The Boomerang Sampler – Supplement

Joris Bierkens [1] **Sebastiano Grazzi** [1] **Kengo Kamatani** [2] **Gareth Roberts** [3]

1. Generator and stationary distribution

1.1. Boomerang Sampler

For simplicity take $\boldsymbol{x}_\star = \boldsymbol{0}$. The generator of the Boomerang Sampler is defined by

$$\begin{aligned}\mathcal{L}\psi(\boldsymbol{x},\boldsymbol{v}) &= \langle \boldsymbol{v}, \nabla_{\boldsymbol{x}}\psi(\boldsymbol{x},\boldsymbol{v})\rangle - \langle \boldsymbol{x}, \nabla_{\boldsymbol{v}}\psi(\boldsymbol{x},\boldsymbol{v})\rangle \\ &\quad + \lambda(\boldsymbol{x},\boldsymbol{v})\left(\psi(\boldsymbol{x},\boldsymbol{R}(\boldsymbol{x})\boldsymbol{v}) - \psi(\boldsymbol{x},\boldsymbol{v})\right) \\ &\quad + \lambda_{\mathrm{refr}}\left(\int_{\mathbb{R}^d}\psi(\boldsymbol{x},\boldsymbol{w})\phi(\boldsymbol{w})\,\mathrm{d}\boldsymbol{w} - \psi(\boldsymbol{x},\boldsymbol{v})\right),\end{aligned}$$

for any compactly supported differentiable function ψ on S, where ϕ is the probability density function of $\mathcal{N}(\boldsymbol{0},\boldsymbol{\Sigma})$.

Taking $\lambda(\boldsymbol{x},\boldsymbol{v})$ and $\boldsymbol{R}(\boldsymbol{x})$ as in Eqs. (2) and (3) of the paper respectively, we will now verify that $\int_S \mathcal{L}\psi\, d\mu = 0$ for all such functions ψ, and for μ being the measure on S with density $\exp(-U(\boldsymbol{x}))$ relative to μ_0. This then establishes that the Boomerang Sampler has stationary distribution μ. A complete proof also requires verification that the compactly supported, differentiable functions form a core for the generator, which is beyond the scope of this paper. For a discussion of this topic for archetypal PDMPs see (Holderrieth, 2019).

First we consider the terms involving the partial derivatives of ψ. By partial integration, we find

$$\begin{aligned}&\int_S \langle \boldsymbol{v}, \nabla_{\boldsymbol{x}}\psi(\boldsymbol{x},\boldsymbol{v})\rangle - \langle \boldsymbol{x}, \nabla_{\boldsymbol{v}}\psi(\boldsymbol{x},\boldsymbol{v})\rangle\,\mu(\mathrm{d}\boldsymbol{x},\mathrm{d}\boldsymbol{v}) \\ &= \int_S \psi(\boldsymbol{x},\boldsymbol{v})\langle \boldsymbol{v}, \nabla U(\boldsymbol{x})\rangle\,\mu(\mathrm{d}\boldsymbol{x},\mathrm{d}\boldsymbol{v})\end{aligned}$$

Next we inspect the term representing the switches occurring at rate $\lambda(\boldsymbol{x},\boldsymbol{v})$. By Eq. (5) of the paper, the coordinate transform $\boldsymbol{w} = \boldsymbol{R}(\boldsymbol{x})\boldsymbol{v}$ (for fixed $\boldsymbol{x}$) leaves the measure $\mathcal{N}(\boldsymbol{0},\boldsymbol{\Sigma})$ over the velocity component invariant. Using this observation, we find that

$$\begin{aligned}&\int_S \lambda(\boldsymbol{x},\boldsymbol{v})(\psi(\boldsymbol{x},\boldsymbol{R}(\boldsymbol{x})\boldsymbol{v}) - \psi(\boldsymbol{x},\boldsymbol{v}))\,\mu(\mathrm{d}\boldsymbol{x},\mathrm{d}\boldsymbol{v}) \\ &= \int_S \lambda(\boldsymbol{x},\boldsymbol{R}(\boldsymbol{x})\boldsymbol{w})\psi(\boldsymbol{x},\boldsymbol{w})\,\mu(d\boldsymbol{x},d\boldsymbol{w}) \\ &\quad - \int_S \lambda(\boldsymbol{x},\boldsymbol{v})\psi(\boldsymbol{x},\boldsymbol{v})\,\mu(\mathrm{d}\boldsymbol{x},\mathrm{d}\boldsymbol{v}) \\ &= \int_S [\lambda(\boldsymbol{x},\boldsymbol{R}(\boldsymbol{x})\boldsymbol{v}) - \lambda(\boldsymbol{x},\boldsymbol{v})]\psi(\boldsymbol{x},\boldsymbol{v})\,\mu(\mathrm{d}\boldsymbol{x},\mathrm{d}\boldsymbol{v}).\end{aligned}$$

