end, we introduce the function

$$\mathbf{g}\colon \Omega \times \mathbf{R}^d \to \mathbf{R}^d, \quad \mathbf{g}(\omega, \mathbf{x}) = \mathbf{x} - \mathbf{u_h^{n-1}}(\omega) - k\big(\mathbf{f_h}(\xi_n(\omega)) - \mathbf{A_h}(\xi_n(\omega), \mathbf{x})\big).$$

Define $q = q(\omega) := \|U_h^{n-1}(\omega)\|_H + k\|f(\xi_n(\omega))\|_H$. Observe that for almost every $\omega \in \Omega$ we have $q(\omega) \in [0, \infty)$. In the following, we consider an arbitrary but fixed $\omega \in \Omega$ with this property. Next, we introduce

$$R = R(\omega) = \frac{q}{2} + \sqrt{\frac{q^2}{4} + k\frac{M^2}{4\mu}} \in (0, \infty).$$

For all $\mathbf{x} \in \mathbf{R}^d$ with $\|\mathbf{x}\|_{\mathbf{M_h}} = R$, we then obtain

$$\begin{aligned}
&(\mathbf{g}(\omega, \mathbf{x}), \mathbf{x})_{\mathbf{M_h}} \\
&\quad = (\mathbf{x} - \mathbf{u_h^{n-1}}(\omega) - k\big(\mathbf{f_h}(\xi_n(\omega)) - \mathbf{A_h}(\xi_n(\omega), \mathbf{x})\big), \mathbf{x})_{\mathbf{M_h}} \\
&\quad = \|\mathbf{x}\|^2_{\mathbf{M_h}} - (\mathbf{u_h^{n-1}}(\omega), \mathbf{x})_{\mathbf{M_h}} - k(\mathbf{f_h}(\xi_n(\omega)), \mathbf{x})_{\mathbf{M_h}} + k(\mathbf{A_h}(\xi_n(\omega), \mathbf{x}), \mathbf{x})_{\mathbf{M_h}} \\
&\quad \geq R^2 - R\|U_h^{n-1}(\omega)\|_H - kR\|f(\xi_n(\omega))\|_H + k\langle \mathcal{A}(\xi_n(\omega))v_{\mathbf{x}}, v_{\mathbf{x}}\rangle_{V^*,V}
\end{aligned} \tag{6.7}$$

where $v_{\mathbf{x}} = \sum_{i=1}^d \mathbf{x}_i \psi_i$. Using Assumption 6.1 (ii) and (iv), the last summand of (6.7) can be estimated by

$$\begin{aligned}
&\langle \mathcal{A}(\xi_n(\omega))v_{\mathbf{x}}, v_{\mathbf{x}}\rangle_{V^*,V} \\
&\quad = \langle \mathcal{A}(\xi_n(\omega))v_{\mathbf{x}} - \mathcal{A}(\xi_n(\omega))0, v_{\mathbf{x}} - 0\rangle_{V^*,V} + \langle \mathcal{A}(\xi_n(\omega))0, v_{\mathbf{x}}\rangle_{V^*,V} \\
&\quad \geq \mu\|v_{\mathbf{x}}\|_V^2 - \|\mathcal{A}(\xi_n(\omega))0\|_{V^*}\|v_{\mathbf{x}}\|_V \\
&\quad \geq \mu\|v_{\mathbf{x}}\|_V^2 - M\|v_{\mathbf{x}}\|_V \\
&\quad \geq \mu\|v_{\mathbf{x}}\|_V^2 - \frac{M^2}{4\mu} - \mu\|v_{\mathbf{x}}\|_V^2 = -\frac{M^2}{4\mu}.
\end{aligned} \tag{6.8}$$

Therefore, after inserting $R$, we obtain

$$(\mathbf{g}(\omega, \mathbf{x}), \mathbf{x})_{\mathbf{M_h}} \geq R^2 - R\big(\|U_h^{n-1}(\omega)\|_H + k\|f(\xi_n(\omega))\|_H\big) - k\frac{M^2}{4\mu} = 0.$$

Since $\mathbf{A_h}(\xi_n(\omega), \cdot)$ is continuous in the second argument due to Assumption 6.1 (iii), this allows us to apply Lemma 4.1 and Remark 4.2. Thus, for almost every $\omega \in \Omega$ we obtain the existence of an element $\mathbf{x} = \mathbf{x}(\omega) \in \mathbf{R}^d$ such that $\mathbf{g}(\omega, \mathbf{x}) = 0$ holds. To prove that this root is unique, assume that there exist $\mathbf{x}, \mathbf{y} \in \mathbf{R}^d$ such that

$$\mathbf{g}(\omega, \mathbf{x}) = 0 \quad \text{and} \quad \mathbf{g}(\omega, \mathbf{y}) = 0$$

is fulfilled. Then, inserting the definition of the function $\mathbf{g}$ leads to

$$\begin{aligned}
\|\mathbf{x} - \mathbf{y}\|^2_{\mathbf{M_h}} &= -k(\mathbf{A_h}(\xi_n(\omega), \mathbf{x}) - \mathbf{A_h}(\xi_n(\omega), \mathbf{y}), \mathbf{x} - \mathbf{y})_{\mathbf{M_h}} \\
&= -k(\mathcal{A}_h(\xi_n(\omega))v_{\mathbf{x}} - \mathcal{A}_h(\xi_n(\omega))v_{\mathbf{y}}, v_{\mathbf{x}} - v_{\mathbf{y}})_H \leq 0.
\end{aligned}$$

This implies $\mathbf{x} = \mathbf{y}$. An application of Lemma 4.3 then yields that the mapping

$$\mathbf{u_h^n} \colon \Omega \to \mathbf{R}^d, \quad \omega \mapsto \mathbf{u_h^n}(\omega) = \big(\alpha_{h,i}^n(\omega)\big)_{i \in \{1,\ldots,d\}}$$

is $\mathcal{F}_n$-measurable, where $\mathbf{u_h^n}(\omega) = \mathbf{x}(\omega)$ is the unique root of $\mathbf{g}(\omega, \cdot)$. To sum up,

$$U_h^n = \sum_{i=1}^d \alpha_{h,i}^n \psi_i, \quad n \in \{0, \ldots, N\},$$

is the well-defined solution to the scheme (6.5). Since $\{\psi_1, \dots, \psi_d\}$ is a basis of $V_h$, this implies that $U_h^n(\omega) \in V_h$ for almost every $\omega \in \Omega$ and all $n \in \{0, \dots, N\}$. □

**Lemma 6.6** *Let Assumption 6.1 be satisfied. Then, for every $f \in L^2(0, T; H)$, every initial value $U_h^0 \in V_h$, and every step size $k = \frac{T}{N}$, $N \in \mathbf{N}$, the unique solution $(U_h^n)_{n\in\{0,\dots,N\}}$ to the implicit scheme (6.5) satisfies the a priori bound*

$$\max_{n\in\{0,\dots,N\}} \mathbf{E}\big[\|U_h^n\|_H^2\big] + \sum_{j=1}^N \mathbf{E}\big[\|U_h^j - U_h^{j-1}\|_H^2\big] + k\mu \sum_{j=1}^N \mathbf{E}\big[\|U_h^j\|_V^2\big]$$
$$\leq C\big(T + \|U_h^0\|_H^2 + \|f\|_{L^2(0,T;H)}^2\big),$$

*where the constant $C$ only depends on $M$, $\mu$, and the embedding $V \hookrightarrow H$.*

