Randomised one-step time integration methods for deterministic operator differential equations

Han Cheng Lie[1] · **Martin Stahn[1]** · **T. J. Sullivan[2,3]**

Abstract

Uncertainty quantification plays an important role in problems that involve inferring a parameter of an initial value problem from observations of the solution. Conrad et al. (Stat Comput 27(4):1065–1082, 2017) proposed randomisation of deterministic time integration methods as a strategy for quantifying uncertainty due to the unknown time discretisation error. We consider this strategy for systems that are described by deterministic, possibly time-dependent operator differential equations defined on a Banach space or a Gelfand triple. Our main results are strong error bounds on the random trajectories measured in Orlicz norms, proven under a weaker assumption on the local truncation error of the underlying deterministic time integration method. Our analysis establishes the theoretical validity of randomised time integration for differential equations in infinite-dimensional settings.

Keywords Time integration · Operator differential equations · Randomisation · Uncertainty quantification

Mathematics Subject Classification 65M15 · 65M75 · 37H10

✉ Han Cheng Lie
hanlie@uni-potsdam.de

Martin Stahn
martin.stahn@uni-potsdam.de

T. J. Sullivan
t.j.sullivan@warwick.ac.uk

1 Institut für Mathematik, Universität Potsdam, Campus Golm, Haus 9, Karl-Liebknecht-Straße 24–25, 14476 Potsdam OT Golm, Germany

2 Mathematics Institute and School of Engineering, University of Warwick, Coventry CV4 7AL, UK

3 Alan Turing Institute, 96 Euston Road, London NW1 2DB, UK

1 Introduction

The numerical solution of deterministic dynamical systems is an important task in many applications where the dynamical system is a spatiotemporal field that satisfies a partial differential equation (PDE). In this case, the field can be viewed as a function u mapping to an infinite-dimensional real separable Banach space $(V, |\cdot|_V)$, and the dynamical system is described by a deterministic operator differential equation initial value problem on a finite time interval $[0, T]$ for some initial condition ϑ:

$$u(0) = \vartheta, \quad u'(t) = f(t, u(t)), \quad t \in [0, T].$$

Operator differential equations have been applied in peridynamics and elastic materials, e.g. [13, 25]. The purpose of this paper is to analyse the error of randomised time integration methods for solving such initial value problems. The methods are of the form

$$U_{k+1} := \psi(h, U_k) + \xi_k(h), \quad k \in \{0, \dots, N-1\},$$

where $\psi(h, U_k)$ represents the output of a deterministic time integration method with time step h corresponding to the input U_k, and $\xi_k(h)$ is a V-valued random variable whose distribution depends on h. Our motivation for considering these methods comes from Bayesian inverse problems.

In many applications, the initial value problem depends on a parameter θ^*—for example, the initial condition ϑ, or a parameter appearing in the vector field f—and it is of interest to infer the value of θ^* given some observational data y, where y results from some fixed measurement process. Let Θ and $\mathcal{Y}$ denote the set of feasible parameter values and the set of feasible data values respectively. We assume that Θ is a Banach space and $\mathcal{Y}$ is a Hilbert space. Let S denote the solution operator that maps every $\theta' \in \Theta$ to the solution of the corresponding initial value problem, and let O denote the observation operator that maps every continuous trajectory in V to the corresponding output $\tilde{y} \in \mathcal{Y}$ of the fixed measurement process. Then the inference problem is to determine the value of the unknown true parameter θ^* given noisy data of the form

$$y = O{\circ}S(\theta^*) + \eta,$$

where η is often assumed to be a centred Gaussian random variable with known, positive-definite covariance operator Γ. In general, the inverse problem is ill-posed, and one can apply deterministic or statistical approaches to solving the inverse problem.

In the Bayesian approach to inverse problems, one assumes that Θ can be equipped with a probability measure μ_0, called the 'prior'. Let $G := O{\circ}S : \Theta \to \mathcal{Y}$ denote the parameter-to-observable map. The Bayesian solution to the inverse problem is given by the 'posterior' probability measure μ^y on Θ, which satisfies

$$\mu^y(\mathrm{d}\theta') = \frac{1}{Z(y)} \exp\left(-\frac{1}{2}\|y - G(\theta')\|_\Gamma^2\right)\mu_0(\mathrm{d}\theta')$$

where $\|x\|_\Gamma^2 = \left\langle x, \Gamma^{-1}x \right\rangle_y$, and $Z(y)$ is a normalisation constant. The posterior is important because one can use it to perform uncertainty quantification for the unknown parameter θ^*. See [34, Section 2.4] for a presentation of the Bayesian approach to inverse problems posed on vector spaces.

For many differential equations arising in applications, one must approximate the exact solution operator S using another operator $\tilde{S}$ that results from a discretisation of the initial value problem. This leads to an approximation $\tilde{G} := O \circ \tilde{S}$ of the parameter-to-observable map, which in turn leads to an approximation $\tilde{\mu}^y$ of the exact posterior μ^y defined above. For a fixed data vector y and prior μ_0, the error in $\tilde{S}$ is propagated via Bayes' theorem to an error in $\tilde{\mu}^y$. Since the posterior is fundamental for performing inference on the unknown parameter θ^*, one seeks a principled way to take into account the discretisation error in $\tilde{S}$.

Under some assumptions, a bound on the error $G - \tilde{G}$ with respect to some appropriate norm can be used to prove a bound on the error in the posterior, as measured by the Hellinger metric, e.g. [34, Corollary 4.9]. Stability bounds of this type ensure that the approximate posterior $\tilde{\mu}^y$ converges in the Hellinger metric to the exact posterior μ^y, in the limit as the discretisation error vanishes. While this property ensures that we can ignore the error in the posterior in the limit of increasingly finer discretisations, it does not indicate how to treat the error in the posterior for a fixed discretisation.

One approach is to ignore the discretisation error. This approach is not ideal from the point of view of statistical inference, because the approximate posterior $\tilde{\mu}^y$ can be tightly concentrated around the wrong parameter values, even in the small-noise limit. This phenomenon of 'overconfidence' is undesirable for uncertainty quantification. See Sect. 1.1 below.

The approach presented in [10] approaches the problem of accounting for the discretisation error, by applying the standard procedure of using random variables as proxies for unknown quantities. Let $\psi(h, v)$ denote the output of applying a time integration method for time step h to the state v, for a fixed time step $h = T/N > 0$, $N \in \mathbb{N}$. Consider the error $u(h) - \psi(h, u(0))$ between the exact solution and the numerical solution, incurred over one time step. Since the one-step error is unknown, we model it using a random variable $\xi_0(h)$. Thus,

$$u(h) \approx \psi(h, u(0)) + \xi_0(h) =: U_1.$$

If we model the one-step error for subsequent steps in a similar way, then this leads to the randomised time integration methods stated at the beginning of this section.

1.1 Illustration of overconfidence phenomenon

Consider the standard heat equation on a bounded domain $D \subset \mathbb{R}^d$ with homogeneous Dirichlet boundary conditions, written as the operator differential equation

$$u(0) = \vartheta \in H, \quad u'(t) + Au(t) = 0, \quad t \in [0, h],$$

where A is the negative Laplacian, $H = L^2(D)$, and $h > 0$. In [34, Section 3.5], one considers the inverse problem of inferring the initial condition ϑ from a noisy

observation of the solution at a later time. We shall use the assumptions and the approach stated there. The parameter-to-observable map is $G : H \to H, v \mapsto e^{-hA}v$. The data y is a realisation of the random variable

$$Y = G(\vartheta) + \delta^{1/2}\eta = G\vartheta + \delta^{1/2}\eta$$

where the noise η is a Gaussian random variable with distribution $\mathcal{N}(0, \Gamma_{obs})$. The noise scaling δ is assumed to be known, and the small noise limit corresponds to $\delta \to 0$. For the unknown parameter ϑ, we use the Gaussian prior $\mu_0 = \mathcal{N}(m_0, \Gamma_0)$. The positive-definite covariance operators Γ_{obs} and Γ_0 are chosen so that 1) draws from $\mathcal{N}(0, \Gamma_{obs})$ and from μ_0 are H-valued, almost surely; and 2) Γ_0 is an appropriate negative fractional power of A. Applying [34, Theorem 6.20] to the jointly Gaussian random variable $(U, G(U) + \delta^{1/2}\eta)$ with $U \sim \mu_0$ yields the Gaussian posterior measure μ^y with mean and covariance

$$m = m_0 + \Gamma_0 G(\delta\Gamma_{obs} + G\Gamma_0 G)^{-1}(y - Gm_0)$$
$$\mathcal{C} = \Gamma_0 - \Gamma_0 G(\delta\Gamma_{obs} + G\Gamma_0 G)^{-1}G\Gamma_0.$$

In the $\delta \to 0$ limit, $y \to G\vartheta$. Using this fact and the assumptions on Γ_0, it follows that $\mathcal{C} \to 0$ and $m \to \vartheta$ in the $\delta \to 0$ limit. Since Gaussian measures are completely characterised by their mean and covariance, the convergence of $\mathcal{C}$ and m implies the weak convergence (in the sense of probability measures) of the posterior measure to the Dirac measure at the true initial condition ϑ as $\delta \to 0$. This convergence captures the concentration of the posterior μ^y around the true unknown ϑ, and validates the Bayesian approach to the inverse problem.

Now suppose we approximate G using the map $\tilde{G}$ defined by a single step of the implicit Euler method, $\tilde{G} : H \to H, v \mapsto (I + hA)^{-1}v$. Applying [34, Theorem 6.20] as we did earlier with $\tilde{G}$ instead of G yields the associated approximate posterior $\tilde{\mu}^y$, which is Gaussian with mean and covariance

$$\tilde{m} = m_0 + \Gamma_0 \tilde{G}(\delta\Gamma_{obs} + \tilde{G}\Gamma_0\tilde{G})^{-1}(y - \tilde{G}m_0)$$
$$\tilde{\mathcal{C}} = \Gamma_0 - \Gamma_0 \tilde{G}(\delta\Gamma_{obs} + \tilde{G}\Gamma_0\tilde{G})^{-1}\tilde{G}\Gamma_0.$$

In the $\delta \to 0$ limit, $\tilde{\mathcal{C}} \to 0$, but $\tilde{m} \to \tilde{G}^{-1}G\vartheta \neq \vartheta$. Thus, the approximate posterior $\tilde{\mu}^y$ converges weakly in the small noise limit to a biased Dirac measure. This demonstrates the overconfidence phenomenon. The bias $\tilde{G}^{-1}G\vartheta - \vartheta$ in the limiting Dirac measure is the local truncation error of the implicit Euler method.

To address the overconfidence phenomenon, we use a random variable as a proxy for the unknown bias. Consider the randomised implicit Euler method given by $\widehat{G}(v) := \tilde{G}v + h^{p+1}\zeta$, where $\zeta \sim \mathcal{N}(0, \Gamma_1)$ is independent of the observation noise η, and Γ_1 is chosen so that draws from $\mathcal{N}(0, \Gamma_1)$ are H-valued almost surely. By rewriting $\widehat{G}(U) + \delta^{1/2}\eta = \tilde{G}U + (h^{p+1}\zeta + \delta^{1/2}\eta)$ and applying [34, Theorem 6.20], it follows that the associated deterministic posterior $\widehat{\mu}^y$ is Gaussian, with mean and covariance

$$\hat{m} = m_0 + \Gamma_0 \tilde{G}(\delta\Gamma_{\mathrm{obs}} + h^{2p+2}\Gamma_1 + \tilde{G}\Gamma_0\tilde{G})^{-1}(y - \tilde{G}m_0)$$
$$\hat{C} = \Gamma_0 - \Gamma_0\tilde{G}(\delta\Gamma_{\mathrm{obs}} + h^{2p+2}\Gamma_1 + \tilde{G}\Gamma_0\tilde{G})^{-1}\tilde{G}\Gamma_0.$$

In the $\delta \to 0$ limit, $\hat{C}$ does not converge to zero, because of the additional $h^{2p+2}\Gamma_1$ term. However, in the limit as $h, \delta \to 0$, the bias $\tilde{G}^{-1}G\vartheta - \vartheta$ associated to $\tilde{G}$ vanishes. Hence $\hat{m} \to \vartheta$ and $\hat{C} \to 0$. For fixed $h > 0$, the additional $h^{2p+2}\Gamma_1$ term ensures that the deterministic approximate posterior $\hat{\mu}^y$ associated to the randomised implicit Euler method $\hat{G}$ is more 'spread out' than the approximate posterior $\tilde{\mu}^y$ associated to the non-randomised implicit Euler method $\tilde{G}$. In this way, the problem of overconfidence is mitigated.

