HAMILTONIAN SAMPLER WITH SOFT CONSTRAINTS

1 First method:

Let $\xi : \mathbb{R}^d \to \mathbb{R}^m$, where $m < d$. Consider the level surface

$$S_z = \{q \in \mathbb{R}^d : \xi(q) = z\}. \tag{1.1}$$

We want to sample from a distribution supported supported on $\mathbb{R}^d$ defined as

$$\nu_\varepsilon(dq) = Z_\varepsilon^{-1} e^{-(V(q)+V_\varepsilon(q))}dq, \quad \text{where} \quad V_\varepsilon(q) = \frac{|\xi(q)|^2}{2\varepsilon^2}, \tag{1.2}$$

and $V(q)$ is some given potential. We want to exploit the energy preserving property of the Hamiltonian flow to take large steps along the level surface S_z. To do this, we need augment the state space with momentum (denoted as $p \in \mathbb{R}^d$) and sample from a distribution supported on phase space. For this, we first need to specify the tangent and co-tangent spaces for S_z, which we denote by T_qS_z and TS_z respectively. The tangent space of S_z at point $q \in S_z$ is the orthogonal complement to the space spanned by the columns of $\nabla\xi(q)$, that is

$$T_qS_z = \{p \in \mathbb{R}^d : \; \nabla\xi(q)^T p = 0\},$$

while the co-tangent space for S_z is defined as

$$TS_z = \{(q,p) \in \mathbb{R}^{2d} : \; \xi(q) = z \; \text{ and } \; \nabla\xi(q)^T p = 0\}.$$

In words, the co-tangent space of S_z consists of points (q,p) in phase space such that $q \in S_z$ and $p \in T_qS_z$. The phase space is the set of $(q,p) \in \mathbb{R}^{2d}$ with the restriction that p must lie in the tangent space of the level surface $S_{\xi(q)}$, that is $p \in T_qS_{\xi(q)}$. For notational convenience, given a position $q \in \mathbb{R}^d$, we use $T_q \equiv T_qS_{\xi(q)}$. In what follows, we use $\sigma_{T_q}(dp)$ to denote the $n = d - m$ Hausdorff measure on $T_qS_{\xi(q)}$. The phase space distribution is then given by

$$\begin{aligned}\mu(dq,dp) &= Z^{-1}e^{-H(q,p)}\sigma_{T_q}(dp)dq, \\ &= \nu_\varepsilon(dq)\alpha_q(dp) \quad \text{where} \quad H(q,p) = V(q) + V_\varepsilon(q) + \frac{|p|^2}{2},\end{aligned} \tag{1.3}$$

and

$$\alpha_q(dp) = (2\pi)^{-n/2} e^{-\frac{|p|^2}{2}} \sigma_{T_q}(dp) \qquad \text{is supported on } T_qS_{\xi(q)}.$$

From (1.3) we see that the q-marginal of $\mu(dq,dp)$ is just the target distribution $\nu_\varepsilon(dq)$. In what follows, we will use $\sigma_z(dq)$ to denote the n-dimensional Hausdorff measure on S_z. By the co-area formula, for any test function φ we have

$$\int_{\mathbb{R}^d}\int_{T_q}\varphi(q,p)\mu(dq,dp) = Z^{-1}\int_{\mathbb{R}^m} e^{-\frac{|z|^2}{2\varepsilon^2}}\int_{S_z} e^{-V(q)}r(q)\int_{T_q}\varphi(q,p)e^{-\frac{|p|^2}{2}}\sigma_{T_q}(dp)\sigma_z(dq)dz, \tag{1.4}$$

where $r(q) = |\nabla\xi^T(q)\nabla\xi(q)|^{-1/2}$. Therefore, we conclude that

$$\mu(dq,dp) = \kappa_\varepsilon(dz)\nu_z(dq)\alpha_q(dp), \tag{1.5}$$

where

$$\begin{aligned}\nu_z(dq) &= Z_z^{-1}e^{-V(q)}r(q)\sigma_z(dq)\\ &= Z_z^{-1}e^{-\overline{V}(q)}\sigma_z(dq) \quad \text{supported on } S_z,\end{aligned}$$

with $\overline{V}(q) = V(q) - \log(r(q))$ and

$$\kappa_\varepsilon(dz) = (2\pi)^{-m/2}\varepsilon^{-m}e^{-\frac{|z|^2}{2\varepsilon^2}}dz \quad \text{supported on } \mathbb{R}^m.$$

Another way to write (1.5) is

$$\begin{aligned}&\mu(dq,dp) = \kappa_\varepsilon(dz)\bar{\mu}_z(dq,dp),\\ &\text{where } \ \bar{\mu}_z(dq,dp) = \nu_z(dq)\alpha_q(dp) \quad \text{supported on } TS_z.\end{aligned} \tag{1.6}$$

