Strong convergence rates of probabilistic integrators for ordinary differential equations

T. J. Sullivan[1]
with H. C. Lie[1] and A. M. Stuart[2]

Probabilistic Scientific Computing
ICERM, Providence, RI, US

5–9 June 2017

[1]Free University of Berlin / Zuse Institute Berlin, DE
[2]California Institute of Technology, US

Outline

Introduction

Setup

High-order integration of Lipschitz flows

Euler integration of locally Lipschitz dissipative vector fields

Closing Remarks

Details in

arXiv:1703.03680

Introduction

Motivation

- The recent work of Conrad et al. (2016) proposed the use of probabilistic solvers for a trajectory $[0, T] \ni t \mapsto u(t) \in \mathbb{R}^n$ satisfying an ODE/IVP of the form

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} u(t) &= f(u(t)), && \text{for } t \geq 0, \\ u(0) &= u_0, \end{aligned} \tag{1}$$

- Stochasticity is a way to systematically introduce and probe the model error that has been introduced by the discretisation, enabling exploration of possible responses of the system to inputs.
- Such ideas have wide application in forward uncertainty quantification, inverse problems (Kaipio and Somersalo, 2005; Stuart, 2010), and data assimilation (Law et al., 2015; Reich and Cotter, 2015).
- Just as with classical numerical analysis of deterministic integration schemes, we can analyse the accuracy and convergence properties of probabilistic solvers for (1).

- Conrad et al. (2016, Theorem 2.2) gave a convergence result for the error between the random values U_k of a discrete-time numerical solution at discrete times $t_k := k\tau$, $\tau > 0$, and the corresponding values $u_k := u(k\tau)$ of the exact solution:

$$\max_{0 \leq k\tau \leq T} \mathbb{E}\left[\|u_k - U_k\|^2\right] \leq C\tau^{2p \wedge 2q},$$

along with an analogous result in continuous time with the same exponent but possibly different constant.
- Loosely speaking, τ^q is the global order of accuracy of a deterministic method underlying U_k and the variance of a Gaussian model ξ_k for the truncation error over a time horizon $[t_k, t_{k+1}]$ of length τ scales like τ^{1+2p}.

Slogan

"The choice $p = q$ introduces the maximum amount of solution uncertainty consistent with preserving the order of accuracy of the underlying deterministic integrator."

What's New? The Take-Home Message

- We extend the setting of the IVP (1) from $\mathbb{R}^n$ to a Hilbert space $\mathcal{H}$.
- We relax the assumption that all deviations are Gaussian, and work directly with conditions on polynomial moments.
- We bring the time supremum inside the expectation to yield

$$\mathbb{E}\left[\max_{0\leq k\tau\leq T}\|u_k-U_k\|^2\right]\leq C\tau^{2p\wedge 2q},$$

  so that the mode of convergence is strengthened to mean square convergence in the uniform norm on path space, but with the same rate — very useful for later application to inverse problems (Stuart, 2010).
- The assumption that the vector field f is globally Lipschitz is weakened in two ways: for integrators of arbitrary order, we consider Lipschitz flows; for Euler integrators, which have $q=1$, we consider dissipative vector fields with polynomially-growing locally Lipschitz constant.

Setup

Some Notation

- $\mathcal{H}$ is a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$.
- $(\Omega, \mathcal{F}, \mathbb{P})$ is a rich enough probability space; $\mathbb{E}$ denotes expectation (integration over Ω) with respect to $\mathbb{P}$.
- C, C', etc. will denote non-negative constants whose value may change from one occurence to the next, but will always be independent of any time step $\tau > 0$ used to numerically solve the ODE of interest.
- For real numbers a and b, $a \wedge b$ denotes their minimum.
- $\mathrm{Lip}(\Phi)$ denotes the minimal Lipschitz constant of a function Φ, defined on a subset of $\mathcal{H}$ and taking values in $\mathcal{H}$, i.e. $\mathrm{Lip}(\Phi)$ is the least $L \geq 0$ such that

$$\|\Phi(x) - \Phi(y)\| \leq L\|x - y\| \text{ for all } x, y \in \mathrm{domain}(\Phi).$$