***Proof*** Due to the definition of scheme (6.5), we can write for all $j \in \{1, \dots, N\}$

$$\frac{1}{k}\big(U_h^j - U_h^{j-1}\big) + \mathcal{A}_h(\xi_j)U_h^j = f(\xi_j).$$

We test this equation with $U_h^j$ in the $H$ inner product and apply the polarization identity

$$\frac{1}{2}\big(\|U_h^j\|_H^2 - \|U_h^{j-1}\|_H^2 + \|U_h^j - U_h^{j-1}\|_H^2\big) = (U_h^j - U_h^{j-1}, U_h^j)_H.$$

In addition, recall from (6.4) and (6.8) that

$$\big(\mathcal{A}_h(\xi_j)U_h^j, U_h^j\big)_H = \big\langle \mathcal{A}(\xi_j)U_h^j, U_h^j\big\rangle_{V^*,V} \geq \mu\big\|U_h^j\big\|_V^2 - \big\|\mathcal{A}(\xi_j)0\big\|_{V^*}\big\|U_h^j\big\|_V.$$

From this and (6.5) as well as from Assumption 6.1 (ii), we obtain that

$$\begin{aligned}
&\frac{1}{2k}\big(\|U_h^j\|_H^2 - \|U_h^{j-1}\|_H^2 + \|U_h^j - U_h^{j-1}\|_H^2\big) + \mu\|U_h^j\|_V^2 \\
&\quad\leq \frac{1}{k}(U_h^j - U_h^{j-1}, U_h^j)_H + (\mathcal{A}_h(\xi_j)U_h^j, U_h^j)_H + \big\|\mathcal{A}(\xi_j)0\big\|_{V^*}\big\|U_h^j\big\|_V \\
&\quad= (f(\xi_j), U_h^j)_H + M\big\|U_h^j\big\|_V \\
&\quad\leq \|f(\xi_j)\|_H\|U_h^j\|_H + M\|U_h^j\|_V \\
&\quad\leq C\big(1 + \|f(\xi_j)\|_H^2\big) + \frac{\mu}{2}\|U_h^j\|_V^2,
\end{aligned}$$

where the constant $C$ only depends on $M$, $\mu$, and the embedding $V \hookrightarrow H$. Next we sum up with respect to $j$ from 1 to $n$ and obtain

$$\begin{aligned}&\|U_h^n\|_H^2 + \sum_{j=1}^{n} \|U_h^j - U_h^{j-1}\|_H^2 + k\mu \sum_{j=1}^{n} \|U_h^j\|_V^2 \\ &\quad \le \|U_h^0\|_H^2 + k2C \sum_{j=1}^{n} \big(1 + \|f(\xi_j)\|_H^2\big).\end{aligned}$$

Taking the expectation, we further obtain for the term containing the inhomogeneity $f$ that

$$\mathbf{E}\Big[k \sum_{j=1}^{n} \|f(\xi_j)\|_H^2\Big] = \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \|f(t)\|_H^2 \,\mathrm{d}t \le \|f\|_{L^2(0,T;H)}^2$$

holds. This completes the proof. □

After these preparatory results, we can now state the abstract convergence result for the numerical method (6.5). For its formulation, we define for each $v \in H$

$$\mathrm{dist}_H(v, V_h) := \inf_{v_h \in V_h} \|v - v_h\|_H. \tag{6.9}$$

Similarly, if $v \in V$ we set

$$\mathrm{dist}_V(v, V_h) := \inf_{v_h \in V_h} \|v - v_h\|_V. \tag{6.10}$$

With (6.9), we therefore measure how well a given element $v \in H$ can be approximated by elements from $V_h$. Since $V_h$ is finite dimensional, it is clear that $P_h v \in V_h$ has the best approximation properties with respect to the $H$-norm, that is

$$\|P_h v - v\|_H = \mathrm{dist}_H(v, V_h), \quad \text{for all } v \in H. \tag{6.11}$$

In the same way, if we define $Q_h \colon V \to V_h$ as the orthogonal projection onto $V_h$ with respect to the inner product $(\cdot,\cdot)_V$, then it holds true that

$$\|Q_h v - v\|_V = \mathrm{dist}_V(v, V_h), \quad \text{for all } v \in V. \tag{6.12}$$

Since we consider a general Galerkin method in this section, we will not quantify the best approximation property of $(V_h)_{h\in(0,1)}$ at this point.

We also mention that the error estimate in Theorem 6.7 requires the boundedness of $\|P_h\|_{\mathcal{L}(V)}$. However, one cannot expect that $\sup_{h\in(0,1]} \|P_h\|_{\mathcal{L}(V)} < \infty$ is fulfilled in general. For a discussion of the stability of the orthogonal projector $P_h$ in case of the Galerkin finite element method, we refer to [6–8,11].

**Theorem 6.7** *Let Assumption 6.1 be satisfied. Then, for a given inhomogeneity $f \in L^2(0,T;H)$ and initial value $u_0 \in V$ let $u$ be the unique weak solution to the abstract problem (6.1). In addition, we assume that there exists $\gamma \in (0,1)$ with*

$$u \in C^{\gamma}([0,T];V) \tag{6.13}$$

*as well as*

$$\int_0^T \|\mathcal{A}(t)u(t)\|_H^2 \,\mathrm{d}t < \infty. \tag{6.14}$$

*Then, there exists a constant $C$ only depending on $L$, $\mu$, and $T$ such that for every step size $k = \frac{T}{N}$, $N \in \mathbf{N}$, and $h \in (0,1)$ we have*

$$\begin{aligned}
&\max_{n\in\{0,\dots,N\}} \|U_h^n - u(t_n)\|_{L^2(\Omega;H)} + \Big( k \sum_{n=1}^{N} \|U_h^n - u(t_n)\|^2_{L^2(\Omega;V)} \Big)^{\frac{1}{2}} \\
&\le C \Bigg( k^{\frac{1}{2}} \big( \|f\|_{L^2(0,T;H)} + \|\mathcal{A}(\cdot)u(\cdot)\|_{L^2(0,T;H)} \big) + k^{\gamma} \|u\|_{C^{\gamma}([0,T];V)} \\
&\quad + \max_{n\in\{0,\dots,N\}} \mathrm{dist}_H(u(t_n), V_h) \\
&\quad + \big(1 + \|P_h\|_{\mathcal{L}(V)}\big) \Big( k \sum_{n=1}^{N} \mathrm{dist}_V(u(t_n), V_h)^2 \Big)^{\frac{1}{2}} \Bigg),
\end{aligned}$$

*where $(U_h^n)_{n\in\{0,\dots,N\}} \subset L^2(\Omega; V_h)$ is given by the scheme (6.5).*

***Proof*** Throughout the proof, we consider an arbitrary but fixed finite-dimensional subspace $V_h$, $h \in (0,1)$, of $V$. Moreover, we denote the error of the scheme (6.5) at the time $t_n$ by $E^n$, i.e., $E^n := U_h^n - u(t_n)$ for each $n \in \{0,\dots,N\}$. Note that for every $n \ge 1$ we have $E^n \in L^2(\Omega; V)$ since $U_h^n \in L^2(\Omega; V_h)$ by Lemmata 6.5 and 6.6. In addition, due to (6.13) we have $u(t) \in V$ for every $t \in [0,T]$.

