MCMC Methods on Path Space

Jochen Voss

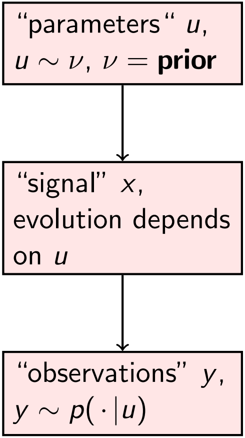
University of Leeds, Statistics Department

9th April 2009

Joint work with Martin Hairer and Andrew Stuart (University of Warwick)

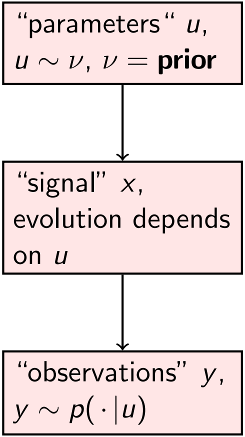
Outline

1. Bayesian Inference for Signal Processing



Many problems can be formulated in a Bayesian framework:

- signal processing/filtering (*e.g.* unknown parameters),
- data assimilation (*e.g.* unknown initial condition),
- the oil-reservoir problem from David White’s talk later today,
- ...



We consider the following situation:

- we are given the values of observations y
- we want to generate samples from the **posterior** distribution μ_y of u, *i.e.* from the conditional distribution of u given the observations y.

In this talk we assume that the posterior μ_y is of the form

$$\frac{d\mu_y}{d\mu_0}(u) = \frac{1}{Z}\exp\bigl(-\Phi(u;y)\bigr)$$

where μ_0 is some Gaussian reference measure.

Example 1: Sampling the initial condition

Assume the following situation:

- the signal x solves an ODE in $\mathbb{R}^d$:

$$\frac{dx(t)}{dt} = f\big(x(t)\big), \qquad x(0) = u \sim \nu.$$

- we have discrete, noisy observations:

$$y_k = g\big(x(t_k)\big) + \eta_k \qquad \forall k = 1, \dots, K$$

If u and η_k are Gaussian, this example fits into the given framework: we have

$$y \sim \mathcal{N}\big(\mathcal{G}(u), \Sigma\big)$$

and thus ...

... the density of observations is

$$p(y|u) \propto \exp\Big(-\frac{1}{2}\big|\mathcal{G}(u) - y\big|^2_{\Sigma^{-1}}\Big) =: \exp\big(-\Phi(u;y)\big).$$

We can use Bayes' rule to get

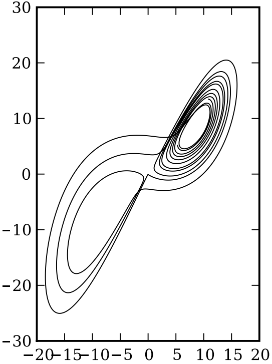
$$p(u|y) = \frac{p(y|u)p(u)}{p(y)} \propto p(y|u)p(u).$$

Using the prior $p(u)\,du$ aus the reference meassure μ_0 we get the posterior density

$$\frac{d\mu_y}{d\mu_0}(u) = \exp\Big(-\Phi(u;y)\Big).$$

Example: Lorenz system. Consider

$$\frac{dx(t)}{dt} = f\big(x(t)\big), \qquad f(x) = \begin{pmatrix} \sigma(x_2 - x_1) \\ \rho x_1 - x_2 - x_1 x_3 \\ x_1 x_2 - \beta x_3 \end{pmatrix}$$