1.2 Main contributions

In this paper, we rigorously prove strong forward error bounds for randomised one-step time integration methods applied to operator differential equations. Our work builds on the approach for proving the error bounds in L^2 of [10, Theorem 2.2] and the error bounds in L^R—for user-specified $R \in \mathbb{N}$—of [20, Theorem 3.5]. These bounds were stated for initial value problems formulated in $\mathbb{R}^d$, where the associated exact flow maps are globally Lipschitz, and where the randomised time integrators are generated using uniform time grids and numerical methods ψ that satisfy a uniform local truncation error assumption.

The error bounds that we prove in this paper generalise the existing error bounds in multiple aspects. Our bounds are valid for time-dependent vector fields, non-uniform time grids (i.e. variable time steps), and operator differential equations that are formulated on Banach spaces or on Gelfand triples. In Theorem 3.7, we show that one can obtain strong error bounds in L^R for $R > 1$, without the assumption of uniform local truncation error of the numerical method, and without the assumption that the flow map of the initial value problem is globally Lipschitz. In fact, we show that one can obtain strong error bounds in more general Orlicz norms. The bounds that we prove in this paper demonstrate that the paradigm of randomised time integration extends in a natural way to the time integration for PDEs with time-dependent coefficients. Moreover, the proofs we give for our main results are simpler than the proofs of the corresponding results given in [20].

A related but distinct contribution that we make is to consider the setting where the random variables used in the randomisation are independent and centred. We generalise the L^2 uniform error bound [20, Theorem 3.4] for centred and independent randomisation—which was proven in the setting of ODEs in $\mathbb{R}^d$—to the setting of operator differential equations on Gelfand triples, under weaker assumptions on the time integration map ψ. We address the question of whether it is possible to obtain better error bounds under these additional assumptions. This question was implicit in the analysis of [20], but was not addressed there.

1.3 Related work

Randomised time integration methods for differential equations have been studied extensively in the context of 'probabilistic numerics'. For some reviews of research in this area, see [9, 17, 27]. In probabilistic numerics, ODEs have been considered from many perspectives, including structure- or symmetry-preserving methods [1, 40], Bayesian modelling of the unknown solution with Gaussian processes [5, 10, 33, 36, 38, 40], data-based statistical estimation of discretisation error [24, 35], and filtering [19, 38]. The papers [10, 20] cited earlier also belong to this context. For PDEs, methods based on Bayesian inference and Gaussian processes [6, 8, 10, 28, 31, 39], multiscale techniques [29], and random meshes [2] have been studied. The research area of 'information field dynamics' [11, 14] also considers probabilistic simulation schemes for PDEs by using Gaussian processes and information theoretic ideas.

Random approximate posteriors arising from randomised solution operators for differential equations have been studied in [21, Section 5] under a strong assumption of exponentially integrable discretisation error $S - \tilde{S}$, and more recently under a weaker square integrability hypothesis in [15].

Two aspects differentiate the problem we consider from the problems considered in numerical methods for stochastic evolution equations. The most important aspect is that the operator differential equation of interest in this paper is deterministic. Thus, our context is fundamentally different from the context of numerical integration methods for stochastic differential equations and numerical integration methods for random differential equations. The second aspect is that the random variables used in the randomisation need not be constructed using i.i.d. copies of a Wiener process or Lévy process.

1.4 Overview

We introduce notation and some recurring objects in the next section. In Sect. 2, we consider the setting where the initial value problem is formulated on a Banach space. The main result is the strong error bound in Orlicz norm proven in Theorem 2.8 under the assumption of uniform local truncation error of the time integration method ψ.

In Sect. 3, we consider the setting where the initial value problem is formulated on a Gelfand triple, and where ψ satisfies a weaker local truncation error assumption. This setting is considered in the variational approach to PDEs. We prove strong L^2 error bounds for mutually independent and centred randomisation in Sect. 3.1. In Sect. 3.2, we discuss the feasibility of obtaining L^R bounds for $R > 2$ that are of the same order in the time step h, under the same assumptions of independence and centredness. In Sect. 3.3, we state in Theorem 3.7 a strong error bound in Orlicz norm without assuming independence or centredness.

In Sect. 4, we show that the assumptions we make in Sect. 3 are reasonable for a class of operator differential equations that includes the heat equation on a C^2 bounded domain.

We conclude in Sect. 5. In the appendices, we collect material that is useful for the main part of the paper.

1.5 Notation and setup

Below, $(V, |\cdot|_V)$ and $(H, \langle\cdot, \cdot\rangle_H)$ denote a real separable Banach space and a real separable Hilbert space respectively. We write $|\cdot|_H$ for the Hilbert space norm. All integrals are Bochner integrals unless otherwise stated. We define $C^1([0, T]; V) := \{u \in C([0, T]; V) \mid u' \in C([0, T]; V)\}$ and equip it with the norm $\|u\|_{1,\infty} = \|u\|_\infty + \|u'\|_\infty$ where $\|u\|_\infty := \sup_{t \in [0,T]} |u(t)|_V$. We define the space $C^1([0, T]; H)$ analogously.

All random variables will be defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We denote expectation with respect to $\mathbb{P}$ by $\mathbb{E}[\cdot]$ and write $X \sim \mu$ to mean that X has μ as its distribution. For a V-valued random variable X and $R \geq 1$, we shall write $\|X\|_{L^R(\Omega; V)} := \mathbb{E}[|X|_V^R]^{1/R}$. Similarly, if X is H-valued, then $\|X\|_{L^R(\Omega; H)} := \mathbb{E}[|X|_H^R]^{1/R}$. For a Young function $\Psi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, the corresponding Orlicz norm[1] $\|\cdot\|_\Psi$ of a $\mathbb{R}$-valued random variable Z is defined by

$$\|Z\|_\Psi := \inf\{k \in (0, \infty) \ : \ \mathbb{E}[\Psi(|Z|/k)] \leq 1\}.$$

If Z is a V-valued (respectively, H-valued) random variable, then $\|Z\|_{\Psi(\Omega; V)} := \big\|\,|Z|_V\,\big\|_\Psi$ (resp. $\|Z\|_{\Psi(\Omega; H)} := \big\|\,|Z|_H\,\big\|_\Psi$). The $\|\cdot\|_{\Psi(\Omega; V)}$ norm includes as a special case the $\|\cdot\|_{L^R(\Omega; V)}$ norm when $R > 1$, but not when $R = 1$. The analogous statement holds for the $\|\cdot\|_{\Psi(\Omega; H)}$ norm. An important choice of Young function Ψ is given by $\Psi_2(z) := \exp(z^2) - 1$, because finiteness of $\|X\|_{\Psi_2}$ implies that X is sub-Gaussian and hence exponentially square integrable.

We write $p \wedge q = \min\{p, q\}$ for $p, q \in \mathbb{R}$. For $h > 0$, $p \geq 0$, and $a = a(h) \in \mathbb{R}$, we write $a = \mathcal{O}(h^p)$ to mean that $|a| \leq Ch^p$ for some h-independent term $C > 0$. Given $N \in \mathbb{N}, [N] := \{1, \ldots, N\}$ and $[N]_0 := [N] \cup \{0\} = \{0, 1, \ldots, N\}$.

Throughout the paper, we consider the following initial value problem on a deterministic time interval $[0, T]$,

$$u(0) = \vartheta, \quad u'(t) = f(t, u(t)), \quad t \in [0, T] \tag{1.1}$$

for fixed $T > 0$ and suitable initial condition ϑ. We specify the domain and codomain of f in the following sections. We denote by φ the exact flow map associated to (1.1) as follows: for suitable $h \in [0, T]$, $t \in [0, T - h]$, and u_s,

$$\varphi(h, t, u_s) = u_s + \int_t^{t+h} f(\tau, \varphi(\tau, t, u_s)) \, d\tau. \tag{1.2}$$

---

[1] See [3, Chapter 8] for a general introduction to Orlicz spaces and norms.

We equip the time interval $[0, T]$ in (1.1) with a time grid $(t_k)_{k \in [N]_0}$, where

$$0 =: t_0 < t_1 < \cdots < t_N := T, \quad h_k := t_{k+1} - t_k, \quad h := \max_{k \in [N-1]_0} h_k. \tag{1.3}$$

From (1.3) it follows that for any $\tau \geq 0$,

$$\sum_{\ell \in [N-1]_0} h_\ell^{\tau+1} \leq h^\tau \sum_{\ell \in [N-1]_0} h_\ell = h^\tau T. \tag{1.4}$$

Given (1.2), the exact sequence $(u(t_k))_{k \in [N]_0}$ associated with the time grid satisfies

$$u(t_{k+1}) = \varphi(h_k, t_k, u(t_k)), \quad k \in [N-1]_0. \tag{1.5}$$

We denote by ψ the approximate flow map associated to a time integration method, and define a deterministic approximating sequence $(u_k)_{k \in [N]_0}$ by

$$u_{k+1} := \psi(h_k, t_k, u_k), \quad u_0 = \vartheta.$$

Let $(\xi_k)_{k \in \mathbb{N}_0}$ be a sequence of stochastic processes, where each ξ_k is a stochastic process on $[0, \infty)$. In Sect. 2 (respectively, Sect. 3), each ξ_k takes values in the Banach space V (resp. the Hilbert space H). Given the time grid in (1.3), we use $(\xi_k(h_k))_{k \in [N-1]_0}$ as a randomisation sequence in order to define the random approximating sequence $(U_k)_{k \in [N]_0}$ by

$$U_{k+1} := \psi(h_k, t_k, U_k) + \xi_k(h_k), \quad k \in [N-1]_0 \tag{1.6}$$

for a given random variable U_0. The sequence of errors $(e_k)_{k \in [N]_0}$ of the random approximating sequence (1.6) with respect to the exact sequence (1.5) is defined by

$$e_0 = u(0) - U_0, \quad e_{k+1} := u(t_{k+1}) - U_{k+1}, \quad k \in [N-1]_0.$$

By (1.5) and (1.6), we obtain

$$e_{k+1} = \varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k) - \xi_k(h_k), \quad k \in [N-1]_0. \tag{1.7}$$

The equation (1.7) shall be the starting point for our error analysis.

2 Classical setting

In this section, we prove the generalisation of [10, Theorem 2.2] and [20, Theorem 3.5] to the setting of a time-dependent vector field f on an infinite-dimensional, real, separable Banach space V. We assume that the vector field f in (1.1) satisfies $f : [0, T] \times V \to V$. In addition, we assume that for every initial condition $\vartheta \in V$, there exists a unique classical solution $u \in C^1([0, T]; V)$. For example, if f is continuous and uniformly Lipschitz in the second argument, then this assumption is satisfied, and φ exists [12, Satz 7.2.6].

We state the assumptions needed to prove the main result of this section. The first is a Lipschitz continuity assumption on the exact flow map.

Assumption 2.1 The exact flow map φ admits a constant $L_\varphi > 0$ such that for any $t \in [0, T]$, for every $h \geq 0$ such that $t + h \leq T$, and for every $x, y \in V$,

$$|\varphi(h, t, x) - \varphi(h, t, y)|_V \leq (1 + L_\varphi h)|x - y|_V.$$

If f is uniformly Lipschitz in the second argument, then Assumption 2.1 is satisfied [12, Satz 7.3.4].

Ideally, the deterministic sequence $(u_k)_k$ approximates the exact sequence $(u(t_k))_k$ well. We make this precise by introducing the following uniform local truncation error assumption.

Assumption 2.2 The approximate flow map ψ admits constants $0 < h^* < \infty$, $0 < C_{\varphi,\psi} < \infty$, and $q \geq 0$, such that for all $0 < h \leq h^*$,

$$\sup_{\substack{v \in V \\ t \in [0, T - h]}} |\varphi(h, t, v) - \psi(h, t, v)|_V \leq C_{\varphi,\psi} h^{q+1} .$$

The parameter h^* is included in order to account for implicit time integration methods that provide a unique output whenever the time step is small enough. In order to achieve an order of $q \geq 1$ for the truncation error, one usually requires higher regularity of f or equivalently higher regularity for the solution u [16, Section III.2, Theorem 2.4]. For classical one-step methods, the corresponding analysis extends to infinite-dimensional Banach spaces; see Appendix B.

The assumptions above are similar to [10, Assumption 2] and [20, Assumption 3.1, 3.2]. Note that Assumption 2.2 is restrictive, because it requires uniformity in t and v. For example, in [10], the analogous assumption is justified under the assumption that $f : \mathbb{R}^d \to \mathbb{R}^d$ is sufficiently smooth and sufficiently many of its derivatives are uniformly bounded. However, Assumption 2.2 is not satisfied in general. For example, in the setting where the operator differential equation is given by $u'(t) = Au(t) \in H$ for a Hilbert space H and the infinitesimal generator A of an analytic semigroup with domain $\text{Dom}(A)$, and ψ is given by the implicit Euler method, there exists $C > 0$ such that for all $\vartheta \in \text{Dom}(A)$, $n \in \mathbb{N}$, and all sufficiently small $h > 0$,

$$|\varphi(nh, 0, \vartheta) - \psi(nh, 0, \vartheta)|_H \leq Ch|A\vartheta|_H,$$

see [37, Theorem 7.1].

