MCMC, SMC, and IS in High and Infinite Dimensional Spaces 1+2+3
Andrew Stuart
August 28, 2015

There will be there sections:

Probability measures of interest:

$$\mu(du) = \frac{1}{z} e^{-\phi(u)} \mu_0(du), \mu_0 = N(0, C) \tag{1}$$

We want to understand properties of probability measures which have a density with respect to a Gaussian μ_0. The main objective is to understand what the form of ϕ is.

Measure preserving dynamics: $c = (-\Delta)^{-s}, s > \frac{d}{2}$.

$$\frac{du}{dt} = K(-(-\Delta)^s u - D\phi(u)) + \sqrt{2K}\frac{dw}{dt} \tag{2}$$

$$M\frac{d^2u}{dt^2} + (-\Delta)^s u + D\phi(u) = 0$$

Two dynamical systems: Stochastic differential equation for the first equation and Hamiltonian mechanics for the second. We are interested in choices of K and M. For example:

1. If we take $s = 1$ and $K = 1$ the first equation becomes the nonlinear stochastic heat equation.
2. If we take $s = 1$ and $M = I$ then we have a wave equation with nonlinear forcing for the second equation.

Measure preserving dynamics - discrete time (MCMC) We will show how these continuous time dynamical systems play a role in a Monte-Carlo Markov Chain.

1 Probability measures of interest

1.1 Gaussian reference measure

$(\mathcal{H}, \langle\cdot,\cdot\rangle, |\cdot|)$ separable Hilbert (sometimes $|\cdot|$ will be the Euclidean norm).
Mean: $m \in \mathcal{H}$.
Covariance: $c \in \mathcal{L}(\mathcal{H}, \mathcal{H})$ trace-class in $\mathcal{H}$, positive, self-adjoint.

$$c\phi_j = \lambda_j \phi_j, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq 0,\ \lambda_j \to 0$$

$\{\phi_j\}_{j\in\mathbb{N}}$ form a complete orthonormal system for $\mathcal{H}$ and $\mu_0 = N(m, c)$

Lemma 1.1 (Karshunen-Loeve) *$u \sim \mu_0 \Leftrightarrow u = m + \sum_{j=1}^{\infty} \xi_j \sqrt{\lambda_j}\phi_j$ where $\{\xi_j\}_{j\in\mathbb{N}}$ i.i.d $\xi_1 \sim N(0,1)$.*

Corollary 1.0.1 *Let $u_j = \langle u - m, \phi_j\rangle$ then $\frac{1}{N}\sum_{j=1}^{N} \frac{u_j^2}{\lambda_j} \to 1$ as $N \to \infty$ μ_0-a.s.*

Example: $\mathcal{H} = L^2(D;\mathbb{R})$, $D \subset \mathbb{R}^d$ bounded and open.
Assumptions:

- A self-adjoint, invertable, positive definite on $\mathcal{H}$.
- $\{\phi_j\}_{j\in\mathbb{N}}$ be a complete orthonormal system (smooth) for $\mathcal{H}$.
- $A\phi_j = \alpha_j \phi_j$, α_j eigenvalues.
- α_j is upper and lower bounded by $j^{\frac{2}{d}}$.
- $\sup_{j\in\mathbb{N}} \left(\|\phi_j\|_{L^\infty} + \frac{1}{j^{1/d}} \mathrm{Lip}(\phi(j)) \right) < \infty$

If we take $A = -\Delta + I, D(A) = H^2(\mathbb{T}^d)$ then these assumptions are satisfied. More generally:

Theorem 1.1 *Let $c = A^{-s}$. Then for $u \sim \mu_0 = N(0, c)$ a.s., $u \in H^t, u \in c^{\lfloor t \rfloor, t - \lfloor t \rfloor}$ and $t < s - \frac{d}{2}$.*

Example: Brownian Bridge $d = 1$ on $I(0, 1)$. Take $A = -\frac{d^2}{dx^2}, D(A) = H^2(I) \cap H_0^1(I), u \in H^{1/2}, u \in C^{0,1/2}$.

1.2 Measure of interest

$(X, \|\cdot\|)$ a separable Banach Space and assume the Gaussian measure satisfies $\mu_0(X) = 1$ (this is short for saying $u \in X, \mu_0 - a.s.$). Also assume $\phi : X \to \mathbb{R}$ satisfies

- $\phi \geq 0$.
- ϕ is locally Lipschitz.
- $e^{-\phi} \in L^1_{\mu_0}(X, \mathbb{R})$.

These conditions can (and will for a couple examples) be relaxed, but are sufficient for our understanding in the lectures.