In the first step, we split the error into two parts using the orthogonal projection $P_h : H \to V_h$ by

$$E^n = P_h E^n + (I - P_h)E^n =: \Theta^n + \Xi^n.$$

Due to the orthogonality of $P_h$ with respect to the inner product in $H$, we have

$$\|E^n\|_H^2 = \|\Theta^n\|_H^2 + \|\Xi^n\|_H^2$$

for every $n \in \{0,\dots,N\}$. By taking note of (6.11) we obtain

$$\|\Xi^n\|_H = \mathrm{dist}_H(u(t_n), V_h)$$

since $\Xi^n = (I - P_h)E^n = (P_h - I)u(t_n)$. In addition, we have

$$\Big( k \sum_{n=1}^{N} \|E^n\|^2_{L^2(\Omega;V)} \Big)^{\frac{1}{2}} \le \Big( k \sum_{n=1}^{N} \|\Theta^n\|^2_{L^2(\Omega;V)} \Big)^{\frac{1}{2}} + \Big( k \sum_{n=1}^{N} \|\Xi^n\|_V^2 \Big)^{\frac{1}{2}}$$

since $\Xi^n = (P_h - I)u(t_n)$ is deterministic. After adding and subtracting the orthogonal projector $Q_h \colon V \to V_h$, we further obtain the estimate

$$\begin{aligned}\| \Xi^n \|_V &= \big\|(P_h - I)u(t_n)\big\|_V \leq \big\| P_h(I - Q_h)u(t_n)\big\|_V + \big\|(Q_h - I)u(t_n)\big\|_V \\ &\leq \big(1 + \|P_h\|_{\mathcal{L}(V)}\big)\|(Q_h - I)u(t_n)\| = \big(1 + \|P_h\|_{\mathcal{L}(V)}\big)\mathrm{dist}_V(u(t_n), V_h),\end{aligned}$$

due to (6.12). This shows that

$$k \sum_{n=1}^{N} \| \Xi^n \|_V^2 \leq \big(1 + \|P_h\|_{\mathcal{L}(V)}\big)^2 k \sum_{n=1}^{N} \mathrm{dist}_V(u(t_n), V_h)^2. \tag{6.15}$$

Thus, it remains to estimate $\mathbf{E}[\|\Theta^n\|_H^2]$ and $k \sum_{n=1}^{N} \mathbf{E}[\|\Theta^n\|_V^2]$. To this end, we apply the polarization identity

$$\frac{1}{2}\big(\|\Theta^n\|_H^2 - \|\Theta^{n-1}\|_H^2 + \|\Theta^n - \Theta^{n-1}\|_H^2\big) = (\Theta^n - \Theta^{n-1}, \Theta^n)_H, \tag{6.16}$$

which holds for every $n \in \{1, \dots, N\}$. From the orthogonality of $P_h$ with respect to the inner product in $H$, we further have

$$(\Theta^n - \Theta^{n-1}, \Theta^n)_H = (E^n - E^{n-1}, \Theta^n)_H$$

which motivates us to consider the term $E^n - E^{n-1}$ tested with $\Theta^n$ in what follows.

To estimate the difference of the errors $E^n - E^{n-1} = U_h^n - u(t_n) - U_h^{n-1} + u(t_{n-1})$, we insert the definition of the scheme (6.5) and (6.3). This yields

$$\begin{aligned}\big(U_h^n - U_h^{n-1}, \Theta^n\big)_H &= k\big(P_h f(\xi_n) - \mathcal{A}_h(\xi_n)U_h^n, \Theta^n\big)_H \\ &= k\big(f(\xi_n), \Theta^n\big)_H - k\big\langle \mathcal{A}(\xi_n)U_h^n, \Theta^n\big\rangle_{V^*,V}.\end{aligned}$$

Moreover, since the random variable $\Theta^n$ takes values in $V_h \subset V$, we get from the canonical embedding $H \cong H^* \hookrightarrow V^*$ and (6.2) that

$$\begin{aligned}\big(u(t_n) - u(t_{n-1}), \Theta^n\big)_H &= \big\langle u(t_n) - u(t_{n-1}), \Theta^n\big\rangle_{V^*,V} \\ &= \int_{t_{n-1}}^{t_n} \big\langle \dot{u}(s), \Theta^n\big\rangle_{V^*,V} \,\mathrm{d}s \\ &= \int_{t_{n-1}}^{t_n} \big\langle f(s) - \mathcal{A}(s)u(s), \Theta^n\big\rangle_{V^*,V} \,\mathrm{d}s.\end{aligned}$$

Therefore, altogether we obtain the following representation

$$\begin{aligned}(E^n - E^{n-1}, \Theta^n)_H &= -k\big\langle \mathcal{A}(\xi_n)U_h^n - \mathcal{A}(\xi_n)u(t_n), \Theta^n\big\rangle_{V^*,V}\\&\quad - k\big\langle \mathcal{A}(\xi_n)u(t_n) - \mathcal{A}(\xi_n)u(\xi_n), \Theta^n\big\rangle_{V^*,V}\\&\quad - \int_{t_{n-1}}^{t_n} \big\langle \mathcal{A}(\xi_n)u(\xi_n) - \mathcal{A}(s)u(s), \Theta^n\big\rangle_{V^*,V}\,\mathrm{d}s\\&\quad + \int_{t_{n-1}}^{t_n} \big(f(\xi_n) - f(s), \Theta^n\big)_H\,\mathrm{d}s\\&=: \Gamma_1 + \Gamma_2 + \Gamma_3 + \Gamma_4.\end{aligned} \tag{6.17}$$

We give estimates for the four terms $\Gamma_i$, $i \in \{1, \dots, 4\}$, in (6.17) separately. By recalling $\Theta^n = P_h E^n = U_h^n - P_h u(t_n)$, the first term is estimated using Assumption 6.1 (iii) and (iv) as follows:

$$\begin{aligned}\Gamma_1 &= -k\big\langle \mathcal{A}(\xi_n)U_h^n - \mathcal{A}(\xi_n)P_h u(t_n), \Theta^n\big\rangle_{V^*,V}\\&\quad - k\big\langle \mathcal{A}(\xi_n)P_h u(t_n) - \mathcal{A}(\xi_n)u(t_n), \Theta^n\big\rangle_{V^*,V}\\&\le -k\mu\big\|\Theta^n\big\|_V^2 + kL\big\|(I - P_h)u(t_n)\big\|_V\big\|\Theta^n\big\|_V\\&\le -k\mu\big\|\Theta^n\big\|_V^2 + k\frac{L^2}{\mu}\big\|\Xi^n\big\|_V^2 + k\frac{\mu}{4}\big\|\Theta^n\big\|_V^2.\end{aligned} \tag{6.18}$$

Observe that we also applied the weighted Young inequality in the last step.

We similarly obtain an estimate for the second summand in (6.17) of the form

$$\begin{aligned}\Gamma_2 &\le k\big\|\mathcal{A}(\xi_n)u(t_n) - \mathcal{A}(\xi_n)u(\xi_n)\big\|_{V^*}\big\|\Theta^n\big\|_V\\&\le k\frac{L^2}{\mu}\|u(t_n) - u(\xi_n)\|_V^2 + k\frac{\mu}{4}\big\|\Theta^n\big\|_V^2.\end{aligned}$$

Since $|t_n - \xi_n(\omega)| \le k$ for every $\omega \in \Omega$ and $u \in C^\gamma([0, T]; V)$, we therefore conclude

$$\Gamma_2 \le k^{1+2\gamma}\frac{L^2}{\mu}\|u\|_{C^\gamma([0,T];V)}^2 + k\frac{\mu}{4}\big\|\Theta^n\big\|_V^2. \tag{6.19}$$

Concerning the term $\Gamma_3$ in (6.17), let us recall that both $\Theta^n$ and $\xi_n$ are square-integrable random variables which are $\mathcal{F}_n$-measurable. Moreover, $\Theta^n$ takes values in $V_h \subset V$, while $\omega \mapsto \mathcal{A}(\xi_n(\omega))u(\xi_n(\omega))$ takes almost surely values in $H$ due to (6.14).

