

# Strong convergence rates of probabilistic integrators for ordinary differential equations

Han Cheng Lie · A. M. Stuart · T. J. Sullivan

Received: September 28, 2018 / Accepted: December 28, 2018 / Handling Editor: C. J. Oates

This is a post-peer-review, pre-copyedit version of an article published in *Statistics and Computing* (2019). The final authenticated version is available online at: <https://doi.org/10.1007/s11222-019-09898-6>.

**Abstract** Probabilistic integration of a continuous dynamical system is a way of systematically introducing discretisation error, at scales no larger than errors introduced by standard numerical discretisation, in order to enable thorough exploration of possible responses of the system to inputs. It is thus a potentially useful approach in a number of applications such as forward uncertainty quantification, inverse problems, and data assimilation. We extend the convergence analysis of probabilistic integrators for deterministic ordinary differential equations, as proposed by Conrad et al. (*Stat. Comput.*,

2017), to establish mean-square convergence in the uniform norm on discrete- or continuous-time solutions under relaxed regularity assumptions on the driving vector fields and their induced flows. Specifically, we show that randomised high-order integrators for globally Lipschitz flows and randomised Euler integrators for dissipative vector fields with polynomially-bounded local Lipschitz constants all have the same mean-square convergence rate as their deterministic counterparts, provided that the variance of the integration noise is not of higher order than the corresponding deterministic integrator. These and similar results are proven for probabilistic integrators where the random perturbations may be state-dependent, non-Gaussian, or non-centred random variables.

**Keywords** probabilistic numerical methods · ordinary differential equations · convergence rates · uncertainty quantification

**Mathematics Subject Classification (2000)**  
65L20 · 65C99 · 37H10 · 68W20

HCL and TJS are supported by the Freie Universität Berlin within the Excellence Initiative of the German Research Foundation. HCL is supported by the Universität Potsdam. AMS is grateful to DARPA, EPSRC and ONR for funding. This material was based upon work partially supported by the National Science Foundation under Grant DMS-1127914 to the Statistical and Applied Mathematical Sciences Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of these funding agencies and institutions.

Han Cheng Lie  
Institute of Mathematics, Universität Potsdam, Campus Golm, Haus 9, Karl-Liebknecht Str. 24-25, 14476 Potsdam OT Golm, Germany  
<https://orcid.org/0000-0002-6905-9903>  
E-mail: hanlie@uni-potsdam.de

A. M. Stuart  
Department of Computing and Mathematical Sciences, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, United States of America  
E-mail: astuart@caltech.edu

T. J. Sullivan  
Freie Universität Berlin, and Zuse Institute Berlin, Takustrasse 7, 14195 Berlin, Germany  
E-mail: sullivan@zib.de

## 1 Introduction

This article concerns the analysis of probabilistic numerical integrators for deterministic initial value problems of the form

$$\begin{aligned} \frac{d}{dt} u(t) &= f(u(t)), \quad \text{for } 0 \leq t \leq T, \\ u(0) &= u_0 \in \mathbb{R}^d \end{aligned} \quad (1.1)$$

where  $T > 0$ . Let  $\Phi^t: \mathbb{R}^d \rightarrow \mathbb{R}^d$  denote the flow induced by (1.1), so that

$$\Phi^t(u_0) = u_0 + \int_0^t f(\Phi^s(u_0)) ds \quad (1.2)$$

for all  $(t, u_0) \in [0, T] \times \mathbb{R}^d$ . Given an integration time step  $\tau > 0$  such that  $K := T/\tau \in \mathbb{N}$  and the corresponding time mesh

$$t_k := k\tau \text{ for } k \in [K] := \{0, 1, \dots, K\}, \quad (1.3)$$

a deterministic one-step numerical method for the solution of (1.1) is a numerical flow map  $\Psi^\tau: \mathbb{R}^d \rightarrow \mathbb{R}^d$  that generates approximations  $\tilde{u}_k \approx u_k := u(t_k)$  by the recursion  $\tilde{u}_k := \Psi^\tau(\tilde{u}_{k-1})$ ; note that  $u_k = \Phi^\tau(u_{k-1})$ . A key property of the numerical method is its global order of convergence, i.e. the largest  $q \geq 0$  such that, for some constant  $C = C(T)$ , independent of  $\tau$ ,

$$\max_{k \in [K]} \|u_k - \tilde{u}_k\| \leq C\tau^q. \quad (1.4)$$

As a modelling choice, epistemic stochasticity can be introduced into the numerical solution of (1.1) on the basis that, while the exact solution satisfies

$$u_{k+1} = \Phi^\tau(u_k) = u_k + \int_{t_k}^{t_{k+1}} f(u(s)) ds \quad (1.5)$$

for all  $k \in [K-1]$ , the only information available about the values of the solution off the time mesh comes from the numerical solution on the mesh, and so the integrand  $f(u(s))$  is not exactly accessible. This uncertainty is relevant in the setting where, given a large-scale mathematical model, it may be more statistically informative to spend computational resources on solving a differential equation-based model many times on a coarser grid than on solving the same model a few times on a finer grid. This is often the case in forward uncertainty quantification (Smith 2014; Sullivan 2015), inverse problems (Kaipio and Somersalo 2005; Stuart 2010), and data assimilation (Law et al. 2015; Reich and Cotter 2015); the area of multi-level Monte Carlo methods makes particular use of this kind of cost-accuracy tradeoff (Giles 2015). Furthermore, in many such settings, the quantity of interest is often not the solution of a differential equation-based model, but a functional thereof. In all cases, estimates of the off-mesh uncertainty due to the numerical method can and should be fed forward to estimate the uncertainty in the quantity of interest.

This article is motivated by the work of Conrad et al. (2017), in which one seeks to model the off-mesh uncertainty by considering probabilistic solvers. For the same mesh given in (1.3), the probabilistic solver of

Conrad et al. (2017) involves producing a sequence of random variables  $(U_k)_{k \in [K]}$  according to

$$U_{k+1} := \Psi^\tau(U_k) + \xi_k(\tau), \quad U_0 = u_0, \quad (1.6)$$

where  $\Psi^\tau$  is the map associated to the deterministic numerical method, and each  $\xi_k(\tau)$  is an i.i.d. copy of a random variable  $\xi_0(\tau) := \int_0^\tau \chi_0(s) ds$ , where  $\chi_0$  is a stochastic process over the time interval  $[0, \tau]$  that models the off-mesh behaviour of the unknown function  $f(u(s))$  in (1.5). We refer the reader to Conrad et al. (2017, Figure 2) for a pictorial representation of (1.6). The process  $\chi_0$  is introduced so that one can probe the uncertainty induced by the mesh  $(t_k)_{k \in [K]}$  and the underlying solver, and thus explore possible responses of the system to inputs. Comparing (1.5) and (1.6), it follows that the random variable  $\xi_k(\tau)$  is a statistical model for the approximation error  $\Phi^\tau(u_k) - \Psi^\tau(u_k)$ .

We emphasise that the additive, state-independent noise model appearing in (1.6) should be interpreted as providing a prior on the local truncation error (Hairer et al. 2009). A frequent criticism levelled at the field of probabilistic numerical integration is that the statistical properties of the noise  $\xi_k$  that have been imposed in existing published works do not reflect known prior information about local truncation error. Here we address this issue by considerably weakening the assumptions made on the  $\xi_k$ . However we anticipate future work in this direction, especially when specific structure on the vector field  $f$  is used to further inform the prior. Note also that, in the presence of large amounts of data, we expect posterior contraction and forgetting of the prior; see, e.g., Knapik et al. (2011). Posterior contraction for (1.6) was demonstrated numerically on a number of examples by Conrad et al. (2017).

In the spirit of (1.4), the main convergence result (Conrad et al. 2017, Theorem 2.2) yields that, if the vector field  $f$  in (1.1) is globally Lipschitz, if the deterministic numerical method has uniform local truncation of order  $q+1$ , and if  $\chi_0$  is a centred Gaussian process such that the second moment of  $\xi_0(\tau)$  decays as  $\tau^{2p+1}$  for some  $p \geq 1$ , then

$$\max_{k \in [K]} \mathbb{E}[\|u_k - U_k\|^2] \leq C\tau^{2 \min\{p, q\}}. \quad (1.7)$$

This shows that the convergence rate of the probabilistic solver (1.6) is determined by the convergence of the ‘worst-case error’ of the deterministic method  $\Psi^\tau$ , and the convergence of the ‘statistical error’  $\xi_0$ , as described by the parameters  $q$  and  $p$  respectively. Choosing  $\xi_0$  with  $p = q$  introduces the maximum amount of solution uncertainty consistent with preserving the order of accuracy of the original deterministic integrator.

It is important to stress that, despite the apparent similarities between (1.6) and Euler–Maruyama schemes

for stochastic differential equations (SDEs) driven by Brownian motion, the analysis of the latter does not directly apply to probabilistic solvers, even though we will borrow some techniques from that field. This is because the variance of  $\xi_0(\tau)$  for probabilistic solvers of the form (1.6) is assumed to decay to zero strictly faster than  $\tau$ , whereas, for SDEs driven by Brownian motion, the variance is proportional to  $\tau$ . A key aspect of this work is to determine how to scale the noise so that the rate of convergence of the underlying deterministic numerical integrator is not affected, yet uncertainty arising from numerical approximation is accounted for.

### 1.1 Contribution and outline of the paper

The purpose of this paper is to make significant extensions of the convergence analysis of Conrad et al. (2017) for (1.1). We accomplish this by obtaining stronger error estimates (and hence stronger convergence results) under assumptions on both the underlying differential equation and on the noise model for probabilistic numerical integration that are weaker than their counterparts in Conrad et al. (2017). The convergence results of this paper are of the form

$$\mathbb{E} \left[ \max_{k \in [K]} \|u_k - U_k\|^n \right] \leq C \tau^{n \cdot \min\{p-c, q\}}, \quad (1.8)$$

where  $n \in \mathbb{N}$ ,  $q$  is the order of the numerical method  $\Psi^\tau$ ,  $p$  is an exponent of decay in the moments of the random variables  $(\xi_k(\tau))_{k \in [K]}$ , and  $c \geq 0$  is a penalty term in the convergence rate that depends solely on the random variables  $(\xi_k(\tau))_{k \in [K]}$ . Note that, when  $c = 0$  and  $n = 2$ , the convergence rate of  $n \min\{p - c, q\}$  on the right-hand side of (1.8) agrees with that of (1.7) shown by Conrad et al. (2017), so that the right-hand sides of (1.7) and (1.8) differ only in the constant prefactor  $C$ . However, because the time supremum is inside the expectation, (1.8) implies (1.7). Furthermore, by Markov's inequality, (1.8) yields an estimate of the frequentist coverage of the true solution  $u$  by the randomised solutions  $U$ :

$$\mathbb{P} \left[ \max_{k \in [K]} \|u_k - U_k\| \leq r \right] \geq 1 - C \tau^{n \cdot \min\{p-c, q\}} r^{-n};$$

such estimates are useful in the context of forward uncertainty quantification and inverse problems (Lie et al. 2018).

We emphasize that, in addition to strengthening the form of the convergence results so that the supremum is inside the expectation, we also prove the results in this paper under weaker assumptions on the vector field  $f$ , and under weaker assumptions on the noise  $\xi_k$ , than

those employed by Conrad et al. (2017). Specifically we do not assume that  $f$  and its derivatives are globally bounded, and we do not assume that the random variables are Gaussian; furthermore in results generalizing (1.8) we relax the assumption that the random variables are centred, paving the way for future analyses which incorporate specific known structure and bias in the truncation error.

Error estimates like (1.8) show that the randomised numerical solution has convergence properties that are asymptotically no worse than the deterministic numerical solution. This can be interpreted as saying that the trajectories obtained from the randomised numerical integrator are all equally valid approximations to the solution of the original system, modulo the uncertainty induced by solving in discrete time. This can be useful for many purposes, for example in studying limits on predictability in chaotic systems, as shown for the Lorenz-63 system by Conrad et al. (2017).

After introducing some notation and auxiliary results in Section 2, the rest of the paper is organised as follows. In Section 3, Theorem 3.4 yields (1.8) for numerical methods of arbitrary order, for vector fields  $f$  whose induced flow maps  $\Phi^\tau$  are globally Lipschitz — including one-sided Lipschitz vector fields — and for collections  $(\xi_k(\tau))_{k \in [K]}$  of random variables that are independent and centred, but not necessarily Gaussian. Conrad et al. (2017) assumed the vector field  $f$  to be globally Lipschitz, and the random variables  $(\xi_k(\tau))_{k \in [K]}$  were assumed to be i.i.d. centred Gaussian random variables. In Theorem 3.5, we prove a result similar to (1.8) in which we relax the assumption that the  $(\xi_k(\tau))_{k \in [K]}$  are independent and that they are centred; the price we pay for these weaker constraints on the noise is a stronger decay assumption, with respect to the time-step, on the second moments of the  $(\xi_k(\tau))_{k \in [K]}$ . We use this assumption in order to introduce the maximal noise that is consistent with retaining the rate of convergence of the underlying deterministic numerical integrator.

In Section 4, we further weaken the conditions on the vector field  $f$ , by considering locally Lipschitz vector fields that satisfy a polynomial growth condition. In Theorem 4.2, we show that, under the assumption that the  $(\xi_k(\tau))_{k \in [K]}$  are almost surely bounded, we can again obtain (1.8). In Theorem 4.5, we remove the almost-sure boundedness condition, but add the assumption that the vector field  $f$  satisfies a generalised dissipativity condition.

In Section 5 we discuss a continuous-time analogue of (1.6), and show how convergence results of the form (1.8) can be obtained. We also show that there exists a nonempty set of random variables (or more gener-

ally, stochastic processes) that satisfy the regularity assumptions on the random variables  $(\xi_k(\tau))_{k \in [K]}$  used throughout this paper.

Proofs of the results may be found in Appendix A.

## 1.2 Review of probabilistic numerical methods

Continuous relationships such as ODEs and PDEs are commonplace as forward models in uncertainty quantification problems, or as Bayesian likelihoods in modern statistical inverse problems (Kaipio and Somersalo 2005; Stuart 2010), and in particular in data assimilation algorithms with critical everyday applications such as numerical weather prediction (Law et al. 2015; Reich and Cotter 2015). The use of a discretised solver for such forward models is usually unavoidable in practice, but introduces an additional source of uncertainty both into forward propagation of uncertainty and into subsequent inferences. While the solution to the ODE/PDE may not be random in the frequentist sense, it is nonetheless only imperfectly known through the discretised numerical solution. Probability in the subjective or Bayesian sense is one appropriate means of representing this epistemic uncertainty, particularly if the ODE/PDE solution forms part of the forward model in a Bayesian inverse problem. Failure to properly account for discretisation errors and uncertainties can lead to biased, inconsistent, and over-confident inferences (Conrad et al. 2017).