More Notation

- $\Phi^t \colon \mathcal{H} \to \mathcal{H}$ will denote the flow map induced by the ODE (1), i.e. $\Phi^t(u_0) := u(t)$.
- We approximately solve the ODE (1) over $[0, T]$ with uniform time step $\tau > 0$, and write $t_k := k\tau$ and $K := T/\tau \in \mathbb{N}$.
- Let $u_k := u(t_k) \equiv \Phi^\tau(u_{k-1})$ denote the value of the exact solution to (1) at time t_k.
- Discrete-time deterministic approximate solutions are given by a one-step integrator, i.e. a numerical flow map $\Psi^\tau \colon \mathcal{H} \to \mathcal{H}$,

$$U_{k+1} := \Psi^\tau(U_k).$$

- Discrete-time stochastic approximate solutions are given by

$$U_{k+1} := \Psi^\tau(U_k) + \xi_k(\tau)$$

and approximations with continuous-time output 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 suitable stochastic processes $\xi_k \colon [0, \tau] \times \Omega \to \mathcal{H}$.

High-order integration of Lipschitz flows

- In this section we consider a generic one-step integrator Ψ^τ, possibly of high order.
- The focus here is on:
  - relaxing regularity assumptions about f to regularity assumptions about Φ^t;
  - bringing the time supremum inside the expectation.
- The main tools are the Grönwall and Burkholder–Davis–Gundy inequalities.
- The analysis is structurally similar to convergence analysis for Wiener–Itô SDEs (Higham et al., 2002; Mao and Szpruch, 2013), but the noise is smaller.
- The surprise is that we don't lose anything in the convergence rate, only in the constant.

Assumption 1 (Smoothness of the flow)

Suppose that f is smooth enough that, for $|t|$ small enough, its flow map Φ^t is globally Lipschitz with Lipschitz constant $\mathrm{Lip}(\Phi^t) \leq 1 + C|t|$.

- Assumption 1 holds in the 'classical' Conrad et al. (2016) setting of a globally Lipschitz vector field.
- Assumption 1 also holds if f merely satisfies, for some constant $\mu \in \mathbb{R}$, the one-sided Lipschitz condition

$$\langle f(x) - f(y), x - y \rangle \leq \mu \|x - y\|^2 \text{ for all } x, y \in \mathcal{H}, \tag{2}$$

in which case $\mathrm{Lip}(\Phi^t) \leq 1 + 2|\mu||t|$ for small enough $|t|$.

Assumption 2 (Accuracy of the numerical flow)

The numerical flow-map Ψ^τ has uniform local truncation error of order $q+1$: for some constant $C \geq 0$,

$$\sup_{u \in \mathcal{H}} \|\Psi^\tau(u) - \Phi^\tau(u)\| \leq C\tau^{q+1}.$$

- Prime example, q^{th} Runge–Kutta.
- Implied global convergence rate in the absence of noise:

$$\max_{0 \leq k \leq K} \|u_k - U_k\| \leq C\tau^q.$$

Assumptions III

Assumption 3 (Regularity of noise)

The ξ_k are mutually independent and identically distributed mean-zero stochastic processes, and there are constants $C \geq 0$ and $p \geq 1$ such that, for all k and all $t \in [0, \tau]$,

$$\mathbb{E}\left[\|\xi_k(t) \otimes \xi_k(t)\|\right] \leq Ct^{2p+1},$$

and, in particular, $\mathbb{E}\left[\|\xi_k(t)\|^2\right] \leq Ct^{2p+1}$.

- Prime example: ξ modelled on the time integral of Brownian motion, $\xi_0(t) := \tau^{p-1} \int_0^t B(s)\,\mathrm{d}s$, where B denotes a standard $\mathcal{H}$-valued Brownian motion.

Convergence with time supremum outside expectation

The first improvement relative to Conrad et al. (2016) is to obtain the same mode and rate of convergence with Lipschitz flow instead of Lipschitz vector field:

Theorem 4 (Uniform mean-square convergence)

Under Assumptions 1, 2, and 3, there exist constants $C \geq 0$ such that

$$\max_{0 \leq k \leq K} \mathbb{E}\left[\|u_k - U_k\|^2\right] \leq C\tau^{2p \wedge 2q},$$

$$\sup_{0 \leq t \leq T} \mathbb{E}\left[\|u(t) - U(t)\|^2\right] \leq C\tau^{2p \wedge 2q}.$$

