Reflection, Refraction, and Hamiltonian Monte Carlo

Hadi Mohasel Afshar
Dept. Computer Science & Engineering
Australian National University
Canberra, ACT 0200
hadi.afshar@anu.edu.au

Justin Domke
NICTA and Australian National University
Justin.Domke@nicta.com.au

Abstract

Hamiltonian Monte Carlo (HMC) is a successful approach for sampling from continuous densities. However, it has difficulty simulating Hamiltonian dynamics with non-smooth functions, leading to poor performance. This paper is motivated by the behavior of Hamiltonian dynamics in physical systems like optics. We introduce a modification of the Leapfrog discretization of Hamiltonian dynamics on piecewise continuous energies, where intersections of the trajectory with discontinuities are detected, and the momentum is *reflected* or *refracted* to compensate for the change in energy. We prove that this method preserves the correct stationary distribution when boundaries are affine. Experiments show that by reducing the number of rejected samples, this method improves on traditional HMC.

1 Introduction

Markov chain Monte Carlo sampling is among the most general methods for probabilistic inference. When the probability distribution is smooth, Hamiltonian Monte Carlo (HMC) (originally called hybrid Monte Carlo [4]) provides a valuable sampling device since it avoids correlated parameters by simulating the gradient-aware Hamiltonian dynamics and reduces random walk behavior. This often leads to a rapid exploration of the distribution [7, 2]. HMC has recently become popular in Bayesian statistical inference [13], and is often the algorithm of choice.

Some problems display *piecewise* smoothness, where the density is differentiable except at certain boundaries. Firstly, probabilistic models may intrinsically have finite support, being constrained to some region. In Bayesian inference, it might be convenient to state a piecewise prior. More complex and highly piecewise distributions emerge in applications where the distributions are derived from other distributions (e.g. the distribution of the product of two continuous random variables [5]) as well as applications such as preference learning [1], or probabilistic programming [8].

While HMC is motivated by smooth distributions, the inclusion of an acceptance probability means HMC does asymptotically sample from piecewise distributions[1] as well. However, since leapfrog numerical integration of Hamiltonian dynamics (see [9]) relies on the assumption that the corresponding potential energy is smooth, it such cases it leads to high rejection probabilities, and poor performance. Hence, traditional HMC is rarely used for piecewise distributions.

In physical systems that follow Hamiltonian dynamics [6], a discontinuity in the energy can result in two possible behaviors. If the energy decreases across a discontinuity, or the momentum is large enough to overcome an increase, the system will cross the boundary with an instantaneous change in momentum, known as *refraction* in the context of optics [3]. If the change in energy is too large to be overcome by the momentum, the system will *reflect* off the boundary, again with an instantaneous change in momentum.

---

[1]Technically, here we assume the total measure of the non-differentiable points is zero so that, with probability one, none is ever encountered
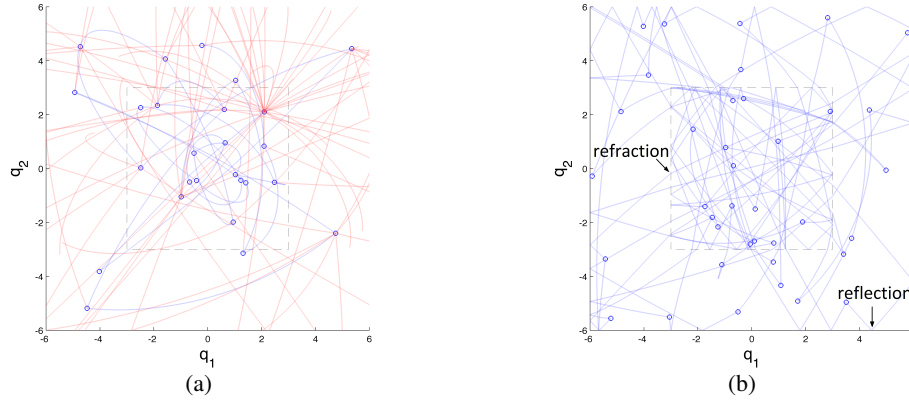
Figure 1: Trajectories of the rejected (red) and accepted (blue) Hamiltonian proposals using (a) baseline and (b) reflective HMC (RHMC) both tuned with leapfrog parameters $L = 100$ and $\epsilon = 0.1$ in the 2D model described in Section 6. In RHMC, the trajectory reflects or refracts on the boundaries of the internal and external polytope boundaries and ends very few proposal rejections (0 for the depicted samples) while most proposals generated by baseline HMC are in the 0-probability region and all rejected.

Recently, Pakman et al. [11, 10] proposed methods for HMC-based sampling from piecewise Gaussian distributions by exactly solving the Hamiltonian equations, and accounting for what we refer to as refraction and reflection above. However, since Hamiltonian equations of motion can rarely be solved exactly, the applications of this method are restricted to distributions whose log-density is piecewise quadratic.

In this paper, we generalize this work to arbitrary piecewise continuous distributions, where each region is a polytope, i.e. is determined by a set of affine boundaries. We introduce a modification to the leapfrog numerical simulation of Hamiltonian dynamics, called reflective HMC (RHMC), by incorporating reflection and refraction via detecting the first intersection of a linear trajectory with a boundary. We prove that our method has the correct stationary distribution, where the main technical difficulty is proving volume preservation of our dynamics to establish detailed balance. Numerical experiments confirm that our method is more efficient than baseline HMC, due to having fewer rejected proposal trajectories, particularly in high dimensions. As mentioned, the main advantage of this method over [11] and [10] is that it can be applied to arbitrary piecewise densities, without the need for a closed-form solution to the Hamiltonian dynamics, greatly increasing the scope of applicability.