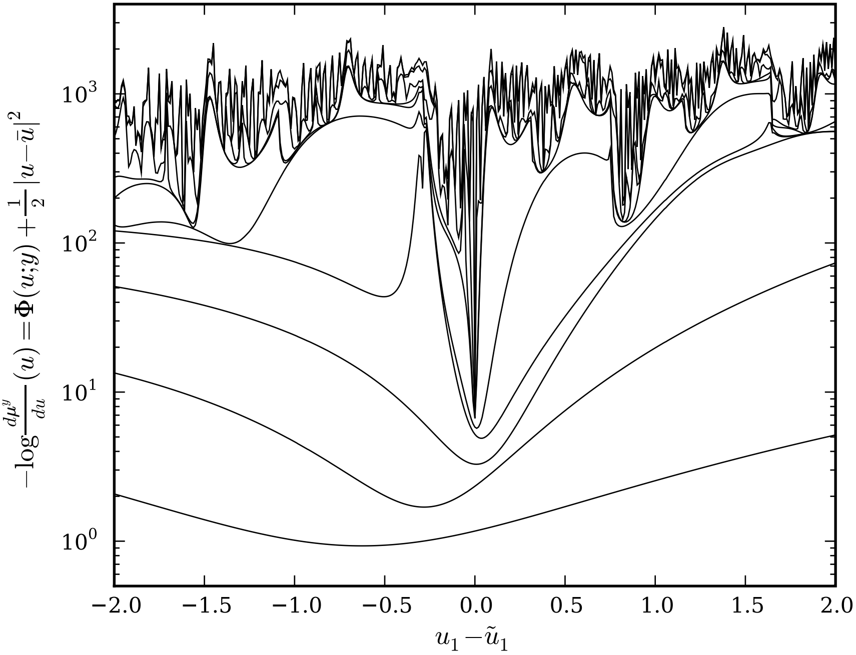
with

$$x(0) = u \sim \mathcal{N}(\bar{u}, 1).$$

The posterior density

$$\frac{d\mu_y}{d\mu_0}(u) = \exp\Big(-\Phi(u; y)\Big).$$

is easily evaluated but may be difficult to sample from.



$-\log \frac{d\mu^y}{du}(u) = \Phi(u;y) + \frac{1}{2}|u - \bar{u}|^2$
10^0
10^1
10^2
10^3
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
$u_1 - \tilde{u}_1$

Example 2: Model Error

Assume the following situation:

- the signal x solves an ODE in $\mathbb{R}^d$:

$$\frac{dx(t)}{dt} = f\big(x(t)\big) + v(t), \qquad x(0) = u \sim \nu,$$

  where v is a stationary stochastic process.

- we have discrete, noisy observations:

$$y_k = g\big(x(t_k)\big) + \eta_k \qquad \forall k = 1, \ldots, K$$

Again, we want to sample from the posterior, *i.e.* from the conditional distribution of $(u, v) \in \mathbb{R}^d \times C\big([0, T], \mathbb{R}^d\big)$ given the observations $y_1, \ldots, y_K$.

As before, the values $x(t_1), \ldots, x(t_k)$ are completely determined by u, v:

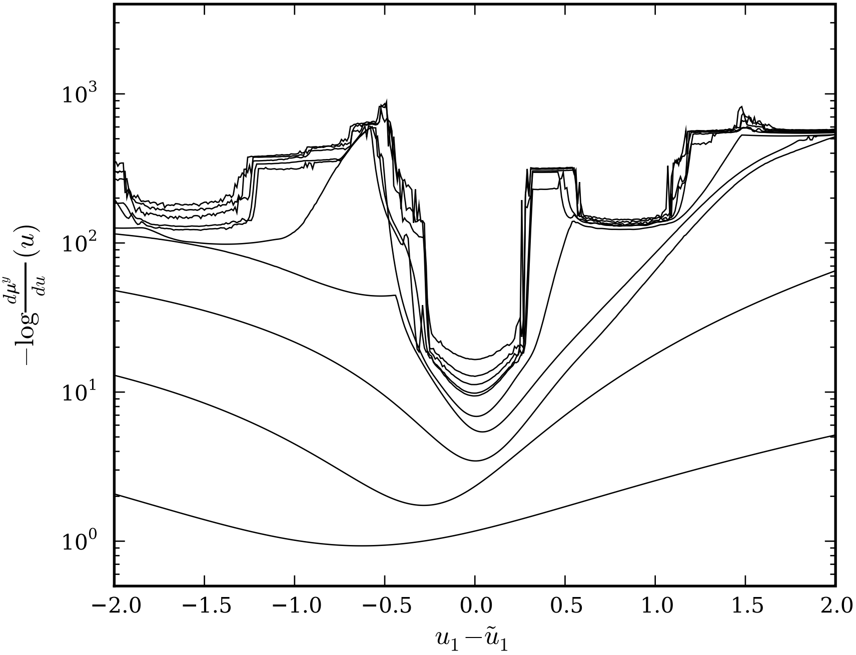
$$p(y|u,v) \propto \exp\Big(-\frac{1}{2}\big|\mathcal{G}(u,v) - y\big|^2_{\Sigma^{-1}}\Big) =: \exp\big(-\Phi(u,v;y)\big).$$

Again, we can use the prior distribution as the reference measure μ_0 to get the posterior density

$$\frac{d\mu_y}{d\mu_0}(u,v) = \exp\Big(-\Phi(u,v;y)\Big)$$

on $\mathbb{R}^d \times C\big([0,T],\mathbb{R}^d\big)$.