For equations of the form (1.1) derived from PDEs and fixed time argument t, the right-hand side f is in many cases not Lipschitz from V to V. Furthermore, one cannot in general expect that (1.1) admits a classical solution $u \in C^1([0, T]; V)$, because a classical solution requires regularity assumptions on the problem data that need not hold in general. In Sect. 3, we will consider vector fields f that do not satisfy the assumptions above. This will lead us to consider variational solutions of (1.1). There are other approaches to generalise the classical setting to problems with less regularity, e.g. mild solutions, but they are outside the scope of this paper.

2.1 Randomisation sequence

Recall the random approximating sequence $(U_k)_k$ defined in (1.6). In this section, we shall assume that each ξ_k is a V-valued stochastic process indexed by $[0, \infty)$, and we shall assume U_0 is a V-valued random variable. Below, we shall impose the following regularity assumption on the $(\xi_k)_{k \in \mathbb{N}_0}$.

For the remainder of Sect. 2, we shall shorten notation and write $\|Z\|_{\Psi}$ instead of $\|Z\|_{\Psi(\Omega;V)}$ for any V-valued random variable Z.

Assumption 2.3 The collection $(\xi_k)_{k \in \mathbb{N}_0}$ admits an Orlicz norm $\|\cdot\|_{\Psi}$ and constants $p \geq 0$ and $0 < C_{\xi} < \infty$, such that for all $k \in \mathbb{N}_0$ and $t > 0$,

$$\big\|\xi_k(t)\big\|_{\Psi} \leq C_{\xi} t^{p+1}.$$

The assumption allows the stochastic processes to be non-Gaussian, to be probabilistically dependent, and to have different distributions and nonzero means. Furthermore, Assumption 2.3 allows for $\xi_k(t)$ to have different orders of integrability. The rates at which the absolute moments decrease to zero as t decreases to zero may differ as well. The function Ψ quantifies the maximal common order of integrability, and the parameter p quantifies the maximal common decay rate with respect to $\|\cdot\|_{\Psi}$.

Assumption 2.3 generalises [20, Assumption 3.3], which in turn generalised [10, Assumption 1]. The latter two assumptions considered the $\|\cdot\|_R$ norm for $R \in \mathbb{N}$ and the $\|\cdot\|_2$ norm of $\mathbb{R}^d$-valued random variables respectively.

We recall the motivation given in [10] for the additive random perturbation in (1.6) and in particular for Assumption 2.3. Comparing (1.5) and (1.6) yields

$$u(t_1) = u(0) + \int_0^{h_0} f(s, u(s))\, ds \approx \psi(h_0, 0, u(0)) + \xi_0(h_0) = U_1.$$

Thus, the random variable $\xi_0(h_0)$ models the uncertainty in the value of the integral term due to the fact that the value of the solution u over the time interval $[0, h_0]$ is known only at time 0, and not at every time s in the interval $[0, h_0]$.

It is desirable that the approximation above is good with high probability. Given that any reasonable choice of ψ must satisfy $\lim_{h_0 \to 0} \psi(h_0, 0, u_0) = u_0$, a necessary condition for the approximation above to be good with high probability is that the law of $\xi_0(h_0)$ concentrates around 0 as $h_0 \to 0$, because the integral term $\int_0^{h_0} f(s, u(s))\, ds \to 0$ as $h_0 \to 0$. Using Assumption 2.3 with Markov's inequality yields that for every $\varepsilon > 0$,

$$\mathbb{P}\big(\big|\xi_k(t)\big|_V \geq \varepsilon\big) \leq \left(\frac{C_{\xi} t^{p+1}}{\varepsilon}\right)^r.$$

The inequality above shows that the parameter p quantifies the maximal common rate at which all the laws $(\mathbb{P} \circ (\big|\xi_k(t)\big|_V)^{-1})_k$ contract around the Dirac measure at zero, as t decreases to zero.

In [10, 20], the parameter p is chosen in order to ensure that the error of the random approximate solution sequence $(U_k)_k$ with respect to the exact sequence $(u(t_k))_k$ decreases with h at the same rate as the error of the deterministic approximate solution sequence $(u(t_k))_k$. This choice is motivated by the goal of showing that probabilistic integrators can have the same convergence rate as the underlying deterministic one-step method.

Recall that if V is a separable Banach space and μ is a Gaussian measure whose support equals V, then the Cameron–Martin space of μ is dense in V, and hence there exists a V-valued Wiener process $(W(t))_{t \geq 0}$ associated to μ such that $W(1) \sim \mu$ [4, Theorem 3.6.1, Proposition 7.2.3].[2] The next lemma shows that there exists a large class of Gaussian processes that satisfies Assumption 2.3.

Lemma 2.4 *Let μ be a Gaussian distribution with support equal to V, and let $(W(t))_{t \geq 0}$ be a Wiener process associated to μ such that $W(1) \sim \mu$. Let ξ be a stochastic process on $[0, \infty)$ defined by $t \mapsto \xi(t) := t^{p+1/2} W(t)$, and let $(\xi_k)_{k \in \mathbb{N}_0}$ be i.i.d. copies of ξ. Then*

$$\|\xi(t)\|_\Psi = \|\xi(1)\|_\Psi t^{p+1}, \tag{2.1}$$

for $\|\cdot\|_\Psi = \|\cdot\|_R, R > 1, or \|\cdot\|_\Psi = \|\cdot\|_{\Psi_2}, \Psi_2(z) := \exp(z^2) - 1$.

Proof For $t > 0$, we have $\|\xi(t)\|_\Psi = t^{p+1/2} \|W(t)\|_\Psi = t^{p+1} \|W(1)\|_\Psi$. The first equation follows from the definition of $\xi(t)$, and the second equation follows from the scaling property of the Wiener process, i.e. that $W(t) = t^{1/2} W(1)$ in distribution for every $t > 0$. The conclusion follows since $W(1) = \xi(1)$ as random variables, and because Gaussian random variables are exponentially square integrable by Fernique's theorem. $\square$

Remark 2.5 The preceding discussion shows that a collection of i.i.d. copies of the standard Wiener process W satisfies Assumption 2.3 with $p = -1/2$, in which case we may set $\xi_k(h_k)$ in (1.6) to be a centred Gaussian random variable with variance proportional to h_k. This choice yields a time integration method that resembles methods for stochastic differential equations. However, for the error bound in Theorem 2.8 below to imply convergence in probability of $(U_n)_n$ to the exact solution sequence $(u(t_k))_k$, we need $p > 0$. This observation highlights an important difference between the type of time integration methods that we analyse in this paper and time integration methods for stochastic differential equations.

2.2 Error bounds

Recall from (1.7) that

$$e_{k+1} = \varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k) - \xi_k(h_k), \quad k \in [N-1]_0.$$

The following bound is the generalisation of [10, Theorem 2] to our setting.

Lemma 2.6 *Suppose that*

- Assumption 2.1 holds with parameters L_φ,
- Assumption 2.2 holds with parameters h^*, $C_{\varphi,\psi}$ and q,
- Assumption 2.3 holds with parameters $\|\cdot\|_\Psi$, p, and C_ξ, and
- the initial state U_0 satisfies $\|U_0\|_\Psi < \infty$.

Then for any time grid $(t_k)_k$ such that $0 < h \leq h^*$, the corresponding error sequence $(e_k)_k$ satisfies

$$\max_k \|e_k\|_\Psi \leq \exp(L_\varphi T)\|e_0\|_\Psi + \frac{C_{\varphi,\psi} + C_\xi}{L_\varphi}\big(\exp(L_\varphi T) - 1\big)h^{p \wedge q}.$$

In particular, if $\|e_0\|_\Psi = 0$, *then* $\max_k \|e_k\|_\Psi = \mathcal{O}(h^{p \wedge q})$.

Proof It suffices to prove the first statement. Let $k \in [N-1]_0$. From (1.7) we have

$$
\begin{aligned}
|e_{k+1}|_V &\leq |\varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k)|_V + |\xi_k(h_k)|_V \\
&\leq |\varphi(h_k, t_k, u(t_k)) - \varphi(h_k, t_k, U_k)|_V + |\varphi(h_k, t_k, U_k) - \psi(h_k, t_k, U_k)|_V \\
&\quad + |\xi_k(h_k)|_V \\
&\leq (1 + L_\varphi h_k)|e_k|_V + C_{\varphi,\psi} h_k^{q+1} + |\xi_k(h_k)|_V
\end{aligned}
\tag{2.2}
$$

where (2.2) follows from Assumptions 2.1 and 2.2. By taking the $\|\cdot\|_\Psi$ norm of both sides of (2.2), using the triangle inequality, Assumption 2.3, and the bound $h_k \leq h$ from (1.3), we obtain

$$\|e_{k+1}\|_\Psi \leq (1 + L_\varphi h)\|e_k\|_\Psi + (C_{\varphi,\psi} + C_\xi)h^{(p \wedge q)+1}.$$

Applying the discrete Gronwall inequality in Lemma C.1 completes the proof.   □

Remark 2.7 In addition to bounds on the strong error $\|e_k\|_\Psi$, one can prove bounds on the weak error, i.e. bounds of the form

$$|\mathbb{E}[\Phi(U_n^h)] - \Phi(u_n)| \leq Ch^w,$$

for all sufficiently smooth $\mathbb{R}$-valued functions Φ. Such bounds were proven in [10, Theorem 2.4] and [1, Section 3], for example. We focus on strong error bounds in this paper.

To prove Lemma 2.6, we take expectations via the $\|\cdot\|_\Psi$ norm before applying the discrete Gronwall inequality in Lemma C.1 to conclude. By reversing the order of these operations and by using a different discrete Gronwall inequality, we

can bound $\left\| \max_k |e_k|_V \right\|_\Psi$. This yields the result below, which extends [20, Theorem 3.5] to our setting. On one hand, this bound has worse constants than the bound in Lemma 2.6. On the other hand, the bound is stronger, because

$$\max_k \|e_k\|_\Psi \leq \left\| \max_k |e_k|_V \right\|_\Psi, \tag{2.3}$$

and because the bound has the same order in h as Lemma 2.6.

Theorem 2.8 *Suppose the hypotheses of Lemma 2.6 hold. Then for any time grid $(t_k)_k$ with $0 < h \leq h^*$, the corresponding error sequence $(e_k)_k$ satisfies*

$$\left\| \max_k |e_k|_V \right\|_\Psi \leq \left(\|e_0\|_\Psi + C_{\varphi,\psi} h^q T + C_\xi h^p T \right) \exp\left(L_\varphi T \right),$$

In particular, if $\|e_0\|_\Psi = 0$, then $\left\| \max_k |e_k|_V \right\|_\Psi = \mathcal{O}(h^{p \wedge q})$.

Remark 2.9 When $\Psi(z) = \exp(z^2) - 1$, then the strong error bound given in Theorem 2.8 implies the exponential square integrability of the pathwise error $\max_k |e_k|_V^2$. The exponential square integrability of the pathwise error was used in [21, Section 5] to establish local Lipschitz continuity of random approximate posteriors—measured in the Hellinger metric—with respect to the expected error of the randomised time integrator. In [20], exponential integrability was obtained by considering $\left\| \max_k |e_k|_V \right\|_R$ for all $R \in \mathbb{N}$ and using the series representation of the exponential function. The use of Orlicz norms allows us to exploit the fact that the random approximating sequence $(U_k)_k$ inherits the integrability properties of the collection $(\xi_k)_k$. This leads to a simpler proof of exponential integrability.

Proof of Theorem 2.8 Using (2.2) and applying the discrete Gronwall inequality in Lemma C.3, we obtain for every $k \in [N-1]_0$ that

$$|e_{k+1}|_V \leq \left(|e_0|_V + \sum_{k \in [N-1]_0} \left(C_{\varphi,\psi} h_k^{q+1} + |\xi_k(h_k)|_V \right) \right) \exp\left(\sum_{0 \leq j \leq k} L_\varphi h_j \right).$$

Since the sum in the exponential increases with k, setting $k = N - 1$ above and using (1.3) to obtain $\sum_{j \in [N-1]_0} h_j = T$ yields the 'pathwise' bound

$$\max_k |e_k|_V \leq \left(|e_0|_V + C_{\varphi,\psi} h^q T + \sum_{k \in [N-1]_0} |\xi_k(h_k)|_V \right) \exp\left(L_\varphi T \right). \tag{2.4}$$

By taking the $\|\cdot\|_\Psi$ norm of both sides of (2.4), the triangle inequality, Assumption 2.3, and (1.4), we obtain

$$\left\| \max_k |e_k|_V \right\|_{\Psi} \leq \left(\|e_0\|_{\Psi} + C_{\varphi,\psi} h^q T + \sum_{k \in [N-1]_0} \|\xi_k(h_k)\|_{\Psi} \right) \exp\left(L_{\varphi} T\right)$$

$$\leq \left(\|e_0\|_{\Psi} + C_{\varphi,\psi} h^q T + C_{\xi} h^p T \right) \exp\left(L_{\varphi} T\right),$$

which completes the proof. □

Remark 2.10 Under the assumption that $V = \mathbb{R}^d$ and under the assumption that the randomisation sequence $(\xi_k(h_k))_k$ consists of centred, independent random variables, [10, Theorem 2] and [20, Theorem 3.4] consider the special case where $\|\cdot\|_{\Psi} = \|\cdot\|_2$ in Lemma 2.6 and Theorem 2.8, and establish $\mathcal{O}(h^{q \wedge (p+1/2)})$ bounds on the strong error respectively. The order in these bounds is better than the bounds we proved above. However, both the proofs of these results exploit both the inner product structure of $\mathbb{R}^d$ and the fact that linear functionals of the ξ_k appear in the expansion of $|e_{k+1}|_{\mathbb{R}^d}^2$. In the key inequality (2.2), we cannot exploit an inner product even if it were available, because we only consider $|e_{k+1}|_V$. In Sect. 3, we shall generalise [10, Theorem 2] and [20, Theorem 3.4] from $\mathbb{R}^d$ to general Hilbert spaces.