Using Eq. (2) and (4) of the paper, and the identity $(-a)_+ - (a)_+ = -a$, it follows that this expression is equal to

$$\begin{aligned}&\int_S [\langle \boldsymbol{R}(\boldsymbol{x})\boldsymbol{v}, \nabla U(\boldsymbol{x})\rangle_+ - \langle \boldsymbol{v}, \nabla U(\boldsymbol{x})\rangle_+]\,\psi(\boldsymbol{x},\boldsymbol{v})\,\mu(\mathrm{d}\boldsymbol{x},\mathrm{d}\boldsymbol{v}) \\ &= -\int_S \langle \boldsymbol{v}, \nabla U(\boldsymbol{x})\rangle\psi(\boldsymbol{x},\boldsymbol{v})\,\mu(\mathrm{d}\boldsymbol{x},\mathrm{d}\boldsymbol{v}).\end{aligned}$$

Finally by changing the order of integration, it can be shown that

$$\int_S \lambda_{\mathrm{refr}}\left(\int_{\mathbb{R}^d}\psi(\boldsymbol{x},\boldsymbol{v})\phi(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v} - \psi(\boldsymbol{x},\boldsymbol{v})\right)\mu_0(\mathrm{d}\boldsymbol{x},\mathrm{d}\boldsymbol{v}) = 0.$$

Adding all terms yields that $\int_S \mathcal{L}\psi\,\mathrm{d}\mu = 0$.

1.2. Factorised Boomerang Sampler

The Factorised Boomerang Sampler has generator

$$\begin{aligned}\mathcal{L}\psi(\boldsymbol{x},\boldsymbol{v}) &= \langle \boldsymbol{v}, \nabla_{\boldsymbol{x}}\psi(\boldsymbol{x},\boldsymbol{v})\rangle - \langle \boldsymbol{x}, \nabla_{\boldsymbol{v}}\psi(\boldsymbol{x},\boldsymbol{v})\rangle \\ &\quad + \sum_{i=1}^d \lambda_i(\boldsymbol{x},\boldsymbol{v})(\psi(\boldsymbol{x},\boldsymbol{F}_i(\boldsymbol{v})) - \psi(\boldsymbol{x},\boldsymbol{v})) \\ &\quad + \lambda_{\mathrm{refr}}\left(\int \psi(\boldsymbol{x},\boldsymbol{w})\phi(\boldsymbol{w})\,d\boldsymbol{w} - \psi(\boldsymbol{x},\boldsymbol{v})\right).\end{aligned}$$

Verifying stationarity of μ is done analogously to the case of the non-factorised Boomerang Sampler, but now has to be carried out componentwise.

2. Computational bounds

Suppose $(\boldsymbol{x}_t,\boldsymbol{v}_t)$ satisfies the Hamiltonian dynamics ODE of Eq. (1) in the paper, starting from $(\boldsymbol{x}_0,\boldsymbol{v}_0)$ in $\mathbb{R}^d\times\mathbb{R}^d$. Throughout we assume $U:\mathbb{R}^d\to\mathbb{R}$ is a twice continuously differentiable function with Hessian matrix $\nabla^2 U$. Furthermore we assume without loss of generality that $\boldsymbol{x}_\star = \boldsymbol{0}$. First we consider bounds for switching intensities of the form $\lambda(\boldsymbol{x},\boldsymbol{v}) = \langle \boldsymbol{v}, \nabla U(\boldsymbol{x})\rangle_+$. For a matrix $\boldsymbol{A}\in\mathbb{R}^{d\times d}$ we use $\|\boldsymbol{A}\|$ to denote the matrix norm induced by the Euclidean metric.

Lemma 2.1 (Constant bound)**.** *Suppose there exists a constant $M>0$ such that for all $\boldsymbol{x}\in\mathbb{R}^d$ we have the global bound*

$$\|\nabla^2 U(\boldsymbol{x})\| \le M.$$

Define $m := |\nabla U(\mathbf{0})|$. *Then for all* $t \geq 0$,

$$\lambda(\boldsymbol{x}_t, \boldsymbol{v}_t) \leq \frac{M}{2}(|\boldsymbol{x}_0|^2 + |\boldsymbol{v}_0|^2) + m\sqrt{|\boldsymbol{x}_0|^2 + |\boldsymbol{v}_0|^2}. \quad (1)$$

Proof. We have the following estimate on the switching intensity.

$$\begin{aligned}\lambda(\boldsymbol{x}, \boldsymbol{v}) &= \langle \boldsymbol{v}, \nabla U(\boldsymbol{x})\rangle_+ \\ &\leq \langle \boldsymbol{v}, \nabla U(\mathbf{0})\rangle_+ + \int_0^1 |\langle \boldsymbol{v}, \nabla^2 U(\boldsymbol{x}s)\boldsymbol{x}\rangle| \,\mathrm{d}s.\end{aligned}$$

