Probing Probability Measures In High Dimensions

Andrew M Stuart[1]

[1]Mathematics Institute and
Centre for Scientific Computing
University of Warwick

ECM2012
Krakow, Poland, July 3*rd* 2012

Funded by EPSRC, ERC and ONR

Outline

Outline

General Setting

- Unknown $u \in X$, Hilbert space.
- Prior $\mu_0(du) = \mathbb{P}(du)$ on u : $\mu_0 = N(0, \mathcal{C}_0), \quad \mu_0(X) = 1.$
- Given data $y = \mathcal{G}(u) + \eta, \quad \eta \sim N(0, \Gamma).$
- Potential: $\Phi(u) := \frac{1}{2}\Big\|\Gamma^{-\frac{1}{2}}\big(y - \mathcal{G}(u)\big)\Big\|^2.$
- Posterior $\mu(du) = \mathbb{P}(du|y)$ on u :

$$\mathbb{P}(u|y) \propto \mathbb{P}(y|u)\mathbb{P}(u)$$

$$\frac{d\mu}{d\mu_0}(u) \propto \exp\Big(-\Phi(u)\Big)$$

Finite Dimensional Approximation

- Karhunen-Loeve basis: $\mathcal{C}_0\varphi_j = \lambda_j^2\varphi_j$.
- Finite-dimensionalization: $X \supset X^N = \mathrm{span}\{\varphi_j\}_{j=1}^N$ and $P^N : X \to X^N$ orthogonal projection.
- Approximation: $\Phi^N = \Phi \circ P^N$.
- Approximate Posterior on u :

$$\frac{d\mu^N}{d\mu_0}(u) \propto \exp\Big(-\Phi^N(u)\Big)$$

Assumptions

- Karhunen-Loeve Eigenvalues: $\lambda_j \asymp j^{-k}$, $k > \frac{1}{2}$.
- Hilbert-scale: X^s space with norm $\|\cdot\|_s := \|\mathcal{C}_0^{-\frac{s}{2k}} \cdot \|$.
- Potential Assumptions I: $\exists M \geq 0$ and $s \in [0, k - 1/2)$ such that $\Phi : H^s \to \mathbb{R}^+$ and, $\forall u \in H^s, N \in \mathbb{N}$,

$$\|\partial^2 \Phi(u)\|_{\mathcal{L}(X^s, X^{-s})} + \|\partial^2 \Phi^N(u)\|_{\mathcal{L}(X^s, X^{-s})} \leq M.$$

- Potential Assumptions II: behaviour out at infinity.

Approximation Error

$$\mu(du) \propto \exp\big(-\Phi(u)\big)\mu_0(du),$$
$$\mu^N(du) \propto \exp\big(-\Phi^N(u)\big)\mu_0(du).$$

Theorem

Cotter, Dashti and AMS, SINUM, 2010.
Assume that

$$|\Phi(u) - \Phi^N(u)| \leq K \exp\big(\epsilon\|u\|_X^2\big)\psi(N)$$

where $\psi(N) \to 0$ *as* $N \to \infty$. *Then there is a constant* C*, independent of* N*, and such that*

$$d_{Hell}(\mu, \mu^N) \leq C\psi(N).$$

See also *Marzouk/Xiu CCP 2009.*

Implications

- Mean:

$$\left\|\mathbb{E}^{\mu} u - \mathbb{E}^{\mu^N} u\right\|_X \le C\psi(N).$$

- Covariance

$$\left\|\mathbb{E}^{\mu} u \otimes u - \mathbb{E}^{\mu^N} u \otimes u\right\|_{X \to X} \le C\psi(N)$$

Outline

Old Random Walk Algorithm

Metropolis et al. Chem. Phys. 1953.

- Set $k = 0$ and Pick $u^{(0)}$.
- Propose $v^{(k)} = u^{(k)} + \beta\xi^{(k)}, \quad \xi^{(k)} \sim N(0, \mathcal{C}_0)$.
- Set $u^{(k+1)} = v^{(k)}$ with proability $a(u^{(k)}, v^{(k)})$.
- Set $u^{(k+1)} = u^{(k)}$ otherwise.
- $k \to k + 1$.

Here

$$a(u, v) = \min\{1, \exp\big(I(u) - I(v)\big)\}.$$

$$I(w) = \frac{1}{2}\|\mathcal{C}_0^{-\frac{1}{2}} w\|^2 + \Phi(w).$$

New Random Walk Algorithm

Neal, unpublished, 1999.
Beskos, Roberts, AMS and Voss Stoch. and Dyn., 2010.
Cotter, Roberts, AMS and White, arXiv 2012.

- Set $k = 0$ and Pick $u^{(0)}$.
- Propose $v^{(k)} = \sqrt{(1-\beta^2)}u^{(k)} + \beta\xi^{(k)}, \quad \xi^{(k)} \sim N(0, \mathcal{C}_0)$.
- Set $u^{(k+1)} = v^{(k)}$ with proability $a(u^{(k)}, v^{(k)})$.
- Set $u^{(k+1)} = u^{(k)}$ otherwise.
- $k \to k+1$.

Here $a(u, v) = \min\{1, \exp\big(\Phi(u) - \Phi(v)\big)\}$.

Outline

Langevin Equation

M. Hairer, AMS and J. Voss: Ann. App. Prob. 2007.

Define the Langevin SDE:

$$\frac{du}{dt} = -u + \mathcal{C}_0 D\Phi(u) + \sqrt{2}\frac{dW}{dt}.$$

Theorem

The Langevin equation has a global X^s-valued strong solution which is $\mu-$reversible and satisfies

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \varphi(u(t))\,dt = \int_{H^s} \varphi(u)\mu(du)$$

in probability for every $u(0)$ in the support of μ and every bounded $\varphi : X^s \to \mathbb{R}$ with bounded derivative.

