Ensemble Hamiltonian Monte Carlo

Yixuan Wang*

Caltech
roywang@caltech.edu

*with Ziming Liu and Andrew Stuart

June 1, 2022

Section 1

Bayesian Inverse Problem

Setting and Objective

Bayesian inverse problem:

$$y = \mathcal{G}(u) + \eta \,. \tag{1}$$

$\mathcal{G}$: (possibly nonlinear) forward mapping, η: Gaussian noise.
Known: Law of η, observation y, forward map $\mathcal{G}$.
Objective: distribution of u, given prior, without resorting to derivative $D\mathcal{G}$.

Inverting as a Sampling Problem

Prior $\eta \sim \mathrm{N}(0, \Gamma)$.
Nonlinear least squares functional as the objective for optimization:

$$\Phi(u) = \frac{1}{2}\|y - \mathcal{G}(u)\|_{\Gamma}^{2}. \tag{2}$$

Posterior corresponding to the Gaussian prior $\pi_0(u) \sim \mathrm{N}(0, \Gamma_0)$:

$$\pi(u) \propto \exp(-\Phi(u))\pi_0(u) = \exp\left(-\Phi_R(u)\right). \tag{3}$$

For $R(u) = \frac{1}{2}\|u\|_{\Gamma_0}^2$ and $\Phi_R(u) = \Phi(u) + R(u)$.

Section 2

Hamiltonian Monte Carlo and Langevin Dynamics

Langevin Dynamics and Preconditioning

Langevin dynamics:

$$\dot{u} = -\nabla \Phi_R(u) + \sqrt{2}\dot{\mathbf{W}}\,, \tag{4}$$

with the desired posterior distribution as the invariant distribution.
Ensemble preconditioner (acceleration) with covariance:

$$\dot{u}^{(j)} = -C(U)\nabla \Phi_R\left(u^{(j)}\right) + \sqrt{2C(U)}\dot{\mathbf{W}}^{(j)}\,. \tag{5}$$

Boost convergence and use approximate difference in lieu of derivative.

HMC and Second-order Langevin Dynamics

Hamiltonian:

$$\mathrm{H}(q,p) = \frac{1}{2}\left\langle p, \mathcal{M}^{-1}p \right\rangle + \Phi_R(q)\,. \tag{6}$$

For a general preconditioner independent of the variables p, q, we can split it into its diagonal and skew-symmetric form $\mathcal{K}$, $\mathcal{J}$, and obtain the following SOL-HMC in $\mathbb{R}^{2N}$ with preconditioning:

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \mathcal{J}D\mathrm{H}(z) - \mathcal{K}D\mathrm{H}(z) + \sqrt{2\mathcal{K}}\frac{\mathrm{d}W}{\mathrm{d}t}, \tag{7}$$

This Langevin dynamics has the posterior distribution automatically as the invariant distribution; jointly for the state space variable q and the velocity space variable p.

Choice of Preconditioner

$$\mathcal{J} = \begin{pmatrix} 0 & \mathcal{J}_1 \\ -\mathcal{J}_1 & 0 \end{pmatrix}, \tag{8}$$

$$\mathcal{K} = \begin{pmatrix} \mathcal{K}_1 & 0 \\ 0 & \mathcal{K}_2 \end{pmatrix}, \tag{9}$$

$\mathcal{M} = \mathcal{J}_1 = \mathcal{K}_2 = \mathcal{C}(\rho_q)$, $\mathcal{K}_1 = 0$.
$\mathcal{C}(\rho_q)$ is the covariance of the distribution of the position q.
Now

$$\begin{aligned} \frac{\mathrm{d}q}{\mathrm{d}t} &= p\,, \\ \frac{\mathrm{d}p}{\mathrm{d}t} &= -\mathcal{C}(\rho_q)(\Gamma_0^{-1}q + D\Psi(q)) - p + \sqrt{2\mathcal{C}(\rho_q)}\frac{\mathrm{d}W_2}{\mathrm{d}t}\,. \end{aligned} \tag{10}$$

Nonlinear FP Equation

Nonlinear Fokker-Planck equation

$$\partial_t \rho = \nabla^T \cdot ((\mathcal{K} - J)(\rho \nabla H + \nabla \rho)) \,. \tag{11}$$

The preconditioner is only dependent on the density distribution (covariance) ρ and not-dependent on particle z.