Therefore, after taking expectation we obtain

$$\begin{aligned}\mathbf{E}\big[\Gamma_3\big] &= \mathbf{E}\Big[\int_{t_{n-1}}^{t_n} \big\langle \mathcal{A}(\xi_n)u(\xi_n) - \mathcal{A}(s)u(s),\, \Theta^n\big\rangle_{V^*,V}\,\mathrm{d}s\Big]\\ &= \mathbf{E}\Big[\int_{t_{n-1}}^{t_n} \big(\mathcal{A}(\xi_n)u(\xi_n) - \mathcal{A}(s)u(s),\, \Theta^n\big)_H\,\mathrm{d}s\Big]\\ &= \mathbf{E}\Big[\int_{t_{n-1}}^{t_n} \big(\mathcal{A}(\xi_n)u(\xi_n) - \mathcal{A}(s)u(s),\, \Theta^n - \Theta^{n-1}\big)_H\,\mathrm{d}s\Big]\\ &\quad + \mathbf{E}\Big[\int_{t_{n-1}}^{t_n} \big(\mathcal{A}(\xi_n)u(\xi_n) - \mathcal{A}(s)u(s),\, \Theta^{n-1}\big)_H\,\mathrm{d}s\Big].\end{aligned} \tag{6.20}$$

Standard arguments then directly yield a bound for the first summand of the form

$$\begin{aligned}&\mathbf{E}\Big[\int_{t_{n-1}}^{t_n} \big(\mathcal{A}(\xi_n)u(\xi_n) - \mathcal{A}(s)u(s),\, \Theta^n - \Theta^{n-1}\big)_H\,\mathrm{d}s\Big]\\ &\le \Big(\mathbf{E}\Big[\int_{t_{n-1}}^{t_n} \|\mathcal{A}(\xi_n)u(\xi_n) - \mathcal{A}(s)u(s)\|_H^2\,\mathrm{d}s\Big]\Big)^{\frac{1}{2}} \big(k\mathbf{E}\big[\|\Theta^n - \Theta^{n-1}\|_H^2\big]\big)^{\frac{1}{2}}\\ &\le k\mathbf{E}\Big[\int_{t_{n-1}}^{t_n} \|\mathcal{A}(\xi_n)u(\xi_n) - \mathcal{A}(s)u(s)\|_H^2\,\mathrm{d}s\Big] + \frac{1}{4}\mathbf{E}\big[\|\Theta^n - \Theta^{n-1}\|_H^2\big]\\ &\le 4k\int_{t_{n-1}}^{t_n} \|\mathcal{A}(s)u(s)\|_H^2\,\mathrm{d}s + \frac{1}{4}\mathbf{E}\big[\|\Theta^n - \Theta^{n-1}\|_H^2\big],\end{aligned}$$

where the last step follows from

$$\mathbf{E}\big[k\|\mathcal{A}(\xi_n)u(\xi_n)\|_H^2\big] = \int_{t_{n-1}}^{t_n} \|\mathcal{A}(s)u(s)\|_H^2\,\mathrm{d}s.$$

To estimate the second summand in (6.20), we make use of the tower property for conditional expectations and the fact that $\Theta^{n-1}$ is $\mathcal{F}_{n-1}$-measurable. This yields

$$\begin{aligned}&\mathbf{E}\Big[\int_{t_{n-1}}^{t_n} \big(\mathcal{A}(\xi_n)u(\xi_n) - \mathcal{A}(s)u(s),\, \Theta^{n-1}\big)_H\,\mathrm{d}s\Big]\\ &= \mathbf{E}\Big[\int_{t_{n-1}}^{t_n} \mathbf{E}\big[\big(\mathcal{A}(\xi_n)u(\xi_n) - \mathcal{A}(s)u(s),\, \Theta^{n-1}\big)_H \big|\mathcal{F}_{n-1}\big]\,\mathrm{d}s\Big]\\ &= \mathbf{E}\Big[k\mathbf{E}\big[\big(\mathcal{A}(\xi_n)u(\xi_n),\, \Theta^{n-1}\big)_H\big|\mathcal{F}_{n-1}\big] - \int_{t_{n-1}}^{t_n} \big(\mathcal{A}(s)u(s),\, \Theta^{n-1}\big)_H\,\mathrm{d}s\Big] = 0,\end{aligned}$$

where the last step follows from

$$\begin{aligned}k\mathbf{E}\big[\big(\mathcal{A}(\xi_n)u(\xi_n),\, \Theta^{n-1}\big)_H\big|\mathcal{F}_{n-1}\big] &= \big(k\mathbf{E}\big[\mathcal{A}(\xi_n)u(\xi_n)\big|\mathcal{F}_{n-1}\big],\, \Theta^{n-1}\big)_H\\ &= \big(k\mathbf{E}\big[\mathcal{A}(\xi_n)u(\xi_n)\big],\, \Theta^{n-1}\big)_H\\ &= \int_{t_{n-1}}^{t_n} \big(\mathcal{A}(s)u(s),\, \Theta^{n-1}\big)_H\,\mathrm{d}s\end{aligned}$$

due to the independence of $\xi_n$ from $\mathcal{F}_{n-1}$. Altogether this shows

$$\mathbf{E}\big[\Gamma_3\big] \le 4k \int_{t_{n-1}}^{t_n} \|\mathcal{A}(s)u(s)\|_H^2 \,\mathrm{d}s + \frac{1}{4}\mathbf{E}\big[\|\Theta^n - \Theta^{n-1}\|_H^2\big]. \tag{6.21}$$

The same steps with $f(\cdot)$ in place of $\mathcal{A}(\cdot)u(\cdot)$ also yield an estimate for $\Gamma_4$. Therefore,

$$\mathbf{E}\big[\Gamma_4\big] \le 4k \int_{t_{n-1}}^{t_n} \|f(s)\|_H^2 \,\mathrm{d}s + \frac{1}{4}\mathbf{E}\big[\|\Theta^n - \Theta^{n-1}\|_H^2\big]. \tag{6.22}$$

In summary, after taking expectation and inserting (6.18), (6.19), (6.21), and (6.22) into (6.16) we obtain

$$\begin{aligned}
&\frac{1}{2}\mathbf{E}\big[\|\Theta^n\|_H^2 - \|\Theta^{n-1}\|_H^2 + \|\Theta^n - \Theta^{n-1}\|_H^2\big]\\
&\quad = \mathbf{E}\big[\Gamma_1 + \Gamma_2 + \Gamma_3 + \Gamma_4\big]\\
&\quad \le -\frac{1}{2}k\mu\mathbf{E}\big[\big\|\Theta^n\big\|_V^2\big] + k\frac{L^2}{\mu}\big\|\Xi^n\big\|_V^2 + k^{1+2\gamma}\frac{L^2}{\mu}\|u\|_{C^\gamma([0,T];V)}^2\\
&\qquad + 4k\int_{t_{n-1}}^{t_n} \|\mathcal{A}(s)u(s)\|_H^2\,\mathrm{d}s + 4k\int_{t_{n-1}}^{t_n}\|f(s)\|_H^2\,\mathrm{d}s\\
&\qquad + \frac{1}{2}\mathbf{E}\big[\|\Theta^n - \Theta^{n-1}\|_H^2\big].
\end{aligned}$$