2 Exact Hamiltonian Dynamics

Consider a distribution $P(\mathbf{q}) \propto \exp(-U(\mathbf{q}))$, where U is the *potential energy*. HMC [9] is based on considering a joint distribution on momentum and position space $P(\mathbf{q}, \mathbf{p}) \propto \exp(-H(\mathbf{q}, \mathbf{p}))$, where $H(\mathbf{q}, \mathbf{p}) = U(\mathbf{q}) + K(\mathbf{p})$, and K is a quadratic, meaning that $P(\mathbf{p}) \propto \exp(-K(\mathbf{p}))$ is a normal distribution. If one could exactly simulate the dynamics, HMC would proceed by (1) iteratively sampling $\mathbf{p} \sim P(\mathbf{p})$, (2) simulating the Hamiltonian dynamics

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = \frac{\partial H}{\partial p_i} = p_i \qquad (1) \qquad\qquad \frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\partial H}{\partial q_i} = -\frac{\partial U}{\partial q_i} \qquad (2)$$

for some period of time ϵ, and (3) reversing the final value $\mathbf{p}$. (Only needed for the proof of correctness, since this will be immediately discarded at the start of the next iteration in practice.) Since steps (1) and (2-3) both leave the distribution $P(\mathbf{p}, \mathbf{q})$ invariant, so does a Markov chain that alternates between the two steps. Hence, the dynamics have $P(\mathbf{p}, \mathbf{q})$ as a stationary distribution. Of course, the above differential equations are not well-defined when U has discontinuities, and are typically difficult to solve in closed-form.

3 Reflection and Refraction with Exact Hamiltonian Dynamics

Take a potential function $U(\mathbf{q})$ (over $\mathbb{R}^n$) which is differentiable in all points except at some boundaries of partitions. Suppose that, when simulating the Hamiltonian dynamics, $(\mathbf{q}, \mathbf{p})$ evolves over time as in the above equations whenever these equations are differentiable. However, when the state reaches a boundary, decompose the momentum vector $\mathbf{p}$ into a component $\mathbf{p}_\perp$ perpendicular to the boundary and a component $\mathbf{p}_\parallel$ parallel to the boundary. Let ΔU be the (signed) difference in potential energy on the two sides of the discontinuity. If $\|\mathbf{p}_\perp\|^2 > 2\Delta U$ then $\mathbf{p}_\perp$ is instantaneously replaced by $\mathbf{p}'_\perp := \sqrt{\|\mathbf{p}_\perp\|^2 - 2\Delta U} \cdot \frac{\mathbf{p}_\perp}{\|\mathbf{p}_\perp\|}$. That is, the discontinuity is passed, but the momentum is changed in the direction perpendicular to the boundary (refraction). (If ΔU is positive, the momentum will decrease, and if it is negative, the momentum will increase.) On the other hand, if $\|\mathbf{p}_\perp\|^2 \leq 2\Delta U$, then $\mathbf{p}_\perp$ is instantaneously replaced by $-\mathbf{p}_\perp$. That is, if the particle energy is insufficient to climb the potential boundary, it bounces back by reversing the momentum component which is perpendicular to the boundary.

Pakman and Paninski [11, 10] present an algorithm to exactly solve these dynamics for quadratic U. However, for non-quadratic U, the Hamiltonian dynamics rarely have a closed-form solution, and one must resort to numerical integration, the most succesful method for which is known as the *leapfrog* dynamics.

4 Reflection and Refraction with Leapfrog Dynamics

Informally, HMC with leapfrog dynamics iterates three steps. (1) Sample $\mathbf{p} \sim P(\mathbf{p})$. (2) Perform leapfrog simulation, by discretizing the Hamiltonian equations into L steps using some small step-size ϵ. Here, one interleaves a *position step* $\mathbf{q} \leftarrow \mathbf{q} + \epsilon\mathbf{p}$ between two half *momentum* steps $\mathbf{p} \leftarrow \mathbf{p} - \epsilon\nabla U(\mathbf{q})/2$. (3) Reverse the sign of $\mathbf{p}$. If $(\mathbf{q}, \mathbf{p})$ is the starting point of the leapfrog dynamics, and $(\mathbf{q}', \mathbf{p}')$ is the final point, *accept* the move with probability $\min(1, \exp(H(\mathbf{p}, \mathbf{q}) - H(\mathbf{p}', \mathbf{q}')))$ See Algorithm 1.

It can be shown that this baseline HMC method has detailed balance with respect to $P(\mathbf{p})$, even if $U(\mathbf{q})$ is discontinuous. However, discontinuities mean that large changes in the Hamiltonian may occur, meaning many steps can be rejected. We propose a modification of the dynamics, namely, *reflective Hamiltonian Monte Carlo* (RHMC), which is also shown in Algorithm 1.

The only modification is applied to the position steps: In RHMC, the first intersection of the trajectory with the boundaries of the polytope that contains $\mathbf{q}$ must be detected [11, 10]. The position step is only taken up to this boundary, and reflection/refraction occurs, depending on the momentum and change of energy at the boundary. This process continues until the entire amount of time ϵ has been simulated. Note that if there is no boundary in the trajectory to time ϵ, this is equivalent to baseline HMC. Also note that several boundaries might be visited in one position step.

