Bayesian Analysis (2016) **11**, Number 4, pp. 1269–1273

Comment on Article by Chkrebtii, Campbell, Calderhead, and Girolami*

Martin Lysy[†,‡]

Abstract. The authors present an ingenious probabilistic numerical solver for deterministic differential equations (DEs). The true solution is progressively identified via model interrogations, in a formal framework of Bayesian updating. I have attempted to extend the authors' ideas to stochastic differential equations (SDEs), and discuss two challenges encountered in this endeavor: (i) the non-differentiability of SDE sample paths, and (ii) the sampling of diffusion bridges, typically required of solutions to the SDE inverse problem.

Keywords: stochastic differential equations, probabilistic solution, diffusion bridge sampling.

1 Motivation

A stochastic differential equation (SDE) for a one-dimensional continuous Markov process $X_t = X(t)$ is written as

$$\mathrm{d}X_t = \mu(X_t, \boldsymbol{\theta})\,\mathrm{d}t + \sigma(X_t, \boldsymbol{\theta})\,\mathrm{d}B_t, \qquad X_0 = x_0, \tag{1}$$

where B_t is Brownian motion. Much like for deterministic differential equations (DEs), the transition density $p(X_{t+s} \,|\, X_t, \boldsymbol{\theta})$ for the SDE (1) is rarely available in closed form.

The SDE forward problem consists of sampling a path X_t solving (1) on the interval $t \in [0, T]$. Perhaps the simplest approach to this is with an Euler-type scheme,

$$\Delta X_t \approx \mu(X_t, \boldsymbol{\theta})\Delta t + \sigma(X_t, \boldsymbol{\theta})\Delta B_t, \tag{2}$$

where $\Delta X_t = X_{t+\Delta t} - X_t$ and $\Delta B_t = B_{t+\Delta t} - B_t \sim \mathcal{N}(0, \Delta t)$. The Euler approximation (2) becomes arbitrarily accurate as $\Delta t \to 0$.

In a Bayesian setting, the SDE inverse problem consists of sampling from the posterior distribution $p(\boldsymbol{\theta} \,|\, \boldsymbol{X}) \propto p(\boldsymbol{X} \,|\, \boldsymbol{\theta})\pi(\boldsymbol{\theta})$ of the unknown parameters $\boldsymbol{\theta}$, given discrete observations $\boldsymbol{X} = \big(X(t_0), \ldots, X(t_n)\big)$. It is commonly approached by way of data augmentation (Section 3).

In Chkrebtii et al. (2016), referred to hereafter as UQDE, the authors present an ingenious probabilistic method for solving deterministic DEs, which I have attempted to apply to the SDE inverse problem. In the following I describe two challenges encountered along the way:

1. Non-differentiability of SDE sample paths (Section 2).

*Main article DOI: 10.1214/16-BA1017.

†Department of Statistics and Actuarial Science, University of Waterloo, mlysy@uwaterloo.ca

‡Research supported by NSERC-DG grant RGPIN-2014-04225.

 DOI: 10.1214/16-BA1036

2. Solving the SDE forward problem conditioned on both endpoints: $X_0 = x_0$ and $X_T = x_T$ (Section 3). This is called *diffusion bridge sampling*, features prominently in most numerical approaches to the SDE inverse problem, and is a notoriously difficult endeavor (e.g., Bladt et al., 2016, and the references therein).

2 Non-Differentiability of SDE Sample Paths

Consider the SDE forward problem of generating a probabilistic solution X_t to (1) on the interval $t \in [0, T]$. Since the Brownian motion B_t can be pre-generated, in some sense the problem reduces to finding a deterministic solution to (1) for a fixed path $\boldsymbol{B_t} = \{B_t : t \in [0, T]\}$. One then could apply directly the methodology of the authors – if X_t were differentiable in the traditional sense. As it is not, instead I have tried to work with the increment process $\Delta X_t = X_{t+\Delta t} - X_t$.