After canceling the last term from both sides of the inequality, we sum over $n \in \{1, \ldots, j\}$ for some arbitrary $j \in \{1, \ldots, N\}$. Moreover, since $U_h^0 = P_h u_0$ we also have $\Theta^0 = 0$. Hence, we obtain

$$\begin{aligned}
&\mathbf{E}\big[\|\Theta^j\|_H^2\big] + \frac{1}{2}k\mu\sum_{n=1}^{j}\mathbf{E}\big[\|\Theta^n\|_V^2\big]\\
&\quad \le k\frac{L^2}{\mu}\sum_{n=1}^{j}\big\|\Xi^j\big\|_V^2 + k^{2\gamma}T\frac{L^2}{\mu}\|u\|_{C^\gamma([0,T];V)}^2\\
&\qquad + 4k\big(\|\mathcal{A}(\cdot)u(\cdot)\|_{L^2(0,T;H)}^2 + \|f\|_{L^2(0,T;H)}^2\big).
\end{aligned}$$

The proof is completed by taking the maximum over $j \in \{1, \ldots, N\}$ and an application of (6.15). □

***Remark 6.8*** Let us briefly discuss the additional regularity conditions (6.13) and (6.14) in Theorem 6.7. First note that since $f \in L^2(0, T; H)$, the condition (6.14) is essentially equivalent to

$$\dot{u} \in L^2(0, T; H).$$

A sufficient condition for (6.13) is then to additionally require

$$\dot{u} = f - \mathcal{A}(\cdot)u(\cdot) \in L^q(0, T; V)$$

with $q = \frac{1}{1-\gamma}$. In Sect. 7, we will discuss more explicit classes of linear and semilinear evolution equations, whose solutions enjoy the required regularity.

## 7 Regularity of Non-autonomous Evolution Equations

To prove a rate of convergence in Sect. 6, we had to impose additional assumptions on the regularity of the exact solution $u$. In the following, we will discuss cases where this particular regularity can be expected.

We begin by considering a class of linear problems that fulfills the regularity conditions imposed in Theorem 6.7. As in Sect. 6, we consider the real, separable Hilbert spaces $(V, (\cdot,\cdot)_V, \|\cdot\|_V)$ and $(H, (\cdot,\cdot)_H, \|\cdot\|_H)$ that form the Gelfand triple

$$V \overset{c,d}{\hookrightarrow} H \cong H^* \overset{c,d}{\hookrightarrow} V^*.$$

Further, we state the following assumption to obtain a suitable evolution operator.

**Assumption 7.1** For all $t \in [0, T]$ let $a_0(t; \cdot,\cdot)\colon V \times V \to \mathbf{R}$ be a bilinear form that fulfills the following conditions:

(i) For every $v_1, v_2 \in V$ the mapping $a_0(\cdot; v_1, v_2)\colon [0, T] \to \mathbf{R}$ is measurable.
(ii) There exists $\beta \in (0, \infty)$ such that for all $t \in [0, T]$ it holds true that

$$|a_0(t; v_1, v_2)| \le \beta \|v_1\|_V \|v_2\|_V, \quad \text{for all } v_1, v_2 \in V.$$

(iii) There exists $\mu \in (0, \infty)$ such that for all $t \in [0, T]$ it holds true that

$$a_0(t; v, v) \ge \mu \|v\|_V^2, \quad \text{for all } v \in V.$$

For every $t \in [0, T]$, let $\mathcal{A}_0(t)\colon V \to V^*$ and $A_0(t)\colon \mathrm{dom}(A_0(t)) \subset H \to H$ denote the associated operators to the bilinear form $a_0(t; \cdot,\cdot)$ from Assumption 7.1. More precisely, the linear operator $\mathcal{A}_0(t)$ is uniquely determined by

$$\left\langle \mathcal{A}_0(t)v_1, v_2 \right\rangle_{V^*,V} = a_0(t; v_1, v_2), \quad \text{for all } v_1, v_2 \in V.$$

Moreover, we set $\mathrm{dom}(A_0(t)) := \{v \in V : \mathcal{A}_0(t)v \in H\}$ and define $A_0(t)$ as the restriction of $\mathcal{A}_0(t)$ to the domain $\mathrm{dom}(A_0(t))$. Note that $\mathrm{dom}(A_0(t))$ becomes a Banach space if endowed with the graph norm. For given $u_0 \in V$ and $f \in L^2(0, T; H)$, we consider the following linear and non-autonomous problem

$$\begin{cases} \dot{u}(t) + \mathcal{A}_0(t)u(t) = f(t), & t \in (0, T], \\ u(0) = u_0. \end{cases} \tag{7.1}$$

For the formulation of the regularity result, we first recall that the initial value problem (7.1) is said to have *maximal $L^p$-regularity in $H$* for some $p \in [2, \infty)$ if for all $f \in L^p(0, T; H)$ we have a unique solution $u \in \mathcal{W}(0, T)$ with $\dot{u} \in L^p(0, T; H)$ and $\mathcal{A}_0(\cdot)u(\cdot) \in L^p(0, T; H)$.

**Theorem 7.2** *For every $t \in [0, T]$, let $a_0(t; \cdot, \cdot): V \times V \to \mathbf{R}$ fulfill Assumption 7.1. Suppose there exists $M \geq 0$ such that $\|A_0(0)A_0^{-1}(t)\|_{\mathcal{L}(H)} \leq M$ holds true for all $t \in [0, T]$, that the embedding $\mathrm{dom}(A_0(0)) \hookrightarrow H$ is compact, and that $\mathrm{dom}(A_0(0)^{\frac{1}{2}}) = V$ (Kato's square root property). Fix $u_0 \in V$ and assume that the Cauchy problem (7.1) has maximal $L^p$-regularity in $H$ for some $p \in (2, \infty)$. Then, $u \in C^\gamma([0, T]; V)$ for $\gamma < \frac{1}{2} - \frac{1}{p}$.*

***Proof*** Since (7.1) has maximal $L^p$-regularity, it follows that $u \in W^{1,p}(0, T; H)$ and that $\int_0^T \|A_0(t)u(t)\|_H^p \, \mathrm{d}t < \infty$. As $\|A_0(0)A_0^{-1}(t)\|_{\mathcal{L}(H)} \leq M$ for all $t \in [0, T]$, we have that $u \in L^p(0, T; \mathrm{dom}(A_0(0)))$. Let $0 < \epsilon < \frac{1}{2} - \frac{1}{p}$ and consider the continuous embeddings of real interpolation spaces (for the precise definition of an interpolation space, see, for example, [41, Sec. 1.3.2.])

$$(H, \mathrm{dom}(A_0(0)))_{\frac{1}{2}+\epsilon, p} \hookrightarrow (H, \mathrm{dom}(A_0(0)))_{\frac{1}{2}, 1} \hookrightarrow \mathrm{dom}(A_0(0)^{\frac{1}{2}}) = V,$$

where a proof for the first embedding can be found in, for example, [41, Sec. 1.3.3.] and in [41, Sec. 1.15.2.] for the second. Then, it follows from [2, Thm 5.2] that for all $0 \leq \gamma < \frac{1}{2} - \epsilon - \frac{1}{p}$ we have $u \in C^\gamma([0, T]; V)$ and the proof is complete. □

A sufficient condition for maximal $L^p$-regularity for (7.1) is found in [18]: If $a_0(t; \cdot, \cdot): V \times V \mapsto \mathbf{R}$ fulfills Assumption 7.1 and

$$|a_0(s; u, v) - a_0(t, u, v)| \leq \omega(|s - t|)\|u\|_V \|v\|_V,$$

for all $u, v \in V$ and $s, t \in [0, T]$ where $\omega: [0, T] \to [0, \infty)$ is a non-decreasing function such that

$$\int_0^T \left(\frac{\omega(t)}{t}\right)^p \mathrm{d}t < \infty \tag{7.2}$$

for some $p \in (2, \infty)$ and $u_0 \in (H, \mathrm{dom}(A_0(0)))_{1-\frac{1}{p}, p}$, then the initial value problem (7.1) has maximal $L^p$-regularity in $H$.