As with baseline HMC, there are two alternating steps, namely drawing a new momentum variable $\mathbf{p}$ from $P(\mathbf{p}) \propto \exp(-K(\mathbf{p}))$ and proposing a move $(\mathbf{p}, \mathbf{q}) \rightarrow (\mathbf{p}', \mathbf{q}')$ and accepting or rejecting it with a probability determined by a Metropolis-Hastings ratio. We can show that both of these steps leave the joint distribution P invariant, and hence a Markov chain that also alternates between these steps will also leave P invariant.

As it is easy to see, drawing $\mathbf{p}$ from $P(\mathbf{p})$ will leave $P(\mathbf{q}, \mathbf{p})$ invariant, we concentrate on the second step i.e. where a move is proposed according to the piecewise leapfrog dynamics shown in Alg. 1. Firstly, it is clear that these dynamics are time-reversible, meaning that if the simulation takes state $(\mathbf{q}, \mathbf{p})$ to $(\mathbf{q}', \mathbf{p}')$ it will also take state $(\mathbf{q}', \mathbf{p}')$ to $(\mathbf{q}, \mathbf{p})$. Secondly, we will show that these dynamics are volume preserving. Formally, if $\mathcal{D}$ denotes the leapfrog dynamics, we will show that the absolute value of the determinant of the Jacobian of $\mathcal{D}$ is one. These two properties together show that the probability density of proposing a move from $(\mathbf{q}, \mathbf{p})$ to $(\mathbf{q}', \mathbf{p}')$ is the same of that proposing a move from $(\mathbf{q}', \mathbf{p}')$ to $(\mathbf{q}, \mathbf{p})$. Thus, if the move $(\mathbf{q}, \mathbf{p}) \rightarrow (\mathbf{q}', \mathbf{p}')$ is accepted according to the standard Metropolis-Hastings ratio, $R\left((\mathbf{q}, \mathbf{p}) \rightarrow (\mathbf{q}', \mathbf{p}')\right) = \min(1, \exp(H(\mathbf{q}, \mathbf{p}) - H(\mathbf{q}', \mathbf{p}')))$, then detailed balance will be satisfied. To see this, let Q denote the proposal distribution, Then, the usual proof of correctness for Metropolis-Hastings applies, namely that

$$\frac{P(\mathbf{q},\mathbf{p})Q\left((\mathbf{q},\mathbf{p})\to(\mathbf{q}',\mathbf{p}')\right)R\left((\mathbf{q},\mathbf{p})\to(\mathbf{q}',\mathbf{p}')\right)}{P(\mathbf{q}',\mathbf{p}')Q((\mathbf{q}',\mathbf{p}')\to(\mathbf{q},\mathbf{p}))R((\mathbf{q}',\mathbf{p}')\to(\mathbf{q},\mathbf{p}))} = \frac{P(\mathbf{q},\mathbf{p})\min(1,\exp(H(\mathbf{q},\mathbf{p})-H(\mathbf{q}',\mathbf{p}')))}{P(\mathbf{q}',\mathbf{p}')\min(1,\exp(H(\mathbf{q}',\mathbf{p}')-H(\mathbf{q},\mathbf{p})))} = 1.$$

(The final equality is easy to establish, considering the cases where $H(\mathbf{q},\mathbf{p}) \geq H(\mathbf{q}',\mathbf{p}')$ and $H(\mathbf{q}',\mathbf{p}') \leq H(\mathbf{q},\mathbf{p})$ separately.) This means that detailed balance holds, and so P is a stationary distribution.

The major difference in the analysis of RHMC, relative to traditional HMC is that showing conservation of volume is more difficult. With standard HMC and leapfrog steps, volume conservation is easy to show by observing that each part of a leapfrog step is a shear transformation. This is not the case with RHMC, and so we must resort to a full analysis of the determinant of the Jacobian, as explored in the following section.

Algorithm 1: BASELINE & REFLECTIVE HMC ALGORITHMS

input : $\mathbf{q}_0$, current sample; U, potential function,
L, number of leapfrog steps; ϵ, leapfrog step size
output: next sample

1 **begin**
2     $\mathbf{q} \leftarrow \mathbf{q}_0$; $\mathbf{p} \sim \mathcal{N}(0,1)$
3     $H_0 \leftarrow \|\mathbf{p}\|^2/2 + U(\mathbf{q})$
4   **for** $l = 1$ **to** L **do**
5         $\mathbf{p} \leftarrow \mathbf{p} - \epsilon\nabla U(\mathbf{q})/2$ *# Half-step evolution of momentum*
6   *# Full-step evolution of position*:
7   **if** BASELINEHMC **then**
8            $\mathbf{q} \leftarrow \mathbf{q} + \epsilon\mathbf{p}$
9   **else**
10   *# (i.e. if* REFLECTIVEHMC*):*
11           $t_0 \leftarrow 0$
12   **while** $\left(\langle \boldsymbol{x}, t_x, \Delta U, \phi\rangle \leftarrow \text{FIRSTDISCONTINUITY}(\boldsymbol{q},\boldsymbol{p},\epsilon-t_0,U)\right) \neq \emptyset$ **do**
13              $\mathbf{q} \leftarrow \mathbf{x}$
14              $t_0 \leftarrow t_0 + t_x$
15              $\langle\mathbf{p}_\perp, \mathbf{p}_\|\rangle = \text{DECOMPOSE}(\mathbf{p},\phi)$ *# perpendicular/ parallel to boundary plane ϕ*
16   **if** $\|\boldsymbol{p}_\perp\|^2 > 2\Delta U$ **then**
17                 $\mathbf{p}_\perp \leftarrow \sqrt{\|\mathbf{p}_\perp\|^2 - 2\Delta U} \cdot \frac{\mathbf{p}_\perp}{\|\mathbf{p}_\perp\|}$ *# refraction*
18   **else**
19                 $\mathbf{p}_\perp \leftarrow -\mathbf{p}_\perp$ *# reflection*
20
21              $\mathbf{p} \leftarrow \mathbf{p}_\perp + \mathbf{p}_\|$
22           $\mathbf{q} \leftarrow \mathbf{q} + (\epsilon - t_0)\mathbf{p}$
23
24        $\mathbf{p} \leftarrow \mathbf{p} - \epsilon\nabla U(\mathbf{q})/2$ *# Half-step evolution of momentum*
25    $\mathbf{p} \leftarrow -\mathbf{p}$ *# not required in practice; for proof of reversibility*
26    $H \leftarrow \|\mathbf{p}\|^2/2 + U(\mathbf{q})$; $\Delta H \leftarrow H - H_0$
27   **if** $s \sim \mathcal{U}(0,1) < e^{-\Delta H}$ **return q else** **return** $\mathbf{q}_0$
28 **end**