Diffusion Limit for Old Random Walk

J. Mattingly, N. Pillai and AMS 2011

Let $\delta = \beta^2/2$.

$$u^\delta(t) := u^{(k)} + \frac{1}{\delta}(t - k\delta)\left(u^{(k+1)} - u^{(k)}\right), \quad t \in \left[k\delta, (k+1)\delta\right).$$

Theorem

The Old Random Walk Markov chain is μ^N-reversible on X^N and, if $\delta = O(N^{-1})$, and $u^{(0)} \sim \mu^N$ then $u^\delta \Rightarrow u$ in $C([0,T]; X^s)$ as $N \to \infty$ (and hence $\delta \to 0$).

Number of MCMC steps is $\mathcal{O}(N)$.

Diffusion Limit for New Random Walk

N. Pillai, AMS and A. Thiery 2011

Let $\delta = \beta^2/2$.

$$u^{\delta}(t) := u^{(k)} + \frac{1}{\delta}(t - k\delta)\big(u^{(k+1)} - u^{(k)}\big), \quad t \in \big[k\delta, (k+1)\delta\big).$$

Theorem

The New Random Walk Markov chain is μ^N−reversible on X^N and, for any fixed $u^{(0)} \in X^N$, $u^{\delta} \Rightarrow u$ in $C([0,T]; X^s)$ as $\delta \to 0$.

Number of MCMC steps is $\mathcal{O}(1)$.

Outline

L^2 Spectral Gap

- $L^2_\mu = \{f : X \to \mathbb{R} : \|f\|_2^2 := \mathbb{E}^\mu |f(u)|^2 < \infty\}.$
- $L^2_0 = \{f \in L^2_\mu : \mu(f) = 0.\}$
- Define the Markov kernel $(Pf)(u) = \mathbb{E}\Big(f\big(u^{(1)}\big)|u^{(0)} = u\Big).$
- $\|P\|_{L^2_0 \to L^2_0} := \sup_{f \in L^2_0} \frac{\|Pf\|_2^2}{\|f\|_2^2}.$
- We have $L^2_\mu-$ spectral gap γ if $\|P\|_{L^2_0 \to L^2_0} < 1 - \gamma.$
- $\gamma \in (0, 1).$
- Larger γ implies faster rate of convergence.

Standard RWM Theorem

The standard method behaves poorly as $N \to \infty$:

Theorem

(Hairer, AMS, Vollmer, arXiv 2012.)
For the standard Random walk algorithm:

- *If $\beta = N^{-a}$ with $a \in [0, 1)$ then the spectral gap is bounded above by $C_p N^{-p}$ for any positive integer p.*
- *If $\beta = N^{-a}$ with $a \in [1, \infty)$ then the spectral gap is bounded above by $CN^{-\frac{a}{2}}$.*

Hence spectral gap is bounded above by $CN^{-\frac{1}{2}}$.

New RWM Theorem

The new method behaves well as $N \to$ *infty*:

Theorem

(Hairer, AMS, Vollmer, arXiv 2012.)
For the new Random walk algorithm the spectral gap is bounded below independently of N. Hence CLT and, for $u^{(0)} \sim \nu$ and C independent of N,

$$\mathbb{E}^{\nu}\left|\frac{1}{K}\sum_{k=1}^{K} f(u^{(k)}) - \mathbb{E}^{\mu^N} f\right|^2 \le CK^{-1}.$$

Outline

What We Have Shown

We have shown that:

- **Applications:** Many inverse problems in differential equations can be formulated in the framework of Bayesian statistics on function space.
- **Common Structure:** These problems share a common mathematical structure leading to *well-posed* inverse problems for measures.
- **Approximation:** This well-posedness leads to a transfer of approximation properties from the forward problem to the inverse problem, in the Hellinger metric.
- **Algorithms:** MCMC methods can be defined on function space. Results in new algorithms robust to discretization.

http : //www.maths.warwick.ac.uk/ ~ masdr/

- A.M. Stuart. "Inverse Problems: A Bayesian Perspective." Acta Numerica **19**(2010).
- S.L.Cotter, M. Dashti, A.M.Stuart. "Approximation of Bayesian Inverse Problems." SIAM Journal of Numerical Analysis **48**(2010) 322–345.
- S.L. Cotter , G.O. Roberts, A.M. Stuart and D. White. "MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster." `http://arxiv.org/abs/1202.0709`
- M. Hairer, A.M.Stuart and S. Vollmer. "Spectral Gaps for a Metropolis-Hastings Algorithm in Infinite Dimensions." `http://arxiv.org/abs/1112.1392`

$http://www.maths.warwick.ac.uk/\sim masdr/$

- M. Hairer, A.M.Stuart and J. Voss. "Analysis of SPDEs Arising in Path Sampling. Part 2: The Nonlinear Case." Ann. Appl. Prob. 17(2007), 1657–1706.
- A. Beskos, G. Roberts, A.M.Stuart and J. Voss. "An MCMC method for diffusion bridges." Stochastics and Dynamics 8 (2008), 319-350.
- J.C. Mattingly, N. Pillai and A.M. Stuart, "Diffusion limits of random walk Metropolis algorithms in high dimensions." (To appear Ann. Appl. Prob.). `http://arxiv.org/abs/1003.4306`
- N. Pillai, A.M. Stuart and A. Thiery "Optimal proposal design for random walk type Metropolis algorithms with Gaussian random field priors." `http://arxiv.org/abs/1108.1494`