Conditional Path Sampling of SDEs and Data Assimilation

Andrew Stuart,
Mathematics Institute,
University of Warwick.
http://www.maths.warwick.ac.uk/~stuart/

collaboration with
Martin Hairer (Warwick)and Jochen Voss (Warwick)
& Gareth Roberts (Lancaster) and Petter Wiberg
(Goldman-Sachs).

Data Assimilation Workshop,
IPAM, UCLA
February 21st–25th, 2005

Overview

- Lagrangian Data Assimilation
- Sampling and The Langevin Method
- Infinite Dimensional Sampling and SPDEs
- Theoretical Background
- Simulations
- Optimal Algorithms
- Conclusions

Warning

$\boxed{\text{u is time!}}$

$\boxed{\text{t is algorithmic time}}$

The Real Problem

We are **given** Lagrangian particle tracers y_i subject to molecular diffusion:

$$\boxed{\frac{dy_i}{du} = V(y_i, u) + \sigma \frac{dB}{du}.}$$

We want to **find** information about $V(y, u)$. For example to find the Fourier co-efficients $x_i^{(k)}(u)$ in an expression

$$\boxed{V(y, u) = \sum_k \{\sin(y) x_1^{(k)}(u) + \cos(y) x_2^{(k)}(u)\}.}$$

A Model Problem

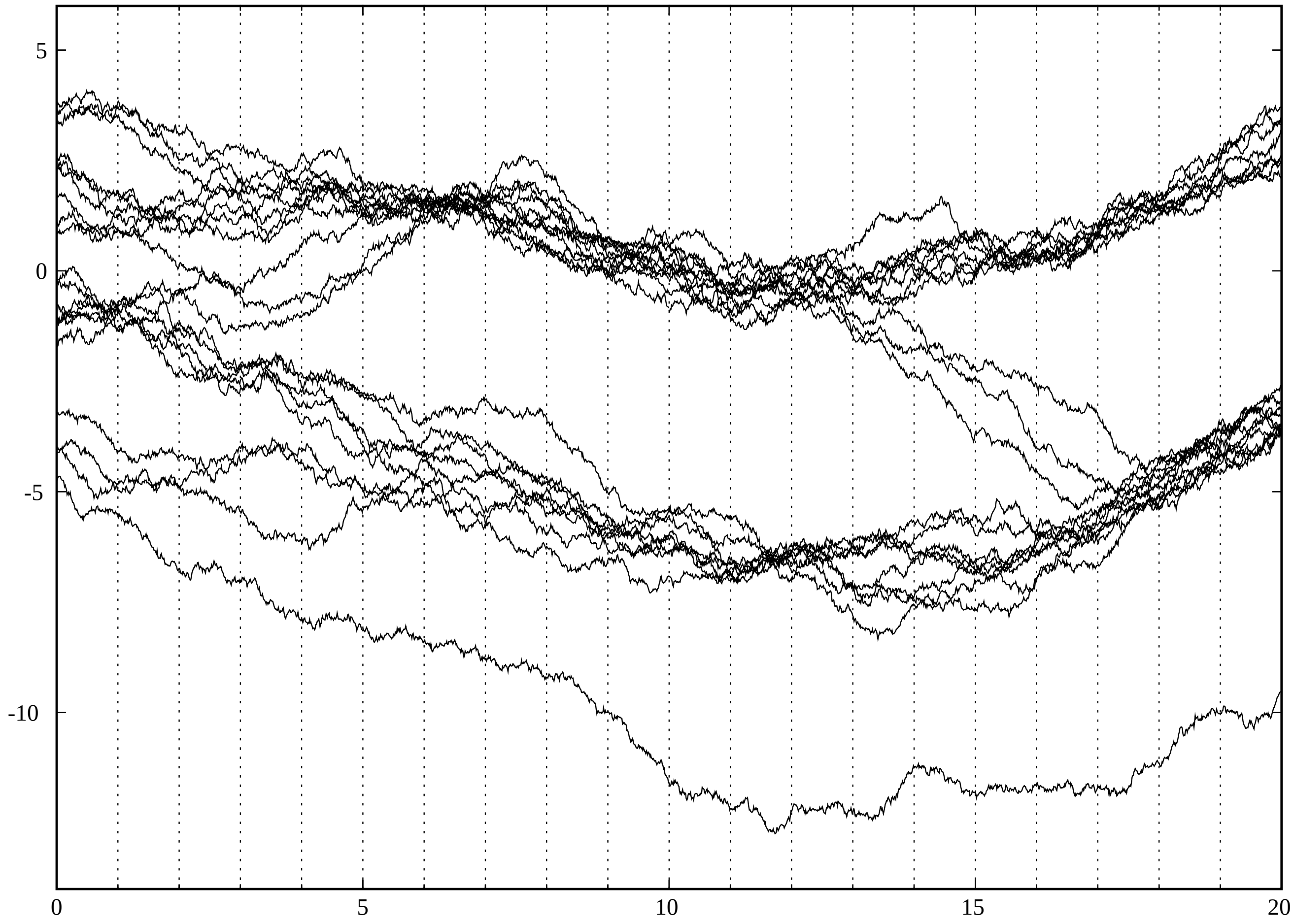
We are **given** the 20 paths of particle tracers y_i solving

$$\boxed{\frac{dy_i}{du} = x_1 + \sin(y)x_2 + \cos(y)x_3 + \sigma\frac{dB_i}{du}.}$$