---

note : FIRSTDISCONTINUITY$(\cdot)$ returns **x**, the position of the first intersection of a boundary plain with line segment $[\mathbf{q}, \mathbf{q} + (\epsilon - t_0)\mathbf{p}]$; t_x, the time it is visited; $\Delta U = U(\mathbf{x}^+)$ - $U(\mathbf{x}^-)$ and ϕ, the visited partition boundary. If no such point exists, $\emptyset$ is returned.

5 Volume Conservation

5.1 Refraction

In our first result, we assume without loss of generality, that there is a boundary located at the hyperplane $q_1 = c$. This Lemma shows that, in the refractive case, volume is conserved.

Lemma 1. *Let $\mathcal{T} : \langle \boldsymbol{q}, \boldsymbol{p} \rangle \to \langle \boldsymbol{q}', \boldsymbol{p}' \rangle$ be a transformation in $\mathbb{R}^n$ that takes a unit mass located at $\boldsymbol{q} := (q_1, \ldots, q_n)$ and moves it with constant momentum $\boldsymbol{p} := (p_1, \ldots, p_n)$ till it reaches a plane $q_1 = c$ (at some point $\boldsymbol{x} := (c, x_2, \ldots, x_n)$ where c is a constant). Subsequently the momentum is changed to $\boldsymbol{p}' = \left(\sqrt{p_1^2 - 2\Delta U(\boldsymbol{x})}, p_2, \ldots, p_n \right)$ (where $\Delta U(\cdot)$ is a function of $\boldsymbol{x}$ s.t. $p_1^2 > 2\Delta U(\boldsymbol{x})$). The move is carried on for the total time period τ till it ends in $\boldsymbol{q}'$. For all $n \in \mathbb{N}$, $\mathcal{T}$ satisfies the volume preservation property.*

Proof. Since for $i > 1$, the momentum is not affected by the collision, $q_i' = q_i + \tau \cdot p_i$ and $p_i' = p_i$. Thus,

$$\forall j \in \{2, \ldots, n\} \, s.t. \, j \neq i, \qquad \frac{\partial q_i'}{\partial q_j} = \frac{\partial q_i'}{\partial p_j} = \frac{\partial p_i'}{\partial q_j} = \frac{\partial p_i'}{\partial p_j} = 0.$$

Therefore, if we explicitly write out the Jacobian of $\mathcal{T}$, it is

$$|J| = \begin{vmatrix} \frac{\partial q_1'}{\partial q_1} & \frac{\partial q_1'}{\partial p_1} & \cdots & \frac{\partial q_1'}{\partial p_{k-1}} & \frac{\partial q_1'}{\partial q_k} & \frac{\partial q_1'}{\partial p_k} \\ \frac{\partial p_1'}{\partial q_1} & \frac{\partial p_1'}{\partial p_1} & \cdots & \frac{\partial p_1'}{\partial p_{k-1}} & \frac{\partial p_1'}{\partial q_k} & \frac{\partial p_1'}{\partial p_k} \\ \frac{\partial q_2'}{\partial q_1} & \frac{\partial q_2'}{\partial p_1} & \cdots & \frac{\partial q_2'}{\partial p_{k-1}} & \frac{\partial q_2'}{\partial q_k} & \frac{\partial q_2'}{\partial p_k} \\ \vdots & & \ddots & \vdots & & \vdots \\ \frac{\partial p_{k-1}'}{\partial q_1} & \frac{\partial p_{k-1}'}{\partial p_1} & \cdots & \frac{\partial p_{k-1}'}{\partial p_{k-1}} & \frac{\partial p_{k-1}'}{\partial q_k} & \frac{\partial p_{k-1}'}{\partial p_k} \\ \frac{\partial q_k'}{\partial q_1} & \frac{\partial q_k'}{\partial p_1} & \cdots & \frac{\partial q_k'}{\partial p_{k-1}} & \frac{\partial q_k'}{\partial q_k} & \frac{\partial q_k'}{\partial p_k} \\ \frac{\partial p_k'}{\partial q_1} & \frac{\partial p_k'}{\partial p_1} & \cdots & \frac{\partial p_k'}{\partial p_{k-1}} & \frac{\partial p_k'}{\partial q_k} & \frac{\partial p_k'}{\partial p_k} \end{vmatrix} = \begin{vmatrix} \frac{\partial q_1'}{\partial q_1} & \frac{\partial q_1'}{\partial p_1} & \cdots & \frac{\partial q_1'}{\partial p_{k-1}} & \frac{\partial q_1'}{\partial q_k} & \frac{\partial q_1'}{\partial p_k} \\ \frac{\partial p_1'}{\partial q_1} & \frac{\partial p_1'}{\partial p_1} & \cdots & \frac{\partial p_1'}{\partial p_{k-1}} & \frac{\partial p_1'}{\partial q_k} & \frac{\partial p_1'}{\partial p_k} \\ 0 & 0 & \cdots & \frac{\partial q_2'}{\partial p_{k-1}} & \frac{\partial q_2'}{\partial q_k} & \frac{\partial q_2'}{\partial p_k} \\ \vdots & & \ddots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & \frac{\partial p_{k-1}'}{\partial q_k} & \frac{\partial p_{k-1}'}{\partial p_k} \\ 0 & 0 & \cdots & 0 & 1 & \frac{\partial q_k'}{\partial p_k} \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{vmatrix} \tag{3}$$