3 Variational setting

For evolution equations originating from PDEs with possibly non-smooth right-hand sides or non-smooth initial conditions, the classical solution theory that we considered in Sect. 2 might not apply, because the requirement that the operator f in (1.1) satisfies $f(t, v) \in V$ for every $v \in V$ and all suitable t might be too strong. For example, this requirement does not hold for the heat equation in Sobolev spaces $W^{k,p}$. There are several settings that extend the classical setting for such problems. In this section, we focus on the variational setting, because it is suitable for numerical time integration methods. In the variational setting, we consider a Gelfand triplet $V \hookrightarrow H \simeq H' \hookrightarrow V'$, which is a sequence of continuous embeddings of a Banach space V into a Hilbert space H that is identified with its dual space H', which is then embedded in the dual space V' of V [41, Proposition 23.13].

In this section, we further specify the operator differential equation (1.1) to be

$$u(0) = \vartheta \in H, \quad u'(t) + A(t, u(t)) = b(t) \in V', \quad t \in [0, T] \tag{3.1}$$

for a given operator $A : [0, T] \times V \to V'$ and $b \in L^{p'}(0, T; V')$. The equation (3.1) is written in the form that is common in PDE theory instead of the form used in (1.1), where the right-hand side would be defined by $f(t, u(t)) := b(t) - A(t, u(t))$. The solution of (3.1) belongs to the space

$$\mathcal{W}^p(0, T) := \left\{ u \in L^p(0, T; V) \,\middle|\, u' \in L^{p'}(0, T; V') \text{ with } \frac{1}{p} + \frac{1}{p'} = 1 \right\},$$

which is continuously embedded into $C([0, T]; H)$ [12, Satz 8.4.1]. We emphasise that a solution of (3.1) must satisfy the equation only for almost every $t \in [0, T]$, and not for every t.

There are several conditions—e.g. Lipschitz or one-sided Lipschitz conditions, strong positivity, monotonicity, or coercivity — that one can impose on A and b in order to guarantee the existence of a unique variational solution $u \in \mathcal{W}^p(0, T) \hookrightarrow C([0, T]; H)$ [41, Prop. 23.23]. Under stronger assumptions, higher regularity of u can be achieved [12, Satz 8.5.1]. In some cases, the flow map is continuous and even Lipschitz; see [41, Theorem 23.A] for linear problems and [41, Corollary 23.26] for the time-dependent case.

Recall the definition (1.5) of the sequence $(u(t_k))_{k \in [N]_0}$ of states of the exact solution:

$$u(t_{k+1}) = \varphi(h_k, t_k, u(t_k)), \quad k \in [N-1]_0,$$

where φ is the flow map associated to the differential equation of interest (3.1). In the variational setting, the flow map φ maps (h, t, u_s) with $h \in [0, T]$, $t \in [0, T-h]$, and $u_s \in H$ to a vector $\varphi(h, t, u_s) \in H$. Next, recall that ψ is the approximate flow map associated to a time integration method, and that according to (1.6), we construct the random approximating sequence $(U_k)_{k \in [N]_0}$ according to

$$U_{k+1} = \psi(h_k, t_k, U_k) + \xi_k(h_k), \quad k \in [N-1]_0.$$

In this section, we shall assume that the initial condition U_0 is a H-valued random variable, and that each ξ_k is a H-valued stochastic process indexed by $[0, \infty)$.

We shall make the following assumptions on ψ.

Assumption 3.1 Let $h^* > 0$, and let $\psi : [0, h^*] \times [0, T] \times H \to V$ satisfy the following conditions:

1. There exists a scalar $q \geq 0$, a function $C_{\varphi,\psi} : [0, T] \times H \to (0, \infty)$ that is bounded on bounded subsets, and a dense subset $\mathcal{D} \subset H$, such that, for every $h \in [0, h^*]$ and for every $(t, x) \in [0, T-h] \times H$ with $x = \varphi(s, 0, \vartheta')$ for some $s \geq 0$ and $\vartheta' \in \mathcal{D}$,

$$|\varphi(h, t, x) - \psi(h, t, x)|_H \leq C_{\varphi,\psi}(t, x) h^{q+1}; \tag{3.2}$$

2. There exists a constant $L_\psi > 0$ such that for all $(h, t) \in [0, h^*] \times [0, T]$ such that $h + t \leq T$ and for any $x, y \in H$,

$$|\psi(h, t, x) - \psi(h, t, y)|_H \leq (1 + L_\psi h)|x - y|_H. \tag{3.3}$$

The first statement of Assumption 3.1 means that the one-step error bound (3.2) holds for any x that lies on some solution $u \in C([0, T]; H)$ of (3.1), where the initial condition $\vartheta' = u(0)$ belongs to the dense subset $\mathcal{D}$. We make the hypothesis of density in order to account for known results concerning error bounds for time integration of PDEs, see e.g. [37, Chapter 7].

The local truncation error (3.2) is a reasonable requirement for any deterministic time integration method ψ and weakens the uniform local truncation error bound of Assumption 2.2. Given (3.2), we define

$$\left\|C_{\varphi,\psi}\right\|_\infty := \left\|C_{\varphi,\psi}\right\|_\infty(\vartheta) := \sup_{t\in[0,T]} C_{\varphi,\psi}(t, u(t)), \tag{3.4}$$

for any solution u of (3.1) with initial condition $\vartheta \in \mathcal{D}$. Since the solution u of (3.1) belongs to $C([0, T]; H)$, it is a bounded set. Hence, the first statement of Assumption 3.1 ensures the finiteness of $\left\|C_{\varphi,\psi}\right\|_\infty$ for any $\vartheta \in \mathcal{D}$. The second statement of Assumption 3.1 describes a global Lipschitz continuity property of the approximate flow map ψ with respect to the third argument of the map ψ. For the error bounds that we prove in this section, the bounds (3.2) and (3.3) shall play the roles of Assumptions 2.2 and 2.1 respectively in the error bounds of Sect. 2.2. Next, we formulate the analogue of Assumption 2.3 for the collection $(\xi_k)_{k\in\mathbb{N}_0}$ of stochastic processes. For the remainder of Sect. 3, we shall simplify notation and write $\|Z\|_\Psi$ instead of $\|Z\|_{\Psi(\Omega;H)}$ for any H-valued random variable Z.

Assumption 3.2 The collection $(\xi_k)_{k\in\mathbb{N}_0}$ admits an Orlicz norm $\|\cdot\|_\Psi$ and constants $p \geq 0$ and $0 < C_\xi < \infty$, such that for all $k \in \mathbb{N}_0$ and $t > 0$,

$$\left\|\xi_k(t)\right\|_\Psi \leq C_\xi t^{p+1}.$$

The only difference between Assumption 3.2 and Assumption 2.3 is that the stochastic processes are H-valued instead of V-valued.

3.1 L^2-error bounds for independent and centred randomisation

In this section, we assume that the $(\xi_k)_k$ are mutually independent and centred stochastic processes. In particular, for any time grid (1.3), the corresponding random variables $(\xi_k(h_k))_{k\in[N-1]_0}$ are mutually independent and centred. We shall generalise the L^2-error bounds from [10, Theorem 2] and [20, Theorem 3.4] to the variational setting.

For any time grid $(t_k)_{k\in[N]_0}$ and $k \in [N-1]_0$, let $\mathcal{F}_k := \sigma(\xi_j(h_j) : j \in [k]_0)$, i.e. $(\mathcal{F}_k)_{k\in[N-1]_0}$ is the filtration generated by the randomisation sequence $(\xi_j(h_j))_{j\in[N-1]_0}$.

The following lemma only requires mutual independence of the $(\xi_\ell)_\ell$.

Lemma 3.3 *Suppose that Assumption* 3.1 *holds. Let* $(t_k)_{k\in[N]_0}$ *be an arbitrary time grid. Then for* $j \in [N-1]_0$, U_{j+1} *is a measurable function of* U_0 *and* $\{\xi_\ell(h_\ell) : \ell \in [j]_0\}$. *In particular, if the* $(\xi_\ell)_\ell$ *are mutually independent, then for every* $j \in [N-1]$, $\xi_j(h_j)$ *and* U_j *are independent, and* $\xi_j(h_j)$ *is independent of* $\mathcal{F}_j$.

Proof It follows from (3.3) in Assumption 3.1 that, for arbitrary (h, t), $\psi(h, t, z)$ is globally Lipschitz continuous with respect to $z \in H$. Hence, U_{j+1} is a measurable function of U_j and $\xi_j(h_j)$, for every $j \in [K-1]_0$. This proves the first statement. The second statement follows from the first and the definition of $\mathcal{F}_j$. $\qquad\square$

The following result is the generalisation of [10, Theorem 2.2], which considered the case $H = \mathbb{R}^d$ for $d \in \mathbb{N}$.

Lemma 3.4 *Suppose the following statements are true*:

- Assumption 3.1 holds with parameters h^*, q, $C_{\varphi,\psi}$, $\mathcal{D}$, and L_ψ,
- Assumption 3.2 holds with parameters $\|\cdot\|_\psi := \|\cdot\|_2$, p, and C_ξ,
- the $(\xi_j)_j$ are mutually independent and centred, and
- the initial condition ϑ of (3.1) belongs to $\mathcal{D}$, and $\|U_0\|_2 < \infty$.

Then there exists a $L'_\psi > 0$ depending only on L_ψ, such that for any time grid $(t_k)_k$ satisfying $0 < h \leq 1 \wedge h^*$, the associated error sequence $(e_k)_k$ satisfies

$$\max_k \|e_k\|_2^2 \leq \left(\|e_0\|_2^2 + 3T \left\| C_{\varphi,\psi} \right\|_\infty^2 Th^{2q} + C_\xi^2 Th^{2p+1} \right) \exp\left(L'_\psi T \right).$$

In particular, if $\|e_0\|_2 = 0$, *then* $\max_{k\in[N]_0} \|e_k\|_2 = \mathcal{O}(h^{q\wedge(p+1/2)})$.

We state the proof below, even though it is very similar to the proof of [10, Theorem 2.2]. This is because the proof will be useful later in Sect. 3.2, where we discuss the feasibility of bounding $\max_k \|e_k\|_R$ for $R > 2$ under similar assumptions as Lemma 3.4. An important difference between our proof and the proof of [10, Theorem 2.2] is that the latter assumes uniform truncation error, e.g. as in Assumption 2.2. Instead, we use Assumption 3.1.