We want to **find** the distribution of x_1, x_2 which we assume solve

$$\begin{aligned}\frac{dx_1}{dt} &= -\alpha x_1 + \sqrt{\lambda}\frac{dW_1}{dt},\\ \frac{dx_2}{dt} &= -\alpha x_2 + \sqrt{\lambda}\frac{dW_2}{dt},\\ \frac{dx_3}{dt} &= -\alpha x_3 + \sqrt{\lambda}\frac{dW_3}{dt}.\end{aligned}$$

In demo we have $\sigma = 0.4, \lambda = 0.5, \alpha = 0.2$.



The twenty particles tracers $y_i(u)$.

Abstraction

$\boxed{\text{To sample from paths } \{x(u)\}_{u\in[0,1]}\text{, conditional on some observations}}$

- This is an infinite dimensional sampling problem.
- Metropolis adjusted Langevin algorithms are known to be good sampling methods in high dimensions. [**Roberts et al** 1990s].
- We seek to generalize this methodology.

The Langevin SDE

Assume that we know $q : \mathbb{R}^N \to \mathbb{R}$ where $\rho(x) = Cq(x)$, and $\rho(x)$ is a pdf from which we wish to sample. The basic idea of the Langevin algorithm is to generate paths of the SDE

$$\boxed{\frac{dx}{dt} = \nabla \log q(x) + \sqrt{2}\frac{dW}{dt}}.$$

Provided the SDE is ergodic (a condition on the tails of q):

$$\boxed{\frac{1}{T}\int_0^T \phi(x(t))dt \to \int_{\mathbb{R}^N} \phi(x)\rho(x)dx \;\; \text{as} \;\; T \to \infty.}$$

The Langevin MCMC Method

The SDE is the basis for an MCMC method in which, given x_n, the proposal move is

$$\boxed{x^* = x^n + \Delta t \nabla \log q(x^n) + \sqrt{\{2\Delta t\}}\mathcal{N}(0, I)}$$

(or some other discretization of the SDE) and then

$$x^{n+1} = \left\{ \begin{array}{ccc} x^* & w.p. & p^* \\ x^n & w.p. & 1 - p^* \end{array} \right\}$$

and p^* is the Metropolis-Hastings acceptance probability.

We generalize these ideas to situations where the distribution to be sampled is infinite dimensional.

Infinite Dimensional Applications

We study sampling of paths of SDEs, conditional on observations. Applications include:

- Signal Processing;
- Data Assimilation;
- Transition Path Sampling;
- Interpolating Discrete Time Data by SDEs (for model identification);

Bridge Path Sampling

In some applications (econometrics, transition path sampling) it is important to be able to generate paths of

$$\frac{dx}{du} = -\nabla F(x) + \gamma \frac{dB}{du}$$

subject to

$$x(0) = X^- \quad \& \quad x(1) = X^+.$$

Note that $x(u; \{W\})$ and that the observation of $x(1; \{W\})$ conditions the random variable W, and hence x.

Bridge Path Sampling

By generalizing the Langevin method we obtain the following SPDE for $x(u,t)$:

$$\frac{\partial x}{\partial t} = \frac{1}{\gamma^2}\{\frac{\partial^2 x}{\partial u^2} - \nabla \mathcal{F}(x)\} + \sqrt{2}\frac{\partial W}{\partial t},$$

$$x = X^-, \quad u = 0,$$

$$x = X^+, \quad u = 1,$$

$$x = x_0, \quad t = 0.$$

Here

$$\mathcal{F}(x) = \frac{1}{2}|\nabla F|^2 - \frac{\gamma^2}{2}\Delta F(x).$$

and $\frac{\partial W}{\partial t}$ is space time white noise. [**Stuart, Voss and Wiberg, 2004**, **Reznikoff and Vanden-Eijnden, 2005**].

Nonlinear Filter/Smoother

In some applications (signal processing, data assimilation) it is important to be able to generate paths of

$$\frac{dx}{du} = -\nabla F(x) + \gamma \frac{dB_1}{du}, \quad X(0) \sim \mathcal{N}(a, \delta^2)$$

subject to observation of y solving

$$\frac{dy}{du} = Ax + \sigma \frac{dB_2}{du}, \quad Y(0) = 0.$$

That is, to sample from the *distribution* of

$$x(u)|\{y(s)\}_{0 \le s \le T}, \quad 0 \le u \le T.$$

Note that $x(u; \omega, \{B_1\})$ and $y(u; \omega, \{B_1\}, \{B_2\})$ and that observation of y conditions the random variable $(\omega, \{B_1\})$, and hence x.