Probabilistic numerical solutions of problems such as the solution of ODEs have a long history. Modern foundations for this field were laid by the work of Diaconis (1988), O’Hagan (1992), and Skilling (1992) under the term of “Bayesian numerical analysis”. More recently, such ideas have received renewed attention under the term “probabilistic numerics” (Hennig et al. 2015; Cockayne et al. *SIAM Rev.*, to appear): the discussion of probabilistic numerical methods for ordinary differential equations given by Schober et al. (2014); Conrad et al. (2017); Chkrebtii et al. (2016), and Teymur et al. (2018) is particularly relevant here. Also of interest in the field of probabilistic numerics, but not directly relevant to the present work, are probabilistic numerical methods for linear algebra (Hennig 2015), optimisation (Gonzalez et al. 2016), partial differential equations (Cockayne et al. 2017; Owhadi 2015, 2017; Wang et al. 2018), and quadrature (Briol et al. 2015). In particular, Cockayne et al. (*SIAM Rev.*, to appear) sets out some axiomatic foundations for probabilistic numerical methods broadly conceived, and in particular what it means for a probabilistic numerical method to be “Bayesian”.

Randomised solutions of ODEs have also been studied in the context of stochastic or rough differential equations. In the case of non-autonomous ODEs driven by Carathéodory vector fields — i.e. vector fields that are locally integrable in time and continuous in the state space — it has been observed that randomised Euler and Runge–Kutta methods outperform their deterministic counterparts: see e.g. Stengle (1990); Jentzen and Neuenkirch (2009), and Kruse and Wu (2017) and the references therein.

We note that analysing the convergence properties of numerical solutions to (1.1) in terms of the approximation error for the solution, as in (1.7) and (1.8), is very much in the spirit of classical numerical analysis. For uncertainty quantification of the discretised solution of (1.1) as a stand-alone forward problem, this viewpoint is often sufficient. However, for applications to inverse problems and data assimilation, in which the numerical solution of the (1.1) is used to (approximately) evaluate the data misfit or likelihood, an alternative paradigm is to directly examine the impact of discretisation upon the quality of later inferences using e.g. Bayes factors (Capistrán et al. 2016; Christen 2017). There is also the well-established literature of information-based complexity and average-case analysis, with its greater emphasis on algorithmic aspects such as computational cost and optimal accuracy for given classes of information (Novak 1988; Ritter 2000; Traub and Woźniakowski 1980; Traub et al. 1983).

## 2 Setup and notation

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space sufficiently rich to serve as a common domain of definition for all the random variables and processes under consideration, and let  $\mathbb{E}$  denote expectation with respect to  $\mathbb{P}$ . The space of  $s^{\text{th}}$ -power integrable random variables over  $(\Omega, \mathcal{F}, \mathbb{P})$  will be denoted  $L_{\mathbb{P}}^s$ . The scalars  $C, C'$ , etc. denote non-negative constants whose value may change from occurrence to occurrence, but are independent of the time step  $\tau > 0$ .  $\text{Lip}(\Phi)$  denotes the best Lipschitz constant of  $\Phi: \mathbb{R}^d \rightarrow \mathbb{R}^d$ :

$$\text{Lip}(\Phi) := \min\{L \geq 0 \mid \|\Phi(x) - \Phi(y)\| \leq L\|x - y\|\}$$

for all  $x, y \in \mathbb{R}^d$ . We let  $\mathbb{N}$  denote the natural numbers beginning with 1, and  $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ . We shall sometimes abuse notation and write  $[K] := \{0, 1, \dots, K - 1\}$  or  $[K] := \{1, 2, \dots, K\}$ , and we shall write  $u_k := u(t_k) \equiv \Phi^\tau(u_{k-1})$  for the value of the exact solution to (1.1) at time  $t_k$ . We denote the minimum of a pair of real numbers  $a$  and  $b$  by  $a \wedge b = \min\{a, b\}$ .

It will be assumed throughout that  $T > 0$  is a fixed, deterministic time, and that  $f$  in (1.1) is sufficiently

smooth such that (1.1) has a unique solution for every initial condition  $u_0$ . The flow map  $\Phi^t$  associated to (1.1) is defined in (1.2), and the output of a one-step deterministic numerical integration method for a given  $x$  and time step  $\tau$  will be given by  $\Psi^\tau(x)$ . This setting encompasses many of the time-stepping methods in common use, such as Runge–Kutta methods of all orders.

The analysis of this paper will make repeated use of several useful inequalities, which are collected here for reference. First, recall Young’s inequality: for any  $\delta > 0$  and any pair of Hölder conjugate exponents  $r, r^* > 1$ ,

$$ab \leq \frac{\delta}{r} a^r + \frac{1}{r^* \delta^{r^*/r}} b^{r^*}, \quad \text{for all } a, b \geq 0. \quad (2.1)$$

Combining that inequality for  $r = r^* = 2$  with the Cauchy–Schwarz inequality in  $\mathbb{R}^d$  yields

$$\|x - y\|^2 \leq (1 + \delta)\|x\|^2 + (1 + \delta^{-1})\|y\|^2, \quad (2.2)$$

which will often be used either with  $\delta = 1$  or  $\delta = \tau$ .

The following discrete-time version of Grönwall’s inequality (Holte 2009) will also be useful:

**Theorem 2.1** *Let  $(x_k)_{k \in \mathbb{N}_0}$ ,  $(\alpha_k)_{k \in \mathbb{N}_0}$ , and  $(\beta_k)_{k \in \mathbb{N}_0}$  be non-negative sequences. If, for all  $k \in \mathbb{N}_0$ ,*

$$x_k \leq \alpha_k + \sum_{0 \leq j < k} \beta_j x_j \quad \text{and} \quad \alpha_k \leq A,$$

*then  $x_k \leq A \exp\left(\sum_{0 \leq j < k} \beta_j\right)$  for all  $k \in \mathbb{N}_0$ .*

For completeness, we state the following lemma.

**Lemma 2.1** *Let  $x, y \geq 0$ ,  $n \in \mathbb{N}$ , and  $\delta > 0$ . Then*

$$(x + y)^n \leq x^n(1 + \delta 2^{n-1}) + y^n(1 + (2/\delta)^{n-1}). \quad (2.3)$$

We shall also use the following inequality, which is valid for arbitrary  $N \in \mathbb{N}$  and  $m \geq 1$ : for all  $\{s_j\}_{j \in [N]} \in \mathbb{R}^N$ ,

$$\left| \sum_{j=1}^N s_j \right|^m \leq N^{m-1} \sum_{j=1}^N |s_j|^m. \quad (2.4)$$

This follows from

$$\begin{aligned} \left| \sum_{j=1}^N s_j \right|^m &\leq N^m \left( \frac{1}{N} \sum_{j=1}^N |s_j| \right)^m \leq N^m \left( \frac{1}{N} \sum_{j=1}^N |s_j|^m \right) \\ &= N^{m-1} \sum_{j=1}^N |s_j|^m, \end{aligned}$$

where we used Jensen’s inequality in the second inequality.

### 3 High-order integration of Lipschitz flows

The purpose of this section is to establish, given the initial value problem (1.1), the strong convergence result (1.8) for probabilistic solvers of the form (1.6), under the following assumptions.

**Assumption 3.1** *The vector field  $f$  admits  $0 < \tau^* \leq 1$  and  $C_\Phi \geq 1$ , such that for  $0 < \tau < \tau^*$ , the flow map  $\Phi^\tau$  defined by (1.2) is globally Lipschitz with Lipschitz constant  $\text{Lip}(\Phi^\tau) \leq 1 + C_\Phi \tau$ .*

As is well known, Assumption 3.1 holds if the generating vector field  $f$  is itself globally Lipschitz. However, Assumption 3.1 holds if, for instance,  $f$  merely satisfies the one-sided Lipschitz inequality

$$\langle f(x) - f(y), x - y \rangle \leq \mu \|x - y\|^2, \quad \text{for all } x, y \in \mathbb{R}^d,$$

for some constant  $\mu \in \mathbb{R}$ ; in this case, a calculation of  $\frac{d}{dt} \|u(t) - v(t)\|^2$  for trajectories  $u$  and  $v$  starting at initial conditions  $u_0, v_0 \in \mathbb{R}^d$  and an application of the differential version of Grönwall’s inequality shows that  $\|u(t) - v(t)\| \leq \exp(\mu|t|) \|u_0 - v_0\|$ , so that  $\text{Lip}(\Phi^t) \leq 1 + 2|\mu|t$  for small  $|t|$ .

**Assumption 3.2** *The numerical method  $\Psi^\tau$  has uniform local truncation error of order  $q + 1$ : for some constant  $C_\Psi \geq 1$  that does not depend on  $\tau$ ,*

$$\sup_{u \in \mathbb{R}^d} \|\Psi^\tau(u) - \Phi^\tau(u)\| \leq C_\Psi \tau^{q+1}.$$

Assumption 3.2 holds, in particular, for single- and multi-step methods derived from a  $q$ -times continuously differentiable vector field  $f$  with bounded  $q^{\text{th}}$  derivatives (Hairer et al. 2009, Section III.2). Imposing global bounds on the derivatives of  $f$ , and therefore on those of  $\Phi^\tau$ , forces us to consider a smaller class of flow maps  $\Phi^\tau$  than the class of flow maps that satisfy Assumption 3.1. We may alleviate this problem by weakening Assumption 3.2 to a bound of the form

$$\|\Phi^\tau(u) - \Psi^\tau(u)\| \leq C'(u) \tau^{q+1}, \quad (3.1)$$

with the consequence that the dependence of  $C'(u)$  on  $u$  must be specified; this dependence will vary according to the chosen numerical method  $\Psi^\tau$ . Moreover, whenever we apply (3.1) in place of Assumption 3.2 with a random variable  $U_k$  in place of a deterministic  $u_k$  — as we do below, e.g. in deriving (A.2) — we will need to ensure that  $\mathbb{E}[C'(U_k)]$  is finite, and of the correct order in  $\tau$  if necessary. In Section 4, we consider the implicit Euler method for a class of *locally* Lipschitz flow maps  $\Phi^\tau$ , obtain an expression for  $C'(U_k)$ , and with this expression obtain a bound of the form

$$\mathbb{E}[\|\Phi^\tau(U_k) - \Psi^\tau(U_k)\|^n] \leq C \tau^{n(q+1)}$$

where  $U_k$  denotes the output of the randomised numerical integrator according to (1.6),  $n \in \mathbb{N}$ , and  $C > 0$  does not depend on  $\tau$  or on  $k$ ; see Proposition 4.2. Note that there is no supremum inside the expectation in the inequality above. However, in this section, we shall apply Assumption 3.2 instead of (3.1), in order to avoid lengthy analyses that are specific to the choice of numerical method. We make no assumptions about how the integrator  $\Psi^\tau$  has been derived and treat it as a ‘black box’ satisfying Assumption 3.2.

**Assumption 3.3** *The random variables  $(\xi_k(\tau))_{k \in \mathbb{N}}$  admit parameters  $p \geq 1$ ,  $R \in \mathbb{N} \cup \{+\infty\}$ , and  $C_{\xi,R} \geq 1$ , independent of  $k$  and  $\tau$ , such that for all  $1 \leq r \leq R$  and all  $k \in \mathbb{N}$ ,*

$$\mathbb{E}[\|\xi_k(\tau)\|^r] \leq \left(C_{\xi,R} \tau^{p+1/2}\right)^r.$$

Note that we do not assume that the  $(\xi_k(\tau))_{k \in [K]}$  are identically distributed nor that they are centred. However we will impose these two additional assumptions in Theorem 3.4. The parameter  $p$  determines the decay rate of the  $r^{\text{th}}$  moments of the  $(\xi_k(\tau))_{k \in [K]}$ , for  $1 \leq r \leq R$ , while  $R$  determines the highest order moment for which the same decay behaviour holds.

Since Assumption 3.3 does not assume that the  $\xi_k$  are identically distributed or mutually independent, it can hold for the following variant of (1.6):

$$U_{k+1} := \Psi^\tau(U_k) + \xi_k(\tau, U_k), \quad \text{for all } k \in [K].$$

In this setting, we interpret Assumption 3.3 as the condition that the dependence of the moments of  $\xi_k$  on the state  $U_k$ , can be uniformly controlled by the constant  $C_{\xi,R}$ . We leave a more extensive investigation of state-dependent noise models for future work.

It follows from (2.4) and Assumption 3.3 that, for  $v, w \in \mathbb{N}$ ,

$$\mathbb{E} \left[ \left( \sum_{i=1}^{T/\tau} \|\xi_i(\tau)\|^w \right)^v \right] \leq \left( TC_{\xi,R}^w \tau^{w(p+1/2)-1} \right)^v. \quad (3.2)$$

This is because

$$\begin{aligned} \mathbb{E} \left[ \left( \sum_{i=1}^{T/\tau} \|\xi_i(\tau)\|^w \right)^v \right] &\leq \left( \frac{T}{\tau} \right)^{v-1} \sum_{i=1}^{T/\tau} \left( C_{\xi,R} \tau^{p+1/2} \right)^{wv} \\ &\leq \left( \frac{T}{\tau} \right)^v \left( C_{\xi,R} \tau^{p+1/2} \right)^{wv} \\ &= \left( TC_{\xi,R}^w \tau^{w(p+1/2)-1} \right)^v, \end{aligned}$$

where we used (2.4) and Assumption 3.3 for the first and second inequality respectively.

As noted in the introduction, the focus of this paper is on the convergence rate of the error  $e_k := u_k - U_k$  and

not on, say, the covariance operator of  $e_k$ , though that information is also important in applications. Note that if  $\xi_k(\tau)$  in Assumption 3.3 does not belong to  $L_{\mathbb{P}}^2$ , then  $\xi_k(\tau)$  does not admit a covariance operator. Accordingly, Assumption 3.3 and similar assumptions later in the paper are only upper bounds, and we do not actually work with the covariance operator of  $\xi_k$ . The precise construction of stochastic models for discretisation and truncation error is an interesting topic in its own right at the interface of numerical analysis and probability, upon which this paper only starts to touch; we anticipate that there will be further research concerning this question.

Given  $e_k = u_k - U_k$ , it follows from (1.5) and (1.6) that

$$e_{k+1} = (\Phi^\tau(u_k) - \Phi^\tau(U_k)) - (\Psi^\tau(U_k) - \Phi^\tau(U_k)) - \xi_k(\tau). \quad (3.3)$$

We shall use the decomposition (3.3) throughout this article.