***Example 7.3*** Let $\mathcal{D} \subset \mathbf{R}^d$ be a bounded domain with either a smooth boundary or a polygonal boundary if $\mathcal{D}$ is also convex. We set $H = L^2(\mathcal{D})$ and $V = H_0^1(\mathcal{D})$. Then, we consider the bilinear form $a_0(t; \cdot, \cdot): V \times V \mapsto \mathbf{R}$ given by

$$a_0(t; u, v) = \int_{\mathcal{D}} \alpha(t, x)\nabla u(x) \cdot \nabla v(x) \, \mathrm{d}x.$$

The coefficient function $\alpha\colon [0,T]\times\mathcal{D}\to\mathbf{R}$ is assumed to satisfy uniform bounds

$$0<\alpha_1\le\alpha(t;x)\le\alpha_2,\quad \text{for all } t,x\in[0,T]\times\mathcal{D}.$$

In addition, we assume that $\alpha(t,\cdot)$ is sufficiently smooth with respect to $x$ for every $t\in[0,T]$. Further, we suppose that

$$\|\alpha(t,\cdot)-\alpha(s,\cdot)\|_{L^\infty(\mathcal{D})}\le\omega(|t-s|)$$

with a mapping $\omega\colon[0,T]\to[0,\infty)$ as in (7.2). Then, $a_0(t;\cdot,\cdot)$ satisfies Assumption 7.1. Moreover, the domain $\mathrm{dom}(A_0(t))=H^2(\mathcal{D})\cap H_0^1(\mathcal{D})$ is constant in $t\in[0,T]$ and we have maximal $L^p$-regularity for any $p\in(2,\infty)$ with (7.2) (see [18]). Theorem 7.2 then yields that $u\in C^\gamma([0,T];V)$ for $\gamma<\frac{1}{2}-\frac{1}{p}$ and all regularity conditions of Theorem 6.7 are satisfied. A sufficient condition for the initial value is, for example, to choose $u_0\in H^2(\mathcal{D})\cap H_0^1(\mathcal{D})$.

***Remark 7.4*** For further regularity results of linear problems, see [1, Chap. II.1] or [32, Chap. 6]. There, non-autonomous linear evolution equations are considered and the existence of a Hölder continuous solution is proved.

The following assumption states sufficient conditions on a nonlinear perturbation of $\mathcal{A}_0$ such that the regularity results can be extended from the linear case to a class of semilinear problems.

**Assumption 7.5** Let $\mathcal{B}\colon[0,T]\times H\to H$ fulfill the following conditions:

(i) For every $v_1,v_2\in H$ the mapping $(\mathcal{B}(\cdot)v_1,v_2)_H\colon[0,T]\to\mathbf{R}$ is measurable.
(ii) There exists a constant $M\ge0$ such that $\|\mathcal{B}(t)0\|_H\le M$ for every $t\in[0,T]$.
(iii) For every $t\in[0,T]$ and $v_1,v_2\in H$, it holds true that

$$(\mathcal{B}(t)v_1-\mathcal{B}(t)v_2,v_1-v_2)_H\ge0.$$

(iv) There exists $L>0$ such that

$$\|\mathcal{B}(t)v_1-\mathcal{B}(t)v_2\|_H\le L\|v_1-v_2\|_H$$

holds for every $t\in[0,T]$ and $v_1,v_2\in H$.

***Remark 7.6*** Applying Remark 6.2, we can weaken Assumption 7.5 (iii) to

(iii′) There exists $\kappa\in[0,\infty)$ such that for every $t\in[0,T]$ and $v_1,v_2\in H$ it holds true that

$$(\mathcal{B}(t)v_1-\mathcal{B}(t)v_2,v_1-v_2)_H\ge-\kappa\|v_1-v_2\|_H^2.$$

at the cost that the constants in the error estimate grow exponentially with growing $T$.

***Example 7.7*** In the situation of Example 7.3, let $b\colon[0,T]\times\mathcal{D}\times\mathbf{R}\to\mathbf{R}$ be a mapping satisfying the following conditions:

(i) For every $z \in \mathbf{R}$ the mapping $b(\cdot, \cdot, z)\colon [0, T] \times \mathcal{D} \to \mathbf{R}$ is measurable.
(ii) There exists a constant $m \in [0, \infty)$ such that $|b(t, x, 0)| \le m$ for every $t \in [0, T]$, $x \in \mathcal{D}$.
(iii) For every $t \in [0, T]$ and $x \in \mathcal{D}$ the mapping $b(t, x, \cdot)\colon \mathbf{R} \to \mathbf{R}$ is non-decreasing and globally Lipschitz continuous.

Then, the Nemytskii operator $\mathcal{B}\colon [0, T] \times L^2(\mathcal{D}) \to L^2(\mathcal{D})$ defined by $(t, v) \mapsto b(t, \cdot, v(\cdot))$ satisfies Assumption 7.5.

Assumptions 7.1 and 7.5 in mind, we now consider the nonlinear problem

$$\begin{cases} \dot{u}(t) + \mathcal{A}_0(t)u(t) + \mathcal{B}(t)u(t) = f(t), & t \in (0, T], \\ u(0) = u_0. \end{cases} \tag{7.3}$$

A simple insertion of Assumptions 7.1 and 7.5 proves that the sum of the operators $\mathcal{A} = \mathcal{A}_0 + \mathcal{B}$ fulfills Assumption 6.1. Further, we obtain the same regularity result for the perturbed problem (7.3) as for the linear problem (7.1).

**Theorem 7.8** *Let the assumptions of Theorem 7.2 be satisfied for some $p \in (2, \infty)$ and let $\mathcal{B}$ fulfill Assumption 7.5. Then, the solution $u$ to (7.3) belongs to $C^\gamma([0, T]; V)$ for $\gamma < \frac{1}{2} - \frac{1}{p}$.*

***Proof*** The proof for the regularity follows a similar idea as presented in [33, Thm 2.9]. To this end, let $u \in C([0, T]; H)$ be the unique solution of (7.3) (see Proposition 6.3). We consider the function $g = f - \mathcal{B}u$ which fulfills

$$\begin{aligned} \|g\|_{L^p(0,T;H)} &\le \|f\|_{L^p(0,T;H)} + \|\mathcal{B}(\cdot)u(\cdot) - \mathcal{B}(\cdot)0\|_{L^p(0,T;H)} + \|\mathcal{B}(\cdot)0\|_{L^p(0,T;H)} \\ &\le \|f\|_{L^p(0,T;H)} + L\|u\|_{L^p(0,T;H)} + T^{\frac{1}{p}} M. \end{aligned}$$

Thus, $g \in L^p(0, T; H)$ and the solution $v$ of the linear problem

$$\begin{cases} \dot{v}(t) + \mathcal{A}_0(t)v(t) = g(t), & t \in (0, T], \\ v(0) = u_0. \end{cases} \tag{7.4}$$

is an element of the space $C^\gamma([0, T]; V)$ due to Theorem 7.2.

