Hybrid Monte Carlo on Hilbert spaces

J. M. Sanz-Serna
Universidad de Valladolid, Spain

Joint work with A. Beskos, F. Pinski, A.M. Stuart

I. A REVIEW OF THE STANDARD HMC ON $\mathbb{R}^N$

• Aim: obtain samples $q^{(n)}$, $n = 0, 1, \dots$, from a probability density $\pi \propto \exp\big(-V(q)\big)$ in $\mathbb{R}^N$.

• Write $V(q) = \frac{1}{2}\langle q, Lq\rangle + \Phi(q)$, L sym. pos. semdef. (perhaps $L = 0$), so that $\pi \propto \exp\big(\frac{1}{2}\langle q, Lq\rangle + \Phi(q)\big)$.

• HMC (Duane *et al* 1987) includes three ingredients:

A Hamiltonian flow in $\mathbb{R}^{2N}$.

A numerical integrator for that flow.

An accept/reject rule.

Hamiltonian flow in $\mathbb{R}^{2N}$

• Intoduce Hamiltonian $H(q,p) = \frac{1}{2}\langle p, M^{-1}p\rangle + \frac{1}{2}\langle q, Lq\rangle + \Phi(q)$. M is a user specified sym. pos. def. ('mass') matrix, p auxiliary variable ('momentum'), H total energy (kinetic+potential).

• Associated canonical ODEs are

$$\frac{dq}{dt} = \frac{\partial H}{\partial p} = M^{-1}p, \quad \frac{dp}{dt} = -\frac{\partial H}{\partial q} = -Lq + f(q), \quad f = -\nabla \Phi.$$

• For any fixed t, the t-flow Ξ^t of the system

$$\Xi^t(q(0), p(0)) = (q(t), p(t))$$

preserves the volume-element $dq\, dp$ and the value of H.

• Therefore Ξ^t also preserves the measure in the phase space $\mathbb{R}^{2N}$ with density $\Pi(q,p) \propto \exp(-H(q,p))$, or

$$\Pi(q,p) \propto \exp\Big(-\tfrac{1}{2}\langle p, M^{-1}p\rangle\Big)\ \exp\Big(-\tfrac{1}{2}\langle q, Lq\rangle - \Phi(q)\Big)$$

. . . and by implication the q-marginal, ie our target π. (The marginal of p is $N(0,M)$.)

• Hence if $q(0)$ is distributed according to target π and we draw $p(0) \sim N(0,M)$, then $q(t)$ will also be distributed according to π.

• This suggests the following . . .

- *'Idea for an algorithm'* (T and M have been fixed.)

— Given $q^{(n)}$, draw $p^{(n)} \sim N(0, M)$.

— Find $(q^*, p^*) = \Xi^T(q^{(n)}, p^{(n)})$.

— Set $q^{(n+1)} = q^*$, discard p^*, $n \leftarrow n + 1$.

- The transition $q^{(n)} \mapsto q^{(n+1)}$ defines a Markov chain that has the target π as invariant distribution.

- The chain is reversible. Transitions are non-local.

A numerical integrator for the canonical equations

• In practice analytic expression of Ξ^t not available and one has to resort to numerical approximations.

• Integrator must be volume preserving and reversible.

• Verlet is method of choice. Convenient to see it here as a splitting algorithm $\Psi_h = \Xi_1^{h/2} \circ \Xi_2^h \circ \Xi_1^{h/2}$, where Ξ_1^t and Ξ_2^t are the flows of the canonical systems with Hamiltonian functions

$$H_1 = \frac{1}{2}\langle q, Lq\rangle + \Phi(q)\ , \quad H_2 = \frac{1}{2}\langle p, M^{-1}p\rangle, \quad H = H_1 + H_2.$$

Accept/reject

• Since Verlet does not conserve H exactly, numerical solution does not preserve Π.