Section 3

Ensemble HMC

Ensemble Approximation

Gradient-free algorithm by using an ensemble approximation of $\mathcal{C}(\rho_q)$.

$$\mathcal{C}(Q) = \frac{1}{J}\sum_{k=1}^{J}\left(q^{(k)} - \bar{q}\right) \otimes \left(q^{(k)} - \bar{q}\right), \tag{12}$$

where

$$\bar{q} = \frac{1}{J}\sum_{j=1}^{J} q^{(j)}. \tag{13}$$

Now the ensemble-based algorithm is

$$\begin{aligned} \dot{q}^{(j)} &= p^{(j)}, \\ \dot{p}^{(j)} &= -\mathcal{C}(Q)\Gamma_0^{-1}q^{(j)} - B^j - p^{(j)} + \sqrt{2\mathcal{C}(Q)}\dot{W}^{(j)}, \end{aligned} \tag{14}$$

$$B^j = \frac{1}{J}\sum_{k=1}^{J}\langle \mathcal{G}(q^{(k)}) - \frac{\sum_{i=1}^{J}\mathcal{G}(q^{(i)})}{J}, \mathcal{G}(q^{(j)}) - y\rangle_\Gamma q^{(k)}. \tag{15}$$

Theoretical Properties of our Ensemble HMC

1. Correction term: finite particle approximation of interacting equations also preserve the invariant distribution.
2. Affine invariance: same convergence rate upon affine transformations.
3. Linear case: our algorithm preserves Gaussian, and convergence speed is independent of the operator.

Section 4

Numerical Results

1D Elliptic Boundary Value Problem

$$-\frac{d}{dx}(\exp(u_1)\frac{d}{dx}p(x)) = 1, \quad x \in [0,1]\,. \tag{16}$$

$p(0) = 0$ and $p(1) = u_2$, forward map $\mathcal{G}(u) = \begin{pmatrix} p(x_1) \\ p(x_2) \end{pmatrix}$, with explicit solution

$$p(x) = u_2 x + \exp(-u_1)(-\frac{x^2}{2} + \frac{x}{2})\,. \tag{17}$$

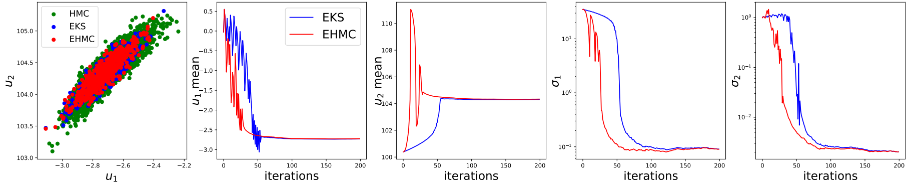

Figure: The low dimensional parameter space example. Left to right: samples; mean u_1; mean u_2; the first singular value σ_1; the second singular value σ_2.

Darcy Flow

$$-\nabla \cdot (a(x)\nabla p(x)) = f(x), x \in D = [0,1]^2. \\ p(x) = 0, x \in \partial D\,. \tag{18}$$

Given (noisy) measurements $p(x)$, infer $a(x;u)$ or u. We model $a(x;u)$ as a log-Gaussian field with precision operator defined as $\mathcal{C}^{-1} = (-\Delta + \tau^2\mathcal{I})^\alpha$.

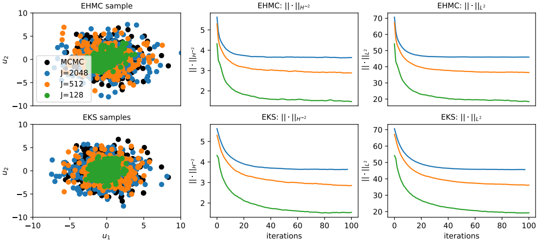

Figure: Darcy flow. Left: samples obtained from EHMC and EKS, compared with MCMC. Middle & right: Evolution of $||u||_{H^{-2}}$ and $||u||_{L^2}$ for EHMC and EKS for different $J = 128, 512, 1024$.

Convergence in a Big Picture

How many convergence we have? Three.

1. The numerical method where we use a discretization in time to approximate the Ensemble HMC. Error estimate in time: Ref to J. Lu on the numerical error of SOL.
2. The ensemble approximation to the preconditioned SOL where we use a finite number of particles to sample the empirical covariance. Finite approximation to mean field equation: Ref to Q. Li.
3. The convergence to equilibrium, i.e. to our desired posterior distribution for the preconditioned SOL. Hypocoercivity: Ref to C. Villani.