The next result is stronger than Conrad et al. (2017, Theorem 2.2), as the discrete time supremum is inside the expectation, and as it does not require the vector field  $f$  to be globally Lipschitz nor  $\xi$  to be Gaussian:

**Theorem 3.4** *Suppose Assumptions 3.1 and 3.2 hold, and fix  $u_0 = U_0$ . Furthermore, if it holds that  $X \in L_{\mathbb{P}}^2 \implies \Psi^\tau(X) \in L_{\mathbb{P}}^2$ , and if the  $(\xi_k(\tau))_{k \in [K]}$  have zero mean, are mutually independent, and satisfy Assumption 3.3 for  $R = 2$  and  $p \geq 1$ , then there exists  $C > 0$  that does not depend on  $\tau$  such that*

$$\mathbb{E} \left[ \max_{k \in [K]} \|e_k\|^2 \right] \leq C \tau^{2p \wedge 2q}. \quad (3.4)$$

In contrast to Theorem 3.4, which required that the  $(\xi_k(\tau))_{k \in [K]}$  be independent and centred in order to construct a martingale, we make no independence or centredness assumptions on the  $(\xi_k(\tau))_{k \in [K]}$  for the rest of this article. The following result should be compared to Theorem 3.4 by considering the case  $R = n = 2$ . Then for the randomised method to have the same order as the deterministic method on which it is based, we need that  $p \geq q + \frac{1}{2}$ . In other words, if we remove the assumptions on the  $(\xi_k(\tau))_{k \in [K]}$  of independence and centredness, then we require that the second moments of the  $(\xi_k(\tau))_{k \in [K]}$  decay to zero with time-step  $\tau$  at a faster rate than in Theorem 3.4, since the lower bound  $q + \frac{1}{2}$  on  $p$  implied by Theorem 3.5 is larger than the lower bound  $q$  on  $p$  implied by Theorem 3.4.

**Theorem 3.5** *Let  $n \in \mathbb{N}$ . Suppose that Assumptions 3.1, 3.2, and 3.3 hold with  $\tau^* \leq 1$ ,  $q \geq 1$ ,  $p \geq 1$ , and*

$R$ , and that  $u_0 = U_0$ . Then, there exists a  $\bar{C} > 0$  that does not depend on  $\tau$  such that for  $0 < \tau < \tau^*$ ,

$$\mathbb{E} \left[ \max_{\ell \in [K]} \|e_\ell\|^n \right] \leq \bar{C} \tau^{n(q \wedge (p-1/2))}. \quad (3.5)$$

where

$$\bar{C} := 2T \max\{(4C_\Psi)^n, (2C_{\xi,R})^n\} \exp(TC_\Phi(n, \tau^*)) \quad (3.6)$$

and  $C_\Phi(n, \tau^*)$  is defined according to (A.5).

We shall show that if we strengthen Assumption 3.3 by allowing for arbitrarily large  $R \in \mathbb{N}$ , then the moment generating function of  $\max_{\ell \in [K]} \|e_\ell\|^n$  is finite on  $\mathbb{R}$ .

**Corollary 3.1** *Fix  $n \in \mathbb{N}$ . Suppose that Assumptions 3.1 and 3.2 hold, and that Assumption 3.3 holds with  $R = +\infty$  and  $p \geq 1/2$ . Then, for all  $0 < \tau < \tau^*$  and all  $\rho \in \mathbb{R}$ ,*

$$\mathbb{E} \left[ \exp \left( \rho \max_{\ell \in [K]} \|e_\ell\|^n \right) \right] < \infty. \quad (3.7)$$

Hence, by Markov's inequality, the distribution of  $\max_{\ell \in [K]} \|e_\ell\|^n$  concentrates exponentially about its mean.

We close this section by noting that, while we have made no attempt to find the optimal constants in Theorem 3.4 and Theorem 3.5, the convergence orders in these results cannot be improved at the present level of generality. This is because the convergence order of the randomised solution cannot exceed that of the underlying deterministic solver, unless the random variables  $\xi_k(\tau)$  used to model the error  $\Phi^\tau(u_k) - \Psi^\tau(u_k)$  at each time step  $t_k$  are chosen to achieve this effect. We leave the construction of such randomised solvers for future work.

#### 4 Integration for locally Lipschitz vector fields

This section considers the numerical integration of vector fields  $f$  that satisfy the following polynomial growth condition.

**Assumption 4.1** *The vector field  $f$  is continuously differentiable, and both  $f$  and the associated map  $\Phi^\tau$  defined by (1.2) admit  $0 < \tau^* \leq 1$ ,  $C_\Phi \geq 1$ , and  $s \geq 1$ , such that the following inequalities hold for all  $a, b \in \mathbb{R}^d$  and all  $0 < \tau < \tau^*$ :*

$$\|f(a) - f(b)\| \leq C_\Phi(1 + \|a\|^s + \|b\|^s)\|a - b\| \quad (4.1a)$$

$$\|\Phi^\tau(a) - \Phi^\tau(b)\| \leq (1 + \tau C_\Phi(1 + \|a\|^s + \|b\|^s))\|a - b\|. \quad (4.1b)$$

The inequality (4.1a) implies

$$\|f(a)\| \leq \|f(a) - f(0)\| + \|f(0)\| \leq C_\Phi(1 + \|a\|^s)\|a\| + \|f(0)\|. \quad (4.2)$$

By Taylor's theorem, the remainder term  $R^\tau(a)$  in the first-order Taylor expansion (A.11) of  $\Phi^\tau(a)$  is given by the derivatives of  $f$ , evaluated at some  $a \in \mathbb{R}^d$  for some  $0 \leq t \leq \tau$ . The condition (4.1b) means that for some  $\tau^* > 0$  that is sufficiently small, the norm of the difference between two remainder terms can be controlled. The growth condition (4.1a) is not new; see for example Higham et al. (2002, Assumption 4.1).

The following result is analogous to Theorem 3.5. It states that we can replace Assumption 3.1 with Assumption 4.1 and obtain the same result as Theorem 3.5, provided that the  $(\xi_k(\tau))_{k \in [K]}$  are  $\mathbb{P}$ -a.s. bounded.

**Theorem 4.2** *Suppose that Assumptions 4.1, 3.2, and 3.3 hold for  $p$  and  $R$  as in Theorem 3.5. Suppose that  $u_0 = U_0$ . If the  $(\xi_k(\tau))_{k \in [K]}$  are  $\mathbb{P}$ -a.s. uniformly bounded over all  $k$  by a positive scalar that is  $O(\tau)$ , then the conclusions of Theorem 3.5 hold.*

It is of theoretical interest to determine whether there exists a deterministic numerical method  $\Psi$  such that the randomised version given by (1.6) has the same order even when each  $\xi_k(\tau)$  is not  $\mathbb{P}$ -a.s. bounded. In the remainder of this section, we shall show that for the implicit Euler method  $\Psi^\tau: \mathbb{R}^d \rightarrow \mathbb{R}^d$  defined by

$$\Psi^\tau(a) := a + \tau f(\Psi^\tau(a)), \quad (4.3)$$

the randomised version given by (1.6) has order 1, under the following dissipativity assumption.

**Assumption 4.3** *The function  $f$  admits parameters  $\alpha \geq 0$  and  $\beta \in \mathbb{R}$  such that*

$$\langle f(v), v \rangle \leq \alpha + \beta \|v\|^2 \text{ for all } v \in \mathbb{R}^d. \quad (4.4)$$

Assumption 4.3 is more general than the usual dissipativity property found in Humphries and Stuart (1994, Equation (1.2)) because  $\beta$  may assume positive values. The sign of  $\beta$  in (4.4) plays an important role in the behaviour of the solution  $u$  of (1.1), as well as in numerical methods for solving for  $u$ . For example, if  $\beta$  is positive, then the problem (1.1) may be stiff. In this paper, we study only the rate of convergence, and leave the issue of stiffness for future work. In particular, allowing for positive  $\beta$  poses no problem for establishing moment bounds, as we show in Lemma 4.1.

Recent studies in numerical methods for stochastic differential equations consider constraints on the drift that feature the same right-hand side as (4.4), e.g. Fang

and Giles (2016) and Mao and Szpruch (2013). We reiterate, however, that the analysis of numerical methods for stochastic differential equations cannot be applied to probabilistic solvers of the form (1.6), because of the different behaviour in the additive noise (see e.g. Assumption 3.3).

**Assumption 4.4** *Let  $\tau^* \leq 1$  be as in Assumption 4.1 and  $\beta \in \mathbb{R}$  be as in Assumption 4.3. Then there exists some  $0 < \tau' \leq \min\{\tau^*, (2|\beta|)^{-1}\}$  such that there exists a solution  $\Psi^\tau(a)$  to the implicit equation (4.3) for every  $0 \leq \tau \leq \tau'$ , such that the solution  $\Psi^\tau(a)$  varies continuously as a function of  $\tau$  in the interval  $0 \leq \tau \leq \tau'$ , and such that  $\Psi^\tau|_{\tau=0}(a) = a$ .*

Note that Assumption 4.4 is weaker than assuming unique solvability of (4.3) for every  $a \in \mathbb{R}^d$  over a sufficiently small time interval.

Unless otherwise specified, we shall assume hereafter that  $0 < \tau < \tau'$ .

#### 4.1 Moment bounds for implicit Euler

**Lemma 4.1** *Suppose that Assumptions 4.1, 4.3, and 4.4 hold, and let  $n \in \mathbb{N}$  be arbitrary. Given a fixed, deterministic  $U_0$ , the following holds uniformly in  $\omega \in \Omega$ :*

$$\max_{i \in [T/\tau]} \|U_i\|^{2n} \leq (2C_2)^n \left[ 1 + \tau^{-n} \left( \sum_{i=1}^{T/\tau} \|\xi_i(\tau)\|^2 \right)^n \right], \quad (4.5)$$

for  $C_2$  given in (A.10) below.

Note that Lemma 4.1 is the only statement for which we directly use Assumption 4.3. The following results depend on Assumption 4.3 only insofar as they depend on the conclusions of Lemma 4.1.

**Proposition 4.1** *Suppose that Assumptions 4.1, 4.4, and 4.3 hold, and let  $n \in \mathbb{N}$  be arbitrary. If Assumption 3.3 holds for some  $R \geq 2n$  and some  $p \geq 1$ , then*

$$\mathbb{E} \left[ \max_{i \in [K]} \|U_i\|^{2n} \right] \leq (2C_2)^n \left( 1 + (TC_{\xi,R}^2 \tau^{2p-1})^n \right), \quad (4.6)$$

for  $C_2$  defined in (A.10), and  $C_{\xi,R}$  in Assumption 3.3.

*Proof* The statement follows directly from the conclusion (4.5) of Lemma 4.1 and (3.2) with  $w = 2$  and  $v = n$ .

**Corollary 4.1** *Suppose that Assumptions 4.1, 4.4, 4.3, and 3.3 hold with  $R = +\infty$  and  $p \geq 1/2$ . Then*

$$\mathbb{E} \left[ \exp \left( \rho \max_{i \in [K]} \|U_i\|^2 \right) \right] < \infty, \quad \text{for all } \rho \in \mathbb{R}.$$

*Proof* The result follows from Proposition 4.1, the series expansion of the exponential, and the dominated convergence theorem.

Lemma 4.1 shows that whenever Assumption 4.3 holds, then regardless of the growth behaviour of  $f$ , the randomised implicit Euler method has the property that if  $X \in L_{\mathbb{P}}^R$  for some  $R \in \mathbb{N}$ , then  $\Psi^\tau(X) \in L_{\mathbb{P}}^R$  as well; cf. the hypothesis on  $\Psi^\tau$  in Theorem 3.4.

#### 4.2 Convergence in discrete time for implicit Euler

**Proposition 4.2** *Let  $n \in \mathbb{N}$ , and suppose that Assumptions 4.1 and 3.3 hold for some  $R \geq 2n(2s+1)$  and some  $p \geq 1$ . Then there exists a scalar  $C_\Psi > 0$  that does not depend on  $\tau$  or  $k \in [K]$ , such that for all  $k \in [K]$ ,*

$$\mathbb{E} [\|\Psi^\tau(U_k) - \Phi^\tau(U_k)\|^{2n}] \leq C_\Psi \tau^{4n} \quad (4.7)$$

with  $C_\Psi$  as in (A.16).

Proposition 4.2 shows that when  $f$  satisfies the polynomial growth condition and  $\Psi$  is the implicit Euler method, then the local truncation error at step  $k$  of the randomised numerical integrator satisfies a bound analogous to that in Assumption 3.2, provided that the random variables  $(\xi_k(\tau))_{k \in [K]}$  are sufficiently regular.

**Theorem 4.5** *Let  $n \in \mathbb{N}$ , and let  $\Psi^\tau$  be given by (4.3). Suppose that Assumptions 4.1, 4.3, and 4.4 hold, with parameters  $s \geq 1$  and  $\tau' > 0$ . Suppose that Assumption 3.3 holds with  $R \geq 2n(2s+1)$  and  $p \geq \frac{3}{2}$ . Then there exists some  $C > 0$  that does not depend on  $\tau$  such that for  $0 < \tau < \tau'$ ,*

$$\mathbb{E} \left[ \max_{\ell \in [K]} \|e_\ell\|^{2n} \right] \leq C \tau^{2n}.$$

Note that the condition  $p \geq \frac{3}{2}$  is the same condition  $p \geq q + \frac{1}{2}$  on  $p$  in Theorem 3.5, since the implicit Euler method has order  $q = 1$ .

#### 4.3 Alternative decomposition of the error

The decomposition (3.3) of the error  $e_{k+1}$  was used to derive the convergence results above. One might consider instead using the decomposition

$$e_{k+1} = (\Phi^\tau(u_k) - \Psi^\tau(u_k)) + (\Psi^\tau(u_k) - \Psi^\tau(U_k)) - \xi_k(\tau)$$

with the goal of using some stability properties of the implicit Euler method. However, this approach leads to a convergence result that is weaker, either because it requires exponential integrability of  $\|U_k\|$ , or because



the convergence is uniform only on a proper subset  $\Omega_\tau$  of the event space  $\Omega$ . Recall that we do not assume any of the  $\xi_k(\tau)$  to be a.s. bounded.