That is, suppose we wish to evaluate a given solution to the SDE forward problem at times $t_0, \ldots, t_N$, with $t_i = i\Delta t$ and $\Delta t = T/N$. An adaptation of the authors' approach is presented in Algorithm 1, using the following notation: $X_i = X(i\Delta t)$, $\Delta X_i = X_{i+1} - X_i$, $\boldsymbol{X} = (X_1, \ldots, X_N)$, $\Delta\boldsymbol{X} = (\Delta X_0, \ldots, \Delta X_{N-1})$, $\boldsymbol{X_t} = \big\{X_t : t \in [0, T]\big\}$, and similarly for B_i, ΔB_i, $\boldsymbol{B}$, $\Delta\boldsymbol{B}$, and $\boldsymbol{B_t}$. For the model interrogations $\mathrm{f}_1, \ldots, \mathrm{f}_N$, let $\mathbf{f}_i = (\mathrm{f}_1, \ldots, \mathrm{f}_i)$.

Algorithm 1 Probabilistic solution to the SDE forward problem (1).

Input. The initial value $X_0 = x_0$, and pre-generated increments $\Delta B_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \Delta t)$.

Prior. A natural candidate is $\pi_0(\boldsymbol{X_t}) \iff \boldsymbol{X_t} \overset{\mathcal{L}}{=} \boldsymbol{B_t}$, which captures both the Markov property and correct "smoothness" of the true solution.

First Update. The model interrogation is $\mathrm{f}_1 = \mu(x_0, \boldsymbol{\theta})\Delta t + \sigma(x_0, \boldsymbol{\theta})\Delta B_0$, which corresponds to the Euler step (2) for ΔX_0. Thus, the likelihood $p(\mathrm{f}_1 \,|\, \boldsymbol{X_t})$ is defined by the density $\mathrm{f}_1 \,|\, \boldsymbol{X_t} \sim \mathcal{N}(\Delta X_0, \Delta t)$, noting that Δt is the predictive variance for ΔX_0 under π_0. This density together with the prior $\pi_0(\boldsymbol{X_t})$ leads to the posterior distribution

$$\pi_1(\boldsymbol{X_t}) = p(\boldsymbol{X_t} \,|\, \mathrm{f}_1) \propto p(\mathrm{f}_1 \,|\, \boldsymbol{X_t})\pi_0(\boldsymbol{X_t}).$$

Subsequent Updates. Given $\mathbf{f}_i$ and $\pi_i(\boldsymbol{X_t}) = p(\boldsymbol{X_t} \,|\, \mathbf{f}_i)$, sample $x_i \sim \pi_i(X_i)$, i.e., from the prior on $X(t_i)$, and set $\mathrm{f}_{i+1} = \mu(x_i, \boldsymbol{\theta})\Delta t + \sigma(x_i, \boldsymbol{\theta})\Delta B_i$. The likelihood contribution $p(\mathrm{f}_{i+1} \,|\, \mathbf{f}_i, \boldsymbol{X_t})$ is defined by $\mathrm{f}_{i+1} \,|\, \mathbf{f}_i, \boldsymbol{X_t} \sim \mathcal{N}(\Delta X_i, \Delta t)$, where Δt is the predictive variance of ΔX_i under π_i. This density together with the prior $\pi_i(\boldsymbol{X_t})$ leads to the posterior distribution

$$\pi_{i+1}(\boldsymbol{X_t}) = p(\boldsymbol{X_t} \,|\, \mathbf{f}_{i+1}) \propto p(\mathrm{f}_{i+1} \,|\, \mathbf{f}_i, \boldsymbol{X_t})\pi_i(\boldsymbol{X_t}).$$

Output. A realization of the solution at discrete time points, $\boldsymbol{X} \sim \pi_N(\boldsymbol{X}) = p(\boldsymbol{X} \,|\, \mathbf{f}_N)$.

The final step of Algorithm 1 can be obtained by drawing $\Delta\boldsymbol{X} \sim \pi_N(\Delta\boldsymbol{X})$, which is $\Delta X_i \overset{\text{ind}}{\sim} \mathcal{N}\big(\tfrac{1}{2}\mathrm{f}_{i+1}, \tfrac{1}{2}\Delta t\big)$. Note that this draw is conditional on the pre-generated Brownian increments, $\Delta\boldsymbol{B}$. Unfortunately, upon integrating over $\Delta\boldsymbol{B}$, Algorithm 1 produces

$$\Delta X_0 \sim \mathcal{N}\Big(\tfrac{1}{2}\mu(x_0,\boldsymbol{\theta})\Delta t, \big(\tfrac{1}{2}+\tfrac{1}{4}\sigma^2(x_0,\boldsymbol{\theta})\big)\Delta t\Big),$$

which for $X_0 = x_0$, does not give the correct diffusion limits

$$\lim_{\Delta t\to 0}\frac{E[X_{\Delta t}-X_0]}{\Delta t} = \mu(x_0,\boldsymbol{\theta}), \qquad \lim_{\Delta t\to 0}\frac{\mathrm{var}(X_{\Delta t}-X_0)}{\Delta t} = \sigma^2(x_0,\boldsymbol{\theta}).$$