Proof of Lemma 3.4 Let $k \in [N-1]_0$. By the definition (1.7) of the error sequence $(e_k)_{k\in[N]_0}$,

$$\begin{aligned}
|e_{k+1}|_H^2 &= \left| \varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k) \right|_H^2 + |\xi_k(h_k)|_H^2 \\
&\quad + 2\langle \varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k), \xi_k(h_k) \rangle_H.
\end{aligned} \tag{3.5}$$

Recall the term $\left\| C_{\varphi,\psi} \right\|_\infty$ from (3.4). We obtain

$$\begin{aligned}
&\left| \varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k) \right|_H^2 \\
&= \left| \varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k) \right|_H^2 \\
&\leq \left(1 + \left(\frac{2}{h_k} \right) \right) C_{\varphi,\psi}(t_k, u(t_k))^2 h_k^{2q+2} + (1 + 2h_k)(1 + L_\psi h_k)^2 |e_k|_H^2 \\
&\leq 3 \left\| C_{\varphi,\psi} \right\|_\infty^2 h_k^{2q+1} + (1 + 2h_k)(1 + L_\psi h_k)^2 |e_k|_H^2.
\end{aligned} \tag{3.6}$$

The first inequality follows from the hypothesis that the initial condition ϑ of (3.1) belongs to $\mathcal{D}$, since we can then apply the local truncation error bound (3.2) of Assumption 3.1 and Young's inequality. The second inequality follows from the fact that $h_k \leq h \leq 1$. By the same fact, there exists $L'_\psi > 0$ that depends only on L_ψ such that $(1 + 2h_k)(1 + L_\psi h_k)^2 \leq 1 + L'_\psi h_k$. Using this inequality in (3.6) yields

$$\left| \varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k) \right|_H^2 \leq 3 \left\| C_{\varphi,\psi} \right\|_\infty^2 h_k^{2q+1} + (1 + L'_\psi h_k) |e_k|_H^2. \tag{3.7}$$

Substituting (3.7) into the bound (3.5) on $|e_{k+1}|_H^2$ yields

$$|e_{k+1}|_H^2 \leq \left(3 \left\| C_{\varphi,\psi} \right\|_\infty^2 h_k^{2q+1} + (1 + L'_\psi h_k)|e_k|_H^2 \right) + |\xi_k(h_k)|_H^2 \tag{3.8}$$
$$+ 2\langle \varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k), \xi_k(h_k) \rangle_H.$$

By mutual independence of the $(\xi_j(h_j))_{j \in [N-1]_0}$, it follows from the second statement of Lemma 3.3 that the arguments of the inner product are independent. By taking expectations of (3.8) and centredness of the $(\xi_j(h_j))_{j \in [N-1]_0}$, the expectation of the inner product vanishes. By Assumption 3.2, we have

$$\left\| e_{k+1} \right\|_2^2 \leq (1 + L'_\psi h_k) \left\| e_k \right\|_2^2 + 3 \left\| C_{\varphi,\psi} \right\|_\infty^2 h_k^{2q+1} + C_\xi^2 h_k^{2p+2}.$$

Using the discrete Gronwall inequality in Lemma C.3 and (1.4) completes the proof. $\square$

We shall use the next result, Lemma 3.5, to prove Proposition 3.6 below. A similar result to Lemma 3.5 was established in the proof of [20, Theorem 3.4], under the assumption that ψ preserves square integrability of random variables, i.e. that $\psi(Z) \in L^2(\Omega; \mathbb{R}^d)$ for every $Z \in L^2(\Omega; \mathbb{R}^d)$. Lemma 3.5 removes this assumption, by using Lemma 3.4.

Lemma 3.5 *Suppose the hypotheses of Lemma 3.4 hold. Then for any time grid $(t_j)_{j \in [N]_0}$ with $h > 0$, the stochastic process $(M_k)_{k \in [N-1]_0}$ defined by*

$$M_k := \sum_{j=0}^{k} \left\langle \varphi(h_j, t_j, u(t_j)) - \psi(h_j, t_j, U_j), \xi_j(h_j) \right\rangle_H \tag{3.9}$$

is a $\mathbb{R}$-valued, square-integrable martingale with respect to $(\mathcal{F}_k)_{k \in [N-1]_0}$. If in addition the time grid $(t_j)_{j \in [N]_0}$ satisfies $h \leq 1 \wedge h^*$, then there exists a universal constant $\kappa > 0$ such that for every $k \in [N-1]_0$,

$$\mathbb{E}\left[\max_{j \in [k]_0} |M_k| \right] \leq \left\| C_{\varphi,\psi} \right\|_\infty^2 h^{2q+1} + \frac{1}{4} \mathbb{E}\left[\max_{j \in [k]_0} \left| e_j \right|_H^2 \right] + \kappa^2 (1 + L'_\psi) T C_\xi^2 h^{2p+1}, \tag{3.10}$$

for the same L'_ψ given in Lemma 3.4.

Proof See Sect. D.1 for the proof. $\square$

Next, we use Lemma 3.5 to prove the following error bound, which is stronger than the bound given in Lemma 3.4 because of (2.3).

Proposition 3.6 *Suppose the hypotheses of Lemma 3.4 hold. Then for any time grid $(t_k)_k$ with $0 < h \leq 1 \wedge h^*$, the corresponding error sequence $(e_k)_k$ satisfies*

$$\left\|\max_k \left|e_k\right|_H\right\|_2^2$$

$$\leq 2\left(\left\|e_0\right\|_2^2 + 4\left\|C_{\varphi,\psi}\right\|_\infty^2 h^{2q}T + C_\xi^2 Th^{2p+1}(1 + \kappa^2(1 + L'_\psi))\right)\exp\left(2L'_\psi T\right),$$

for the universal constant κ in (3.10) and the constant L'_ψ given in Lemma 3.4. In particular, if $\left\|e_0\right\|_2 = 0$, then $\left\|\max_k \left|e_k\right|_H\right\|_2 = \mathcal{O}(h^{q\wedge(p+1/2)})$.

Proof See Sect. D.2 for the proof. □

3.2 Error bounds of higher integrability order for independent and centred randomisation

It is natural to ask if one can prove the analogues of Lemma 3.4 or Proposition 3.6 where we use $\|\cdot\|_R$, $R > 2$, while keeping the same order in h. Suppose that we wish to prove the analogue of Lemma 3.4 for $R = 3$. It follows from the triangle inequality and the definition (1.7) that

$$\left|e_{k+1}\right|_H \leq \left|\varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k)\right|_H + \left|\xi_k(h_k)\right|_H.$$

Thus

$$\left|e_{k+1}\right|_H^3 \leq \left|e_{k+1}\right|_H^2\left(\left|\varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k)\right|_H + \left|\xi_k(h_k)\right|_H\right)$$

and substituting (3.5) results in an upper bound on $\left|e_{k+1}\right|_H^3$ containing the mixed product of an inner product term and a norm term,

$$\langle\varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k), \xi_k(h_k)\rangle_H\left|\varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k)\right|_H.$$

In general, this product will not vanish in expectation, because one can no longer exploit the commutativity of the inner product with the expectation operator. The same assertion is valid for $R \geq 3$. This is the important difference between the $R = 2$ case that was proven in Lemma 3.4 and the case $R \geq 3$. This difference implies that we must use the Cauchy–Schwarz inequality to bound products. Using the Cauchy–Schwarz inequality yields

$$\left|e_{k+1}\right|_H^3 \leq \sum_{i=0}^3 \binom{3}{i}\left|\varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k)\right|_H^i\left|\xi_k(h_k)\right|_H^{3-i}.$$

We can obtain the same bound by applying the binomial theorem to the bound $\left|e_{k+1}\right|_H \leq \left|\varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k)\right|_H + \left|\xi_k(h_k)\right|_H$.

If the stochastic processes $(\xi_k)_k$ are mutually independent, then we may use the second statement of Lemma 3.3. Assuming that $e_0 = 0$ almost surely and taking expectations of the summand for $i = 2$ yields

$$\mathbb{E}\left[\left|\varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k)\right|_H^2 |\xi_k(h_k)|_H\right]$$

$$= \mathbb{E}\left[\left|\varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k)\right|_H^2\right]\mathbb{E}\left[\left|\xi_k(h_k)|_H\right]\right] \qquad \text{by independence}$$

$$\leq \left(\mathcal{O}(h_k^{2q+1}) + (1 + L'_\psi h_k)\mathbb{E}\left[|e_k|_H^2\right]\right)C_\xi h_k^{p+1} \qquad \text{by (3.7), Assumption 3.2}$$

$$\leq \left(\mathcal{O}(h_k^{2q+1}) + \mathcal{O}(h^{2q}) + \mathcal{O}(h^{2p+1})\right)C_\xi h_k^{p+1} \qquad \text{by Lemma 3.4.}$$

This yields a bound on $\left\|e_{k+1}\right\|_3^3$ by a term that is $\mathcal{O}(h^{(2q)\wedge(2p+1)+p+1})$. Applying a discrete Gronwall inequality produces a bound on $\max_{k\in[N]}\left\|e_k\right\|_3^3$ that is $\mathcal{O}(h^{(2q)\wedge(2p+1)+p})$. Since this upper bound on the exponent arises from the mixed product mentioned above, and since such mixed products will arise in any expansion of $|e_{k+1}|_H^R$, we cannot expect to prove that $\max_k \left\|e_k\right\|_R = \mathcal{O}(h^{q\wedge(p+1/2)})$ for $R > 2$ using the techniques that we applied earlier, even if the $(\xi_k)_k$ are mutually independent and centred.

For the L^3 analogue of Proposition 3.6, the fact that terms involving inner products do not vanish in expectation also poses a problem. This is because the proof of the L^2 case in Proposition 3.6 relies on the bound (3.10) in Lemma 3.5 on the martingale $(M_k)_k$. This bound in turn follows from the Burkholder–Davis–Gundy inequality for martingales [32, Chapter IV, §4, Theorem (4.1)]. For the L^3 case, the expectations of products containing an inner product term do not vanish, because one can no longer exploit commutativity of the inner product with the expectation operator, due to the mixed product. As a result, the martingale $(M_k)_k$ does not appear, and one cannot apply the Burkholder–Davis–Gundy inequality to prove a bound similar to (3.10). Instead, one must apply the Cauchy–Schwarz inequality or the binomial theorem, as we did above. This results in a bound on $\left\|\max_k |e_k|_H\right\|_3$ that is worse than $\mathcal{O}(h^{q\wedge(p+1/2)})$.

3.3 Error bounds of higher integrability order without independence or centredness assumptions

In this section, we prove a strong error bound for a general Orlicz norm instead of for the $\|\cdot\|_2$-norm. We use the same hypotheses as for Lemma 3.4 and Proposition 3.6, except that we do not assume mutual independence or centredness of the stochastic processes $(\xi_k)_{k\in\mathbb{N}_0}$.

Theorem 3.7 *Suppose the following statements are true*:

– Assumption 3.1 holds with parameters h^*, q, $C_{\varphi,\psi}$, $\mathcal{D}$, and L_ψ,
– Assumption 3.2 holds with parameters $\|\cdot\|_\Psi$, p, and C_ξ, and
– the initial condition ϑ of (3.1) belongs to $\mathcal{D}$, and $\left\|U_0\right\|_\Psi < \infty$.

Then for any time grid $(t_k)_k$ with $0 < h \leq h^*$, the corresponding error sequence $(e_k)_k$ satisfies

$$\left\| \max_k |e_k|_H \right\|_\Psi \le \left(\|e_0\|_\Psi + \left\| C_{\varphi,\psi} \right\|_\infty h^q T + C_\xi h^p T \right) \exp\left(L_\psi T \right).$$

In the results from Sect. 3.1, we required that the maximal time step h associated to the time grid satisfies $h \le 1 \wedge h^*$. In Theorem 3.7, we only require that $h \le h^*$. The discussion of exponential integrability in Remark 2.9 also applies to Theorem 3.7.

Proof of Theorem 3.7 Recall (1.7):

$$e_{k+1} = \varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k) - \xi_k(h_k), \quad k \in [N-1]_0.$$

By the triangle inequality, and by (3.2) and (3.3) from Assumption 3.1,

$$
\begin{aligned}
&\left| \varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k) \right|_H \\
&\le \left| \varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, u(t_k)) \right|_H + \left| \psi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k) \right|_H \\
&\le \left\| C_{\varphi,\psi} \right\|_\infty h_k^{q+1} + (1 + L_\psi h_k)|e_k|_H.
\end{aligned}
$$

From this it follows that

$$\left| e_{k+1} \right|_H \le \left\| C_{\varphi,\psi} \right\|_\infty h_k^{q+1} + (1 + L_\psi h_k)|e_k|_H + \left| \xi_k(h_k) \right|_H. \tag{3.11}$$

Applying Lemma C.3 and using the same arguments that yielded (2.4), we obtain the analogous pathwise bound

$$\max_k |e_k|_H \le \left(|e_0|_H + \left\| C_{\varphi,\psi} \right\|_\infty h^q T + \sum_{k \in [N-1]_0} \left| \xi_k(h_k) \right|_H \right) \exp\left(L_\psi T \right).$$

Taking the $\|\cdot\|_\Psi$ norm of both sides and applying Assumption 3.2 completes the proof. □

Remark 3.8 The inequality (3.11) in the proof of Theorem 3.7 closely resembles the inequality (2.2), which we used to prove Theorem 2.8. The key difference results from adding $0 = \psi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, u(t_k))$ before applying the triangle inequality to derive (3.11); for (2.2), we added $0 = \varphi(h_k, t_k, U_k) - \varphi(h_k, t_k, U_k)$ instead. The decomposition we use for (3.11) enables us to exploit the weaker local truncation error bound (3.2) in Assumption 3.1 instead of the uniform local truncation error bound in Assumption 2.2.