By (2.2) and the fact that implicit Euler has order one (i.e. Assumption 3.2)

$$\begin{aligned} \|e_{k+1}\|^2 &\leq (\|\Phi^\tau(u_k) - \Psi^\tau(u_k)\| \\ &\quad + \|\Psi^\tau(u_k) - \Psi^\tau(U_k) - \xi_k(\tau)\|)^2 \\ &\leq (1 + \tau^{-1})\|\Phi^\tau(u_k) - \Psi^\tau(u_k)\|^2 \\ &\quad + (1 + \tau)\|\Psi^\tau(u_k) - \Psi^\tau(U_k) - \xi_k(\tau)\|^2 \\ &\leq (1 + \tau^{-1})(C\tau^2)^2 \\ &\quad + (1 + \tau)\|\Psi^\tau(u_k) - \Psi^\tau(U_k) - \xi_k(\tau)\|^2, \end{aligned} \quad (4.8)$$

where one can show, using the proof of Proposition 4.2, that  $C > 0$  in (4.8) depends on  $\|u_k\|^s$  but not on  $\tau$ . By (2.2) we obtain

$$\begin{aligned} &\|\Psi^\tau(u_k) - \Psi^\tau(U_k) - \xi_k(\tau)\|^2 \\ &\leq (1 + \tau)\|\Psi^\tau(u_k) - \Psi^\tau(U_k)\|^2 + (1 + \tau^{-1})\|\xi_k(\tau)\|^2. \end{aligned}$$

Substituting the result above into (4.8), and assuming that  $\tau < 1$ , we obtain

$$\begin{aligned} \|e_{k+1}\|^2 &\leq C\tau^3 + (1 + \tau)^2\|\Psi^\tau(u_k) - \Psi^\tau(U_k)\|^2 \\ &\quad + 4\tau^{-1}\|\xi_k(\tau)\|^2. \end{aligned} \quad (4.9)$$

The definition (4.3) of the implicit Euler method and (2.2) yield

$$\begin{aligned} &\|\Psi^\tau(u_k) - \Psi^\tau(U_k)\|^2 \\ &= \|u_k - U_k + \tau f(\Psi^\tau(u_k)) - \tau f(\Psi^\tau(U_k))\|^2 \\ &\leq (1 + \tau)\|u_k - U_k\|^2 \\ &\quad + (1 + \tau^{-1})\tau^2\|f(\Psi^\tau(u_k)) - f(\Psi^\tau(U_k))\|^2 \\ &\leq (1 + \tau)\|e_k\|^2 + (1 + \tau)\tau D^2\|\Psi^\tau(u_k) - \Psi^\tau(U_k)\|^2 \times \\ &\quad [1 + \|\Psi^\tau(u_k)\|^s + \|\Psi^\tau(U_k)\|^s]^2, \end{aligned}$$

by Assumption 4.1. Rearranging the above yields

$$(1 + \tau)\|e_k\|^2 \geq \|\Psi^\tau(u_k) - \Psi^\tau(U_k)\|^2(1 - (1 + \tau)\tau D^2 \hat{M})$$

where  $\hat{M} := [1 + \|\Psi^\tau(u_k)\|^s + \|\Psi^\tau(U_k)\|^s]^2$  is a random variable. Analogously, define the random variable  $M$  by

$$M := \left[ 1 + \max_{k \in [K]} \|\Psi^\tau(u_k)\|^s + \max_{k \in [K]} \|\Psi^\tau(U_k)\|^s \right]^2.$$

Suppose that  $u_0 = U_0$  are fixed, and define

$$\Omega_\tau := \{\omega \in \Omega \mid 1 - (1 + \tau)\tau D^2 M(\omega) > 0\}. \quad (4.10)$$

Since it is not the case that all of the random variables  $(\xi_k(\tau))_{k \in [K]}$  are a.s.-bounded, it follows that  $\Omega_\tau$  is a proper subset of  $\Omega$ , for every  $\tau > 0$ . In what follows, we assume that  $\Omega_\tau$  is nonempty, and that  $\omega \in \Omega_\tau$ ;

we suppress the  $\omega$ -dependence of all random variables. Define  $\tilde{C} > 0$  by

$$\begin{aligned} (1 - (1 + \tau)\tau D^2 M)^{-1} &= \sum_{n=0}^{\infty} [(1 + \tau)\tau D^2 M]^n \\ &=: 1 + \tilde{C}\tau, \end{aligned} \quad (4.11)$$

Using (4.11), we have

$$\|\Psi^\tau(u_k) - \Psi^\tau(U_k)\|^2 \leq (1 + \tau)(1 + \tilde{C}\tau)\|e_k\|^2,$$

and substituting the above into (4.9) yields

$$\|e_{k+1}\|^2 \leq C\tau^3 + (1 + \tau)^3(1 + \tilde{C}\tau)\|e_k\|^2 + 4\tau^{-1}\|\xi_k(\tau)\|^2.$$

Proceeding as in the proof of Theorem 4.5, we use a telescoping sum, Grönwall's theorem, and Assumption 3.3 with  $p \geq q + 1/2$  to obtain

$$\mathbb{E} \left[ 1_{\Omega_\tau} \max_{k \in [K]} \|e_k\|^2 \right] \leq \mathbb{E} [1_{\Omega_\tau} \exp(T\kappa)] C\tau^2, \quad (4.12)$$

where  $\kappa$  depends on  $\tau$  according to

$$\begin{aligned} \kappa(\tau) &:= \tau^{-1} \left[ (1 + \tau)^3(1 + \tilde{C}\tau) - 1 \right] \\ &= \tilde{C}\tau^3 + (3\tilde{C} + 1)\tau^2 + 3(\tilde{C} + 1)\tau + (3 + \tilde{C}). \end{aligned}$$

For any  $\tau > 0$ , it follows from the definition of  $\kappa$ , and considering the zeroth order term  $3 + \tilde{C}$  above that

$$\kappa(\tau) > 3 + (1 + \tau)D^2 M.$$

From (4.10), it follows that, for all  $\omega \in \Omega_\tau$ , we have

$$(1 + \tau)D^2 M < \tau^{-1}.$$

where the right-hand side increases to infinity as  $\tau$  decreases to zero. Thus, it need not be true that the quantity  $\mathbb{E}[1_{\Omega_\tau} \exp(T\kappa)]$  is finite. One way to ensure that  $\mathbb{E}[1_{\Omega_\tau} \exp(T\kappa)]$  is finite for  $0 < \tau < \tau'$  would be to require that  $\mathbb{E}[\exp(T\kappa)]$  is finite on the same range. By the inequality for  $\kappa$  above, a necessary condition for  $\mathbb{E}[\exp(T\kappa)]$  to be finite is exponential integrability of  $\max_{k \in [K]} \|\Psi^\tau(U_k)\|^{2s}$ . In many cases, a necessary condition for this would be exponential integrability of  $\max_{k \in [K]} \|U_k\|^{2s}$ . By Corollary 4.1, in order to guarantee exponential integrability of  $\max_{k \in [K]} \|U_k\|^{2s}$ , we would need to impose much stronger regularity conditions on the  $(\xi_k(\tau))_{k \in [K]}$  than those in Theorem 4.5. Finally, we also remark that if the  $(\xi_k)_{k \in [K]}$  are not  $\mathbb{P}$ -a.s. uniformly bounded, then for any  $\tau > 0$ , (4.12) is a weaker convergence result than Theorem 4.5, since in this case for any  $\tau > 0$   $\Omega_\tau$  will be a proper subset of  $\Omega$ .

## 5 Additional results

### 5.1 Convergence for continuous-time interpolant

Recall (1.6) defines the discrete-time process  $(U_k)_{k \in [K]}$ ; in many applications, it is often useful to have a numerical method that provides continuous output, e.g. an inverse problem or data assimilation that requires comparison between the numerical solution and an observation that is not on the time grid  $(t_k)_{k \in [K]}$  defined in (1.3). Given this time grid  $(t_k)_{k \in [K]}$ , we may define a continuous-time process  $U$  by

$$U(t) := \Psi^{t-t_k}(U_k) + \xi_k(t-t_k) \quad \text{for } t \in [t_k, t_{k+1}).$$

For the above definition to work, we assume that each  $\xi_k$  is a stochastic process defined on the time interval  $[0, \tau]$ . In addition, to ensure that the process  $U$  has  $\mathbb{P}$ -almost surely continuous paths, we require that  $\mathbb{P}(\xi_k(0) = 0) = 1$ . The corresponding notion of the error at time  $0 \leq t \leq T$  is given by  $e(t) := u(t) - U(t)$ , where  $u(t) = \Phi^t(u_0)$ . We emphasise that the continuous-time process  $(U(t))_{0 \leq t \leq T}$  described above will in general differ from the continuous-time process obtained by linear interpolation of  $(U_k)_{k \in [K]}$ .

We now demonstrate how one can obtain a convergence result for the continuous-time process from a discrete-time convergence result by strengthening the assumption on the noise, using Theorem 3.5 as an example. Consider the following version of Assumption 3.3:

**Assumption 5.1** Fix  $\tau > 0$ . The collection  $(\xi_k)_{k \in \mathbb{N}}$  of stochastic processes  $\xi_k: \Omega \times [0, \tau] \rightarrow \mathbb{R}^d$  satisfies  $\mathbb{P}(\xi_k(0) = 0) = 1$  and admits  $p \geq 1$ ,  $R \in \mathbb{N} \cup \{+\infty\}$  and some  $C_{\xi, R} \geq 1$  that do not depend on  $k \in \mathbb{N}$  or  $\tau$ , such that for all  $1 \leq r \leq R$  and for all  $k \in \mathbb{N}$ ,

$$\mathbb{E} \left[ \sup_{0 < t \leq \tau} \|\xi_k(t)\|^r \right] \leq \left( C_{\xi, R} \tau^{p+1/2} \right)^r.$$

Recall that we do not assume that the  $\xi_k$  are independent, identically distributed, or centred.

**Theorem 5.2** Let  $n \in \mathbb{N}$ , and suppose that Assumptions 3.1, 3.2, and 5.1 hold with parameters  $\tau^*$ ,  $C_{\Phi}$ ,  $C_{\Psi}$ ,  $q$ ,  $C_{\xi, R}$ ,  $p$ , and  $R$ . Then for all  $0 < \tau < \tau^*$ ,

$$\mathbb{E} \left[ \sup_{0 \leq t \leq T} \|e(t)\|^n \right] \leq 3^{n-1} \left( (1 + C_{\Phi} \tau^*)^n \bar{C} + C_{\Psi}^n (\tau^*)^n + T C_{\xi, R}^n \right) \times \tau^{n(q \wedge (p-1/2))}, \quad (5.1)$$

where  $\bar{C}$  is defined in (3.6).

The next result follows from Theorem 5.2 in the same way that Corollary 3.1 follows from Theorem 3.5.

**Corollary 5.1** Fix  $n \in \mathbb{N}$ . Suppose that Assumptions 3.1 and 3.2 hold, and that Assumption 5.1 holds with  $R = +\infty$  and  $p \geq 1/2$ . Then, for all  $0 < \tau < \tau^*$ ,

$$\mathbb{E} \left[ \exp \left( \rho \sup_{0 \leq t \leq T} \|e(t)\|^n \right) \right] < \infty, \quad \text{for all } \rho \in \mathbb{R}. \quad (5.2)$$

*Proof* The proof follows by the series representation of the exponential and the dominated convergence theorem; see the proof of Corollary 3.1.

### 5.2 Existence of processes that satisfy the $(p, R)$ -regularity condition

The lemma below shows that there exist random variables that are not  $\mathbb{P}$ -a.s. bounded, and that satisfy Assumption 3.3 and, more generally, Assumption 5.1 for  $R = +\infty$ .

**Lemma 5.1** Let  $\tau > 0$  and  $p \geq 1$  be arbitrary, and let  $(B_t)_{0 \leq t \leq \tau}$  be  $\mathbb{R}^d$ -valued Brownian motion. Then

$$\xi_0(t) := \tau^{p-1} \int_0^t B_s \, ds$$

satisfies

$$\mathbb{E} \left[ \sup_{t \leq \tau} \|\xi_0(t)\|^r \right] \leq 4\tau^{r(p+1/2)} \quad \text{for all } r \in \mathbb{N}. \quad (5.3)$$

Note that variants of the integrated Brownian motion process have been used for modelling local truncation error in other works (Schober et al. 2014; Conrad et al. 2017). However, the point of Lemma 5.1 is *not* to suggest that the local truncation error behaves as an integrated Brownian motion, nor even that the integrated Brownian motion process is a suitable model for the local truncation error. The point of Lemma 5.1 is simply to show that there exist processes that satisfy Assumption 5.1 with  $R = +\infty$ . The construction of models that better reflect known properties of the truncation error, for specific classes of vector fields  $f$ , is an interesting task that we leave for future work.

## References

- F.-X. Briol, C. Oates, M. Girolami, and M. A. Osborne. Frank–Wolfe Bayesian quadrature: Probabilistic integration with theoretical guarantees. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 1162–1170. Curran Associates, Inc., 2015.
- M. A. Capistrán, J. A. Christen, and S. Donnet. Bayesian analysis of ODEs: solver optimal accuracy and Bayes factors. *SIAM/ASA J. Uncertain. Quantif.*, 4(1):829–849, 2016. doi: 10.1137/140976777.