- Cf. implied global convergence rate in the absence of noise:

$$\max_{0 \leq k \leq K} \|u_k - U_k\| \leq C\tau^{q}.$$

Recurrence for the error $e_k := u_k - U_k$:

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

Hence:

$$\|e_{k+1}\|^2 - \|e_k\|^2 \le C\tau\|e_k\|^2 + C\tau^{1+2q} + \|\xi_k(\tau)\|^2 + 2\langle \Phi^\tau(u_k) - \Psi^\tau(U_k), \xi_k(\tau)\rangle.$$

Take expectations and apply Assumptions 1–3:

$$\left|\mathbb{E}\left[\|e_{k+1}\|^2 - \|e_k\|^2\right]\right| \le C\tau^{1+(2p\wedge 2q)} + \sum_{j=0}^{k-1}\left|\mathbb{E}\left[\|e_{j+1}\|^2 - \|e_j\|^2\right]\right|.$$

Then apply Grönwall:

$$\left|\mathbb{E}\left[\|e_{k+1}\|^2 - \|e_k\|^2\right]\right| \le C\tau^{1+(2p\wedge 2q)}\exp(kC\tau) \le C\tau^{1+(2p\wedge 2q)},$$

Theorem 5 (Mean-square uniform convergence)

Under Assumptions 1, 2, and 3, there exists $C \geq 0$ such that

$$\mathbb{E}\left[\max_{0\leq k\leq K} \|u_k - U_k\|^2\right] \leq C\tau^{2p\wedge 2q}.$$

If, in addition, Assumption 3 is strengthened to

$$\mathbb{E}\left[\sup_{0\leq t\leq \tau} \|\xi_0(t)\|^2\right] \leq C\tau^{1+2p},$$

then

$$\mathbb{E}\left[\sup_{0\leq t\leq T} \|u(t) - U(t)\|^2\right] \leq C\tau^{2p\wedge 2q}.$$

Sketch proof

Applying $\mathbb{E}\left[\max_{k\leq\ell}\cdot\right]$, where $\ell \leq K = T/\tau$, to $\|e_k\|^2 - \|e_0\|^2 = \sum_{j=0}^{k-1}(\|e_{j+1}\|^2 - \|e_j\|^2)$ gives

$$\mathbb{E}\left[\max_{k\leq\ell}\|e_k\|^2\right] \leq \mathbb{E}\left[\max_{k\leq\ell}\sum_{j=0}^{k-1}\left(\tau C\|e_j\|^2 + C\tau^{1+2q} + \|\xi_j(\tau)\|^2\right)\right] + 2\mathbb{E}\left[\max_{k\leq\ell}\left\|\sum_{j=0}^{k-1}\langle\Phi^\tau(u_j) - \Psi^\tau(U_j), \xi_j(\tau)\rangle\right\|\right].$$

Burkholder–Davis–Gundy inequality (Peškir, 1996; Ren, 2008) gives

$$\begin{aligned}
\mathbb{E}\left[\max_{k\leq\ell}\|e_k\|^2\right] &\leq \mathbb{E}\left[\sum_{j=0}^{\ell-1}(\tau C\|e_j\|^2 + C\tau^{1+2q} + \|\xi_j(\tau)\|^2)\right] + C\mathbb{E}\left[[\langle\Phi^\tau(u_\bullet) - \Psi^\tau(U_\bullet), \xi_\bullet(\tau)\rangle]_{\ell-1}^{1/2}\right] \\
&\leq \tau C\sum_{j-0}^{\ell-1}\mathbb{E}\left[\|e_j\|^2\right] + CT\tau^{2p\wedge 2q} + \frac{1}{2}\mathbb{E}\left[\max_{k\leq\ell}\|e_k\|^2\right] \\
&\leq \tau C\sum_{j-0}^{\ell-1}\mathbb{E}\left[\max_{k\leq j}\|e_k\|^2\right] + CT\tau^{2p\wedge 2q} + \frac{1}{2}\mathbb{E}\left[\max_{k\leq\ell}\|e_k\|^2\right].
\end{aligned}$$

Followed by Grönwall to complete the argument.

The strengthened Assumption 3 holds for at least one reasonable model of truncation error, namely one that corresponds to the assumption that the true solution u has one continuous derivative and a second derivative can be modelled as Brownian white noise.