We want our MCMC sampler to be defined as a composition of transition kernels such that each kernel preserves the phase space distribution $\mu(dq,dp)$. If this holds, then the sampler (as a composition of measure-preserving kernels) also preserves $\mu(dq,dp)$. Intuitively, we would like to compose (1) a simple Gaussian Metropolis move to resample position q with (2) a momentum move in which we draw a new p from $\alpha_q(dp)$ with 3) a RATTLE move[1] on TS_z, were $z = \xi(q)$. The Gaussian Metropolis move should change the level "z" and make the Markov chain irreducible (and thus ergodic). The Metropolis move (1) and momentum resampling (2) obviously preserve $\mu(dq,dp)$. The RATTLE move (3) preserves $\bar{\mu}_z$ defined in (1.6), which is supported on the co-tangent bundle TS_z. In fact, (3) also preserves the phase space measure μ. To see why, let $R_z\big((q,p),(dq',dp')\big)$ be the transition kernel representing the RATTLE step on the co-tangent bundle TS_z. For any fixed $z \in \mathbb{R}^m$, R_z has to satisfy the following two requirements. First, $TS_z \ni (q,p) \mapsto R_z\big((q,p),A\big)$ has to define a measurable function on TS_z for any measurable set $A \subset TS_z$. In addition, for any $(q,p) \in TS_z$, we must have that $R_z\big((q,p),(dq',dp')\big)$ defines measure on the co-tangent bundle TS_z. With this in mind, given any function on phase space φ, we use $R_z\varphi$ to denote a new function on TS_z defined as

$$R_z f(q,p) = \int_{\mathbb{R}^{2d}} R_z\big((q,p),(dq',dp')\big)f(q',p').$$

Suppose that R_z preserves $\bar{\mu}_z$ in the sense that for any test function φ

$$\mathbb{E}_{\bar{\mu}_z}R_z\varphi = \mathbb{E}_{\bar{\mu}_z}\varphi \tag{1.7}$$

[1]Throughout this paper, by "RATTLE move" or "RATTLE step" we mean one execution of steps (*ii*) and (*iii*) of Algorithm 2 in [1], page 407.

Then, R_z also preserves the phase space distribution $\mu(dq, dp)$ since

$$\begin{aligned}\mathbb{E}_\mu R_z \varphi &= \int_{\mathbb{R}^m} \kappa_\varepsilon(dz) \int_{S_z} \int_{T_q} R_z \varphi(q,p) \alpha_q(dp) \nu_z(dq) \\ &= \int_{\mathbb{R}^m} \kappa_\varepsilon(dz) \mathbb{E}_{\bar{\mu}_z} R_z \varphi \\ &= \int_{\mathbb{R}^m} \kappa_\varepsilon(dz) \mathbb{E}_{\bar{\mu}_z} \varphi \\ &= \int_{\mathbb{R}^m} \kappa_\varepsilon(dz) \int_{S_z} \int_{T_q} \varphi(q,p) \alpha_q(dp) \nu_z(dq) \\ &= \mathbb{E}_\mu \varphi,\end{aligned}$$

where in the third equality we used (1.7). Thus, the RATTLE move (3) also preserves the phase space target. We can write

$$\bar{\mu}_z(dq, dp) = \overline{Z}_z^{-1} e^{-\overline{H}(q,p)} \sigma_z(dq) \sigma_{T_q}(dp)$$

$$\text{where} \quad \overline{H}(q,p) = \overline{V}(q) + \frac{|p|^2}{2} = V(q) - \log(r(q)) + \frac{|p|^2}{2}.$$

Overall, the algorithm consists in the following sequence of moves. Given a current state $(q^n, p^n) \in TS_{z^n} \subset \mathbb{R}^{2d}$, where $z^n = \xi(q^n)$, we do the following:

(1) Sample $q \in \mathbb{R}^d$ from $\nu_\varepsilon(dq)$ via isotropic Gaussian Metropolis;
(2) Sample $p \in T_q$ from the distribution $\alpha_q(dp)$;
(3) Use one step of RATTLE dynamics with momentum reversal and reverse projection step

$$(\tilde{q}, \tilde{p}) = \Psi_{\Delta t}^{\mathrm{rev}}(q, p);$$

(4) Accept the proposal $(\tilde{q}, \tilde{p})$ with probability A^n given by the Metropolis-Hastings ratio

$$\begin{aligned}A^n &= \exp\Big(- \overline{H}(\tilde{q}, \tilde{p}) + \overline{H}(q,p) \Big) \\ &= \exp\Big(- V(\tilde{q}) + V(q) - \frac{1}{2}|\tilde{p}|^2 + \frac{1}{2}|p|^2 \Big) \frac{r(\tilde{q})}{r(q)} \\ &= \exp\Big(- V(\tilde{q}) + V(q) - \frac{1}{2}|\tilde{p}|^2 + \frac{1}{2}|p|^2 \Big) \frac{\det(\nabla\xi^T(q)\nabla\xi(q))^{1/2}}{\det(\nabla\xi^T(\tilde{q})\nabla\xi(\tilde{q}))^{1/2}};\end{aligned} \tag{1.8}$$

To do this, draw a random varialbe $U^n \sim \mathrm{Unif}[0,1]$
- if $U^n \le A^n$, accept and set $(q^{n+1}, p^{n+1}) = (\tilde{q}, \tilde{p})$;
- otherwise, reject and set $(q^{n+1}, p^{n+1}) = (q, p)$.