- O. A. Chkrebtii, D. A. Campbell, B. Calderhead, and M. A. Girolami. Bayesian solution uncertainty quantification for differential equations. *Bayesian Anal.*, 11(4):1239–1267, 2016. doi: 10.1214/16-BA1017.
- J. A. Christen. Posterior distribution existence and error control in Banach spaces, 2017. arXiv:1712.03299.
- J. Cockayne, C. Oates, T. J. Sullivan, and M. Girolami. Probabilistic numerical methods for PDE-constrained Bayesian inverse problems. In G. Verdoolaege, editor, *Proceedings of the 36<sup>th</sup> International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume 1853 of *AIP Conference Proceedings*, pages 060001–1–060001–8, 2017. doi: 10.1063/1.4985359.
- J. Cockayne, C. Oates, T. J. Sullivan, and M. Girolami. Bayesian probabilistic numerical methods, *SIAM Rev.*, to appear. arXiv:1702.03673.
- P. R. Conrad, M. Girolami, S. Särkkä, A. M. Stuart, and K. C. Zygalakis. Statistical analysis of differential equations: introducing probability measures on numerical solutions. *Stat. Comput.*, 27(4):1065–1082, 2017. ISSN 0960-3174. doi: 10.1007/s11222-016-9671-0.
- P. Diaconis. Bayesian numerical analysis. In *Statistical Decision Theory and Related Topics, IV, Vol. 1 (West Lafayette, Ind., 1986)*, pages 163–175. Springer, New York, 1988.
- W. Fang and M. B. Giles. Adaptive Euler–Maruyama method for SDEs with non-globally Lipschitz drift: Part I, finite time interval, 2016. arXiv:1609.08101.
- M. B. Giles. Multilevel Monte Carlo methods. *Acta Numer.*, 24:259–328, 2015. doi: 10.1017/S096249291500001X.
- J. Gonzalez, M. Osborne, and N. Lawrence. GLASSES: Relieving the myopia of Bayesian optimisation. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, pages 790–799. 2016. <http://jmlr.org/proceedings/papers/v51/gonzalez16b.html>.
- E. Hairer, S. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems*, volume 8 of *Springer Series in Computational Mathematics*. Springer-Verlag, New York, 2009. doi: 10.1007/978-3-540-78862-1. First softcover printing.
- P. Hennig. Probabilistic interpretation of linear solvers. *SIAM J. Optim.*, 25(1):234–260, 2015. doi: 10.1137/140955501.
- P. Hennig, M. A. Osborne, and M. Girolami. Probabilistic numerics and uncertainty in computations. *P. Roy. Soc. Lond. A Mat.*, 471(2179):20150142, 2015. doi: 10.1098/rspa.2015.0142.
- D. J. Higham, X. Mao, and A. M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM J. Numer. Anal.*, 40(3):1041–1063, 2002. doi: 10.1137/S0036142901389530.
- J. M. Holte. Discrete Gronwall lemma and applications, 2009. <http://homepages.gac.edu/~holte/publications/GronwallLemma.pdf>.
- A. R. Humphries and A. M. Stuart. Runge–Kutta methods for dissipative and gradient dynamical systems. *SIAM J. Numer. Anal.*, 31(5):1452–1485, 1994. doi: 10.1137/0731075.
- A. Jentzen and A. Neuenkirch. A random Euler scheme for Carathéodory differential equations. *J. Comput. Appl. Math.*, 224(1):346–359, 2009. doi: 10.1016/j.cam.2008.05.060.
- J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*, volume 160 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2005. doi: 10.1007/b138659.
- B. T. Knapik, A. W. van der Vaart, and J. H. van Zanten. Bayesian inverse problems with Gaussian priors. *Ann. Statist.*, 39(5):2626–2657, 2011. ISSN 0090-5364. doi: 10.1214/11-AOS920.
- 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. ISSN 1609-4840. doi: 10.1515/cmam-2016-0048.
- K. Law, A. Stuart, and K. Zygalakis. *Data Assimilation: A Mathematical Introduction*, volume 62 of *Texts in Applied Mathematics*. Springer, 2015. doi: 10.1007/978-3-319-20325-6.
- H. C. Lie, T. J. Sullivan, and A. L. Teckentrup. Random forward models and log-likelihoods in Bayesian inverse problems. *SIAM/ASA J. Uncertain. Quantif.*, 6(4):1600–1629, 2018. doi: 10.1137/18M1166523.
- X. Mao and L. Szpruch. Strong convergence and stability of implicit numerical methods for stochastic differential equations with non-globally Lipschitz continuous coefficients. *J. Comp. Appl. Math.*, 238:14 – 28, 2013. doi: 10.1016/j.cam.2012.08.015.
- E. Novak. *Deterministic and Stochastic Error Bounds in Numerical Analysis*, volume 1349 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1988. doi: 10.1007/BFb0079792.
- A. O’Hagan. Some Bayesian numerical analysis. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics, 4: Proceedings of the Fourth Valencia International Meeting: Dedicated to the memory of Morris H. DeGroot, 1931–1989: April 15–20, 1991*, pages 345–363. Clarendon Press, 1992.
- H. Owhadi. Bayesian numerical homogenization. *Multiscale Model. Simul.*, 13(3):812–828, 2015. doi: 10.1137/140974596.
- H. Owhadi. Multigrid with rough coefficients and multiresolution operator decomposition from hierarchical information games. *SIAM Rev.*, 59(1):99–149, 2017. doi: 10.1137/15M1013894.
- G. Peškir. On the exponential Orlicz norms of stopped Brownian motion. *Studia Math.*, 117(3):253–273, 1996. doi: 10.4064/sm-117-3-253-273.
- S. Reich and C. Cotter. *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge University Press, New York, 2015. doi: 10.1017/CBO9781107706804.
- K. Ritter. *Average-Case Analysis of Numerical Problems*, volume 1733 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2000. doi: 10.1007/BFb0103934.
- M. Schober, D. K. Duvenaud, and P. Hennig. Probabilistic ODE solvers with Runge–Kutta means. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 739–747. Curran Associates, Inc., 2014.
- J. Skilling. Bayesian solution of ordinary differential equations. In C. R. Smith, G. J. Erickson, and P. O. Neudorfer, editors, *Maximum Entropy and Bayesian Methods*, volume 50 of *Fundamental Theories of Physics*, pages 23–37. Springer, 1992. doi: 10.1007/978-94-017-2219-3.
- R. C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*, volume 12 of *Computational Science & Engineering*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2014.
- G. Stengle. Numerical methods for systems with measurable coefficients. *Appl. Math. Lett.*, 3(4):25–29, 1990. doi: 10.1016/0893-9659(90)90040-I.

- A. M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19:451–559, 2010. doi: 10.1017/S0962492910000061.
- T. J. Sullivan. *Introduction to Uncertainty Quantification*, volume 63 of *Texts in Applied Mathematics*. Springer, 2015. doi: 10.1007/978-3-319-23395-6.
- O. Teymur, H. C. Lie, T. J. Sullivan, and B. Calderhead. Implicit probabilistic integrators for ODEs. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31 (NIPS 2018)*. 2018. URL <http://papers.nips.cc/paper/7955-implicit-probabilistic-integrators-for-odes>.
- J. F. Traub and H. Woźniakowski. *A General Theory of Optimal Algorithms*. ACM Monograph Series. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1980.
- J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski. *Information, Uncertainty, Complexity*. Addison-Wesley Publishing Company, Advanced Book Program, Reading, MA, 1983.
- J. Wang, J. Cockayne, and C. Oates. On the Bayesian solution of differential equations, 2018. arXiv:1805.07109.

## A Proofs

*Proof (Proof of Lemma 2.1)* The assertion (2.3) holds immediately for  $n = 1$ , so let  $n \in \mathbb{N} \setminus \{1\}$ , and recall the binomial formula: for  $x, y \in \mathbb{R}$  and  $n \in \mathbb{N} \setminus \{1\}$ ,

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} = x^n + y^n + \sum_{k=1}^{n-1} \binom{n}{k} x^k y^{n-k}.$$

Fix  $\delta > 0$ . By (2.1), for any  $1 \leq k \leq n - 1$ ,

$$\begin{aligned} x^k y^{n-k} &\leq \delta \frac{k}{n} x^n + \frac{1}{\delta^{k/(n-k)}} \frac{n-k}{n} y^n \\ &\leq \delta \frac{k}{n} x^n + \frac{1}{\delta^{n-1}} \frac{n-k}{n} y^n, \end{aligned}$$

where the second inequality follows from  $-\frac{k}{n-k} \geq -(n-1)$ . Therefore,

$$\begin{aligned} (x + y)^n &\leq x^n \left( 1 + \delta \sum_{k=1}^{n-1} \binom{n}{k} \frac{k}{n} \right) \\ &\quad + y^n \left( 1 + \frac{1}{\delta^{n-1}} \sum_{k=1}^{n-1} \binom{n}{k} \frac{n-k}{n} \right), \end{aligned} \quad (\text{A.1})$$

and the proof is complete upon observing that

$$\begin{aligned} \sum_{k=1}^{n-1} \binom{n}{k} \frac{k}{n} &= \sum_{k=1}^{n-1} \binom{n-1}{k-1} = \sum_{j=0}^{n-1} \binom{n-1}{j} - \binom{n-1}{n-1} \\ &= (1+1)^{n-1} - 1 \leq 2^{n-1} \end{aligned}$$

and bounding the other binomial sum in a similar way.  $\square$

*Proof (Proof of Theorem 3.4)* By (3.3),

$$\begin{aligned} \|e_{k+1}\|^2 &= \|(\Phi^\tau(u_k) - \Phi^\tau(U_k)) - (\Psi^\tau(U_k) - \Phi^\tau(U_k))\|^2 \\ &\quad + \|\xi_k(\tau)\|^2 + 2\langle \Phi^\tau(u_k) - \Psi^\tau(U_k), \xi_k(\tau) \rangle. \end{aligned}$$

By (2.2) with  $\delta = \tau$ , by Assumption 3.1 and Assumption 3.2, and using that  $\tau < \tau^* \leq 1$ ,

$$\begin{aligned} &\|\Phi^\tau(u_k) - \Psi^\tau(U_k)\|^2 \\ &= \|(\Phi^\tau(u_k) - \Phi^\tau(U_k)) - (\Psi^\tau(U_k) - \Phi^\tau(U_k))\|^2 \\ &\leq (1 + \tau) \|\Phi^\tau(u_k) - \Phi^\tau(U_k)\|^2 \\ &\quad + (1 + \tau^{-1}) \|\Psi^\tau(U_k) - \Phi^\tau(U_k)\|^2 \\ &\leq (1 + \tau)(1 + C_\Phi \tau)^2 \|e_k\|^2 + 2C_\Psi^2 \tau^{1+2q}. \end{aligned} \quad (\text{A.2})$$

Observe that  $[(1 + \tau)(1 + C_\Phi \tau)^2 - 1]\tau^{-1}$  equals a quadratic polynomial in  $\tau$  with coefficients  $a_0$ ,  $a_1$ , and  $a_2$ . Calculating these coefficients and defining

$$C_1 = C_1(C_\Phi, \tau^*) := 1 + 2C_\Phi + C_\Phi(2 + C_\Phi)\tau^* + C_\Phi^2(\tau^*)^2 \quad (\text{A.3})$$

then yields that  $[(1 + \tau)(1 + C_\Phi \tau)^2 - 1]\tau^{-1} \leq C_1$  for all  $0 < \tau < \tau^*$ .

Combining the preceding estimates yields

$$\begin{aligned} \|e_{k+1}\|^2 - \|e_k\|^2 &\leq C_1 \tau \|e_k\|^2 + 2C_\Psi^2 \tau^{1+2q} + \|\xi_k(\tau)\|^2 \\ &\quad + 2\langle \Phi^\tau(u_k) - \Psi^\tau(U_k), \xi_k(\tau) \rangle. \end{aligned} \quad (\text{A.4})$$

Using (A.4) in the telescoping sum

$$\|e_k\|^2 - \|e_0\|^2 = \sum_{j=0}^{k-1} (\|e_{j+1}\|^2 - \|e_j\|^2),$$

the fact that  $e_0 = u_0 - U_0 = 0$  and  $K = T/\tau$ , we obtain

$$\begin{aligned} \|e_k\|^2 &\leq \sum_{j=0}^{k-1} \left[ C_1 \tau \|e_j\|^2 + C_\Psi \tau^{1+2q} + \|\xi_j(\tau)\|^2 \right. \\ &\quad \left. + 2\langle \Phi^\tau(u_k) - \Psi^\tau(U_k), \xi_k(\tau) \rangle \right] \\ &\leq C_1 \tau \sum_{j=0}^{k-1} \|e_j\|^2 + \sum_{j=0}^{K-1} (2C_\Psi^2 \tau^{1+2q} + \|\xi_j(\tau)\|^2) \\ &\quad + 2 \left\| \sum_{j=0}^{k-1} \langle \Phi^\tau(u_k) - \Psi^\tau(U_k), \xi_k(\tau) \rangle \right\| \\ &\leq C_1 \tau \sum_{j=0}^{k-1} \|e_j\|^2 + 2TC_\Psi^2 \tau^{2q} + \sum_{j=0}^{K-1} \|\xi_j(\tau)\|^2 \\ &\quad + 2 \left\| \sum_{j=0}^{k-1} \langle \Phi^\tau(u_k) - \Psi^\tau(U_k), \xi_k(\tau) \rangle \right\|. \end{aligned}$$

It follows from the last inequality that

$$\begin{aligned} \max_{\ell \leq k} \|e_\ell\|^2 &\leq C_1 \tau \sum_{j=0}^{k-1} \|e_j\|^2 + 2TC_\Psi^2 \tau^{2q} + \sum_{j=0}^{K-1} \|\xi_j(\tau)\|^2 \\ &\quad + 2 \max_{\ell \leq k} \left\| \sum_{j=0}^{\ell-1} \langle \Phi^\tau(u_k) - \Psi^\tau(U_k), \xi_k(\tau) \rangle \right\|. \end{aligned}$$

Now replace  $\|e_j\|^2$  on the right-hand side with  $\max_{\ell \leq j} \|e_\ell\|^2$  and take expectations of both sides of the inequality. Since Assumption 3.3 holds with  $R = 2$ ,

$$\mathbb{E} \left[ \sum_{j=0}^{K-1} \|\xi_j(\tau)\|^2 \right] \leq \frac{T}{\tau} (C_{\xi, R} \tau^{p+1/2})^2 = TC_{\xi, R}^2 \tau^{2p}.$$

Next, define for every  $k \in [K]$  the  $\sigma$ -algebra  $\mathcal{F}_j$  generated by  $\xi_0(\tau), \dots, \xi_j(\tau)$ . Then the sequence  $(\mathcal{F}_j)_{j \in [K]}$  forms a filtration. Define  $(M_k)_{k \in [K]}$  by

$$M_k := \sum_{j=0}^k \langle \Phi^\tau(u_j) - \Psi^\tau(U_j), \xi_j(\tau) \rangle.$$

We want to show that this process is a martingale with respect to  $(\mathcal{F}_j)_{j \in [K]}$ . By (1.6),  $U_j$  is measurable with respect to  $\mathcal{F}_{j-1}$ , so  $M_k$  is measurable with respect to  $\mathcal{F}_k$ . Hence  $(M_k)_{k \in [K]}$  is adapted with respect to  $(\mathcal{F}_k)_{k \in [K]}$ . Observe that

$$\begin{aligned} \mathbb{E}[\|M_k\|] &\leq \sum_{j=0}^k \mathbb{E}[\|\langle \Phi^\tau(u_j) - \Psi^\tau(U_j), \xi_j(\tau) \rangle\|] \\ &\leq \sum_{j=0}^k (\mathbb{E}[\|\Phi^\tau(u_j) - \Psi^\tau(U_j)\|^2] + \mathbb{E}[\|\xi_j(\tau)\|^2]) \\ &\leq 2 \sum_{j=0}^k (\|\Phi^\tau(u_j)\|^2 + \mathbb{E}[\|\Psi^\tau(U_j)\|^2] + \|\xi_j(\tau)\|^2). \end{aligned}$$