This now enables us to apply a bootstrap argument for the regularity of the solution $u$ of (7.3). Both (7.3) and (7.4) are uniquely solvable, and an insertion of $u$ in (7.4) shows that $u$ also solves the linear initial value problem. Therefore, $u = v$ holds and we obtain that $u \in C^\gamma([0, T]; V)$. □

***Remark 7.9*** The verification of the regularity conditions in Theorem 6.7 for general nonlinear PDEs can be quite challenging. However, besides the linear and semilinear problems discussed in this section, there are further classes of nonlinear problems that yield Hölder continuous solutions. For more general regularity results of semilinear problems, we refer the reader to [32, Chap. 7]. In [4], some quasi-linear problems are considered. They prove maximal $p$-regularity for these problems, which could potentially be extended to fit our setting as well. A further class of nonlinear problems is considered in [35], where regularity results from [32] are used. Here, a rather strong temporal regularity condition is imposed on the coefficients which would also lead to higher-order convergence results of the classical backward Euler method. But, as it can be seen from our numerical examples in Sects. 5 and 8, the randomized schemes (4.1) and (6.5) might still offer more reliable results in comparison with their deterministic counterparts if, for instance, the coefficients are smooth but highly oscillating.

## 8 Numerical Experiment with a Non-autonomous PDE

In this section, we finally illustrate the usability of the randomized backward Euler method (6.5) for the numerical solution of evolution equations. To this end, we follow a similar approach as for ODEs presented in Sect. 5. Here, we consider a nonlinear PDE of the form

$$\begin{cases} u_t(t,x) - u_{xx}(t,x) + b(u(t,x)) = f(t,x), & (t,x) \in (0,1)^2, \\ u(t,0) = u(t,1) = 0, & t \in (0,1), \\ u(0,x) = u_0(x), & x \in (0,1), \end{cases} \tag{8.1}$$

where we choose the function $b$ given by

$$b\colon \mathbf{R} \to \mathbf{R}, \quad b(x) = \begin{cases} |x|^{\tilde{p}-2}x, & \text{for } |x| \le R, \\ R^{\tilde{p}-2}x, & \text{for } |x| > R, \end{cases} \tag{8.2}$$

for a fixed $R \in (0,\infty)$ and $\tilde{p} \in [2,\infty)$ as well as suitable functions $f$ and $u_0$ which are specified further below. Using $H = L^2(0,1)$ and $V = H_0^1(0,1)$, (8.1) fits into the setting of Sect. 7, where

$$a_0(t; v_1, v_2) = \langle \mathcal{A}_0 v_1, v_2\rangle_{V\times V^*} = \int_\Omega v_1'(x) v_2'(x)\,\mathrm{d}x, \quad \text{for all } v_1, v_2 \in V,$$

fulfills Assumption 7.1 and

$$(\mathcal{B}v_1, v_2)_H = \int_\Omega b(v_1(x)) v_2(x)\,\mathrm{d}x, \quad \text{for all } v_1, v_2 \in H,$$

fulfills Assumption 7.5. For a function $u_0 \in V$ and $f \in L^p(0,1;H)$ with $p \in [2,\infty)$ (8.1) has maximal $L^p$-regularity. Furthermore, the assumptions of Theorem 7.8 are fulfilled such that the solution $u$ is an element of $C^\gamma([0,T];V)$ for $\gamma < \frac{1}{2} - \frac{1}{p}$.

In our numerical example, we consider a highly oscillating function $w$. For $P = 2^{-K}$, $K \in \mathbf{N}$, $w$ is the continuous, piecewise linear function determined by

$$w(iP) = \begin{cases} iP^2, & \text{for } i \in \{0,\dots,2^K\} \text{ odd}, \\ 0, & \text{for } i \in \{0,\dots,2^K\} \text{ even}, \end{cases}$$

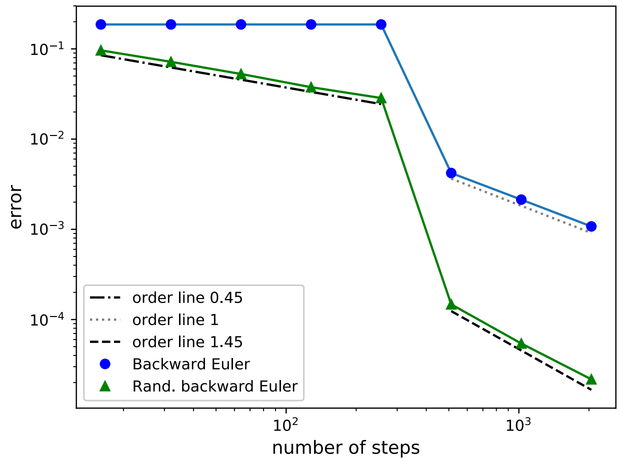


**Fig. 2** $L^2(\Omega; L^2(0,1))$-errors of the classical backward Euler method and scheme (6.5) for Eq. (8.1)

and the affine linear interpolation of these values for all other $t \in (0,1)$. The function $w$ is then weakly differentiable with derivative $\dot{w}$ given by

$$\dot{w}(t) = \begin{cases} iP, & \text{for } t \in [(i-1)P, iP),\ i \in \{1,\dots,2^K\} \text{ odd}, \\ -(i-1)P, & \text{for } t \in [(i-1)P, iP),\ i \in \{1,\dots,2^K\} \text{ even}. \end{cases}$$

For the functions

$$\begin{aligned} f(t,x) &= (x^2 - x^3)\dot{w}(t) - (2-6x)w(t) + \sin(\pi x) \\ &\quad + b((x^2 - x^3)w(t) + \pi^{-2}\sin(\pi x)), \\ u_0(x) &= \pi^{-2}\sin(\pi x) \end{aligned}$$

the solution is given by

$$u(t,x) = (x^2 - x^3)w(t) + \pi^{-2}\sin(\pi x)$$

as can be seen by a simple insertion.

Then, we see that $f \in L^\infty(0,1;H)$, $u_0 \in V$, $u \in C^\gamma([0,1];V)$ for every $\gamma \in (0,1)$ and (6.14) holds true, and thus, the assumptions of Theorem 6.7 are fulfilled.

The numerical behavior of this problem is very similar to the ODE example in Sect. 5. The right-hand side $f$ is highly oscillating. Thus, the classical backward Euler method needs a step size smaller than $P$ in order to give an accurate numerical approximation. The randomized scheme (6.5), on the other hand, yields much better approximations of the solution for larger values of the step size.

In our numerical test displayed in Fig. 2, we considered $R = 10$ and $\tilde{p} = 4$ in (8.2), $P = 2^{-9}$ and step sizes $k = 2^n$ with $n \in \{4,\dots,11\}$. To approximate the $L^2(\Omega; L^2(0,1))$-norm of the error, we used 200 Monte Carlo iterations. Since we are only interested in demonstrating the temporal convergence, we use a fixed finite element space with 500 degrees of freedom based on a uniform mesh in order to keep the spatial error on a negligible level for all considered temporal step sizes. For the implementation, we used the finite element software package FEniCS [31].

The results are well comparable to the results for the ODE example in Sect. 5. When the step size is larger than the value $P$, we can recognize a convergence rate of 0.45 for the randomized scheme. On the other hand, the error of the classical backward Euler method does not decrease for these step sizes. The errors of both schemes improve significantly when the step size is sufficiently small to resolve the oscillations. After that, we see the classical rate of 1 for the deterministic scheme and a rate of 1.45 in our randomized scheme.

**Acknowledgements** The authors like to thank Wolf-Jürgen Beyn for very helpful comments on non-autonomous evolution equations and Rico Weiske for good advice on programming. Also we like to thank two anonymous referees for their valuable suggestions. This research was partially carried out in the framework of MATHEON supported by Einstein Foundation Berlin. ME would like to thank the Berlin Mathematical School for the financial support. RK also gratefully acknowledges financial support by the German Research Foundation (DFG) through the research unit FOR 2402—Rough paths, stochastic partial differential equations and related topics—at TU Berlin.