Sampling from the posterior is now an infinite dimensional problem, but the presence of the model error term v makes the distribution a lot smoother. Sometimes this may be advantageous!



10^3
10^2
10^1
10^0
$-\log \frac{d\mu^y}{du}(u)$
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
$u_1 - \tilde{u}_1$

2. Sampling on Path Space

We have seen how posterior distributions on path space may arise.

Question. How to sample from these infinite dimensional distributions?

There are several generic methods available.

- Langevin sampling: construct a continuous time stochastic process with values in $C\big([0, T], \mathbb{R}^d\big)$ which has the posterior as its stationary distribution.
- Metropolis sampling: use a rejection algorithm to modify a discrete time Markov chain to have the required stationary distribution.
- Combinations of both methods.

Langevin Sampling.

- Find a stochastic process u with values in $C\big([0, T], \mathbb{R}^d\big)$ whose stationary distribution coincides with the target distribution μ_y. Typically, the process u will be given as the solution to a **Stochastic Partial Differential Equation (SPDE)**.
- Simulate this sampling SPDE on a computer.
- Assuming ergodicity, we can probe all statistical properties of μ using ergodic averages:

$$\int_{C\left([0,T],\mathbb{R}^d\right)} \varphi(u)\, d\mu_y(u) = \lim_{S\to\infty} \frac{1}{S} \int_0^S \varphi\big(u(\tau)\big)\, d\tau.$$

Illustration: sampling Brownian bridges

The stochastic heat equation

$$\partial_\tau u(\tau,t) = \partial_t^2 u(\tau,t) + \sqrt{2}\,\partial_\tau w(\tau,t)$$

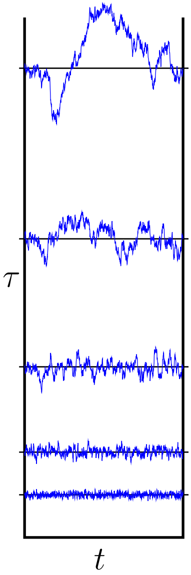
with Dirichlet boundary conditions

$$u(\tau,0) = 0, \qquad u(\tau,T) = 0$$

has the distribution of a Brownian bridge as its stationary distribution.

- $\partial_\tau w$ is space-time white noise
- $t \in [0,T]$ is *physical time* ("space" of the SPDE, time of the Brownian bridge)
- $\tau \in [0,\infty)$ is *algorithmic time* (time of the SPDE)

Adding a drift to the SPDE allows to sample from more interesting distributions.



Metropolis Sampling.

Result. Let $P(u, dv)$ be the transition kernel of a Markov chain on $C\big([0, T], \mathbb{R}^d\big)$. Construct a new Markov chain $(u_n)_{n\in\mathbb{N}}$ as follows: for each $n > 1$

- construct a *proposal* $v_n \sim P(u_{n-1}, \cdot\,)$, and
- let

$$u_n = \begin{cases} v_n & \text{with probability } \alpha(u_{n-1}, v_n) \\ u_{n-1} & \text{else.} \end{cases}$$

Then the Markov chain $(u_n)_{n\in\mathbb{N}}$ has stationary distribution μ_y.

Here the acceptance probability α is given by

$$\alpha(u, v) = \min\Big(1, \frac{\mu_y(dv)P(v, du)}{\mu_y(du)P(u, dv)}(u, v)\Big).$$

Remarks.

- The method only works if the measures $\mu_y(dv)P(v,du)$ and $\mu_y(du)P(u,dv)$ are equivalent so that the density in the construction of α exists.
- Efficiency of the method depends on the average acceptance probabilities obtained. This can be controlled by the choice of the proposal distribution $P(u,dv)$.
- If the proposal distribution is symmetric, then

$$\begin{aligned}\alpha(u,v) &= \min\Big(1, \frac{\mu_y(dv)P(v,du)}{\mu_y(du)P(u,dv)}(u,v)\Big)\\ &= \min\Big(1, \exp\big(\Phi(v;y)-\Phi(u;y)\big)\Big)\end{aligned}$$

- Good proposals can be constructed by taking one step of a discretised Langevin equation.

3. Conclusions

Conclusions

- Many applied problems can be written as sampling problems on a function space.
- In some situations an infinite dimensional method may provide more regularity and thus may be easier to use.
- There are various methods available to solve the resulting sampling problems.