Using the assumption that  $X \in L_{\mathbb{P}}^2 \implies \Psi^\tau(X) \in L_{\mathbb{P}}^2$ , (1.6), Assumption 3.3, and the fact that  $U_0 = u_0$  is fixed, it follows that  $U_j$  and  $\Psi^\tau(U_j)$  belong to  $L_{\mathbb{P}}^2$ ; thus  $M_k$  belongs to  $L_{\mathbb{P}}^1$  for every  $k \in [K]$ . We now use the assumption that  $\mathbb{E}[\xi_j(\tau)] = 0$  for every  $j \in [K]$ , and that the  $(\xi_k(\tau))_{k \in [K]}$  are mutually independent, in order to establish the martingale property:

$$\begin{aligned} \mathbb{E}[M_k - M_{k-1} | \mathcal{F}_{k-1}] \\ = \mathbb{E}[\langle \Phi^\tau(u_k) - \Psi^\tau(U_k), \xi_k(\tau) \rangle | \mathcal{F}_{k-1}], \end{aligned}$$

and the right-hand side vanishes since  $U_k$  is measurable with respect to  $\mathcal{F}_{k-1}$  as noted earlier. Since  $(M_k)_{k \in [K]}$  is a martingale, we may apply the Burkholder–Davis–Gundy inequality (Peškir 1996, Equation (2.2)). Letting  $[Y]_\ell$  denote the quadratic variation up to time  $\ell$  of a process  $Y_k$ , we have

$$\begin{aligned} \mathbb{E} \left[ \max_{k \leq \ell} \left\| \sum_{j=0}^{k-1} \langle \Phi^\tau(u_j) - \Psi^\tau(U_j), \xi_j(\tau) \rangle \right\|^2 \right] \\ \leq 3\mathbb{E} \left[ \left[ \langle \Phi^\tau(u_\bullet) - \Psi^\tau(U_\bullet), \xi_\bullet(\tau) \rangle \right]_{\ell-1}^{1/2} \right] \leq 3\mathbb{E}[ab] \end{aligned}$$

where we define  $b := \sqrt{\sum_{j=1}^{\ell-1} \|\xi_j(\tau) - \xi_{j-1}(\tau)\|^2}$  and  $a := \sqrt{\max_{j \leq \ell} \|\Phi^\tau(u_j) - \Psi^\tau(U_j)\|^2}$ . Using (2.1) with the same  $a$  and  $b$ ,  $r = r^* = 2$ , and  $\delta = [6(1+\tau)(1+C_\Phi\tau^2)]^{-1}$ , and using (A.2), it follows that

$$\begin{aligned} 3\mathbb{E} \left[ \left( \max_{j \leq \ell} \|\Phi^\tau(u_j) - \Psi^\tau(U_j)\| \right) \sqrt{\sum_{j=1}^{\ell-1} \|\xi_j(\tau) - \xi_{j-1}(\tau)\|^2} \right] \\ \leq \frac{1}{4} \left( \mathbb{E} \left[ \max_{j \leq \ell} \|e_j\|^2 \right] + 2C_\Psi^2 \tau^{1+2q} \right) \\ + 9(1+\tau^*)(1+C_\Phi\tau^*)^2 \sum_{j=1}^{\ell-1} \mathbb{E}[\|\xi_j(\tau) - \xi_{j-1}(\tau)\|^2] \\ \leq \frac{1}{4} \left( \mathbb{E} \left[ \max_{j \leq \ell} \|e_j\|^2 \right] + 2C_\Psi^2 \tau^{1+2q} \right) \\ + 18(1+\tau^*)(1+C_\Phi\tau^*)^2 \sum_{j=1}^{\ell-1} \mathbb{E}[\|\xi_j(\tau)\|^2] \end{aligned}$$

where we applied (2.2) with  $\delta = 1$ ,  $r = r^* = 2$ ,  $a = \xi_j(\tau)$  and  $b = \xi_{j-1}(\tau)$  to obtain the last inequality. Thus by Assumption 3.3 and by using  $\ell - 1 \leq K = T/\tau$ ,

$$\begin{aligned} 2\mathbb{E} \left[ \max_{k \leq \ell} \left\| \sum_{j=0}^{\ell-1} \langle \Phi^\tau(u_j) - \Psi^\tau(U_j), \xi_j(\tau) \rangle \right\|^2 \right] \\ \leq \frac{1}{2} \left( \mathbb{E} \left[ \max_{k \leq \ell} \|e_k\|^2 \right] + 2C_\Psi^2 \tau^{1+2q} \right) \\ + 36(1+\tau^*)(1+C_\Phi\tau^*)^2 \sum_{j=0}^{\ell-1} \mathbb{E}[\|\xi_j(\tau)\|^2] \\ \leq \frac{1}{2} \left( \mathbb{E} \left[ \max_{k \leq \ell} \|e_k\|^2 \right] + 2C_\Psi^2 \tau^{1+2q} \right) \\ + 36(1+\tau^*)(1+C_\Phi\tau^*)^2 T \tau^{-1} \left( C_{\xi, R} \tau^{p+1/2} \right)^2 \end{aligned}$$

Combining the preceding estimates, we obtain

$$\begin{aligned} \mathbb{E} \left[ \max_{\ell \leq k} \|e_\ell\|^2 \right] \\ \leq \tau C_1 \sum_{j=0}^{k-1} \mathbb{E} \left[ \max_{\ell \leq j} \|e_\ell\|^2 \right] + 2TC_\Psi^2 \tau^{2q} + TC_{\xi, R}^2 \tau^{2p} \\ + \frac{1}{2} \left( \mathbb{E} \left[ \max_{k \leq \ell} \|e_k\|^2 \right] + 2C_\Psi^2 \tau^{1+2q} \right) \\ + 36(1+\tau^*)(1+C_\Phi\tau^*)^2 T \tau^{-1} \left( C_{\xi, R} \tau^{p+1/2} \right)^2, \end{aligned}$$

and by rearranging terms and using that  $\tau < \tau^* \leq 1$ , we obtain

$$\begin{aligned} \mathbb{E} \left[ \max_{\ell \leq k} \|e_\ell\|^2 \right] \\ \leq 2\tau C_1 \sum_{j=0}^{k-1} \mathbb{E} \left[ \max_{\ell \leq j} \|e_\ell\|^2 \right] + 4(1+T)C_\Psi^2 \tau^{2q} \\ + 2TC_{\xi, R}^2 (1+36(1+\tau^*)(1+C_\Phi\tau^*)^2) \tau^{2p}. \end{aligned}$$

By the discrete Grönwall inequality (Theorem 2.1) with  $x_k := \mathbb{E}[\max_{\ell \leq k} \|e_\ell\|^2]$  and constant  $\alpha_k$  and  $\beta_j = 2\tau C_1$ , and by using that  $K = T/\tau$ , we obtain

$$\begin{aligned} \mathbb{E} \left[ \max_{\ell \in [K]} \|e_\ell\|^2 \right] \\ \leq \exp(2TC_1) \left[ 4(1+T)C_\Psi^2 \tau^{2q} \right. \\ \left. + 2TC_{\xi, R}^2 (1+36(1+\tau^*)(1+C_\Phi\tau^*)^2) \tau^{2p} \right]. \end{aligned}$$

This establishes (3.4).  $\square$

*Proof (Proof of Theorem 3.5)* Let  $0 \leq k \leq K-1$  and  $n \in \mathbb{N}$ . By applying the triangle inequality, (2.3), Assumptions 3.1 and 3.2, and by using that  $1 + \tau 2^{n-1} \leq 1 + 2^{n-1}$  (since  $\tau \leq 1$ ),

$$\begin{aligned} \|e_{k+1}\|^n \\ \leq (\|\Phi^\tau(u_k) - \Psi^\tau(U_k)\| + \|\xi_k(\tau)\|)^n \\ \leq (1 + \tau 2^{n-1}) \|\Phi^\tau(u_k) - \Psi^\tau(U_k)\|^n \\ + (1 + (2/\tau)^{n-1}) \|\xi_k(\tau)\|^n \\ \leq (1 + \tau 2^{n-1}) ((1 + \tau 2^{n-1}) \|\Phi^\tau(u_k) - \Phi^\tau(U_k)\|^n \\ + (1 + (2/\tau)^{n-1}) \|\Phi^\tau(U_k) - \Psi^\tau(U_k)\|^n) \\ + (1 + (2/\tau)^{n-1}) \|\xi_k(\tau)\|^n \\ \leq (1 + \tau 2^{n-1})^2 (1 + \tau C_\Phi)^n \|e_k\|^n \\ + (1 + (2/\tau)^{n-1}) \left( (1 + 2^{n-1}) C_\Psi^n \tau^{n(q+1)} + \|\xi_k(\tau)\|^n \right). \end{aligned}$$

Observe that, since  $2^{n-1}$  and  $C_\Phi$  are nonnegative, and since  $0 < \tau < \tau^*$ ,

$$C_\Phi(n, \tau) := [(1 + \tau 2^{n-1})^2 (1 + \tau C_\Phi)^n - 1] \tau^{-1}. \quad (\text{A.5})$$

Note that  $C_\Phi(n, \tau) \leq C_\Phi(n, \tau^*)$ .

Since  $n \geq 1$  implies that  $1 + (2/\tau)^{n-1} \leq 2^n \tau^{1-n}$ , we have

$$\begin{aligned} & \|e_{k+1}\|^n - \|e_k\|^n \\ & \leq C_\Phi(m, \tau^*) \tau \|e_k\|^m + \tau^{1-n} (1 + 2^{n-1})^2 C_\Psi^n \tau^{n(q+1)} \\ & \quad + \tau^{1-n} (1 + 2^{n-1}) \|\xi_k(\tau)\|^n \\ & \leq C_\Phi(m, \tau^*) \tau \|e_k\|^n + \tau^{1-n} (4C_\Psi \tau^{q+1})^n \\ & \quad + \tau^{1-n} (2\|\xi_k(\tau)\|)^n. \end{aligned}$$

Decomposing  $\|e_{k+1}\|^n - \|e_0\|^n$  as a telescoping sum, using that  $e_0 = u_0 - U_0 = 0$ , using the nonnegativity of the summands on the right-hand side of the last inequality, and using the relation  $\|e_\ell\|^n \leq \max_{j \leq \ell} \|e_j\|^n$ , we obtain

$$\begin{aligned} \max_{\ell \leq k+1} \|e_\ell\|^n & \leq \left( \tau^{1-n} \sum_{k=0}^{K-1} ((4C_\Psi \tau^{q+1})^n + (2\|\xi_k(\tau)\|)^n) \right) \\ & \quad + C_\Phi(n, \tau^*) \tau \sum_{\ell=0}^k \max_{j \leq \ell} \|e_j\|^n. \end{aligned}$$

Using that  $K = T\tau$  and Grönwall's inequality (Theorem 2.1),

$$\begin{aligned} \max_{\ell \in [K]} \|e_\ell\|^n & \leq (4C_\Psi \tau^q)^n T \exp(TC_\Phi(n, \tau^*)) \\ & \quad + \left( \tau^{1-n} 2^n \sum_{k=0}^{K-1} \|\xi_k(\tau)\|^n \right) \exp(TC_\Phi(n, \tau^*)). \end{aligned} \quad (\text{A.6})$$

Taking expectations, using (3.2) with  $w = n$  and  $v = 1$ , and using that  $K = T/\tau$  yields

$$\begin{aligned} \mathbb{E} \left[ \max_{\ell \in [K]} \|e_\ell\|^n \right] & \leq (4C_\Psi \tau^q)^n T \exp(TC_\Phi(n, \tau^*)) \\ & \quad + 2^n TC_{\xi, R}^n \tau^{n(p-1/2)} \exp(TC_\Phi(n, \tau^*)). \end{aligned}$$

Rearranging the above produces the desired inequality.  $\square$

*Proof (Proof of Corollary 3.1)* Let  $m \in \mathbb{N}$  be arbitrary. Using (A.6), and applying (2.4) twice, we obtain

$$\begin{aligned} \max_{\ell \in [K]} \|e_\ell\|^{nm} & \leq 2^{m-1} e^{(TC_\Phi(n, \tau^*))^m} \left[ ((4C_\Psi \tau^q)^n T)^m \right. \\ & \quad \left. + (\tau^{1-n} 2^n)^m \left( \sum_{k=0}^{K-1} \|\xi_k(\tau)\|^n \right)^m \right]. \end{aligned}$$

Taking expectations and using (3.2) with  $w = n$  and  $v = m$ , we obtain

$$\begin{aligned} \mathbb{E} \left[ \max_{\ell \in [K]} \|e_\ell\|^{nm} \right] & \leq 2^{m-1} \exp(TC_\Phi(n, \tau^*))^m ((4C_\Psi \tau^q)^n T)^m \\ & \quad + (\tau^{1-n} 2^n)^m \left( TC_{\xi, R}^n \tau^{n(p+1/2)-1} \right)^m. \end{aligned}$$

The conclusion follows by the series expansion of the exponential and the dominated convergence theorem.  $\square$

*Proof (Proof of Theorem 4.2)* Recall that the solution map  $\Phi^\tau$  of the initial value problem (1.1) satisfies

$$\Phi^\tau(a) := a + \int_0^\tau f(\Phi^t(a)) dt.$$

For any  $\tau > 0$  and  $a, b \in \mathbb{R}^d$ , Assumption 4.1 and the integral Grönwall–Bellman inequality yield

$$\begin{aligned} & \|\Phi^\tau(a) - \Phi^\tau(b)\| \\ & = \left\| a - b + \int_0^\tau f(\Phi^t(a)) - f(\Phi^t(b)) dt \right\| \\ & \leq \|a - b\| \\ & \quad + D \int_0^\tau (1 + \|\Phi^t(a)\|^s + \|\Phi^t(b)\|^s) \|\Phi^t(a) - \Phi^t(b)\| dt \\ & \leq \|a - b\| \exp \left( D \int_0^\tau (1 + \|\Phi^t(a)\|^s + \|\Phi^t(b)\|^s) dt \right). \end{aligned}$$

Given the boundedness hypothesis on the  $(\xi_k(\tau))_{k \in [K]}$ , we may define a finite constant  $C > 0$  that does not depend on  $\tau$  or  $k$ , such that

$$\begin{aligned} \|\Phi^\tau(u_k) - \Phi^\tau(U_k)\| & \leq \|e_k\| \exp(D\tau(1 + 2C)) \\ & \leq \|e_k\| (1 + C'\tau). \end{aligned}$$

The rest of the proof follows in a similar manner to that of Theorem 3.5.  $\square$

*Proof (Proof of Lemma 4.1)* In what follows, we shall omit the dependence of all random variables on  $\omega$ , with the understanding that  $\omega$  is arbitrary. Let  $n \in [K]$ , where  $K = T/\tau \in \mathbb{N}$ . From (1.6) we have, by (2.1),

$$\|U_{n+1}\|^2 \leq (1 + \tau) \|\Psi^\tau(U_n)\|^2 + (1 + \tau^{-1}) \|\xi_n(\tau)\|^2. \quad (\text{A.7})$$

Taking the inner product of (4.3) with  $\Psi^\tau(U_n)$ , we obtain by (4.4)

$$\begin{aligned} & \|\Psi^\tau(U_n)\|^2 \\ & = \langle \Psi^\tau(U_n), U_n \rangle + \tau \langle f(\Psi^\tau(U_n)), \Psi^\tau(U_n) \rangle \\ & \leq \frac{1}{2} (\|\Psi^\tau(U_n)\|^2 + \|U_n\|^2) + \tau (\alpha + \beta \|\Psi^\tau(U_n)\|^2) \\ & = \|\Psi^\tau(U_n)\|^2 \left( \frac{1}{2} + \beta\tau \right) + \frac{1}{2} \|U_n\|^2 + \alpha\tau. \end{aligned}$$

Thus,

$$\begin{aligned} \|\Psi^\tau(U_n)\|^2 & \leq \frac{1}{1 - 2\beta\tau} (\|U_n\|^2 + 2\alpha\tau) \\ & \leq \frac{1}{1 - 2|\beta|\tau} (\|U_n\|^2 + 2\alpha\tau), \end{aligned} \quad (\text{A.8})$$

where we used the inequality  $1 - 2|\beta|\tau \leq 1 + 2\beta\tau$  for the second inequality. Then (A.7) and (A.8) yield

$$\|U_n\|^2 \leq \frac{1 + \tau}{1 - 2|\beta|\tau} (\|U_{n-1}\|^2 + 2\alpha\tau) + \frac{1 + \tau}{\tau} \|\xi_{n-1}(\tau)\|^2. \quad (\text{A.9})$$

Let  $c_1(\tau) := \frac{1+2|\beta|}{1-2|\beta|\tau}$  and  $c_2(\tau) := \frac{2\alpha}{1-2|\beta|\tau}$ . By (A.9), it follows that

$$\begin{aligned} & \|U_n\|^2 - \|U_{n-1}\|^2 \\ & \leq \tau c_1(\tau) \|U_{n-1}\|^2 + (1 + \tau) (\tau c_2(\tau) + \tau^{-1} \|\xi_{n-1}(\tau)\|^2). \end{aligned}$$