The failure of Algorithm 1 is due perhaps to a poor choice of likelihood $p(\mathrm{f}_{i+1}\,|\,\mathbf{f}_i, \boldsymbol{X_t})$, or to the degenerate relation between X_t and its increment process ΔX_t. However, these issues can be circumvented if one is willing to replace B_t by a continuously differentiable process G_t. In this case, (1) becomes

$$\dot{X}_t = \mu(X_t,\boldsymbol{\theta}) + \sigma(X_t,\boldsymbol{\theta})\dot{G}_t, \tag{3}$$

where $\dot{X}_t = \frac{\mathrm{d}}{\mathrm{d}t}X_t$ and $\dot{G}_t = \frac{\mathrm{d}}{\mathrm{d}t}G_t$. Then, if $\dot{G}_t$ is pre-generated at appropriate locations, (3) reduces to a differentiable deterministic problem, and the authors' method of solution can be applied directly.

One possibility for G_t in (3) is the integrated Ornstein–Uhlenbeck (iOU) process, a stationary Gaussian process with $\mathrm{cov}(\dot{G}_0,\dot{G}_t) = e^{-\lambda|t|}$. As $\lambda\to\infty$, one recovers the original SDE (1), with the caveat that the stochastic integral be defined in the Stratonovich sense (e.g., Van Kampen, 1981). Note that this is also the sense in which (1) can be interpreted pathwise for any continuous G_t, including Brownian motion (Lysy and Pillai, 2013).

3 Diffusion Bridge Sampling

In light of the above, let us consider the simplified problem of drawing a path X_t, $t\in[0,T]$, from a differentiable SDE with "additive" noise:

$$\dot{X}_t = \mu(X_t,\boldsymbol{\theta}) + \sigma(\boldsymbol{\theta})\dot{G}_t, \qquad X_0 = x_0, \quad X_T = x_T, \tag{4}$$

where G_t is an iOU process. The connection between (4) and the SDE inverse problem on $p(\boldsymbol{\theta}\,|\,X_0,X_T)$ owes to the following method of data augmentation.

Carrying forward the notation of Section 2 ($\Delta t = T/N$, $X_i = X(i\Delta t)$, etc.), let $f(\Delta\boldsymbol{G}\,|\,\lambda)$ denote the joint density of the iOU increments, $\Delta\boldsymbol{G} = (\Delta G_0,\ldots,\Delta G_{N-1})$. Then by continuous differentiability of X_t and G_t, the approximate joint density

$$\hat{p}(\boldsymbol{X}\mid X_0,\boldsymbol{\theta}) = f(\Delta\widehat{\boldsymbol{G}}\,|\,\lambda)/\sigma(\boldsymbol{\theta})^N, \qquad \Delta\widehat{G}_i = \big(\Delta X_i - \mu(X_i,\boldsymbol{\theta})\Delta t\big)/\sigma(\boldsymbol{\theta}),$$

converges to the true density $p(\boldsymbol{X}\mid X_0,\boldsymbol{\theta})$ as $N\to\infty$. One then can approach the SDE inverse problem by sampling from the "complete data" posterior distribution

$$\hat{p}(\boldsymbol{X}_{\mathrm{miss}},\boldsymbol{\theta}\,|\,X_0,X_T) \propto \hat{p}(\boldsymbol{X}\,|\,X_0,\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \qquad \boldsymbol{X}_{\mathrm{miss}} = (X_1,\ldots,X_{N-1}), \tag{5}$$

thereby indirectly obtaining samples from the desired parameter distribution

$$\hat{p}(\boldsymbol{\theta}\,|\,X_0,X_T) = \int \hat{p}(\boldsymbol{X}_{\mathrm{miss}},\boldsymbol{\theta}\,|\,X_0,X_T)\,\mathrm{d}\boldsymbol{X}_{\mathrm{miss}} \xrightarrow{N} p(\boldsymbol{\theta}\,|\,X_0,X_T).$$