Proposition 6

The strengthened Assumption 3 holds for ξ modelled on the time integral of Brownian motion, $\xi_0(t) := \tau^{p-1} \int_0^t B(s)\,\mathrm{d}s$, where B denotes a standard $\mathcal{H}$-valued Brownian motion.

Euler integration of locally Lipschitz dissipative vector fields

- We now consider a randomised version of the implicit Euler scheme defined by

$$\Psi^{\tau}(x) = x + \tau f(\Psi^{\tau}(x)).$$

  and merely locally Lipschitz vector fields.
- The proofs heavily use the structure of this particular integrator and a dissipativity condition on the vector field
- Similar but more involved arguments will work for explicit Euler.

Assumption 7 (Generalised dissipativity)

Assume that f satisfies, for constants $\alpha \geq 0$ and $\beta \in \mathbb{R}$, the generalised dissipativity condition that

$$\langle f(v), v\rangle \leq \alpha + \beta \|v\|^2 \text{ for all } v \in \mathcal{H}\,. \tag{3}$$

- We allow positive values of β, so Assumption 7 is more general than the usual dissipativity property found in the literature (Humphries and Stuart, 1994, Eq. (1.2)).
- Recent studies in numerical methods for stochastic differential equations consider constraints on the drift and diffusion of the SDEs that feature the same right-hand side above (Fang and Giles, 2016; Mao and Szpruch, 2013).

- We assume that the vector field satisfies a polynomial growth condition; this condition may be seen as a kind of local Lipschitz property (Higham et al., 2002, Assumption 4.1).

Assumption 8 (Polynomial growth condition)

Constants $D \geq 1$ and $q \in \mathbb{N}_0$ are such that

$$\|f(a) - f(b)\|^2 \leq D(1 + \|a\|^q + \|b\|^q)\|a - b\|^2 \text{ for all } a, b \in \mathcal{H}. \tag{4}$$

- q is now measuring nonlinearity, not the order of accuracy of the integrator, which is just 1 for Euler. The case $q = 0$ is global Lipschitz continuity of f,

Assumption 9 ((p, R)-regularity condition on noise)

We call the process ξ (p, R)-regular if $\xi_0(t) := \int_0^t \chi_0(s)\,\mathrm{d}s$, where $\chi_0 \colon [0, \tau] \times \Omega \to \mathcal{H}$, and there exists $p \geq 1$, $R \in \mathbb{N}$ and $C(R) \geq 1$ such that

$$\mathbb{E}\left[\|\xi_0(\tau)\|^r\right] \leq C(R)\tau^{r(2p+1)/2} \text{ for all } r \in \{1, \dots, R\}.$$

Remark 10

Setting $r = 1$ in the inequality above implies that, in the limit of small τ, the mean of $\xi_0(\tau)$ converges to zero. However, this does not imply that the mean of $\xi_0(\tau)$ itself must be zero.

Theorem 11 (Mean-square uniform convergence in discrete time)

Suppose that Assumptions 7 and 8 hold. Suppose that ξ is (p, R)-regular with parameters $R \geq 2q + 2$ and $p \geq 1$. If

(i) $q = 0$, *or*

(ii) $q \geq 1$ *and* $p \geq \max\left\{\frac{2}{q} + \frac{1}{2}, \frac{3}{2}\right\}$,

then, for τ small enough,

$$\mathbb{E}\left[\max_{k \leq T/\tau} \|u_k - U_k\|^2\right] \leq C\tau^2.$$

In particular, if the noise is (p, R)-regular with $R \geq 2q + 2$ and some $p \geq 5/2$, then the conclusion holds.

Before we state the continuous-time convergence result, we state the following modification of Assumption 9:

Assumption 12 (Strong (p,R)-regularity condition on noise)

We call the process ξ strongly (p,R)-regular if $\xi_0(t) := \int_0^t \chi_0(s)\,\mathrm{d}s$, where $\chi_0\colon [0,\tau]\times\Omega \to \mathcal{H}$, and there exists a $p \geq 1$ and $R \in \mathbb{N}$ such that

$$\mathbb{E}\left[\sup_{t\in[0,\tau]} \|\xi_k(t)\|^r\right] \leq C(R)\tau^{r(2p+1)/2} \text{ for all } r \in \{1,\dots,R\}.$$