Now, using standard properties of the determinant, we have that $|J| = \begin{vmatrix} \frac{\partial q_1'}{\partial q_1} & \frac{\partial q_1'}{\partial p_1} \\ \frac{\partial p_1'}{\partial q_1} & \frac{\partial p_1'}{\partial p_1} \end{vmatrix}$.

We will now explicitly calculate these four derivatives. Due to the significance of the implied result, we carry out the computations in detail. Nonetheless, as this is a largely mechanical process, for brevity, we do not comment on the derivation.

Let t_1 be the time to reach $\mathbf{x}$ and t_2 be the period between reaching $\mathbf{x}$ and the last point $\mathbf{q}'$. Then:

$$t_1 \overset{\text{def}}{=} \frac{c - q_1}{p_1} \quad (4) \qquad\qquad \mathbf{x} = \mathbf{q} + t_1 \mathbf{p} \quad (5) \qquad\qquad t_2 \overset{\text{def}}{=} \tau - t_1 = \tau + \frac{q_1 - c}{p_1} \quad (6)$$

$$q_1' = c + p_1' \cdot t_2 \quad (7) \qquad p_1' \overset{\text{def}}{=} \sqrt{p_1^2 - 2\Delta U(\mathbf{x})} \quad (8) \qquad\qquad \frac{\partial t_2}{\partial q_1} \overset{\text{by (6)}}{=} \frac{1}{p_1} \quad (9)$$

$$\frac{\partial q_1'}{\partial q_1} = \frac{\partial q_1'}{\partial p_1'} \cdot \frac{\partial p_1'}{\partial q_1} + \frac{\partial q_1'}{\partial t_2} \cdot \frac{\partial t_2}{\partial q_1} \overset{(7 \, \& \, 9)}{=} t_2 \cdot \frac{\partial p_1'}{\partial q_1} + p_1' \cdot \frac{1}{p_1} \tag{10}$$

$$\frac{\partial q_1'}{\partial p_1} = \frac{\partial q_1'}{\partial p_1'} \cdot \frac{\partial p_1'}{\partial p_1} + \frac{\partial q_1'}{\partial t_2} \cdot \frac{\partial t_2}{\partial p_1} \overset{(6 \, \& \, 7)}{=} t_2 \cdot \frac{\partial p_1'}{\partial p_1} + p_1' \cdot \frac{c - q_1}{p_1^2} \tag{11}$$

$$\frac{\partial p_1'}{\partial p_1} \overset{(8)}{=} \frac{1}{2\sqrt{p_1^2 - 2\Delta U(\mathbf{x})}} \cdot \frac{\partial \left(p_1^2 - 2\Delta U(\mathbf{x}) \right)}{\partial p_1} = \frac{p_1 - \partial \Delta U(\mathbf{x}) / \partial p_1}{p_1'} \tag{12}$$

$$\frac{\partial p_1'}{\partial q_1} = \frac{1}{2\sqrt{p_1^2 - 2\Delta U(\mathbf{x})}} \cdot \frac{\partial \left(p_1^2 - 2\Delta U(\mathbf{x})\right)}{\partial q_1} = \frac{1}{p_1'} \cdot \frac{-\partial \Delta U(\mathbf{x})}{\partial q_1} \tag{13}$$

$$\frac{\partial \mathbf{x}}{\partial p_1} \stackrel{(4,\,5)}{=} \frac{\partial \left(\mathbf{q} + \frac{c-q_1}{p_1}\mathbf{p}\right)}{\partial p_1} = \frac{\partial \mathbf{q}}{\partial p_1} + (c - q_1) \cdot \frac{\partial(\mathbf{p}/p_1)}{\partial p_1} = (c - q_1)\frac{-1}{p_1^2} \cdot (0, p_2, p_3, \dots, p_n) \tag{14}$$

$$\frac{\partial \mathbf{x}}{\partial q_1} \stackrel{(4,\,5)}{=} \frac{\partial \mathbf{q}}{\partial q_1} + \frac{\mathbf{p}}{p_1} \cdot \frac{\partial(c - q_1)}{\partial q_1} = \frac{\mathbf{q}}{q_1} - \frac{\mathbf{p}}{p_1} = (1, 0, \dots, 0) - (1, \frac{p_2}{p_1}, \dots, \frac{p_n}{p_1}) = \frac{-1}{p_1}(0, p_2, \dots, p_n) \tag{15}$$