Define

$$\mu(du) = \frac{1}{z} e^{-\phi(u)} \mu_0(du), \quad z = \int_x e^{-\phi(u)} \mu_0(du).$$

Since μ is absolutely continuous with respect to μ_0, the same things (corollary 1.0.1) holds for μ a.s.

1.3 Elliptic inverse problems

$$\begin{cases} -\nabla \cdot (\kappa \nabla p) = f, & x \in D \subset \mathbb{R}^2 \\ p = 0, & x \in \partial D \end{cases}$$

Spaces:

- $Z = L^\infty(D; \mathbb{R})$
- $Z^+ = \{\kappa \in Z : \mathrm{essinf}_{x\in D} \kappa > 0\}$
- $V = H_0^1(D)$ (weak formulation)

Proposition 1.1 *If $\kappa \in Z^+$, then $\exists! p \in V$ solving the equation. Thus we may write $p = G(\kappa)$ for some $G : Z^+ \to V$. Furthermore, G is locally Lipschitz.*

Inverse Problem: We have a collection of linear functions $l_j \in V^*,\ j = 1, \dots, J$. Our goal is to find κ from noisy measurements $\{l_j(p)\}_{j=1}^J$.

Probability comes in because of the noisy data as well as noting that we are trying to reconstruct a function $\kappa \in L^\infty$ from a finite set of observations.

Bayesian Inverse Problem: $X = C(D; \mathbb{R})$, $F : X \to Z^+$.

(i) (first choice) $F(u) = e^u$ i.e. $\kappa = e^u$.

(ii) (second choice) $F(u) = \kappa^+ \mathbb{1}_{u \geq 0} + \kappa^- \mathbb{1}_{u<0}$ where $\kappa^+, \kappa^- < 0$.

Now F maps from the place where we will put Gaussians into the space of permabilities. From permabilities, G will map us to p. Then we will map into the finite set of operators. Putting this together:

$y_j = (l_j \circ G \circ F)(u) + \eta_j$, where $\eta \sim N(0, \gamma^2)$ (i.i.d).
$y = \mathcal{G}(u) + \eta,\ \eta \sim (0, \gamma^2 I)$ where $\mathcal{G} : X \to \mathbb{R}^J$

(i) (for first choice) $\mathcal{G}$ is locally Lipschitz. (exponentiation is locally Lipschitz)

(ii) (for second choice) $\mathcal{G}$ is continuous μ_0-a.s.

Now $\phi(u; y) = \dfrac{1}{2\gamma^2}|y - \mathcal{G}|^2$ and $\phi : X \times \mathbb{R}^J \to \mathbb{R}^+$.

We will use two distance in these talks:

$$d_{Hell}(\mu, \nu)^2 = \int_x \left| \sqrt{\frac{d\mu}{d\mu_0}(u)} - \sqrt{\frac{d\nu}{d\mu_0}(u)} \right|^2 \mu_0(du)$$

$u \sim \mu_0$ satisfying above assumptions. (Prior)
$y|u \sim N(\mathcal{G}(u), \gamma^2 I)$ - Likelihood $u|y \sim \mu^y$ (Posterior)

Theorem 1.2 *$\mu^y \ll \mu_0$. Furthermore, $\forall |y_1|, |y_2| < r$, $d_{Hell}(\mu^{y_1}, \mu^{y_2}) \leq C(r)|y_1 - y_2|$.*

Notes for lecture 2 begins below:

1.4 Navier-Stokes equation

First we will start with another construction of ϕ. Below we have Navier-Stokes in two dimensions on a Torus:

$$\begin{cases} \partial_t v + v \cdot \nabla v = \nu \Delta v - \nabla p, & x \in \mathbb{T}^2, t > 0 \\ \text{div } v = 0, & x \in \mathbb{T}^2, t > 0 \\ v = u, & x \in \mathbb{T}^2, t > 0 \end{cases}$$

Two examples of data from which we would like to recover the initial condition:

Problem 1 "Weather-forcasting"

$$y_{j,k} = v(x_j, t_k) + \nu_{j,k},\ \eta_{j,k} \sim N(0, \gamma^2 I)$$
$$y = \mathcal{G}(u) + \eta,\ \eta \sim N(0, \gamma^2 I).$$

Problem 2 "Oceanography"

$$\frac{dz_j}{dt} = v(z_j, t), z_j(0) = z_{j,0}$$
$$y_{j,k} = z_j(t_k) + \eta_{j,k}$$
$$y = \mathcal{G}(u) + \eta$$