Using the telescoping sum

$$\|U_n\|^2 = \|U_0\|^2 + \sum_{i=1}^n (\|U_i\|^2 - \|U_{i-1}\|^2)$$

it follows that

$$\|U_n\|^2 \leq \|U_0\|^2 + \sum_{i=1}^n [\tau c_1(\tau) \|U_{i-1}\|^2 + (1+\tau) (\tau c_2(\tau) + \tau^{-1} \|\xi_{i-1}(\tau)\|^2)].$$

Since  $n \leq K := T/\tau$ , and since the right-hand side of the inequality above is nonnegative,

$$\|U_n\|^2 \leq \|U_0\|^2 + (1+\tau) \left( T c_2(\tau) + \tau^{-1} \sum_{i=1}^{T/\tau} \|\xi_{i-1}(\tau)\|^2 \right) + \tau c_1(\tau) \sum_{i=0}^{n-1} \|U_i\|^2.$$

Applying the Grönwall inequality (Theorem 2.1), yields, for all  $n \in [K]$ ,

$$\begin{aligned} & \max_{i \in [K]} \|U_i\|^2 \\ & \leq \left[ \|U_0\|^2 + (1+\tau) \left( T c_2 + \frac{1}{\tau} \sum_{i=1}^{T/\tau} \|\xi_{i-1}(\tau)\|^2 \right) \right] \exp(T c_1) \\ & \leq (1+\tau) \left[ \|U_0\|^2 + T c_2 + \frac{1}{\tau} \sum_{i=1}^{T/\tau} \|\xi_{i-1}(\tau)\|^2 \right] \exp(T c_1) \\ & \leq C_2 \left( 1 + \tau^{-1} \sum_{i=1}^{T/\tau} \|\xi_{i-1}(\tau)\|^2 \right), \end{aligned}$$

where we define, for  $\tau'$  as in Assumption 4.4, the scalar

$$C_2 = (1+\tau') \max \{1, \|U_0\|^2 + T c_2(\tau')\} \exp(T c_1(\tau')). \quad (\text{A.10})$$

This yields (4.5) for  $n = 1$ . By applying (2.4), we obtain (4.5) for arbitrary  $n \in \mathbb{N}$ .  $\square$

*Proof (Proof of Proposition 4.2)* Recall that in Assumption 4.1, we assume  $f \in C^1(\mathbb{R}^d; \mathbb{R}^d)$ . Therefore, Taylor's theorem applied to the function  $t \mapsto \Phi^t(a)$  yields

$$\Phi^\tau(a) = a + \tau f(a) + \tau R^\tau(a), \quad (\text{A.11})$$

where  $R^\tau(a) \rightarrow 0$  as  $\tau \rightarrow 0$ . Then, by (4.1a), (4.3), and (2.4),

$$\begin{aligned} & \|\Psi^\tau(U_k) - \Phi^\tau(U_k)\|^{2n} \\ & \leq 2^{2n-1} \tau^{2n} (\|f(\Psi^\tau(U_k)) - f(U_k)\|^{2n} + \|R^\tau(U_k)\|^{2n}). \end{aligned} \quad (\text{A.12})$$

By (4.1a), (4.3), (4.2), and (2.4) with the fact that  $C_\Phi \geq 1$  in Assumption 4.1, we obtain

$$\begin{aligned} & \|f(U_k) - f(\Psi^\tau(U_k))\|^{2n} \\ & \leq C_\Phi^{2n} (1 + \|U_k\|^s + \|\Psi^\tau(U_k)\|^s)^{2n} \|U_k - \Psi^\tau(U_k)\|^{2n} \\ & = C_\Phi^{2n} (1 + \|U_k\|^s + \|\Psi^\tau(U_k)\|^s)^{2n} \|\tau f(\Psi^\tau(U_k))\|^{2n} \\ & \leq \tau^{2n} C_\Phi^{2n} (1 + \|U_k\|^s + \|\Psi^\tau(U_k)\|^s)^{2n} \times \\ & \quad (C_\Phi (1 + \|\Psi^\tau(U_k)\|^s) \|\Psi^\tau(U_k)\| + \|f(0)\|)^{2n} \\ & \leq \tau^{2n} 3^{2(2n-1)} C_\Phi^{4n} (1 + \|U_k\|^{2ns} + \|\Psi^\tau(U_k)\|^{2ns}) \times \\ & \quad (\|\Psi^\tau(U_k)\|^{2n} + \|\Psi^\tau(U_k)\|^{2n(s+1)} + \|f(0)\|^{2n}). \end{aligned}$$

From (A.8) and (2.4), it holds that for any  $n$  and  $r$  such that  $nr \geq 1$ ,

$$\begin{aligned} \|\Psi^\tau(U_k)\|^{2nr} & \leq \frac{2^{nr-1}}{(1-2|\beta|\tau)^{nr}} (\|U_k\|^{2nr} + (2\alpha\tau)^{nr}) \\ & \leq \frac{2^{nr-1}}{(1-2|\beta|\tau')^{nr}} (\|U_k\|^{2nr} + (2\alpha\tau')^{nr}), \end{aligned}$$

for  $\tau'$  in Assumption 4.4. Applying the second inequality for the appropriate values of  $r$  and computing exponents yields that, for the polynomials  $\pi_1$ ,  $\pi_2$  and  $\pi$  defined on  $\mathbb{R}$  by

$$\begin{aligned} \pi_1(x) & := \left( 1 + x^{ns} + \frac{2^{ns-1}}{(1-2|\beta|\tau')^{ns}} (x^{ns} + (2\alpha\tau')^{ns}) \right) \\ \pi_2(x) & := \frac{2^{n(s+1)-1}}{(1-2|\beta|\tau')^{n(s+1)}} \times \\ & \quad (x^n + (2\alpha\tau')^n + x^{n(s+1)} + (2\alpha\tau')^{n(s+1)} + \|f(0)\|^{2n}) \end{aligned}$$

and  $\pi(x) := \pi_1(x)\pi_2(x)$ , it follows from Lemma 4.1 that

$$\begin{aligned} & \|f(U_k) - f(\Psi^\tau(U_k))\|^{2n} \\ & \leq \tau^{2n} 3^{2(2n-1)} C_\Phi^{4n} \pi(\|U_k\|^2) \\ & \leq \tau^{2n} 3^{2(2n-1)} C_\Phi^{4n} \pi \left( \max_{i \in [K]} \|U_i\|^2 \right). \end{aligned}$$

Taking expectations, applying Proposition 4.1, and using that  $\tau < \tau'$  to bound the right-hand side of the inequality (4.6) in Proposition 4.1, we may define some  $C_3 = C_3(\alpha, \beta, C_\Phi, \tau', n)$  that does not depend on  $k$  or  $\tau$ , such that

$$\begin{aligned} & \mathbb{E} [\|f(U_k) - f(\Psi^\tau(U_k))\|^{2n}] \\ & \leq \tau^{2n} 3^{2(2n-1)} C_\Phi^{4n} \mathbb{E} \left[ \pi \left( \max_{i \in [K]} \|U_i\|^2 \right) \right] =: \tau^{2n} C_3. \end{aligned} \quad (\text{A.13})$$

By Proposition 4.1, the finiteness of  $C_3$  follows from the hypothesis  $R \geq 2n(2s+1)$  and the observation that  $\pi_1(x^2)$  and  $\pi_2(x^2)$  have degree  $ns$  and  $n(s+1)$  in  $x^2$ , respectively.

Now it remains to show that  $\|R^\tau(U_k)\|^{2n} \in L_{\mathbb{P}}^1$ . From (4.1a), (4.1b), and (A.11), we obtain

$$\begin{aligned} & \tau \|R^\tau(a) - R^\tau(b)\| \\ & = \|\Phi^\tau(a) - a - \tau f(a) - \Phi^\tau(b) - b - \tau f(b)\| \\ & \leq \|\Phi^\tau(a) - \Phi^\tau(b)\| + \|a - b\| + \tau \|f(a) - f(b)\| \\ & = 2(1 + \tau C_\Phi (1 + \|a\|^s + \|b\|^s) \|a - b\|) \end{aligned} \quad (\text{A.14})$$

By the triangle inequality and (A.14),

$$\begin{aligned} & \tau \|R^\tau(U_k)\| \\ & \leq \tau (\|R^\tau(0)\| + 2(1 + \tau C_\Phi (1 + \|U_k\|^s) \|U_k\|)) \\ & \leq \tau (\|R^\tau(0)\| + 2(1 + \tau C_\Phi (\|U_k\| + \|U_k\|^{s+1}))) \\ & \leq \tau \left( \|R^\tau(0)\| + 2 \right. \\ & \quad \left. + 2C_\Phi \left( \max_{k \in [K]} \|U_k\| + \max_{k \in [K]} \|U_k\|^{s+1} \right) \right). \end{aligned}$$

Then by applying (2.4) and Proposition 4.1 with the hypothesis that  $R \geq 2n(2s+1) \geq 2n(s+1)$ , and using the bound  $\tau < \tau'$ , it follows that we may define a positive scalar  $C_4$  that

does not depend on  $k$  or  $\tau$ , such that

$$\begin{aligned} & \mathbb{E} [\tau^{2n} \|R^\tau(U_k)\|^{2n}] \\ & \leq \tau^{2n} \mathbb{E} \left[ \left( \|R^\tau(0)\| + 2 + \right. \right. \\ & \quad \left. \left. 2C_\Phi \left( \max_{k \in [K]} \|U_k\| + \max_{k \in [K]} \|U_k\|^{s+1} \right) \right)^{2n} \right] \\ & =: \tau^{2n} C_4. \end{aligned} \quad (\text{A.15})$$

Therefore, with  $C_3$  and  $C_4$  as in (A.13) and (A.15) above, (A.12) yields

$$\begin{aligned} & \mathbb{E} [\|\Psi^\tau(U_k) - \Phi^\tau(U_k)\|^{2n}] \leq 2^{2n-1} \tau^{4n} (C_3 + C_4) \\ & =: C_\Psi \tau^{4n} \end{aligned} \quad (\text{A.16})$$

as desired.  $\square$

The proof below makes clear that we make absolutely no effort to find optimal constants.

*Proof (Proof of Theorem 4.5)* Let  $n \in \mathbb{N}$ . By (2.3)

$$\begin{aligned} & \|e_{k+1}\|^{2n} \\ & \leq (1 + \tau 2^{2n-1}) [(1 + \tau 2^{2n-1}) \|\Phi^\tau(u_k) - \Phi^\tau(U_k)\|^{2n} \\ & \quad + (1 + (2/\tau)^{2n-1}) \|\Phi^\tau(U_k) - \Psi^\tau(U_k)\|^{2n}] \\ & \quad + (1 + (2/\tau)^{2n-1}) \|\xi_k(\tau)\|^{2n}. \end{aligned}$$

Since  $\tau \leq 1$  and  $n \geq 1$ , it holds that  $1 + \tau^{1-2n} 2^{2n-1} \leq \tau^{1-2n} (1 + 2^{2n-1})$  and  $1 + \tau 2^{2n-1} \leq 1 + 2^{2n-1}$ . Using these inequalities, (2.3), and (4.1b) in the preceding inequality, we obtain

$$\begin{aligned} & \|e_{k+1}\|^{2n} \\ & \leq (1 + \tau 2^{2n-1})^2 (1 + \tau C_\Phi (1 + \|u_k\|^s + \|U_k\|^s))^{2n} \|e_k\|^{2n} \\ & \quad + (1 + 2^{2n-1})^2 \tau^{1-2n} \times \\ & \quad (\|\Phi^\tau(U_k) - \Psi^\tau(U_k)\|^{2n} + \|\xi_k(\tau)\|^{2n}). \end{aligned}$$

Using (2.3) again, we obtain

$$\begin{aligned} & (1 + \tau C_\Phi (1 + \|u_k\|^s + \|U_k\|^s))^{2n} \|e_k\|^{2n} \\ & \leq \left[ \left( 1 + \tau C_\Phi \left( 1 + \max_{\ell \in [K]} \|u_\ell\|^s \right) \right) \|e_k\| + \right. \\ & \quad \left. \tau C_\Phi \max_{\ell \in [K]} \|U_\ell\|^s \|e_k\| \right]^{2n} \\ & \leq (1 + \tau 2^{2n-1}) \left( 1 + \tau C_\Phi \left( 1 + \max_{\ell \in [K]} \|u_\ell\|^s \right) \right)^{2n} \|e_k\|^{2n} \\ & \quad + (1 + \tau^{1-2n} 2^{2n-1}) (\tau C_\Phi)^{2n} \max_{\ell \in [K]} \|U_\ell\|^{2ns} \|e_k\|^{2n} \end{aligned}$$

so that by defining

$$C_5 = C_5(n, s, C_\Phi, u) := \max\{2^{2n-1}, C_\Phi(1 + \|u\|_\infty^s)\}$$

we have

$$\begin{aligned} & (1 + \tau C_\Phi (1 + \|u_k\|^s + \|U_k\|^s))^{2n} \|e_k\|^{2n} \\ & \leq (1 + \tau C_5)^{2n+1} \|e_k\|^{2n} \\ & \quad + \tau^{1-2n} (1 + 2^{2n-1}) \left( \tau C_\Phi \max_{\ell \in [K]} \|U_\ell\|^s \|e_k\| \right)^{2n} \end{aligned}$$

and, therefore,

$$\begin{aligned} & \|e_{k+1}\|^{2n} - \|e_k\|^{2n} \\ & \leq [(1 + \tau C_5)^{2n+1} - 1] \tau^{-1} \tau \|e_k\|^{2n} \\ & \quad + \tau^{1-2n} (1 + 2^{2n-1}) \left( \tau C_\Phi \max_{\ell \in [K]} \|U_\ell\|^s \|e_k\| \right)^{2n} \\ & \quad + (1 + 2^{2n-1})^2 \tau^{1-2n} \times \\ & \quad (\|\Phi^\tau(U_k) - \Psi^\tau(U_k)\|^{2n} + \|\xi_k(\tau)\|^{2n}). \end{aligned}$$

By nonnegativity of  $C_5$ , it follows that  $[(1 + \tau C_5)^{2n+1} - 1] \tau^{-1}$  is a polynomial of degree  $2n$  in  $\tau$  with positive coefficients. In particular, if we recall the definition of  $C_5$  and define  $C_6$  by

$$\begin{aligned} & C_6(C_\Phi, n, s, (u_t)_{0 \leq t \leq T}, \tau') \\ & := \left[ (1 + \tau' \max\{2^{2n-1}, C_\Phi(1 + \|u\|_\infty^s)\})^{2n+1} - 1 \right] (\tau')^{-1}, \end{aligned} \quad (\text{A.17})$$

then  $C_6$  does not depend on  $\tau$ ,  $[(1 + \tau C_5)^{2n+1} - 1] \tau^{-1} \leq C_6$  for all  $0 < \tau < \tau'$ , and

$$\begin{aligned} & \|e_{k+1}\|^{2n} - \|e_k\|^{2n} \\ & \leq C_6 \tau \|e_k\|^{2n} \\ & \quad + \tau^{1-2n} (1 + 2^{2n-1}) \left( \tau C_\Phi \max_{\ell \in [K]} \|U_\ell\|^s \|e_k\| \right)^{2n} \\ & \quad + (1 + 2^{2n-1})^2 \tau^{1-2n} \times \\ & \quad (\|\Phi^\tau(U_k) - \Psi^\tau(U_k)\|^{2n} + \|\xi_k(\tau)\|^{2n}). \end{aligned}$$