• Exact conservation is enforced through the Metropolis-Hastings rule: if $(q^{(n)}, p^{(n)})$ is current state of chain and (q^{*}, p^{*}) ('the proposal') is the numerically computed approximation at the end of a time-interval of length T, then compute

$$a = \min\Big(1, \exp\big(H(q^{(n)}, p^{(n)}) - H(q^{*}, p^{*})\big)\Big) \; .$$

• Set $q^{(n+1)} = q^{\star}$ with probability a ('acceptance'); otherwise set $q^{(n+1)} = q^{(n)}$ ('rejection').

Choice of mass matrix (Crucial for efficiency.)

• From canonical equations:

$$\frac{d^2q}{dt^2} = -M^{-1}Lq + M^{-1}f(q).$$

• Consider case $f \equiv 0$ and L is pos. def. and has some very large eigenvalues (π is normal with some very small variances),

choice $M = I$ implies inefficiency (Verlet step h adjusted to smallest variance of π/largest eigenvalue of L).

choice $M = L$ (large mass in directions of large force) then $d^2q/dt^2 = -q$. Both q and 'velocity' $v = dq/dt$ are slowly varying; v is of the size of q ($p = Mv = Lv$ is large).

• This suggests to use $M = L$ and to write dynamics for Verlet integration in q, v variables:

$$\frac{dq}{dt} = v, \quad \frac{dv}{dt} = -q + L^{-1} f(q).$$

• $v \sim N(0, L^{-1})$ and initial value should be drawn accordingly.

• Also in accept/reject compute H in terms of v:

$$H = \tfrac{1}{2}\langle v, Lv\rangle + \tfrac{1}{2}\langle q, Lq\rangle + \Phi(q).$$

II. THE PROBLEM

• π_0 a non-degenerate (non-Dirac) centred Gaussian measure with covariance operator $\mathcal{C}$ in (infinite-dimensional, separable) Hilbert space $\mathcal{H}$.

[$\mathcal{C}$ is a positive, self-adjoint, nuclear (ie its eigenvalues are summable), its eigenfunctions span $\mathcal{H}$. (eg $\mathcal{C}$ inverse Laplacian in L^2 + bd; corresponding to Brownian motion and Brownian bridge.)]

• Aim is to sample from a probability measure π defined by its density with respect to π_0:

$$\frac{d\pi}{d\pi_0}(q) \propto \exp\Big(-\Phi(q)\Big).$$

(Φ small relative to $\mathcal{C}^{-1}$.)

• Sources: conditioned diffusion; Bayesian approach to inverse problems . . .

• May also consider case where $\mathcal{H}$ is replaced by $\mathbb{R}^N$ with very large N and $\mathcal{C}$ by a sym. pos. def. matrix with some of its eigenvalues close to 0. (High-dimensional Gaussian with some very small variances.)

• Such finite-dimensional version is required for implementation.

• Standard HMC:

— not applicable in Hilbert space setting. (No standard density of target π.)

— applicable in finite-dimensional setting . . .

. . . with performance that rapidly deteriorates as $N \uparrow \infty$.

• Situation similar to that for explicit time integrators for PDE $du/dt = Au$, with A an unbounded operator in $\mathcal{H}$: they only make sense if equation has been first discretized in space and their performance degrades as the discretization is refined.

• Here we wish to construct an algorithm that may be expressed on the Hilbert space setting (and is therefore likely to perform uniformly well as $N \uparrow \infty$ in finite-dimensional approximations).

III. NEW ALGORITHM

Hamiltonian flow in $\mathcal{H} \times \mathcal{H}$

• In finite-dimensional version, reference Gaussian measure π_0 has a density $\propto \exp(\frac{1}{2}\langle q, \mathcal{C}^{-1} q\rangle)$ wrt standard Lebesgue measure. (Note there is no Lebesgue measure on $\mathcal{H}$!)

• Hence target π has standard density $\exp\Big(\frac{1}{2}\langle q, \mathcal{C}^{-1} q\rangle + \Phi(q)\Big)$: a format we considered earlier with the notation $L = \mathcal{C}^{-1}$.