Substituting the appropriate terms into $|J| = |\frac{\partial q_1'}{\partial q_1}\frac{\partial p_1'}{\partial p_1} - \frac{\partial p_1'}{\partial q_1}\frac{\partial q_1'}{\partial p_1}|$, we get that

$$
\begin{aligned}
|J| &\stackrel{(3)}{=} \frac{\partial q_1'}{\partial q_1} \cdot \frac{\partial p_1'}{\partial p_1} - \frac{\partial q_1'}{\partial p_1} \cdot \frac{\partial p_1'}{\partial q_1} \stackrel{(10 \, \& \, 11)}{=} \left(t_2 \frac{\partial p_1'}{\partial q_1} + \frac{p_1'}{p_1}\right) \cdot \frac{\partial p_1'}{\partial p_1} - \left(t_2 \frac{\partial p_1'}{\partial p_1} + p_1' \frac{c - q_1}{p_1^2}\right) \cdot \frac{\partial p_1'}{\partial q_1} \\
&= \frac{p_1'}{p_1}\left(\frac{\partial p_1'}{\partial p_1} + \frac{q_1 - c}{p_1} \cdot \frac{\partial p_1'}{\partial q_1}\right) \stackrel{(12 \, \& \, 13)}{=} \frac{1}{p_1}\left(p_1 - \frac{\partial \Delta U(\mathbf{x})}{\partial p_1} - \frac{q_1 - c}{p_1} \cdot \frac{\partial \Delta U(\mathbf{x})}{\partial q_1}\right) \\
&= 1 - \frac{1}{p_1}\left(\frac{\partial \Delta U(\mathbf{x})}{\partial \mathbf{x}} \cdot \frac{\partial \mathbf{x}}{\partial p_1} + \frac{q_1 - c}{p_1} \cdot \frac{\partial \Delta U(\mathbf{x})}{\partial \mathbf{x}} \cdot \frac{\partial \mathbf{x}}{\partial q_1}\right) \\
&\stackrel{(14 \, \& \, 15)}{=} 1 - \frac{1}{p_1} \cdot \frac{\partial \Delta U(\mathbf{x})}{\partial \mathbf{x}}\left(\frac{q_1 - c}{p_1^2} \cdot (0, p_2, p_3, \dots, p_n) + \frac{q_1 - c}{p_1} \cdot \frac{-1}{p_1}(0, p_2, \dots, p_n)\right) = 1.
\end{aligned}
$$

$\square$

5.2 Reflection

Now, we turn to the reflective case, and again show that volume is conserved. Again, we assume without loss of generality that there is a boundary located at the hyperplane $q_1 = c$.

Lemma 2. *Let $\mathcal{T} : \langle \boldsymbol{q}, \boldsymbol{p} \rangle \to \langle \boldsymbol{q}', \boldsymbol{p}' \rangle$ be a transformation in $\mathbb{R}^n$ that takes a unit mass located at $\boldsymbol{q} := (q_1, \dots, q_n)$ and moves it with the constant momentum $\boldsymbol{p} := (p_1, \dots, p_n)$ till it reaches a plane $q_1 = c$ (at some point $\boldsymbol{x} := (c, x_2, \dots, x_n)$ where c is a constant). Subsequently the mass is bounced back (reflected) with momentum $\boldsymbol{p}' = (-p_1, p_2, \dots, p_n)$ The move is carried on for a total time period τ till it ends in $\boldsymbol{q}'$. For all $n \in \mathbb{N}$, $\mathcal{T}$ satisfies the volume preservation property.*

Proof. Similar to Lemma 1, for $i > 1$, $q_i' = q_i + \tau \cdot p_i$ and $p_i' = p_i$. Therefore, for any $j \in \{2, \dots, n\}$ s.t. $j \neq i$, $\frac{\partial q_i'}{\partial q_j} = \frac{\partial q_i'}{\partial p_j} = \frac{\partial p_i'}{\partial q_j} = \frac{\partial p_i'}{\partial p_j} = 0$. Consequently, by equation (3), and since $p_1' = -p_1$,

$$|J| = \begin{vmatrix} \frac{\partial q_1'}{\partial q_1} & \frac{\partial q_1'}{\partial p_1} \\ \frac{\partial p_1'}{\partial q_1} & \frac{\partial p_1'}{\partial p_1} \end{vmatrix} = \begin{vmatrix} \frac{\partial q_1'}{\partial q_1} & \frac{\partial q_1'}{\partial p_1} \\ 0 & -1 \end{vmatrix} = \frac{-\partial q_1'}{\partial q_1} \tag{16}$$

As before, let t_1 be the time to reach $\mathbf{x}$ and t_2 be the period between reaching $\mathbf{x}$ and the last point $\boldsymbol{q}'$. That is, $t_1 \stackrel{\text{def}}{=} \frac{c-q_1}{p_1}$ and $t_2 \stackrel{\text{def}}{=} \tau - t_1$. It follows that $q_1' \stackrel{\text{def}}{=} c + p_1' \cdot t_2$ is equal to $2c - \tau p_1 - q_1$. Hence, $|J| = 1$. $\square$

5.3 Reflective Leapfrog Dynamics

Theorem 1. *In RHMC (Algorithm 1) for sampling from a continuous and piecewise distribution P which has affine partitioning boundaries, leapfrog simulation preserves volume in $(\boldsymbol{q}, \boldsymbol{p})$ space.*

Proof. We split the algorithm into several atomic transformations $\mathcal{T}_i$. Each transformation is either (a) a momentum step, or (b) a full position step with no reflection/refraction or (c) a full or partial position step where exactly one reflection or refraction occurs.