4 Example: heat equation

Consider the heat equation on a C^2 bounded domain $D \subset \mathbb{R}^d$ with homogeneous Dirichlet boundary conditions

$$u(0) = u_0, \quad \partial_t u - \mathrm{div}(\mathcal{E}\nabla u) = b \text{ on } [0, T] \times D, \tag{4.1}$$

where $\mathcal{E} : [0, T] \times D \to \mathbb{R}^{d \times d}$ is a sufficiently smooth elliptic diffusion tensor. Upon multiplying the PDE by a test function and using integration by parts, the left-hand side of the PDE yields a bilinear form $a(u(t), v)$, which allows us to rewrite the problem above as the operator differential equation

$$u(0) = u_0 \in H, \quad u'(t) + Au(t) = b \in V' \tag{4.2}$$

with spaces $H = L^2(D)$, $V = H_0^1(D)$, and $V' = H^{-1}(D)$. The bounded, linear operator $A : V \to V'$ is induced by the bilinear form $a(\cdot, \cdot)$ on $V \times V$ according to $a(u, v) = \langle Au, v \rangle_{V' \times V}$, where $\langle \cdot, \cdot \rangle_{V' \times V}$ denotes the dual pairing. For the particular PDE considered above, the operator A is strongly positive with constant $\mu > 0$ on $V \times V$.

In this section, we will show that the results that we proved for the variational setting in Sect. 3 are valid for parabolic PDEs and the implicit Euler method, by showing that the conditions (3.2) and (3.3) from Assumption 3.1 are satisfied. We shall consider the more general setting of parabolic PDEs with possibly time-dependent coefficients, because this analysis includes the setting of time-independent coefficients — and hence the heat equation stated above—as a special case.

Let $L(V, V')$ be the set of all linear mappings from V to V'. Consider a mapping $a : [0, T] \times V \times V \to \mathbb{R}$ that is bilinear in the second and third argument. This mapping induces a collection $(A(t))_t \subset L(V, V')$ according to

$$\langle A(t)u(t), v \rangle_{V' \times V} = a(t, u(t), v), \quad \forall v \in V.$$

Now we pose the following standard assumptions on a and state their equivalent formulation in terms of A.

Assumption 4.1

1. For fixed t, $a(t, \cdot, \cdot)$ is a bilinear form, and for fixed $u, v \in V$, $a(\cdot, u, v)$ is measurable. Equivalently, for every t, $A(t) \in L(V, V')$ is linear and $t \mapsto A(t)$ is measurable.
2. There exists $\beta > 0$ such that for every (t, u, v), $a(t, u, v) \leq \beta |u|_V |v|_V$. Equivalently, for every t we have $\|A(t)\|_{L(V, V')} \leq \beta$.
3. A Gårding inequality holds, i.e. there exist $\mu > 0$, $\kappa \geq 0$ such that

   $$a(t, u, u) \geq \mu |u|_V^2 - \kappa |u|_H^2, \quad \forall (t, u) \in [0, T] \times V. \tag{4.3}$$

   Equivalently, for every $t \in [0, T]$, $A(t) + \kappa I \in L(V, V')$ is strongly positive.

For the special case of the heat equation (4.1) where $\mathcal{E}$ is the identity matrix, the first statement of Assumption 4.1 holds since $\mathcal{E}$ is constant. By definition of the bilinear form a and the spaces H and V, the second statement holds with $\beta = 1$, and the third statement holds with equality for $\kappa = 0$ and $\mu = 1$.

Consider the implicit Euler scheme

$$\psi(h, t, v) := (I + h\bar{A}_{h,t})^{-1}(h\bar{b}_{h,t} + v), \tag{4.4}$$

for $0 < h \le h^*$, $0 \le t \le T - h$ and $v \in H$. We specify an interval of suitable values of h^* in Sect. 4.2. Above, $\bar{A}_{h,t}$ and $\bar{b}_{h,t}$ denote Steklov time averages of the linear operators $(A(t))_t$ and the right-hand side b respectively,

$$\bar{A}_{h,t} := \frac{1}{h} \int_t^{t+h} A(s) \, \mathrm{d}s, \quad \bar{b}_{h,t} := \frac{1}{h} \int_t^{t+h} b(s) \, \mathrm{d}s,$$

where the integrals in the definitions of $\bar{A}_{h,t}$ and $\bar{b}_{h,t}$ are Bochner–Lebesgue integrals in $L(V, V')$ and V' respectively. The existence of $\psi(h, t, v) \in V$ for $(h\bar{b}_{h,t} + v) \in V'$ is guaranteed by the Lax–Milgram theorem; see e.g. [7, Section 6.2]. For every suitable (h, t), the operator $\bar{A}_{h,t}$ inherits the properties of A stated in Assumption 4.1.

For the heat equation (4.1), $t \mapsto A(t)$ and $t \mapsto b(t)$ are constant. Therefore, $\bar{A}_{h,t} = A$ and $\bar{b}_{h,t} = b$, and (4.4) simplifies to $\psi(h, t, v) := (I + hA)^{-1}(hb + v)$.

4.1 Local truncation error condition

We verify the local truncation error condition (3.2) in Assumption 3.1, for ψ as given in (4.4). Recall the definition (1.5) of $(u(t_k))_k$ and that $(u_k)_k$ is defined by $u_0 = \vartheta$, $u_{k+1} := \psi(h_k, t_k, u_k)$ for $k \in [N-1]_0$. Under the assumption that $(b - u')' \in L^2(0, T; V')$, the result [12, Satz 8.3.6] yields for any initial condition $\vartheta \in H$

$$\left| u_k - u(t_k) \right|_H^2 + \mu \sum_{j=1}^k h_j \left| u_j - u(t_j) \right|_V^2 \le \frac{h^2}{3\mu} \left| (b - u')' \right|_{L^2(0,T;V')}^2,$$

where μ is the constant from positivity assumption on A (4.3). Thus, (3.2) holds with $q = 0$ and $C_{\varphi,\psi}(t, x) = (3\mu)^{-1/2} |(b - u')'|_{L^2(0,T;V')}$ for all (t, x).

One can obtain numerical methods of higher order q, by assuming higher regularity of the solution. For example, [22, Theorems 4.2, 4.3, 4.4] assume $u, u', u'' \in \mathcal{W}^2(0, T)$, and show the existence of a numerical method ψ that satisfies (3.2) with $q = 1$. For a general result dealing with arbitrary regularity $u^{(k+1)} \in \mathcal{W}^2(0, T)$ and numerical method of order $q = k$, see [23, Theorem 3.2].

4.2 Lipschitz condition on approximate flow map

Next, we verify the Lipschitz condition (3.3) for ψ given in (4.4), and determine an interval of suitable values for the upper bound h^* on the time step of the implicit Euler scheme. Fix $0 < h \le h^*$, $t \in [0, T - h]$, and $u_0, v_0 \in V$. Test $w_1 := \psi(h, t, u_0) - \psi(h, t, v_0) \in V$ with $w_0 := u_0 - v_0 \in H$. Then

$$\frac{1}{2h}(|w_1|_H^2 - |w_0|_H^2) \le \left\langle \frac{w_1 - w_0}{h}, w_1 \right\rangle_H \le -\left\langle \bar{A}_{h,t} w_1, w_1 \right\rangle_{V' \times V}$$
$$\le -\mu |w_1|_V^2 + \kappa |w_1|_H^2.$$

The first inequality follows from rearranging $0 \le |w_1 + w_0|_H^2$. The second inequality follows since (4.4) is equivalent to $h^{-1}(\psi(h, t, u_0) - u_0) = \bar{b}_{h,t} - \bar{A}_{h,t}\psi(h, t, u_0)$. The third inequality holds because $\bar{A}_{h,t}$ inherits the positivity property (4.3) from A.

Using $(2h)^{-1}(|w_1|_H^2 - |w_0|_H^2) \leq -\mu|w_1|_V^2 + \kappa|w_1|_H^2$ and the definitions of w_1 and w_0, we obtain

$$\left|\psi(h, t, u_0) - \psi(h, t, v_0)\right|_H^2 \leq (1 - 2h\kappa)^{-1}\left|u_0 - v_0\right|_H^2.$$

If $\kappa \leq 0$, then $(1 - 2h\kappa)^{-1} \leq (1 + L_\psi h)$ for any $L_\psi > 0$ and $h > 0$. Therefore, suppose that $\kappa > 0$. If the bound above on $\left|\psi(h, t, u_0) - \psi(h, t, v_0)\right|_H^2$ holds for all $0 < h \leq h^*$, then we must have $h^* < (2\kappa)^{-1}$. In fact, if $h^* < (2\kappa)^{-1}$, then $L_\psi := [(2\kappa)^{-1} - h^*]^{-1}$ is equivalent to $h^* = \frac{L_\psi - 2\kappa}{2\kappa L_\psi}$. In this case, $0 < h \leq h^*$ is equivalent to

$$h^2(2\kappa L_\psi) \leq h(L_\psi - 2\kappa) \Leftrightarrow 1 \leq (1 + L_\psi h)(1 - 2h\kappa) \Leftrightarrow (1 - 2h\kappa)^{-1} \leq 1 + L_\psi h.$$

Hence, the implicit Euler scheme (4.4) satisfies condition (3.3) in Assumption 3.1.

5 Conclusion

In this paper, we proved strong error bounds for general Orlicz norms for randomised time integration methods applied to operator differential equations, using possibly non-uniform time grids. Our work builds on the ideas and approaches of [10, 20]. We show that the proof techniques of the key error bounds contained therein can be applied in more general settings, where the differential equation is formulated on a possibly infinite-dimensional Banach or Hilbert space, and the numerical time integration method is applied to a possibly non-uniform time grid. Our work has two additional novel aspects relative to [10, 20].

First, we use a different error decomposition to bound the one-step error. Our error decomposition enables us to replace the strong assumption of uniform local truncation error with a weaker assumption on the local truncation error. This is important, because it is known that the strong assumption of uniform local truncation error is invalid even when the linear operator A in the operator differential equation generates an analytic semigroup [37, Theorem 7.1]. For the implicit Euler method, and for a large class of examples that includes the standard heat equation, we showed that our weaker local truncation error assumption is reasonable.

Second, we consider more general Orlicz norms instead of L^R-norms. Previous results concerning higher-order error bounds — for example, [20, Theorem 3.5]—were less direct: they involved finding bounds on the L^R error for each $R \in \mathbb{N}$ and using the series expansion of the exponential function. The use of Orlicz norms leads to shorter and conceptually simpler proofs of our main results, Theorem 2.8 and Theorem 3.7, by exploiting the fact that the random approximating sequence $(U_k)_k$ inherits the integrability properties of the collection $(\xi_k)_k$.

Taylor expansion in Banach Spaces

The following version of Taylor's theorem in Banach spaces is given in [7, Theorem 7.9-1].

Theorem A.1 *Let V and W be normed vector spaces, let U be an open subset of V, let $[a, a + h]$ be a closed segment contained in U, let $f : U \to W$ be a given mapping, and let $m \in \mathbb{N}$.*

(a) (*Taylor-Young*) *If f is $(m - 1)$ times differentiable on U and m times differentiable at $a \in U$, then*

$$f(a + h) = f(a) + f'(a)h + \cdots + \frac{1}{m!}f^{(m)}(a)h^m + \|h\|_V^m \delta(h)$$

with $\lim_{h \to 0} \delta(h) = 0$.

(b) (*Integral remainder*) *If W is a Banach space and f is m times continuously differentiable on U, then*

$$f(a + h) = f(a) + f'(a)h + \cdots + \frac{1}{(m - 1)!}f^{(m-1)}(a)h^{m-1}$$

$$+ \frac{1}{(m - 1)!} \int_0^1 (1 - t)^{m-1} \left(f^{(m)}(a + th)h^m \right) dt.$$

The noteworthy differences to the standard Taylor theorem in $\mathbb{R}$ are: 1) differentiability of f may be slightly more complicated; 2) the k-th derivative of f is a k-linear mapping from $\Pi_{i=1}^k U$ to W, taking k inputs from U, denoted by $h^k = (h, \dots, h) \in U^k$.

In addition, one can derive a Taylor expansion almost everywhere for weakly differentiable functions.

Additional material for Section 2

In the setting of time integration for ODEs, one usually requires higher regularity of f or equivalently higher regularity for the solution u in order to achieve an order of $q \geq 1$ for the truncation error [16, Section III.2, Theorem 2.4]. The purpose of this section is to show that the same ideas apply in the infinite-dimensional Banach space setting. Below, the function f refers to the vector field in (1.1).

For the next lemma, we consider for a general explicit Euler one-step method the map $\psi : [0, h] \times \mathbb{R} \times V \to V$ in a Banach space $(V, |\cdot|_V)$, associated to some step function Y:

$$\begin{aligned}
u_{k+1} &:= \psi(h_k, t_k, u_k) := u_k + h_k Y(h_k, t_k, u_k) \\
&= u_k + h_k \left(a_1 f(t_k, u_k) + a_2 f \left(t_k + b_1 h_k, t_k, u_k + b_2 h_k f(t_k, u_k) \right) \right)
\end{aligned} \tag{B.1}$$

with $a_1, a_2, b_1, b_2 \geq 0$.