Sampling from the complete data posterior (5) typically requires Markov chain Monte Carlo (MCMC) techniques. A natural candidate for this is the Gibbs sampler which alternately draws from $\hat{p}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\text{miss}}, X_0, X_T)$ and $\hat{p}(\boldsymbol{X}_{\text{miss}} \mid \boldsymbol{\theta}, X_0, X_T)$. Efficient exploration of the latter of these densities relies heavily on bridge sampling. However, the authors' ideas could lead to the construction of *joint* MCMC proposals for $(\boldsymbol{\theta}, \boldsymbol{X}_{\text{miss}})$, as outlined in UQDE, Section 4.1. That is, if $(\boldsymbol{\theta}, \boldsymbol{X}_{\text{miss}})$ is the current value of the sampler, draw $\boldsymbol{\theta}' \sim q(\boldsymbol{\theta}' \mid \boldsymbol{\theta})$ from some proposal distribution, and $\boldsymbol{X}'_{\text{miss}} \sim p(\boldsymbol{X}'_{\text{miss}} \mid \boldsymbol{\theta})$ from the probabilistic solver. The new proposal $(\boldsymbol{\theta}', \boldsymbol{X}'_{\text{miss}})$ is then accepted with probability

$$\rho = \min\left\{ \frac{\hat{p}(\boldsymbol{X}' \mid X_0, \boldsymbol{\theta}') \cdot \pi(\boldsymbol{\theta}')}{\hat{p}(\boldsymbol{X} \mid X_0, \boldsymbol{\theta}) \cdot \pi(\boldsymbol{\theta})} \times \frac{q(\boldsymbol{\theta} \mid \boldsymbol{\theta}')}{q(\boldsymbol{\theta}' \mid \boldsymbol{\theta})}, 1 \right\}.$$

This approach circumvents the Gibbs sampler's well-known arbitrary mixing-time degeneration as $\Delta t \to 0$ (Roberts and Stramer, 2001). This deficiency is currently addressed in the SDE literature using sophisticated particle filtering techniques (e.g., Andrieu et al., 2010), to which the authors' "Metropolis-free" probabilistic solver, if applicable to bridge sampling, stands to offer an attractive alternative.

An adaptation of the authors' methodology to SDE bridge sampling might proceed as follows.

Algorithm 2 Probabilistic solution to the bridge sampling problem (4).

Input. The fixed model parameters $\boldsymbol{\theta}$ and endpoints $X_0 = x_0$ and $X_T = x_T$.

Prior. Define $\pi_0(\boldsymbol{X_t}, \boldsymbol{G_t})$ such that X_t and G_t are independent iOU processes, subject to $(G_0, X_0, X_T) = (0, x_0, x_T)$. Note that this implies a joint Gaussian process distribution between (X_t, G_t) and $A_t = X_t - \sigma(\boldsymbol{\theta})G_t$, along with their (continuous) derivatives.

First Update. Draw $x_1 \sim \pi_0(X_1)$, i.e., from the prior on $X(t_1)$, and let $\mathrm{f}_1 = \mu(x_1, \boldsymbol{\theta})$. To define a likelihood for this first model interrogation, note that $\dot{A}_t = \mu(X_t, \boldsymbol{\theta})$, such that $p(\mathrm{f}_1 \mid \boldsymbol{X_t}, \boldsymbol{G_t})$ is defined via $\mathrm{f}_1 \mid \boldsymbol{X_t}, \boldsymbol{G_t} \sim \mathcal{N}(\dot{A}_1, \mathrm{v}_0)$, where v_0 is the predictive variance of $\dot{A}_1 = \dot{A}(t_1)$ under π_0. This leads to the posterior distribution

$$\pi_1(\boldsymbol{X_t}, \boldsymbol{G_t}) = p(\boldsymbol{X_t}, \boldsymbol{G_t} \mid \mathrm{f}_1) \propto p(\mathrm{f}_1 \mid \boldsymbol{X_t}, \boldsymbol{G_t}) \pi_0(\boldsymbol{X_t}, \boldsymbol{G_t}).$$