To prove that the total algorithm preserves volume, it is sufficient to show that the volume is preserved under each $\mathcal{T}_i$ (i.e. $|J_{\mathcal{T}_i}(\mathbf{q}, \mathbf{p})| = 1$) since:

$$|J_{\mathcal{T}_1 o \mathcal{T}_2 o \cdots o \mathcal{T}_m}| = |J_{\mathcal{T}_1}| \cdot |J_{\mathcal{T}_2}| \cdots |J_{\mathcal{T}_m}|$$

Transformations of kind (a) and (b) are shear mappings and therefore, they preserve the volume [9]. Now consider a (full or part of) position step where a single refraction occurs. If the reflective plane is in form $q_1 = c$, by lemma 1, the volume preservation property holds. Otherwise, as long as the reflective plane is affine, via a rotation of basis vectors, the problem is reduced to the former case. Since volume is conserved under rotation, in this case the volume is conserved also. With a same reasoning, by lemma 2, reflection on a affine reflective boundary preserves volume. Thus, since all RHMC leapfrog simulation transformations preserve volume, the proof is complete. □

Along with the fact that the leapfrog dynamics are time-reversible, this shows that the algorithm satisfies detailed balance, and so has the correct stationary distribution.

6 Experiment

Compared to baseline HMC, we expect that RHMC will simulate Hamiltonian dynamics more accurately and therefore leads to fewer rejected samples. On the other hand, this comes at the expense of slower leapfrog position steps since intersections, reflections and refractions must be computed. To test the trade off we compare the RHMC to baseline HMC [9] and tuned Metroplis-Hastings (MH) with a simple isotropic Normal proposal distribution. MH is automatically tuned after [12] by testing 100 equidistant proposal variances in interval $(0, 1]$ and accepting a variance for which the acceptance rate is closest to 0.24. The baseline HMC and RHMC number of steps L and step size ϵ are chosen to be 100 and 0.1 respectively[2]

The comparison takes place on a heavy tail piecewise model with (non-normalized) negative log probability

$$U(\mathbf{q}) = \begin{cases} \sqrt{\mathbf{q}^\top A \, \mathbf{q}} & \text{if } |q_i| \leq 3 & \text{for } i = 1, \ldots, n \\ 1 + \sqrt{\mathbf{q}^\top A \, \mathbf{q}} & \text{if } 3 < |q_i| \leq 6 & \text{for } i = 1, \ldots, n \\ +\infty, & \text{otherwise} \end{cases}$$

where A is a positive definite matrix. We carry out the experiment on three choices of (position space) dimensionalites, $n = 2, 10$ and 50.

Due to the symmetry of the model, the ground truth expected value of $\mathbf{q}$ is known to be $\mathbf{0}$. Therefore, the absolute error of the expected value (estimated by a chain $\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(k)}$ of MCMC samples) in each dimension $d = 1, \ldots, n$ is the absolute value of the mean of d-th element of the sample vectors. The worst mean absolute error (WMAE) over all dimensions is taken as the error measurement of the chain.

$$\text{WMAE}\left(\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(k)}\right) = \max_{d=1,\ldots n} \left| \sum_{s=1}^{k} q_d{}^s \right|$$

For each algorithm, 20 Markov chains are run and the mean WMAE and 99% confidence intervals (as error bars) versus the number of iterations (i.e. Markov chain sizes) are time (milliseconds) are depicted in figure 2. All algorithms are implemented in java and run on a single thread of a 3.40GHz CPU.

For each of the 20 repetitions, some random starting point is chosen uniformly and used for all three of the algorithms. We use a diagonal matrix for A where, for each repetition, each entry on the main dimension is either $\exp(-5)$ or $\exp(5)$ with equal probabilities.

As the results show, even in low dimensions, the extra cost of the position step is more or less compensated by its higher effective sample size but as the dimensionality increases, the RHMC significantly outperforms both baseline HMC and tuned MH.