The good thing about this algorithm is that the RATTLE dynamics are energy preserving, so it's possible to repeat step (3) any finite number of times before proceeding to step (4). This allows for larger time steps while maintaining a high acceptance probability in (4).

2 RATTLE integrator:

As pointed out in [1], the invariant measure of the Markov chain generated by steps (2)-(3)-(4) above is $\bar{\mu}_z$ whatever the potential function $\overline{V}$ which is used in the RATTLE integrator. In fact, the proof of reversibility of steps (3)-(4) only relies on the fact that $\Psi^{\text{rev}}_{\Delta t}$ is an involution which preserves the constrained (level z) phase space measure $\sigma_z(dq)\sigma_{T_q}(dp)$. Of course, in order to sample $\bar{\mu}_z$ it is fundamental to use the correct $\overline{V}$ in the Metropolis-Hastings ratio (1.8). We can now write down the RATTLE integrator with $\overline{V} = 0$. Given (q,p), where $z = \xi(q)$, use the following system of equations to solve for (q^1, p^1)

$$\begin{aligned} p^{\frac{1}{2}} &= p + \nabla\xi(q)\lambda^{\frac{1}{2}}, \\ q^1 &= q + \Delta t\ p^{\frac{1}{2}}, \\ \xi(q^1) &= z, \\ p^1 &= p^{\frac{1}{2}} + \nabla\xi(q^1)\lambda^1, \\ \Big[\nabla\xi(q^1)\Big]^T p^1 &= 0. \end{aligned} \tag{2.1}$$

The following set of steps uses the RATTLE integrator (2.1) to produce the operator $\Psi^{\text{rev}}_{\Delta t}$, which includes momentum reversal and reverse check. Given (q,p), where $z = \xi(q)$, do the following

(1) First, find $\lambda^{\frac{1}{2}}$ such that

$$\xi\Big(q + \Delta t\ p + \Delta t\ \nabla\xi(q)\lambda^{\frac{1}{2}}\Big) = z; \tag{2.2}$$

(2) If step (1) has a solution $\lambda^{\frac{1}{2}}$ (see below for more details), set

$$\begin{aligned} p^{\frac{1}{2}} &= p + \nabla\xi(q)\lambda^{\frac{1}{2}}, \\ q^1 &= q + \Delta t\ p + \Delta t\ \nabla\xi(q)\lambda^{\frac{1}{2}}; \end{aligned}$$

otherwise, interrupt the algorithm and return (q,p).

(3) Find λ^1 that solves the linear system

$$\nabla\xi^T(q^1)\nabla\xi(q^1)\lambda^1 = -\nabla\xi^T(q^1)p^{\frac{1}{2}}; \tag{2.3}$$

(4) Set

$$p^1 = -\Big(p^{\frac{1}{2}} + \nabla\xi(q^1)\lambda^1\Big),$$

where the minus sign is the momentum reversal.

(5) Given (q^1,p^1), compute the configuration (q^2,p^2) obtained by one step of the RATTLE integrator (2.1) starting from (q^1,p^1). If $(q^2,-p^2) = (q^1,p^1)$, return (q^1,p^1); otherwise return (q,p).

Note that the q^1 obtained from steps (1)-(2) corresponds to a Surface sampler move (take a step in T_q and project back onto the surface S_z in a direction that is orthogonal to T_q itself). Additionally, steps (3)-(4) correspond to projecting $p^{\frac{1}{2}}$ onto the new tangent space T_{q^1}. The momentum projection in step (3) is always well defined, since it is a linear projection onto the vector space T_{q^1}. However, the position projection in step (1) is non-linear and it may have no solution at all. This is why we need the reverse check in step (5) to maintain reversibility.

3 Some notes:

For the RATTLE dynamics we need to have $\nabla\overline{V}(q)$ which involves $\nabla\log(r(q)) = \log(\det(\nabla\xi^T(q)\nabla\xi(q))^{-1/2})$. For a square matrix $G(t)$, we can use Jacobi's formula to compute

$$\frac{d}{dt}\det(G(t)) = \det(G(t))\,\mathrm{Tr}\left(G(t)^{-1}\frac{dG(t)}{dt}\right).$$

In particular, let $G(q) = \nabla\xi^T(q)\nabla\xi(q)$ be our Gram matrix. Then, by Jacobi's formula we have

$$\frac{d}{dq_i}\det(G(q)) = \det(G(q))\,\mathrm{Tr}\left(\big(G(q)\big)^{-1}\frac{dG(q)}{dq_i}\right),$$

where by the product rule

$$\frac{d}{dq_i}G(q) = \frac{d\nabla\xi^T(q)}{dq_i}\cdot\nabla\xi(q) + \nabla\xi^T(q)\cdot\frac{d\nabla\xi(q)}{dq_i}$$

References

[1] Tony Lelièvre, Mathias Rousset, and Gabriel Stoltz. *Hybrid Monte Carlo methods for sampling probability measures on submanifolds.* 2019 (cit. on pp. 2, 4).