Subsequent Updates. Given $\mathbf{f}_i$ and $\pi_i(\boldsymbol{X_t}) = p(\boldsymbol{X_t} \mid \mathbf{f}_i)$, sample $x_{i+1} \sim \pi_i(X_{i+1})$, i.e., from the prior on $X(t_{i+1})$, and let $\mathrm{f}_{i+1} = \mu(x_{i+1}, \boldsymbol{\theta})$. Define the likelihood contribution $p(\mathrm{f}_{i+1} \mid \mathbf{f}_i, \boldsymbol{X_t}, \boldsymbol{G_t})$ via $\mathrm{f}_{i+1} \mid \mathbf{f}_i, \boldsymbol{X_t}, \boldsymbol{G_t} \sim \mathcal{N}(\dot{A}_{i+1}, \mathrm{v}_i)$, where v_i is the predictive variance of $\dot{A}_{i+1} = \dot{A}(t_{i+1})$ under π_i. This leads to the posterior distribution

$$\pi_{i+1}(\boldsymbol{X_t}, \boldsymbol{G_t}) = p(\boldsymbol{X_t}, \boldsymbol{G_t} \mid \mathbf{f}_{i+1}) \propto p(\mathrm{f}_{i+1} \mid \mathbf{f}_i, \boldsymbol{X_t}, \boldsymbol{G_t}) \pi_i(\boldsymbol{X_t}, \boldsymbol{G_t}).$$

Output. A discretized realization of the SDE bridge, $\boldsymbol{X} \sim \pi_N(\boldsymbol{X}) = p(\boldsymbol{X} \mid \mathbf{f}_N)$.

Convergence of Algorithm 2 to the bridge process (4) once again can be related to a deterministic problem. That is, if $\boldsymbol{G_t}$ is treated as a fixed but unknown path – rather than an iOU process – then (4) becomes a deterministic DE with uncountably many solutions. The problem of solution multiplicity is considered by the authors in

UQDE, Section 5.2. Their approach is to put a prior on a set of initial conditions $\mathbf{i_0}$ which uniquely identify a candidate DE solution, but for which only the candidate DEs corresponding to a restricted set $\mathbf{i_0} \in \boldsymbol{\mathcal{I}}$ are actually valid. In connection with the bridge sampling problem, I am eager to solicit the authors' opinion on the following proposition:

Proposition 1. *Consider the authors' probabilistic solver for a deterministic DE, with solution multiplicity and prior on identifying conditions* $\mathbf{i_0} \sim \pi_0(\mathbf{i_0})$*. Then as* $N \to \infty$*, the posterior on* $\mathbf{i_0}$ *converges to the restriction of* $\pi_0(\mathbf{i_0})$ *to valid solutions:*

$$\int p(\mathbf{i_0} \,|\, \mathbf{f}_N) p(\mathbf{f}_N)\, \mathrm{d}\mathbf{f}_N \to \pi_0(\mathbf{i_0} \,|\, \mathbf{i_0} \in \boldsymbol{\mathcal{I}}).$$

If Proposition 1 should hold, then correctness of Algorithm 2 follows upon setting $\mathbf{i_0} = \boldsymbol{G_t}$. Contingent on this, Algorithm 2 perhaps can be developed to accommodate the more complex SDE models used in practice: without the conspicuous restriction to additive noise ($\sigma(X_t, \boldsymbol{\theta}) = \sigma(\boldsymbol{\theta})$), and in a multivariate, latent variable setting.

References

Andrieu, C., Doucet, A., and Holenstein, R. (2010). "Particle Markov chain Monte Carlo Methods." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72: 1–33. MR2758115. doi: http://dx.doi.org/10.1111/j.1467-9868.2009.00736.x. 1272

Bladt, M., Finch, S., and Sørensen, M. (2016). "Simulation of multivariate diffusion bridges." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(2): 343–369. MR3454200. doi: http://dx.doi.org/10.1111/rssb.12118. 1270

Chkrebtii, O. A., Campbell, D. A., Calderhead, B., and Girolami, M. A. (2016). "Bayesian solution uncertainty quantification for differential equations." *Bayesian Analysis*. doi: http://dx.doi.org/10.1214/16-BA1017. 1269

Lysy, M. and Pillai, N. S. (2013). "Statistical Inference for Stochastic Differential Equations with Memory." Technical report, University of Waterloo. 1271

Roberts, G. O. and Stramer, O. (2001). "On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm." *Biometrika*, 88(3): 603–621. MR1859397. doi: http://dx.doi.org/10.1093/biomet/88.3.603. 1272

Van Kampen, N. G. (1981). "Itô versus Stratonovich." *Journal of Statistical Physics*, 24(1): 175–187. MR0601694. doi: http://dx.doi.org/10.1007/BF01007642. 1271