---

[2] Although it is known that HMC performance is highly sensitive to the way its parameters, are tuned [7], we observe that the comparative performance of HMC and RHMC does not change over any tested choice of parameters. Therefore, for brevity the results are depicted for one (possibly suboptimal) choice only.
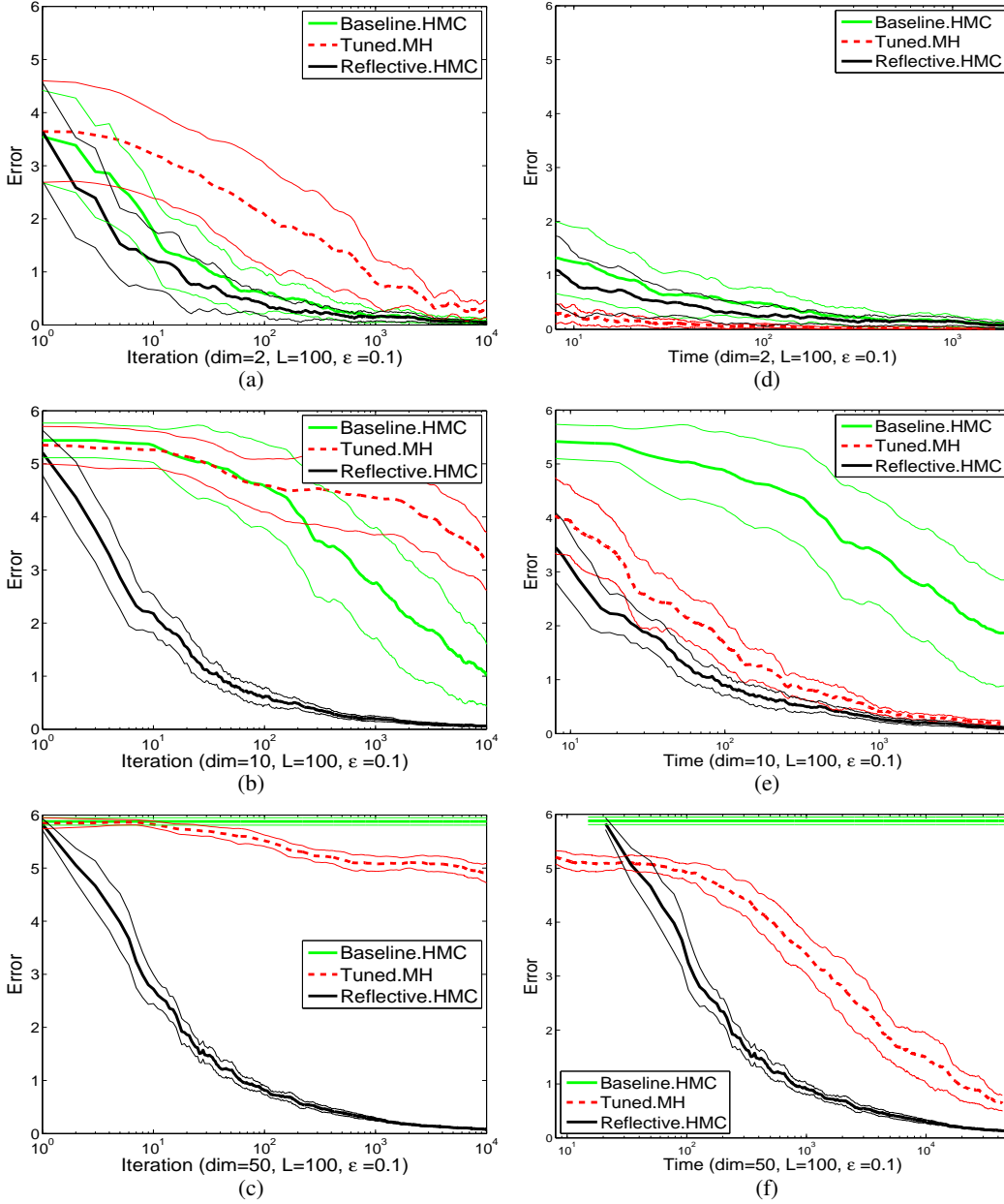
Figure 2: Error (worst mean absolute error per dimension) versus (a-c) iterations and (e-f) time (ms).

7 Conclusion

We have presented a modification of the leapfrog dynamics for Hamiltonian Monte Carlo for piecewise smooth energy functions with affine boundaries (i.e. each region is a polytope), inspired by physical systems. Though traditional Hamiltonian Monte Carlo can in principle be used on such functions, the fact that the Hamiltonian will often be dramatically changed by the dynamics can result in a very low acceptance ratio, particularly in high dimensions. By better preserving the Hamiltonian, *reflective Hamiltonian Monte Carlo* (RHMC) accepts more moves and thus has a higher effective sample size, leading to much more efficient probabilistic inference. To use this method, one must be able to detect the first intersection of a position trajectory with polytope boundaries.

References

[1] Hadi Mohasel Afshar, Scott Sanner, and Ehsan Abbasnejad. Linear-time gibbs sampling in piecewise graphical models. In *Association for the Advancement of Artificial Intelligence*, pages 665–673, 2015.

[2] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of Markov Chain Monte Carlo*. CRC press, 2011.

[3] Hans Adolph Buchdahl. *An introduction to Hamiltonian optics*. Courier Corporation, 1993.

[4] Simon Duane, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth. Hybrid monte carlo. *Physics letters B*, 195(2):216–222, 1987.

[5] Andrew G Glen, Lawrence M Leemis, and John H Drew. Computing the distribution of the product of two continuous random variables. *Computational statistics & data analysis*, 44(3):451–464, 2004.

[6] Donald T Greenwood. *Principles of dynamics*. Prentice-Hall Englewood Cliffs, NJ, 1988.

[7] Matthew D Homan and Andrew Gelman. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *The Journal of Machine Learning Research*, 15(1):1593–1623, 2014.

[8] David Lunn, David Spiegelhalter, Andrew Thomas, and Nicky Best. The bugs project: evolution, critique and future directions. *Statistics in medicine*, 28(25):3049–3067, 2009.

[9] Radford M Neal. Mcmc using hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2, 2011.

[10] Ari Pakman and Liam Paninski. Auxiliary-variable exact hamiltonian monte carlo samplers for binary distributions. In *Advances in Neural Information Processing Systems*, pages 2490–2498, 2013.

[11] Ari Pakman and Liam Paninski. Exact hamiltonian monte carlo for truncated multivariate gaussians. *Journal of Computational and Graphical Statistics*, 23(2):518–542, 2014.

[12] Gareth O Roberts, Andrew Gelman, Walter R Gilks, et al. Weak convergence and optimal scaling of random walk metropolis algorithms. *The annals of applied probability*, 7(1):110–120, 1997.

[13] Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual, Version 2.5.0*, 2014.