In both cases we can define the misfit function :

$$\phi(u; y) = \frac{1}{2\gamma^2}|y - \mathcal{G}(u)|^2$$

$$u \sim \mu_0 = N(0, C) \quad \text{(Prior)}$$
$$C = (-\Delta_{\text{stokes}})^{-s}, s > 1$$
$$y|u \sim N(\mathcal{G}(u), \gamma^2 I)$$
$$u|y \sim \mu^y \text{ in (1)}$$

Theorem 1.3 *$\mu^y \ll \mu_0$ given by (1), then $d_{Hell}(\mu^{y_1}, \mu^{y_2}) \leq c(y_1, y_2)|y_1 - y_2|$.*

Comments: In Problem 1, $\mathcal{G} \in C^1(H; \mathbb{R}^J)$ and in Problem 2, $\mathcal{G} \in C^1(H^t; \mathbb{R}^J), t > 0$.

1.5 Getting information about μ

Can we find a point estimate to maximize μ?

1.5.1 Map Estimators

$\mu_0(x) = 1$ and $\phi \in C(X; \mathbb{R})$. E a Hilbert space is compact in X. Inner product $\langle \cdot, \cdot \rangle_E = \langle C^{-\frac{1}{2}} \cdot, C^{-\frac{1}{2}} \cdot \rangle$ and norm $| \cdot |_E = |C^{-\frac{1}{2}} \cdot |$. Note $u \notin E$ μ_0-a.s. Pretend $H = \mathbb{R}^N$, then to maximize the quantity below, we would minimize $-\frac{1}{2}|u|_E^2$:

$$\mu(du) \propto e^{-\phi(u) - \frac{1}{2}|u|_E^2} du.$$

Consider

$$B_\delta(z) = \{u \in X : \|u - z\|_x < \delta\}$$
$$J^\delta(z) = \mu(B_\delta(z))$$

We are looking for the z value that maximizes $J^\delta(z)$.

Definition 1.1 *$\overline{z}$ is a MAP estimator if*

$$\lim_{\delta \to 0} \frac{J^\delta(\overline{z})}{J^\delta(z^\delta)} = 1$$

Such points exists and can be characterized by:

$I(z) = \frac{1}{2}|z|_E^2 + \phi(z)$ Onsager-Machlup Functional.

Theorem 1.4 (Dosht, Law, Stuart, Voss)

(i) Any MAP estimator is a minimizer of I.

(ii) Any $z^* \in E$ which minimizes I is a map estimator.

1.5.2 Variational characterization

Work by (Pinski, Simpson, Stuart, Weber):

We want to generalize the previous minimization problem. To do this we minimize over Gaussians instead.

$D_{KL}(\nu\|\mu) = E^\nu \log\left(\frac{d\nu}{d\mu}\right)$.
P will be a probability measure on H and
$\frac{d\mu}{d\mu_0} = \frac{1}{z}e^{-\phi(u)}$.
Next we will define a functional J:
$J : P \to \mathbb{R}$ and $J(\nu) = D_{KL}(\nu||\mu_0) + \mathbb{E}^\nu\phi(u)$

Theorem 1.5 $\arg\inf_{\nu\in P} J(\nu) = \mu$

Sketch of proof: $J(\nu) = D_{KL}(\nu\|\mu) +$ constant . Next notice $D_{KL}(\nu\|\mu) \geq 0, D_{KL}(\mu\|\mu) = 0$. Thus the min is attained by setting $\nu = \mu$.

Remark 1.1 *Now minimize $J(\nu)$ over $\mathcal{A} \subset P$. For the Gaussian case,*

$$A = \cup N(m, \Sigma) \text{ and } N(z, \Sigma) \text{ equivalent to } \mu_0$$

The $J(\nu)$ within this class can be written as

$$J(\nu) = \frac{1}{2}|z|_E^2 + \mathbb{E}^{\xi\sim N(0,\Sigma)}\phi(z+\xi) + \frac{1}{2}tr|C^{-1}\Sigma - I) + \ln\left(\frac{\det C}{\det \Sigma}\right)$$

1.5.3 MCMC

Idea is to create a Markov Chain $\{u^{(n)}\}_{n\in\mathbb{N}}$ which is μ-ergodic. Then we have a method (see Jonathan Mattingly's lectures) to show

$$\frac{1}{N}\sum_{n=1}^{N}\phi(u^{(n)}) \to \mathbb{E}^\mu\phi(u)$$

2 μ-preserving dynamics

Goal for the rest of this lecture is to describe the basic ideas of μ-preserving dynamics so we can explain how they relate Markov Chains.