• Earlier discussion suggests to proceed as follows:

— Introduce Gaussian measure Π_0 on $\mathcal{H} \times \mathcal{H}$ given by $\Pi_0(dq, dv) = \pi_0(dq) \otimes \pi_0(dv)$.

— Introduce measure Π on $\mathcal{H} \times \mathcal{H}$ given by $d\Pi/d\Pi_0 \propto \exp\big(-\Phi(q)\big)$. Target π is q-marginal of Π.

— Consider system:

$$\frac{dq}{dt} = v \ , \quad \frac{dv}{dt} = -q + \mathcal{C}\, f(q), \quad f = -D\Phi.$$

- Under natural hypotheses, it may be shown that

— System defines a global flow Ξ^t on $\mathcal{H} \times \mathcal{H}$.

— Ξ^t preserves the measure Π on $\mathcal{H} \times \mathcal{H}$.

— $(q^{(n+1)}, v^{(n+1)}) = \Xi^T(q^{(n)}, v^{(n)}) \ , \ v^{(n)} \sim \pi_0$ defines

via $q^{(n)} \mapsto q^{(n+1)}$, a Markov chain reversible wrt to π.

• System preserves *formally*

$$\mathsf{H}(q,v) = \frac{1}{2}\langle v, \mathcal{C}^{-1}v\rangle + \frac{1}{2}\langle q, \mathcal{C}^{-1}q\rangle + \Phi(q).$$

(which is in the finite-dimensional case is the old energy) and hence $\exp(-\mathsf{H})$.

• However $\langle q, \mathcal{C}^{-1}q\rangle$ and $\langle v, \mathcal{C}^{-1}v\rangle$ are almost surely infinite in an infinite-dimensional context. (If $\mathcal{C}$ is inverse Laplacian in L^2, they are squares of H^1-norms.)

A numerical integrator

- Use Strang splitting $\Psi_h = \Xi_1^{h/2} \circ \Xi_2^{h} \circ \Xi_1^{h/2}$, where

  — Ξ_1 is the flow of

$$\frac{dq}{dt} = 0 \;, \quad \frac{dv}{dt} = \mathcal{C} f(q).$$

  — Ξ_2 is the flow of

$$\frac{dq}{dt} = v \;, \quad \frac{dv}{dt} = -q.$$

- Ξ_1, Ξ_2 available in closed form.

Accept/Reject rule

- The natural candidate for the acceptance probability is

$$a = \min\Big(1, \exp\Big(\mathsf{H}(q^{(n)}, v^{(n)}) - \mathsf{H}(q^*, v^*)\Big)\Big),$$

where H is the invariant we discussed above ...

- ... but, as we noted, H is almost surely infinite in $\mathcal{H}$.

- Remedy is to work a formula for the increment $\mathsf{H}(q^{(n)}, v^{(n)}) - \mathsf{H}(q^*, v^*)$ that does not include the offending almost surely infinite terms.

- The recipe is:

$$\Phi(q_I) - \Phi(q_0) + \frac{h^2}{8}\Big(|\mathcal{C}^{\frac{1}{2}} f(q_0)|^2 - |\mathcal{C}^{\frac{1}{2}} f(q_I)|^2\Big)$$
$$+h \sum_{i=1}^{I-1} \langle f(q_i), v_i \rangle + \frac{h}{2}\Big(\langle f(q_0), v_0 \rangle + \langle f(q_I), v_I \rangle\Big).$$

This makes sense in $\mathcal{H}$ and in the finite-dimensional setting coincides with the energy increment.

- This is discrete analogue of physically meaningful expression:

$$\Phi\Big(q(T)\Big) - \Phi\Big(q(0)\Big) + \int_0^T \langle f\Big(q(t)\Big), v(t) \rangle \, dt.$$

MAIN RESULT

THEOREM: The algorithm defines a Markov chain which is reversible wrt to π.

The proof uses finite-dimensional approximations based on the eigenspaces of $\mathcal{C}$.