Lemma B.1 (Lipschitz property of the numerical flow map) *Let ψ be as in (B.1). If $f : [0, T] \times V \to V$ is Lipschitz continuous in the second argument, then the approximate flow map $\psi(h_k, t_k, \cdot)$ is Lipschitz continuous, uniformly in k.*

Proof We use the Lipschitz property of f multiple times to get

$$|\psi(\tau, t, u) - \psi(\tau, t, v)|_V \leq (1 + L\tau(a_1 L + a_2 L + a_2 b_2 L^2 \tau))|u - v|_V . \quad \text{(B.2)}$$

$\square$

The following theorem is taken from [30, Theorem 7.1.5]. The proof there is only done in one dimension.

Theorem B.2 *Assuming that the right-hand side f belongs to $C^2([0, T] \times V; V)$ and equivalently that the solution u is C^3, it follows that any explicit one-step scheme with step function Y as in (B.1) with*

$$a_1 + a_2 = 1 , \; a_2 b_1 = \frac{1}{2} , \; a_2 b_2 = \frac{1}{2}$$

satisfies assumption 2.2 with $q = 2$.

Proof We assume that $f \in C^2([0, T] \times V; V)$ in order to be able to use Taylor expansion. We first expand the step function Y:

$$Y(h, t, u(t)) = \left[(a_1 + a_2) f(t, u(t)) + h \left(a_2 b_1 \frac{\partial f}{\partial t}(t, u(t)) + a_2 b_2 \frac{\partial f}{\partial u} f(t, u(t)) \right) \right] + \mathcal{O}(h^2).$$

Above, $\frac{\partial f}{\partial u}$ refers to the Gateaux derivative of f with respect to u, and $\frac{\partial f}{\partial u} f(t, u(t))$ denotes the linear mapping $\frac{\partial f}{\partial u}(t, u(t))$ from V to V acting on $f(t, u(t)) \in V$. Expanding u, we have

$$u(t + h) = u(t) + u'(t)h + u''(t)\frac{h^2}{2} + \mathcal{O}(h^3)$$

$$= u(t) + \left[hf(t, u(t)) + \frac{h^2}{2} \left(\frac{\partial f}{\partial t}(t, u(t)) + \frac{\partial f}{\partial u}(t, u(t)) f(t, u(t)) \right) \right] + \mathcal{O}(h^3)$$

$$= u(t) + \left[hY(h, t, u(t)) + \mathcal{O}(h^3) \right] + \mathcal{O}(h^3) .$$

The last equality follows from the conditions on the coefficients a_1, b_1, a_2, b_2. This gives consistency of order 2:

$$u(t + h) - (u(t) + hY(h, t, u(t))) = \varphi(h, t, u(t)) - \psi(h, t, u(t)) = \mathcal{O}(h^3).$$

Convergence follows from the Lipschitz continuity of f with respect to the second argument [30, Theorem 7.10]. $\square$

Discrete Gronwall inequalities

The following statement is given in [26, Lemma 1.6].

Lemma C.1 *Let $T > 0$ be fixed. Let $N \in \mathbb{N}$ and $h = T/N$. Suppose $(y_k)_k \in [0, \infty)^{\mathbb{N}_0}$ is such that for some $A, B \geq 0$ and $p \geq 1$,*

$$y_{k+1} \leq (1 + Ah)y_k + Bh^p, \quad k \in [N - 1]_0.$$

Then

$$y_k \leq e^{AT} y_0 + \frac{B}{A}(e^{AT} - 1)h^{p-1}, \quad k \in [N - 1]_0$$

where for $A = 0$, $A^{-1}(e^{AT} - 1) = 0$.

We restate the "special Gronwall inequality" of [18].

Proposition C.2 *Let $(y_n)_n, (g_n)_n \in [0, \infty)^{\mathbb{N}_0}, c \geq 0$, and $N \in \mathbb{N}$ be arbitrary. If*

$$y_{k+1} \leq c + \sum_{0 \leq j \leq k} g_j y_j, \quad k \in [N - 1]_0$$

then

$$y_{k+1} \leq c \exp\left(\sum_{0 \leq j \leq k} g_j \right), \quad k \in [N - 1]_0.$$

The following lemma is a corollary of Proposition C.2.

Lemma C.3 *Let $T > 0$ and $N \in \mathbb{N}$ be fixed. Let $(h_k)_k, (y_k)_k, (b_k)_k \in [0, \infty)^{\mathbb{N}_0}$ be such that for some $A \geq 0$,*

$$y_{k+1} \leq (1 + Ah_k)y_k + b_k, \quad k \in [N - 1]_0.$$

Then

$$y_{k+1} \leq \left(y_0 + \sum_{\ell \in [N-1]_0} b_\ell \right) \exp\left(\sum_{0 \leq j \leq k} Ah_j \right), \quad k \in [N - 1]_0.$$

Proof Rewriting the upper bound on y_{k+1} yields

$$y_{j+1} - y_j \leq Ah_j y_j + b_j, \quad j \in [N - 1]_0.$$

Summing the differences from $j = 0$ to $j = k \in [N - 1]_0$ yields

$$y_{k+1} \leq y_0 + \sum_{0 \leq j \leq k} (Ah_j y_j + b_j) \leq \left(y_0 + \sum_{0 \leq \ell \leq N-1} b_\ell \right) + \sum_{0 \leq j \leq k} Ah_j y_j, \quad k \in [N - 1]_0.$$

Applying Proposition C.2 completes the proof.   □

Proofs for Section 3.1

Proof of Lemma 3.5

Lemma 3.5 states that the stochastic process $(M_k)_{k \in [N-1]_0}$ defined by (3.9),

$$M_k := \sum_{j=0}^{k} \left\langle \varphi(h_j, t_j, u(t_j)) - \psi(h_j, t_j, U_j), \xi_j(h_j) \right\rangle_H,$$

is a $\mathbb{R}$-valued, square-integrable martingale with respect to the filtration $(\mathcal{F}_k)_{k \in [N-1]_0}$ generated by the $(\xi_k(h_k))_{k \in [N-1]_0}$, and that there exists a universal constant $\kappa > 0$ such that for every $k \in [N-1]_0$, the bound (3.10) holds:

$$\mathbb{E}\left[\max_{j \in [k]_0} |M_k|\right] \leq \left\|C_{\varphi,\psi}\right\|_{\infty}^{2} h^{2q+1} + \frac{1}{4}\mathbb{E}\left[\max_{j \in [k]_0} |e_j|_H^2\right] + \kappa^2 (1 + L_\psi') T C_\xi^2 h^{2p+1}.$$

Proof of Lemma 3.5 For $(M_k)_k$ to satisfy the definition of a martingale, we must show that for every k, $M_k \in L^1(\Omega;\mathbb{R})$ and is $\mathcal{F}_k$-measurable, and that the martingale property $\mathbb{E}[M_{k+1} - M_k | \mathcal{F}_k] = 0$ holds for $k \in [N-2]_0$. The measurability of M_k with respect to $\mathcal{F}_k$ follows from the definition of $\mathcal{F}_k$ and Lemma 3.3. By the triangle inequality, the Cauchy–Schwarz inequality, (3.7), and Assumption 3.2,

$$
\begin{aligned}
\left\|M_k\right\|_{L^2(\Omega;\mathbb{R})} &\leq \sum_{j=0}^{k} \left\|\varphi(h_j, t_j, u(t_j)) - \psi(h_j, t_j, U_j)\right\|_{L^2(\Omega;H)}^{2} \left\|\xi_j(h_j)\right\|_{L^2(\Omega;H)}^{2} \\
&\leq \sum_{j=0}^{k} \left(3\left\|C_{\varphi,\psi}\right\|_{\infty}^{2} h_k^{2q+1} + (1 + L_\psi' h_k)^2 \left\|e_k\right\|_{L^2(\Omega;H)}^{2}\right) C_\xi^2 h^{2p+2}.
\end{aligned}
$$

By Lemma 3.4, $e_k \in L^2(\Omega;H)$ for every k, and thus $M_k \in L^2(\Omega;\mathbb{R}) \subset L^1(\Omega;\mathbb{R})$. Hence, $(M_k)_k$ is a square-integrable martingale.

Next, we prove the martingale property. By Lemma 3.3, $\xi_{k+1}(h_{k+1})$ is independent of U_{k+1} and $\mathcal{F}_k$. By the definition (3.9) of M_k, the tower property of conditional expectation, and the centredness of the $(\xi_k(h_k))_k$,

$$\mathbb{E}[\left\langle \varphi(h_{k+1}, u(t_{k+1})) - \psi(h_{k+1}, U_{k+1}), \xi_{k+1}(h_{k+1}) \right\rangle_H] = 0,$$

and thus $(M_k)_k$ is a $(\mathcal{F}_k)_k$-martingale.

Finally, we prove the second statement. Since $(M_k)_k$ is a square integrable martingale, the Burkholder–Davis–Gundy inequality ensures that for every $k \in [N-1]_0$

$$\mathbb{E}\left[\max_{j \in [k]_0}\left|M_j\right|\right] \leq \kappa \mathbb{E}\left[\langle M \rangle_k^{1/2}\right]$$

where $\kappa > 0$ is the same universal constant appearing in (3.10). The quadratic variation process $\langle M \rangle$ is defined by $\langle M \rangle_0 := 0$ and $\langle M \rangle_k := \sum_{j \in [k]} \mathbb{E}[(M_j - M_{j-1})^2 | \mathcal{F}_{j-1}]$ for $k \in [N-1]$, see e.g. [32, Chapter I, Definition 2.3]. Using (3.9), the measurability of U_j with respect to $\mathcal{F}_{j-1}$ (cf. Lemma 3.3), the Cauchy–Schwarz inequality, and (3.7),

$$\langle M \rangle_k \leq \sum_{j \in [k]} \mathbb{E}\left[\left\langle \varphi(h_j, t_j, u(t_j)) - \psi(h_j, t_j, U_j), \xi_j(h_j) \right\rangle_H^2 \Big| \mathcal{F}_{j-1} \right]$$

$$\leq \max_{j \in [k]} \left| \varphi(h_j, t_j, u(t_j)) - \psi(h_j, t_j, U_j) \right|_H^2 \sum_{j \in [k]} \mathbb{E}\left[\left| \xi_j(h_j) \right|_H^2 \Big| \mathcal{F}_{j-1} \right]$$

$$\leq \max_{j \in [k]} \left(3 \left\| C_{\varphi,\psi} \right\|_\infty^2 h_j^{2q+1} + (1 + L_\psi' h_j) \left| e_j \right|_H^2 \right) \sum_{j \in [k]} \mathbb{E}\left[\left| \xi_j(h_j) \right|_H^2 \Big| \mathcal{F}_{j-1} \right].$$

The hypothesis that $0 < h \leq 1$ was used to obtain (3.7). Using Young's inequality with $s, s' > 1$ such that $s^{-1} + (s')^{-1} = 1$

$$ab \leq \frac{\delta}{s} a^s + \frac{1}{\delta^{s'/s} s'} b^{s'}$$

with $s = 2$ and $\delta = 2\kappa(1 + L_\psi' h)$, and using $h_k \leq h \leq 1$,

$$\langle M \rangle_k^{1/2} \leq \frac{1}{\kappa(4 + 4L_\psi' h)} \max_{j \in [k]} \left(3 \left\| C_{\varphi,\psi} \right\|_\infty^2 h_j^{2q+1} + (1 + L_\psi' h_j) \left| e_j \right|_H^2 \right)$$

$$+ \kappa(1 + L_\psi' h) \sum_{j \in [k]} \mathbb{E}\left[\left| \xi_j(h_j) \right|_H^2 \Big| \mathcal{F}_{j-1} \right]$$

$$\leq \frac{\left\| C_{\varphi,\psi} \right\|_\infty^2}{\kappa} h^{2q+1} + \frac{1}{4\kappa} \max_{j \in [k]_0} |e_j|_H^2 + \kappa(1 + L_\psi') \sum_{j \in [N-1]_0} \mathbb{E}\left[\left| \xi_j(h_j) \right|_H^2 \Big| \mathcal{F}_{j-1} \right].$$

By taking expectations, the tower property removes the conditioning on $\mathcal{F}_{j-1}$ in each summand. Using (1.4) and the Burkholder–Davis–Gundy inequality completes the proof. $\qquad\square$

Proof of Proposition 3.6

Proposition 3.6 states the error bound

$$\left\| \max_{k \in [N]_0} |e_k|_H \right\|_2^2$$

$$\leq 2 \left(\|e_0\|_2^2 + 4 \left\| C_{\varphi,\psi} \right\|_\infty^2 h^{2q} T + C_\xi^2 T h^{2p+1} (1 + \kappa^2 (1 + L_\psi')) \right) \exp\left(2 L_\psi' T \right)$$

for the same universal constant κ in (3.10).