2.1 SDE on $\mathbb{R}^n$

Start with $\mu(du) \propto E^{-\Psi} du$ and $\Psi : \mathbb{R}^n \to \mathbb{R}$. We assume $e^{-\Psi(u)} \in L^2(\mathbb{R}; \mathbb{R}^+)$.

$$\frac{du}{dt} = -k\nabla\Psi(u) + \sqrt{2k}\frac{dw}{dt} \tag{3}$$

$k > 0$, symmetric $\mathbb{R}^{\mathbb{N}\times\mathbb{N}}$ matrix.

Theorem 2.1 (3) *is μ-invariant α-ergodic. (there are more conditions needed, but not listed in lecture)*

Sketch of proof:
Invariance:
$\mathcal{L}\phi =$
$\mathcal{L}^*\phi = \nabla \cdot (J(\phi))$
Get the equation:
$J(\phi) = k\nabla\Psi(u)\phi + \nabla \cdot (k\phi)$
$\frac{d\rho}{dt} = \mathcal{L}^*\rho.$
If $\rho\alpha e^{-\Psi(u)}$ then $J(\rho) = 0$.
Ergodicity:
$\mathcal{L}\phi = \phi - E^\mu\phi.$
Apply the Ito's formula to get: $\frac{d\phi}{dt} = \mathcal{L}\phi + \langle\nabla\phi(u), \sqrt{2k}\frac{dw}{dt}\rangle.$
$\frac{1}{T}\int_0^T \phi(t)dt = \mathbb{E}^\mu\phi + \frac{1}{T}(\phi(T) - \phi(0)) - \frac{1}{T}\int_0^T \langle\nabla\phi(u), \sqrt{2k}\frac{dw}{dt}\rangle.$

Notes for lecture 3 starts below

Recall we are interested in measures given by (1) as well as inverse problems.

We start today with two new μ-reversible dynamical systems. One stochastic and one not:

$$\frac{du}{dt} = -k(C^{-1}u + D\phi(u)) + \sqrt{2k}\frac{dw}{dt} \tag{4}$$

$$\frac{du}{dt} = v, \quad \frac{dv}{dt} = -M^{-1}(C^{-1}u + d\phi(u)) \tag{5}$$

2.2 Lift to H

In the Hilbert setting, (1) can be written as

$$\mu(du) = \frac{1}{z}e^{-I(u)}du$$

where

$$I(u) = \frac{1}{2}|u|_E^2 + \phi(u), \quad |\cdot|_E = |C^{-\frac{1}{2}}\cdot|$$

Using this gives us (4). As in the finite dimensional case, we get the theorem:

Theorem 2.2 (Hairer, Stuart, Voss) *For $k = I, C$ it follows that (2) is μ-reversible, μ-ergodic.*

Recall all the physics is embedded in ϕ so we can pick K however we want.

Comments on the proof for the case $(K = C)$.

$$\frac{du}{dt} = -u - C \cdot D\phi(u) + \sqrt{2C}\frac{dw}{dt}$$

From the point of algorithms this equation is very nice. If we ignore the ϕ-term, then all of the time scales and modes of this equation are the same. In fact we can solve

$$\frac{du}{dt} = -u + \sqrt{2C}\frac{dw}{dt}$$

which gives

$$u(t) = N\left(e^{-t}u(0), \left(1 - e^{-2t}\right)C\right)$$

Call this measure $\mu^t(u_0)$. If we let $t \to \infty$, we get the measure $\mu^\infty = N(0, C)$.
Asymptotics Strong Feller:
$\mu^t(u_0)$ is singular with respect to μ^∞ and
$\mu^t(u_0)$ is singular with respect to $\mu^t(u_0')$ unless $u_0 - u_0' \in E$.

Now we will go on to Hamiltonian mechanics.

2.3 $\mu(du) = e^{-\phi(u)}du$ in $\mathbb{R}^n$

$$\begin{aligned} H(u,p) &= \psi(u) + \frac{1}{2}|M^{-1/2}p|^2 \\ \frac{du}{dt} &= M^{-1}p, \quad \frac{dp}{dt} = -\nabla\Psi(u) \end{aligned} \tag{6}$$

Since the last two equations conserve the Hamiltonian, it's a simple calculation via the Louisville Theorem to show that any function of the Hamiltonian will also be a conserved density for the flow. In particular,

$$\nu(du, dp) = e^{-H(u,p)}dudp$$

From here, it is fairly straightforward to show

Theorem 2.3 *$\nu(du, dp)$ is an invariant measure for the dynamical system* (6).