By the telescoping sum associated to  $\|e_{k+1}\|^{2n} - \|e_k\|^{2n}$ , the fact that  $e_0 = 0$ , the bound  $1 + 2^{2n-1} \leq 2^{2n}$ , the nonnegativity of the terms on the right-hand side of the inequality above, and the bound  $\|e_j\| \leq \max_{\ell \leq j} \|e_\ell\|$ , we obtain

$$\begin{aligned} & \max_{\ell \leq k+1} \|e_\ell\|^{2n} \\ & \leq \tau^{1-2n} 2^{4n} \sum_{\ell=1}^K \left( \left( \tau C_\Phi \max_{j \in [K]} \|U_j\|^s \|e_\ell\| \right)^{2n} + \right. \\ & \quad \left. \|\Phi^\tau(U_\ell) - \Psi^\tau(U_\ell)\|^{2n} + \|\xi_\ell(\tau)\|^{2n} \right) \\ & \quad + C_6 \tau \sum_{\ell=1}^k \max_{j \leq \ell} \|e_j\|^{2n}. \end{aligned}$$

By Lemma 4.1,

$$\begin{aligned} & \left( \tau C_\Phi \max_{j \in [K]} \|U_j\|^s \|e_\ell\| \right)^{2n} \\ & \leq C_\Phi^{2n} (2C_1)^{sn} \times \\ & \quad \left( \tau^{2n} \|e_\ell\|^{2n} + \tau^{n(2-s)} \left( \sum_{i=1}^{T/\tau} \|\xi_i(\tau)\|^2 \right)^{ns} \|e_\ell\|^{2n} \right) \end{aligned}$$

which implies that

$$\begin{aligned} & \max_{\ell \leq k+1} \|e_\ell\|^{2n} \\ & \leq \tau^{1-ns} 2^{n(4+s)} C_\Phi^{2n} C_1^{ns} \sum_{\ell=1}^K \left( \sum_{i=1}^{T/\tau} \|\xi_i(\tau)\|^2 \right)^{ns} \|e_\ell\|^{2n} \\ & \quad + \tau^{1-2n} 2^{4n} \sum_{\ell=1}^K (\|\Phi^\tau(U_\ell) - \Psi^\tau(U_\ell)\|^{2n} + \|\xi_\ell(\tau)\|^{2n}) \\ & \quad + \left( C_6 \tau + \tau^{2n} 2^{n(4+s)} C_\Phi^{2n} C_1^{ns} \right) \sum_{\ell=1}^k \max_{j \leq \ell} \|e_j\|^{2n}. \end{aligned}$$



Define

$$C_7 = C_7(n, s, C_\Phi, C_1) := 2^{n(4+s)} C_\Phi^{2n} C_1^{ns}. \quad (\text{A.18})$$

Since  $C_\Phi, C_1 \geq 1$ , it follows that  $2^{4n} \leq C_7$ , and by Grönwall's inequality (Theorem 2.1) we obtain

$$\begin{aligned} & \max_{\ell \in [K]} \|e_\ell\|^{2n} \\ & \leq \exp(T(C_6 + \tau^{2n-1}C_7)) C_7 \times \\ & \quad \left( \tau^{1-ns} \sum_{\ell=1}^K \left( \sum_{i=1}^{T/\tau} \|\xi_i(\tau)\|^2 \right)^{ns} \|e_\ell\|^{2n} \right. \\ & \quad \left. + \tau^{1-2n} \sum_{\ell=1}^K (\|\Phi^\tau(U_\ell) - \Psi^\tau(U_\ell)\|^{2n} + \|\xi_\ell(\tau)\|^{2n}) \right). \end{aligned}$$

Taking expectations completes the proof, provided that we can ensure each sum is of the right order in  $\tau$ . By Proposition 4.2 with the hypothesis that  $R \geq 2n(2s+1)$ , and by Assumption 3.3,

$$\begin{aligned} & \tau^{1-2n} \sum_{\ell=1}^K \mathbb{E} [\|\Phi^\tau(U_\ell) - \Psi^\tau(U_\ell)\|^{2n} + \|\xi_\ell(\tau)\|^{2n}] \\ & \leq T \left( C_\Psi \tau^{2n} + (C_{\xi, R} \tau^{p-1/2})^{2n} \right). \quad (\text{A.19}) \end{aligned}$$

Thus we need  $p - 1/2 \geq 1$  to hold. Next, using the bound  $\|e_\ell\| \leq \max_{j \in [K]} \|e_j\|$ , Young's inequality (2.1) with  $a = (\sum_{i=1}^K \|\xi_i(\tau)\|^2)^{ns}$ ,  $b = \|e_\ell\|^{2n}$ , and some  $\delta > 0$  and conjugate exponent pair  $(r, r^*) \in (1, \infty)^2$  to be determined later, we obtain with (3.2) that

$$\begin{aligned} & \mathbb{E} \left[ \left( \sum_{i=1}^{T/\tau} \|\xi_i(\tau)\|^2 \right)^{ns} \|e_\ell\|^{2n} \right] \\ & \leq \frac{\delta}{r} \mathbb{E} \left[ \left( \sum_{i=1}^{T/\tau} \|\xi_i(\tau)\|^2 \right)^{nr s} \right] + \frac{1}{\delta^{r^*/r} r^*} \mathbb{E} \left[ \max_{\ell \in [K]} \|e_\ell\|^{2nr^*} \right] \\ & \leq \frac{\delta}{r} (TC_{\xi, R}^2 \tau^{2p})^{nr s} + \frac{1}{\delta^{r^*/r} r^*} \mathbb{E} \left[ \max_{\ell \in [K]} \|e_\ell\|^{2nr^*} \right]. \end{aligned}$$

Since  $R \geq 2n(2s+1)$ , the maximal value of  $r$  compatible with integrability of  $(\sum_{i=1}^K \|\xi_i(\tau)\|^2)^{nr s}$  is  $r = 2 + s^{-1}$ . Since we are not interested in optimal estimates, we shall set  $r = r^* = 2$  and  $\delta = \tau^{-n(2+s)}$ . We thus obtain

$$\begin{aligned} & \tau^{1-ns} \sum_{\ell=1}^K \mathbb{E} \left[ \left( \sum_{i=1}^{T/\tau} \|\xi_i(\tau)\|^2 \right)^{ns} \|e_\ell\|^{2n} \right] \\ & \leq \frac{T}{2} \tau^{-ns} \left( (TC_{\xi, R}^2)^{nr s} \tau^{-n(2+s)+2p(2ns)} \right. \\ & \quad \left. + \tau^{n(2+s)} \mathbb{E} \left[ \max_{\ell \in [K]} \|e_\ell\|^{4n} \right] \right). \end{aligned}$$

For the exponent of  $\tau$  of the first term in the parentheses, we want to ensure that  $-n(2+s) + 2p(2ns) - ns \geq 2n$ , or equivalently that  $p \geq \frac{1}{s} + \frac{1}{2}$ . Comparing this condition with the condition  $p - \frac{1}{2} \geq 1$  that arose from (A.19), and recalling that  $s \geq 1$ , we observe that if  $p \geq \frac{3}{2}$ , then the preceding

estimates yield

$$\begin{aligned} & \mathbb{E} \left[ \max_{\ell \in [K]} \|e_\ell\|^{2n} \right] \\ & \leq \exp(T(C_6 + \tau^{2n-1}C_7)) \frac{C_7 T}{2} \times \\ & \quad \left( (TC_{\xi, R}^2)^{2ns} + \mathbb{E} \left[ \max_{\ell \in [K]} \|e_\ell\|^{4n} \right] \right) \tau^{2n}. \end{aligned}$$

It remains to bound  $\mathbb{E}[\max_{\ell \in [K]} \|e_\ell\|^{4n}]$  by a constant that does not depend on  $\tau$ . By (2.4), Proposition 4.1, and the assumption that  $\tau < \tau'$  for  $\tau'$  in Assumption 4.4, we obtain

$$\begin{aligned} & \mathbb{E} \left[ \max_{\ell \in [K]} \|e_\ell\|^{4n} \right] \\ & \leq 2^{4n} \left( \max_{k \in [K]} \|u_k\|^{4n} + \mathbb{E} \left[ \max_{k \in [K]} \|U_k\|^{4n} \right] \right) \\ & \leq 2^{4n} \left( \|u\|_\infty^{4n} + (2C_2)^{2n} (1 + TC_{\xi, R}^2 \tau^{2p-1})^{2n} \right) \\ & \leq 2^{4n} \left( \|u\|_\infty^{4n} + (2C_2)^{2n} (1 + TC_{\xi, R}^2 (\tau')^{2p-1})^{2n} \right) =: C_8, \quad (\text{A.20}) \end{aligned}$$

where  $C_8 = C_8(C_2, C_{\xi, R}, n, p, \tau', T, (u_t)_{0 \leq t \leq T}) > 0$  does not depend on  $\tau$ . Note that in applying Proposition 4.1, we have used that  $s \geq 1$  for the exponent  $s$  in Assumption 4.1, since this implies that  $2n(2s+1) \geq 4n$ .  $\square$

*Proof (Proof of Theorem 5.2)* Let  $k \in [K]$  and  $t_k < t \leq t_{k+1}$ . Then

$$\begin{aligned} e(t) &= \Phi^{t-t_k}(u_k) - \Psi^{t-t_k}(U_k) - \xi_k(t-t_k) \\ &= \Phi^{t-t_k}(u_k) - \Phi^{t-t_k}(U_k) + \Phi^{t-t_k}(U_k) \\ & \quad - \Psi^{t-t_k}(U_k) - \xi_k(t-t_k), \end{aligned}$$

and given that Assumption 3.1 implies that  $\Phi^{t'}$  is Lipschitz on  $\mathbb{R}^d$  for every  $t' \geq 0$ ,

$$\begin{aligned} \|e(t)\|^n &\leq 3^{n-1} \left( (1 + C_\Phi(t-t_k))^n \|e_k\|^n \right. \\ & \quad \left. + (C_\Psi(t-t_k)^{q+1})^n + \|\xi_k(t-t_k)\|^n \right) \end{aligned}$$

by applying (2.4). Since  $t - t_k \leq \tau$ , it follows from the inequality above that

$$\begin{aligned} & \mathbb{E} \left[ \sup_{0 \leq t \leq T} \|e(t)\|^n \right] \\ &= \mathbb{E} \left[ \max_{k \in [K]} \sup_{t_k < t \leq t_{k+1}} \|e(t)\|^n \right] \\ &\leq 3^{n-1} \left( (1 + C_\Phi \tau)^n \mathbb{E} \left[ \max_{k \in [K]} \|e_k\|^n \right] + (C_\Psi \tau^{q+1})^n \right. \\ & \quad \left. + \mathbb{E} \left[ \max_{0 \leq k \leq K-1} \sup_{t_k < t \leq t_{k+1}} \|\xi_k(t-t_k)\|^n \right] \right). \end{aligned}$$

By Assumption 5.1,

$$\begin{aligned} & \mathbb{E} \left[ \max_{0 \leq k \leq K-1} \sup_{t_k < t \leq t_{k+1}} \|\xi_k(t-t_k)\|^n \right] \\ &\leq \sum_{k=0}^{K-1} \mathbb{E} \left[ \sup_{t_k < t \leq t_{k+1}} \|\xi_k(t-t_k)\|^n \right] \\ &\leq \frac{T}{\tau} \left( C_{\xi, R} \tau^{p+1/2} \right)^n = TC_{\xi, R}^n \tau^{n(p+1/2)-1} \\ &\leq TC_{\xi, R}^n \tau^{n(p-1/2)}. \end{aligned}$$

Note that Assumption 5.1 is stronger than Assumption 3.3. Therefore we may apply Theorem 3.5 to obtain (5.1).  $\square$

*Proof (Proof of Lemma 5.1)* If  $r = 0$ , then the desired statement follows immediately. Therefore, let  $p, r \geq 1$ . Let  $\xi_0$  be the integrated  $\mathbb{P}$ -Brownian motion process scaled by  $\tau^{p-1}$ , so that

$$\begin{aligned} \|\xi_0(t)\|^r &= \tau^{r(p-1)} t^r \left\| \frac{1}{t} \int_0^t B_s \, ds \right\|^r \\ &\leq \tau^{r(p-1)} t^r \left( \frac{1}{t} \int_0^t \|B_s\|^r \, ds \right) \\ &= \tau^{r(p-1)} t^{r-1} \int_0^t \|B_s\|^r \, ds, \end{aligned}$$

where we applied Jensen's inequality to the uniform probability measure on  $[0, t]$ . It follows that

$$\begin{aligned} &\mathbb{E} \left[ \sup_{t \leq \tau} \|\xi_0(t)\|^r \right] \\ &\leq \tau^{r(p-1)} \mathbb{E} \left[ \sup_{t \leq \tau} t^{r-1} \int_0^t \|B_s\|^r \, ds \right] \\ &\leq \tau^{r(p-1)} \tau^{r-1} \mathbb{E} \left[ \sup_{t \leq \tau} \int_0^t \|B_s\|^r \, ds \right] \\ &\leq \tau^{rp-1} \int_0^\tau \mathbb{E} \left[ \sup_{0 \leq t \leq \tau} \|B_t\|^r \right] \, ds \\ &= \tau^{rp} \mathbb{E} \left[ \sup_{0 \leq t \leq \tau} \|B_t\|^r \right]. \end{aligned}$$

Above, we used the Fubini–Tonelli theorem to interchange expectation and integration with respect to  $s$ , and the fact that  $\mathbb{E}[\sup_{t \leq \tau} \|B_t\|^r]$  is constant with respect to the variable of integration  $s$ . For  $r = 1$ , the Burkholder–Davis–Gundy martingale inequality (Peškir 1996, Equation (2.2)) yields

$$\mathbb{E} \left[ \sup_{0 \leq t \leq \tau} \|B_t\|^r \right] \leq \frac{4-r}{2-r} \tau^{r/2},$$

with  $(4-r)/(2-r) = 3$  for  $r = 1$ . For  $r > 1$ , Doob's inequality (Peškir 1996, Equation (2.1)) yields

$$\mathbb{E} \left[ \sup_{0 \leq t \leq \tau} \|B_t\|^r \right] \leq \left( \frac{r}{r-1} \right)^r \tau^{r/2}.$$

Since  $r \mapsto [r/(r-1)]^r$  is continuously differentiable and monotonically decreasing on  $2 < r < \infty$ , the desired conclusion follows.  $\square$