## References

1. H. Amann. *Linear and Quasilinear Parabolic Problems. Vol. I: Abstract Linear Theory*, volume 89 of *Monographs in Mathematics*. Birkhäuser Boston, Inc., Boston, MA, 1995.
2. H. Amann. Compact embeddings of Sobolev and Besov spaces. *Glasnik Matematicki*, 35(55):161–177, 2000.
3. A. Andersson and R. Kruse. Mean-square convergence of the BDF2-Maruyama and backward Euler schemes for SDE satisfying a global monotonicity condition. *BIT Numer. Math.*, 57(1):21–53, 2017.
4. W. Arendt and M. Duelli. Maximal $L^p$-regularity for parabolic and elliptic equations on the line. *J. Evol. Equ.*, 6(4):773–790, 2006.
5. C. Baiocchi and F. Brezzi. Optimal error estimates for linear parabolic problems under minimal regularity assumptions. *Calcolo*, 20(2):143–176, 1983.
6. R. E. Bank and H. Yserentant. On the $H^1$-stability of the $L_2$-projection onto finite element spaces. *Numer. Math.*, 126(2):361–381, 2014.
7. C. Carstensen. Merging the Bramble-Pasciak-Steinbach and the Crouzeix-Thomée criterion for $H^1$-stability of the $L^2$-projection onto finite element spaces. *Math. Comp.*, 71(237):157–163 (electronic), 2002.
8. C. Carstensen. An adaptive mesh-refining algorithm allowing for an $H^1$ stable $L^2$ projection onto Courant finite element spaces. *Constr. Approx.*, 20(4):549–564, 2004.
9. K. Chrysafinos and L. S. Hou. Error estimates for semidiscrete finite element approximations of linear and semilinear parabolic equations under minimal regularity assumptions. *SIAM J. Numer. Anal.*, 40(1):282–306, 2002.
10. D. S. Clark. Short proof of a discrete Gronwall inequality. *Discrete Appl. Math.*, 16(3):279–281, 1987.
11. M. Crouzeix and V. Thomée. The stability in $L_p$ and $W_p^1$ of the $L_2$-projection onto finite element function spaces. *Math. Comp.*, 48(178):521–532, 1987.
12. T. Daun. On the randomized solution of initial value problems. *J. Complexity*, 27(3-4):300–311, 2011.
13. M. Dindoš and V. Toma. Filippov implicit function theorem for quasi-Carathéodory functions. *J. Math. Anal. Appl.*, 214(2):475–481, 1997.
14. E. Emmrich. *Gewöhnliche und Operator-Differentialgleichungen*. Vieweg, Wiesbaden, 2004.
15. E. Emmrich. Two-step BDF time discretisation of nonlinear evolution problems governed by monotone operators with strongly continuous perturbations. *Comput. Methods Appl. Math.*, 9(1):37–62, 2009.
16. L. C. Evans. *Partial Differential Equations*. Graduate studies in mathematics ; 19. American Mathematical Society, 1998.

17. I. Gyöngy. On stochastic equations with respect to semimartingales. III. *Stochastics*, 7(4):231–254, 1982.
18. B. H. Haak and E. M. Ouhabaz. Maximal regularity for non-autonomous evolution equations. *Math. Ann.*, 363(3-4):1117–1145, 2015.
19. S. Haber. A modified Monte-Carlo quadrature. *Math. Comp.*, 20:361–368, 1966.
20. S. Haber. A modified Monte-Carlo quadrature. II. *Math. Comp.*, 21:388–397, 1967.
21. W. Hackbusch. Optimal $H^{p,\,p/2}$ error estimates for a parabolic Galerkin method. *SIAM J. Numer. Anal.*, 18(4):681–692, 1981.
22. E. Hairer and G. Wanner. *Solving Ordinary Differential Equations. Vol. II: Stiff and Differential-Algebraic Problems*, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2010. Second revised edition, paperback.
23. J. K. Hale. *Ordinary Differential Equations*. Robert E. Krieger Publishing Co., Inc., Huntington, N.Y., second edition, 1980.
24. S. Heinrich and B. Milla. The randomized complexity of initial value problems. *J. Complexity*, 24(2):77–88, 2008.
25. L. S. Hou and W. Zhu. Error estimates under minimal regularity for single step finite element approximations of parabolic partial differential equations. *Int. J. Numer. Anal. Model.*, 3(4):504–524, 2006.
26. A. Jentzen and A. Neuenkirch. A random Euler scheme for Carathéodory differential equations. *J. Comput. Appl. Math.*, 224(1):346–359, 2009.
27. B. Z. Kacewicz. Optimal solution of ordinary differential equations. *J. Complexity*, 3(4):451–465, 1987.
28. B. Z. Kacewicz. Almost optimal solution of initial-value problems by randomized and quantum algorithms. *J. Complexity*, 22(5):676–690, 2006.
29. A. Klenke. *Probability Theory. A Comprehensive Course*. Springer, London, 2nd ed. edition, 2014.
30. R. Kruse and Y. Wu. Error analysis of randomized Runge-Kutta methods for differential equations with time-irregular coefficients. *Comput. Methods Appl. Math.*, 17(3):479–498, 2017.
31. A. Logg, K.-A. Mardal, and G. N. Wells. *Automated Solution of Differential Equations by the Finite Element Method: The FEniCS book*. Lecture Notes in Computational Science and Engineering; 84. Springer, Berlin, Heidelberg, 2012 edition, 2012.
32. A. Lunardi. *Analytic Semigroups and Optimal Regularity in Parabolic Problems*. Birkhäuser, Basel, 1995.
33. D. Meidner and B. Vexler. Optimal error estimates for fully discrete Galerkin approximations of semilinear parabolic equations. ArXiv preprint, arXiv:1707.07889v1, 2017.
34. E. Novak. *Deterministic and Stochastic Error Bounds in Numerical Analysis*, volume 1349 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1988.
35. A. Ostermann and M. Thalhammer. Convergence of Runge-Kutta methods for nonlinear parabolic equations. *Appl. Numer. Math.*, 42(1-3):367–380, 2002. Ninth Seminar on Numerical Solution of Differential and Differential-Algebraic Equations (Halle, 2000).
36. A. Prothero and A. Robinson. On the stability and accuracy of one-step methods for solving stiff systems of ordinary differential equations. *Math. Comp.*, 28:145–162, 1974.
37. T. Roubíček. *Nonlinear Partial Differential Equations with Applications*, volume 153 of *International Series of Numerical Mathematics*. Birkhäuser/Springer Basel AG, Basel, second edition, 2013.
38. G. Stengle. Numerical methods for systems with measurable coefficients. *Appl. Math. Lett.*, 3(4):25–29, 1990.
39. G. Stengle. Error analysis of a randomized numerical method. *Numer. Math.*, 70(1):119–128, 1995.
40. J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski. *Information-Based Complexity*. Computer Science and Scientific Computing. Academic Press, Inc., Boston, MA, 1988. With contributions by A. G. Werschulz and T. Boult.
41. H. Triebel. *Interpolation Theory, Function Spaces, Differential Operators*. North-Holland Publishing Company, Amsterdam-New York, 1978.
42. E. Zeidler. *Nonlinear Functional Analysis and its Applications. 2/A, Linear Monotone Operators*. Springer-Verlag, New York, 1990.
43. E. Zeidler. *Nonlinear Functional Analysis and its Applications. 2/B, Nonlinear Monotone Operators*. Springer-Verlag, New York, 1990.