Proof of Proposition 3.6 Since $0 < h \leq 1$, we may use (3.8) to obtain

$$|e_{k+1}|_H^2 - |e_k|_H^2 \leq L'_\psi h_k |e_k|_H^2 + 3\left\|C_{\varphi,\psi}\right\|_\infty^2 h_k^{2q+1} + |\xi_k(h_k)|_H^2$$
$$+ 2\langle \varphi(h_k, t_k, u(t_k)) - \psi(h_k, t_k, U_k), \xi_k(h_k)\rangle_H.$$

Using that $\sum_{j\in[k+1]_0}(|e_{k+1}|_H^2 - |e_k|_H^2) = |e_{k+1}|_H^2 - |e_0|_H^2$, and using the definition (3.9) of the martingale $(M_k)_k$, we obtain

$$|e_{k+1}|_H^2 \leq |e_0|_H^2 + \sum_{j\in[N-1]_0}\left(3\left\|C_{\varphi,\psi}\right\|_\infty^2 h_j^{2q+1} + \left|\xi_j(h_j)\right|_H^2\right)$$
$$+ 2M_k + L'_\psi \sum_{j\in[k]_0} h_j |e_j|_H^2.$$

Since only M_k can attain negative values, the above bound implies

$$\max_{j\in[k+1]_0}|e_j|_H^2 \leq |e_0|_H^2 + \sum_{j\in[N-1]_0}\left(3\left\|C_{\varphi,\psi}\right\|_\infty^2 h_j^{2q+1} + \left|\xi_j(h_j)\right|_H^2\right)$$
$$+ 2\max_{j\in[k]_0}|M_j| + L'_\psi \sum_{j\in[k]_0} h_j \max_{\ell\in[j]_0}|e_\ell|_H^2.$$

Take expectations, apply Assumption 3.2, apply the bound (3.10) from Lemma 3.5, and use that $\mathbb{E}[\max_{j\in[k]_0}|e_j|_H^2] \leq \mathbb{E}[\max_{j\in[k+1]_0}|e_j|_H^2]$ to obtain

$$\mathbb{E}\left[\max_{j\in[k+1]_0}|e_j|_H^2\right] \leq \mathbb{E}\left[|e_0|_H^2\right] + 3\left\|C_{\varphi,\psi}\right\|_\infty^2 Th^{2q} + C_\xi^2 Th^{2p+1}$$
$$+ 2\mathbb{E}\left[\max_{j\in[k]_0}|M_j|\right] + L'_\psi \sum_{j\in[k]_0} h_j\mathbb{E}\left[\max_{\ell\in[j]_0}|e_\ell|_H^2\right]$$
$$\leq \mathbb{E}\left[|e_0|_H^2\right] + 4\left\|C_{\varphi,\psi}\right\|_\infty^2 h^{2q}T + C_\xi^2 Th^{2p+1}(1 + \kappa^2(1 + L'_\psi))$$
$$+ \frac{1}{2}\mathbb{E}\left[\max_{j\in[k+1]_0}|e_j|_H^2\right] + L'_\psi \sum_{j\in[k]_0} h_j\mathbb{E}\left[\max_{\ell\in[j]_0}|e_\ell|_H^2\right].$$

Subtracting $\frac{1}{2}\mathbb{E}[\max_{j\in[k+1]_0}|e_j|_H^2]$ from both sides and applying the discrete Gronwall inequality in Proposition C.2 completes the proof.   □

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Abdulle, A., Garegnani, G.: Random time step probabilistic methods for uncertainty quantification in chaotic and geometric numerical integration. Stat. Comput. **30**(4), 907–932 (2020). https://doi.org/10.1007/s11222-020-09926-w

2. Abdulle, A., Garegnani, G.: A probabilistic finite element method based on random meshes: error estimators and Bayesian inverse problems. Comput. Methods Appl. Mech. Eng. **384**, 113961 (2021). https://doi.org/10.1016/j.cma.2021.113961

3. Adams, R.A., Fournier, J.J.F.: Sobolev spaces, Pure and Applied Mathematics (Amsterdam), vol. 140, second edn. Elsevier/Academic Press, Amsterdam (2003)

4. Bogachev, V.I.: Gaussian Measures, Mathematical Surveys and Monographs, vol. 62. American Mathematical Society, Providence, RI (1998). https://doi.org/10.1090/surv/062

5. Chkrebtii, O.A., Campbell, D.A.: Adaptive step-size selection for state-space probabilistic differential equation solvers. Stat. Comput. **29**(6), 1285–1295 (2019). https://doi.org/10.1007/s11222-019-09899-5

6. Chkrebtii, O.A., Campbell, D.A., Calderhead, B., Girolami, M.A.: Bayesian solution uncertainty quantification for differential equations. Bayesian Anal. **11**(4), 1239–1267 (2016). https://doi.org/10.1214/16-BA1017

7. Ciarlet, P.G.: Linear and Nonlinear Functional Analysis with Applications, vol. 130. Society for Industrial and Applied Mathematics, Philadelphia (2013)

8. Cockayne, J., Oates, C., Sullivan, T.J., Girolami, M.: Probabilistic numerical methods for PDE-constrained Bayesian inverse problems. In: G. Verdoolaege (ed.) Proceedings of the 36th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, AIP Conference Proceedings, vol. 1853, pp. 060001–1–060001–8 (2017). https://doi.org/10.1063/1.4985359

9. Cockayne, J., Oates, C.J., Sullivan, T.J., Girolami, M.: Bayesian probabilistic numerical methods. SIAM Rev. **61**(4), 756–789 (2019). https://doi.org/10.1137/17M1139357

10. Conrad, P.R., Girolami, M., Särkkä, S., Stuart, A.M., Zygalakis, K.C.: Statistical analysis of differential equations: introducing probability measures on numerical solutions. Stat. Comput. **27**(4), 1065–1082 (2017). https://doi.org/10.1007/s11222-016-9671-0

11. Dupont, M., Enßlin, T.: Consistency and convergence of simulation schemes in information field dynamics. Phys. Rev. E **98**, 043307 (2018). https://doi.org/10.1103/PhysRevE.98.043307

12. Emmrich, E.: Gewöhnliche und Operator-Differentialgleichungen. Eine integrierte Einführung in Randwertprobleme und Evolutionsgleichungen für Studierende. Wiesbaden: Vieweg (2004). https://doi.org/10.1007/978-3-322-80240-8

13. Emmrich, E., Weckner, O.: Analysis and numerical approximation of an integro-differential equation modeling non-local effects in linear elasticity. Math. Mech. Solids **12**(4), 363–384 (2007). https://doi.org/10.1177/1081286505059748

14. Enßlin, T.A.: Information field dynamics for simulation scheme construction. Phys. Rev. E **87**, 013308 (2013). https://doi.org/10.1103/PhysRevE.87.013308

15. Garegnani, G.: Sampling methods for Bayesian inference involving convergent noisy approximations of forward maps (2021). ArXiv:2111.03491

16. Hairer, E., Nørsett, S.P., Wanner, G.: Solving ordinary differential equations. I, Springer Series in Computational Mathematics, vol. 8, second edn. Springer-Verlag, Berlin (1993). Nonstiff problems

17. Hennig, P., Osborne, M.A., Girolami, M.: Probabilistic numerics and uncertainty in computations. P. Roy. Soc. Lond. A Mat. **471**(2179), 20150142 (2015). https://doi.org/10.1098/rspa.2015.0142

18. Holte, J.M.: Discrete Gronwall lemma and applications (2009). http://homepages.gac.edu/~holte/publications/GronwallLemma.pdf. Accessed 14-12-2021
19. Kersting, H., Sullivan, T.J., Hennig, P.: Convergence rates of Gaussian ODE filters. Stat. Comput. **30**(6), 1791–1816 (2020). https://doi.org/10.1007/s11222-020-09972-4
20. Lie, H.C., Stuart, A.M., Sullivan, T.J.: Strong convergence rates of probabilistic integrators for ordinary differential equations. Stat. Comput. **29**(6), 1265–1283 (2019). https://doi.org/10.1007/s11222-019-09898-6
21. Lie, H.C., Sullivan, T.J., Teckentrup, A.L.: Random forward models and log-likelihoods in Bayesian inverse problems. SIAM/ASA J. Uncertain. Quantif. **6**(4), 1600–1629 (2018). https://doi.org/10.1137/18M1166523
22. Lubich, C., Ostermann, A.: Linearly implicit time discretization of non-linear parabolic equations. IMA J. Numer. Anal. **15**(4), 555–583 (1995). https://doi.org/10.1093/imanum/15.4.555
23. Lubich, C., Ostermann, A.: Runge-Kutta approximation of quasi-linear parabolic equations. Math. Comp. **64**(210), 601–627 (1995). https://doi.org/10.2307/2153442
24. Matsuda, T., Miyatake, Y.: Estimation of ordinary differential equation models with discretization error quantification. SIAM/ASA J. Uncertain. Quantif. **9**(1), 302–331 (2021). https://doi.org/10.1137/19M1278405
25. Meinlschmidt, H., Meyer, C., Walther, S.: Optimal control of an abstract evolution variational inequality with application to homogenized plasticity. J. Nonsmooth Anal. Optim. **1**, 1–41 (2020). https://doi.org/10.46298/jnsao-2020-5800
26. Milstein, G.N., Tretyakov, M.V.: Stochastic numerics for mathematical physics. Scientific Computation. Springer-Verlag, Berlin (2004). https://doi.org/10.1007/978-3-662-10063-9
27. Oates, C.J., Sullivan, T.J.: A modern retrospective on probabilistic numerics. Stat. Comput. **29**(6), 1335–1351 (2019). https://doi.org/10.1007/s11222-019-09902-z
28. Owhadi, H.: Bayesian numerical homogenization. Multiscale Model. Simul. **13**(3), 812–828 (2015). https://doi.org/10.1137/140974596
29. Owhadi, H., Zhang, L.: Gamblets for opening the complexity-bottleneck of implicit schemes for hyperbolic and parabolic ODEs/PDEs with rough coefficients. J. Comput. Phys. **347**, 99–128 (2017). https://doi.org/10.1016/j.jcp.2017.06.037
30. Plato, R.: Numerische Mathematik kompakt. Grundlagenwissen für Studium und Praxis. Wiesbaden: Vieweg (2004). https://doi.org/10.1007/978-3-8348-9644-5
31. Raissi, M., Perdikaris, P., Karniadakis, G.E.: Numerical Gaussian processes for time-dependent and nonlinear partial differential equations. SIAM J. Sci. Comput. **40**(1), A172–A198 (2018). https://doi.org/10.1137/17M1120762
32. Revuz, D., Yor, M.: Continuous Martingales and Brownian Motion, Grundlehren der mathematischen Wissenschaften, vol. 293, third edn. Springer-Verlag, Berlin (2009). Corrected Third Printing
33. Schober, M., Särkkä, S., Hennig, P.: A probabilistic model for the numerical solution of initial value problems. Stat. Comput. **29**(1), 99–122 (2019). https://doi.org/10.1007/s11222-017-9798-7
34. Stuart, A.M.: Inverse problems: a Bayesian perspective. Acta Numer. **19**, 451–559 (2010). https://doi.org/10.1017/S0962492910000061
35. Teymur, O., Lie, H.C., Sullivan, T.J., Calderhead, B.: Implicit probabilistic integrators for ODEs. In: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (eds.) Advances in Neural Information Processing Systems 31 (NIPS 2018), vol. 31. Curran Associates, Inc. (2018). http://papers.nips.cc/paper/7955-implicit-probabilistic-integrators-for-odes
36. Teymur, O., Zygalakis, K., Calderhead, B.: Probabilistic Linear Multistep Methods. In: D.D. Lee, M. Sugiyama, U.V. Luxburg, I. Guyon, R. Garnett (eds.) Advances in Neural Information Processing Systems 29, pp. 4321–4328. Curran Associates, Inc. (2016). http://papers.nips.cc/paper/6356-probabilistic-linear-multistep-methods.pdf
37. Thomée, V.: Galerkin finite element methods for parabolic problems. Berlin: Springer (2006). https://doi.org/10.1007/3-540-33122-0
38. Tronarp, F., Kersting, H., Särkkä, S., Hennig, P.: Probabilistic solutions to ordinary differential equations as nonlinear Bayesian filtering: a new perspective. Stat. Comput. **29**(6), 1297–1315 (2019). https://doi.org/10.1007/s11222-019-09900-1

39. Wang, J., Cockayne, J., Chkrebtii, O., Sullivan, T.J., Oates, C.: Bayesian numerical methods for nonlinear partial differential equations. Stat. Comput. **31**(5), no. 55, 20pp. (2021). https://doi.org/10.1007/s11222-021-10030-w
40. Wang, J., Cockayne, J., Oates, C.: A role for symmetry in the Bayesian solution of differential equations. Bayesian Anal. **15**(4), 1057–1085 (2020). https://doi.org/10.1214/19-BA1183
41. Zeidler, E.: Nonlinear Functional Analysis and its Applications. II/A. Springer-Verlag, New York (1990). Linear Monotone Operators, Translated from the German by the author and Leo F. Boron. https://doi.org/10.1007/978-1-4612-0985-0