Conditional Path Sampling of SDEs and the Langevin MCMC Method

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

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

Future Challenges in
Multiscale Modeling and Simulation
The IMA, Minnesota,
Thursday, November 18th 2004
OVERVIEW

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

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

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

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

We generalize this idea to situations where the distribution to be sampled is infinite dimensional.
Bridge Path Sampling

In some applications 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 \( x(u, t) \):

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

\[
x = X^-, \quad u = 0,
\]

\[
x = X^+, \quad u = 1,
\]

\[
x = x_0, \quad t = 0.
\]

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. [Betz and Lorinczi, 2002, Stuart, Voss and Wiberg, 2004, Reznikoff and Vanden-Eijnden, 2004].
Nonlinear Filter/Smooother

In other applications 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

\[ x(t) | \{ y(s) \}_{0 \leq s \leq T}, \quad 0 \leq t \leq T. \]

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 $x(u, t)$:

$$
\frac{\partial x}{\partial t} = \epsilon^2 \left\{ \frac{\partial^2 x}{\partial u^2} - \nabla \mathcal{F}(x) \right\} + A^T \left\{ \frac{dy}{du} - Ax \right\} + \sqrt{2\sigma^2} \frac{\partial W}{\partial t},
$$

$$
\frac{\partial x}{\partial u} = -\nabla F(x) + \frac{\gamma^2}{\delta^2} (x - a), \quad u = 0,
$$

$$
\frac{\partial x}{\partial u} = -\nabla F(x), \quad u = 1,
$$

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

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

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 Gaussian Invariant Measures

- For Gaussian processes we need only check that $m(u) = -\mathcal{L}^{-1}h$ is the mean and that the covariance $C(u, v)$ is the Green’s function for $-\mathcal{L}$.

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

- Let $\mathcal{L}$ be the Laplacian. With Dirichlet boundary conditions $M$ is Brownian bridge measure. For Dirichlet (left) and Neumann (right) $M$ is Wiener measure. For Robin (left) we incorporate Gaussian noise in observation of the left end-point.

- This can be used to verify that we sample from the distribution whose mean is the Kalman-Bucy filter/smoother.
THEORETICAL BACKGROUND

Ergodicity and General Invariant Measures

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

• This can be used to verify the sampling properties for nonlinear bridges [Reznikoff and Vanden Eijnden 2004], [Hairer, Stuart, Voss and Wiberg 2004].

• It can also be used to verify the sampling properties for nonlinear filters by writing the measure with respect to the distribution of a Gaussian process whose mean is a Kalman-Bucy filter. See [Hairer, Stuart, Voss and Wiberg 2004].
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).
SIMULATIONS

Nonlinear Filter/Smooother

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

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


**OPTIMAL ALGORITHMS**

**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 positive operator \( \mathcal{G} : \mathcal{H} \to \mathcal{H} \) and considering

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

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

\[
\frac{\partial^2 y}{\partial u^2} = \nabla F(x)
\]

\[
y = X^-, \quad u = 0,
\]

\[
y = X^+, \quad u = 100,
\]

\[
x = x_0, \quad t = 0.
\]

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

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

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 and econometrics;

- evaluation of methods introduced here in comparison with other recently introduced methods (Chib/Pitt/Shepherd, Roberts/Stramer, Beskos/Roberts).