Now we introduce velocity $v = M^{-1}p$ and so (6) becomes

$$\frac{du}{dt} = v, \ \frac{dv}{dt} = -M^{-1}\nabla\Psi(u)$$

The relevant conserved measure:

$$\tilde{\nu}(du, dv) = e^{-\tilde{H}(u,v)}dudv$$

where

$$\tilde{H}(u,v) = \Psi(u) + \frac{1}{2}|M^{\frac{1}{2}}v|^2$$

We can lift this up to the Hilbert space setting.

2.4 Hamiltonian on $\mathcal{H}$

$\tilde{\nu}(du,dv) = \frac{1}{z}e^{-\phi(u)}\mu_0(du)\mu_0(dv)$ provided $M = C^{-1}$. $\mathcal{X} = C([0,T];X)$.

Theorem 2.4 (Beskos, Pinskii, Sanz-Serra, Stuart) *There exists a unique solution to* (5) *in the space* $\mathcal{X}\times\mathcal{X}$*. Furthermore, this solution preserves* $\tilde{\nu}(du,dv)$*.*

Next notice $\int_H \tilde{\nu}(A,dv) = \mu(A)$. If we introduce the flow for this Hamiltonian,

$$\begin{pmatrix} u(t) \\ v(t) \end{pmatrix} = \Xi^t \begin{pmatrix} u(0) \\ v(0) \end{pmatrix}, \quad p\begin{pmatrix} u \\ v \end{pmatrix} = u$$

for $\frac{du}{dt} = v$, $\frac{dv}{dt} = -u - CD\phi(u)$. If ϕ drops out, then this is just a trivial oscillator.

Now we would like an interesting Markov Chain:

$$u^{(n+1)} = P\Xi\begin{pmatrix} u^{(n)} \\ \xi^{(n)} \end{pmatrix},\ \xi^{(n)} \sim N(0,C) \text{ i.i.d.} \tag{7}$$

Corollary 2.4.1 *This Markov chain* $\left\{u^{(n)}\right\}_{n\in\mathbb{Z}^+}$ *is* μ*-reversible.*

This is proved by finite-dimensionalization and passage to the limit.

3 MCMC

3.1 Connect to SDE

$w = (1-\beta^2)^{\frac{1}{2}}u + \beta\xi$, $\xi \sim N(0,C)$. Notice this is a formula for the exact solution of the Ornstein-Uhlenbeck process.

Now we will construct a Markov chain $\left\{u^{(n)}\right\}_{n\in\mathbb{Z}^+}$ as follows:
Let $w^{(n)} = (1-\beta^2)^{\frac{1}{2}}u^{(n)} + \beta\xi^{(n)}$ and $\xi^{(n)} \sim N(0,C)$ i.d.d.
Set $\alpha^{(n)} = 1$ min of $e^{(\phi(u^{(n)})-\phi(w^{(n)}))}$.
Set $u^{(n+1)} = \gamma^{(n)}w^{(n)} + (1-\gamma^{(n)})u^{(n)}$ where $\gamma^{(n)} = \begin{cases} 1 & \text{w.p. } \alpha^{(n)} \\ 0 & \text{otherwise} \end{cases}$.

From the output of the Markov chain a piecewise-linear function with input t and output u_β is constructed. Think of the time step as being $\beta^2/2$.

Theorem 3.1 (Pillai, Stuart, Theiry) u_β *converges weakly to* u *solving the Ornstein-Uhlenbeck process in* $C([0,T];X)$*.*

Theorem 3.2 (Haire, Stuart, Vollmer) *The spectral gap for* $\left\{u^{(n)}\right\}_{\mathbb{Z}^+}$*.*

Another Markov chain (using (7)) similar to the previous one: Let $w^{(n)} = P\Xi^{t,n}\begin{pmatrix} u^{(n)} \\ \xi^{(n)} \end{pmatrix}$ and $\xi^{(n)} \sim N(0,C)$ i.d.d.
Set $\alpha^{(n)} = 1$ min of $e^{\text{``}H(u^{(n)},\xi^{(n)})-H(w^{(n)},\cdot)\text{''}}$.

Set $u^{(n+1)} = \gamma^{(n)} w^{(n)} + (1-\gamma^{(n)}) u^{(n)}$ where $\gamma^{(n)} = \begin{cases} 1 & \text{w.p. } \alpha^{(n)} \\ 0 & \text{otherwise} \end{cases}$. Here's the inverse problem which will appear in the simulations:

$$\frac{du}{dt} = u - u^3 + \frac{dw}{dt}$$

$$y(t) = \int_0^t u(s)ds + B(t)$$

where w, B are standard unit Brownian motions which are independent of each other. The goal is to find $\mathbb{P}(u|y)$.

Several simulations are presented next in the lecture which should be viewed in the video.