Nonlinear Filter/Smoother

From the Langevin method we obtain the following SPDE (after time-rescaling) for $x(u,t)$:

$$
\begin{aligned}
\frac{\partial x}{\partial t} &= \epsilon^2\{\frac{\partial^2 x}{\partial u^2} - \nabla\mathcal{F}(x)\} + A^T\{\frac{dy}{du} - Ax\} + \sqrt{2\sigma^2}\frac{\partial W}{\partial t},\\
\frac{\partial x}{\partial u} &= -\nabla F(x) + \frac{\gamma^2}{\delta^2}(x-a), \quad u = 0,\\
\frac{\partial x}{\partial u} &= -\nabla F(x), \quad u = 1,\\
x &= x_0, \quad s = 0.
\end{aligned}
$$

Here $\epsilon = \sigma/\gamma$, $\mathcal{F}$ as for bridge sampling and $\frac{\partial W}{\partial t}$ is space-time white noise.

The SPDEs as SDEs in Hilbert Space

In the Gaussian case (quadratic F) the SPDEs for sampling can be written as Hilbert space $\mathcal{H}$ valued SDEs of the form

$$\frac{dx}{dt} = \mathcal{L}x + h + \sqrt{2}\frac{dW}{dt} \tag{1}$$

and nonlinear problems (non-quadratic F) can be written as

$$\frac{dx}{dt} = \mathcal{L}x + h + U'(x) + \sqrt{2}\frac{dW}{dt}. \tag{2}$$

Ergodicity and Invariant Measures

- For Gaussian processes we need only check that $m(u) = -\mathcal{L}^{-1}h$ is the mean and that the covariance function $C(u, v)$ is the Green's function for $-\mathcal{L}$. [**Hairer, Stuart, Voss and Wiberg** 2005].

- The Gaussian process (1) is then ergodic and has invariant measure $M(dx)$ in $\mathcal{H}$.

- Under conditions on $U(x)$, equation (2) is ergodic with invariant measure $m(dx) = \exp\{-U(x)\}M(dx)$. [**Zabcyk (1988)**].

- This can be used to verify the sampling properties for nonlinear bridges [**Reznikoff and Vanden Eijnden** 2005], [**Hairer, Stuart and Voss** 2005]. and for nonlinear filters, [**Hairer, Stuart and Voss** 2005].

Bridge Path Sampling

- $f(x) = -F'(x)$
- $F(x) = \frac{(x^2-1)^2}{x^2+1}$
- $\gamma = 1, \quad T = 10^2$
- $X^- = -1, \quad X^+ = 1.$

Red is sample, green is mean (through time-averaging), blue is variance (through time-averaging).

Nonlinear Filter/Smoother

- $f(x) = -F'(x)$
- $F(x) = \frac{(x^2-1)^2}{x^2+1}$
- $\gamma = \sigma = 1, \quad T = 10^2$
- $X^- = -1$, $(a = -1, \delta = 0)$

Red is sample, blue is time average (mean), green is (unobserved) actual path.

Preconditioning

Recall equation (2):

$$\frac{dx}{dt} = \mathcal{L}x + h + U'(x) + \sqrt{2}\frac{dW}{dt}.$$

The invariant measure of is this equation is unchanged by introducing compact postive operator $\mathcal{G} : \mathcal{H} \to \mathcal{H}$ and considering

$$\frac{dx}{dt} = \mathcal{G}\mathcal{L}x + \mathcal{G}h + \mathcal{G}U'(x) + \sqrt{2\mathcal{G}}\frac{dW}{dt}. \tag{3}$$

This leads to some interesting new evolution equations. Optimizing the choice of $\mathcal{G}$ can lead to greater efficiency when Metropolizing.

Based on finite dimensional considerations, it is natural in the context of Metropolizing to choose $\mathcal{G}$ to be a Green's operator proportional to $-\mathcal{L}^{-1}$. We illustrate this for bridge paths.

Preconditioning for Bridge Paths

$$\frac{\partial x}{\partial t} = \frac{1}{\gamma^2}\{-x+y\} + \sqrt{2\mathcal{G}}\frac{\partial W}{\partial t}$$

$$\frac{\partial^2 y}{\partial u^2} = \nabla \mathcal{F}(x)$$

$$y = X^-, \quad u = 0,$$

$$y = X^+, \quad u = 100,$$

$$x = x_0, \quad t = 0.$$

- $f(x) = -F'(x)$
- $F(x) = \frac{(x^2-1)^2}{x^2+1}$
- $\gamma = 1, X^- = -1, \quad X^+ = 1.$

Red is sample, green is mean, blue is variance.

Future Directions

These include:

- continuing to develop a rigorous theory for the sampling properties and ergodicity of the SPDEs described here, and generalizations;
- optimizing pre-conditioning and choice of time-step to improve efficiency in the context of Metropoloizing;
- analysis of the rate of convergence of the SPDEs derived here;
- applications in signal processing, data assimilation and econometrics;
- evaluation of methods introduced here in comparison with other recently introduced methods.