Theorem 13 (Mean-square uniform convergence in continuous time)

Suppose that Assumptions 7 and 8 hold. Suppose that ξ is strongly (p, R)-regular with parameters $R \in \mathbb{N}$ and $p \geq 1$. If

(i) $q = 0$ *and* $R \geq 2$, *or*

(ii) $q \geq 1$, $R \geq 2q + 2$, *and* $p \geq \max\left\{\frac{2}{q} + \frac{1}{2}, \frac{3}{2}\right\}$,

then, for τ *small enough,*

$$\mathbb{E}\left[\sup_{0 \leq t \leq T} \|u(t) - U(t)\|^2\right] \leq C\tau^2.$$

In particular, if ξ is strongly (p, R)-regular with $R \geq 2q + 2$ and some $p \geq 5/2$, then the conclusion holds.

Sketch proof

- Establish uniform bounds on the numerical solution up to time T:

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

- Use (p,R)-regularity to control the moments of these sums of norms. Hence show that the maximum of the numerical solution has finite MGF.
- Bound the (differences of) displacements $V(x,y) := \left(\Phi^\tau(x)-x\right) - \left(\Psi^\tau(y)-y\right)$ as

$$|\langle x-y, V(x,y)\rangle| \leq 2\tau D\left(1+\|x\|^q+\|y\|^q\right)\|x-y\|^2 + C_4(x,y)\tau^3,$$
$$\|V(x,y)\|^2 \leq 2\tau^2 D\left(1+\|x\|^q+\|y\|^q\right)\|x-y\|^2 + 2C_4(x,y)\tau^4.$$

and, for $V_{k-1} := V(u_{k-1}, U_{k-1})$,

$$\mathbb{E}\left[|\langle e_{k-1}, V_{k-1}\rangle|\right] \leq C'(k-1,q)\left(\tau\mathbb{E}\left[\|e_{k-1}\|^2\right]+\tau^3\right)$$
$$\mathbb{E}\left[\|V_{k-1}\|^2\right] \leq C'(k-1,q)\left(\tau^2\mathbb{E}\left[\|e_{k-1}\|^2\right]+\tau^4\right).$$

- Use these to build even more moment bounds, and finish with a Grönwall argument.

Lemma 14 (Integrated Brownian motion satisfies regularity condition)

Let $\tau > 0$ be fixed, $0 \leq t \leq \tau$, $p \geq 1$ be arbitrary, and $(B_t)_t$ be $\mathcal{H}$-valued Brownian motion. Then $\xi_0(t) := \tau^{p-1} \int_0^t B_s \,\mathrm{d}s$ satisfies

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

Closing Remarks

- Randomised integrators for ODEs really do enjoy the same convergence rate as the underlying time-stepper, provided that the noise is not 'too large', in a very strong sense, under weakened regularity assumptions.
- When we use these integrators as forward solvers for Bayesian inverse problems, we expect the forward convergence rate to transfer to the BIP (Stuart, 2010).
- Indeed, one could analyse the impact of solver accuracy on the BIP in term of Bayes factors; Capistrán et al. (2016) observe the same convergence rate.
- Although the presentation here looks like an argument about randomly perturbed dynamical systems, the results can be read as saying something about any path measure with well-understood mean.

- Mathematical aesthetics: can the proofs for locally Lipschitz flows be shortened?
- From constant time step to adaptive time steps?
- Active calibration of the noise structure at runtime?
    - **Oksana Chkrebtii** Wed 09:00
    - **Hans Kersting** Fri 11:30

Thank You

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.

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.*, 2016. doi:10.1007/s11222-016-9671-0.

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.

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.

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.

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.

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.

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.

G. Peškir. On the exponential Orlicz norms of stopped Brownian motion. *Studia Math.*, 117(3):253–273, 1996. `http://eudml.org/doc/216255`.

S. Reich and C. Cotter. *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge University Press, New York, 2015. doi:10.1017/CBO9781107706804.

Y.-F. Ren. On the Burkholder–Davis–Gundy inequalities for continuous martingales. *Statist. Probab. Lett.*, 78(17):3034–3039, 2008. doi:10.1016/j.spl.2008.05.024.

A. M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19:451–559, 2010. doi:10.1017/S0962492910000061.