We may bound the inner product in the integrand as follows.

$$\begin{aligned}|\langle \boldsymbol{v}, \nabla^2 U(\boldsymbol{y})\boldsymbol{x}\rangle| &\leq \|\nabla^2 U(\boldsymbol{x})\| \; |\boldsymbol{v}| \; |\boldsymbol{x}| \\ &\leq M \left(\frac{|\boldsymbol{v}|^2 + |\boldsymbol{x}|^2}{2}\right)\end{aligned}$$

by the Cauchy–Schwarz inequality. Also

$$|\langle \boldsymbol{v}, \nabla U(\mathbf{0})\rangle| \leq m|\boldsymbol{v}| \leq m\sqrt{|\boldsymbol{x}|^2 + |\boldsymbol{v}|^2}.$$

Combining these estimates and the fact that $|\boldsymbol{x}_t|^2 + |\boldsymbol{v}_t|^2$ is invariant under the dynamics of Eq. (1) in the paper yields the stated result. □

Lemma 2.2 (Affine bound)**.** *Suppose* $\|\nabla^2 U(\boldsymbol{x})\| \leq M$ *for all* $\boldsymbol{x} \in \mathbb{R}^d$, *and let* $m = |\nabla U(\mathbf{0})|$. *Then for a solution* $(\boldsymbol{x}_t, \boldsymbol{v}_t)$ *to Eq. (1) of the paper with* $\lambda(\boldsymbol{x}, \boldsymbol{v}) = \langle \boldsymbol{v}, \nabla U(\boldsymbol{x})\rangle_+$, *we have for all* $t \geq 0$

$$\lambda(\boldsymbol{x}_t, \boldsymbol{v}_t) \leq (a(\boldsymbol{x}_0, \boldsymbol{v}_0) + t b(\boldsymbol{x}_0, \boldsymbol{v}_0))_+ \,,$$

where

$$\begin{aligned}a(\boldsymbol{x}, \boldsymbol{v}) &= \langle \boldsymbol{v}, \nabla U(\boldsymbol{x})\rangle_+, \quad \textit{and} \\ b(\boldsymbol{x}, \boldsymbol{v}) &= M\left(|\boldsymbol{x}|^2 + |\boldsymbol{v}|^2\right) + m\sqrt{|\boldsymbol{x}|^2 + |\boldsymbol{v}|^2}.\end{aligned}$$

Proof. By the Hamiltonian dynamics,

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\langle \boldsymbol{v}_t, \nabla U(\boldsymbol{x}_t)\rangle \\ &= -\langle \boldsymbol{x}_t, \nabla U(\boldsymbol{x}_t)\rangle + \langle \boldsymbol{v}_t, \nabla^2 U(\boldsymbol{x}_t)\boldsymbol{v}_t\rangle \\ &= -\langle \boldsymbol{x}_t, \nabla U(\mathbf{0})\rangle - \int_0^1 \langle \boldsymbol{x}_t, \nabla^2 U(s\boldsymbol{x}_t)\boldsymbol{x}_t\rangle \,\mathrm{d}s \\ &\qquad + \langle \boldsymbol{v}_t, \nabla^2 U(\boldsymbol{x}_t)\boldsymbol{v}_t\rangle \\ &\leq |\boldsymbol{x}_t||\nabla U(\mathbf{0})| + M\left(|\boldsymbol{x}_t|^2 + |\boldsymbol{v}_t|^2\right).\end{aligned}$$

Using that $|\boldsymbol{x}_t|^2 + |\boldsymbol{v}_t|^2$ is invariant under the dynamics yields the stated result. □

Lemma 2.3. *Suppose* $|\nabla U(\boldsymbol{y})| \leq C$ *for all* $\boldsymbol{y} \in \mathbb{R}^d$. *Then, for all trajectories* $(\boldsymbol{x}_t, \boldsymbol{v}_t)$ *satisfying Eq. (1) of the paper we have*

$$\lambda(\boldsymbol{x}_t, \boldsymbol{v}_t) \leq C\sqrt{|\boldsymbol{x}_0|^2 + |\boldsymbol{v}_0|^2}.$$

Proof. We have

$$\lambda(\boldsymbol{x}, \boldsymbol{v}) \leq C|\boldsymbol{v}| \leq C\sqrt{|\boldsymbol{x}|^2 + |\boldsymbol{v}|^2},$$

and the latter expression is constant along trajectories. □

Analogously we have the following useful bound for the Factorized Boomerang Sampler.

Lemma 2.4. *Suppose* $U : \mathbb{R}^d \to \mathbb{R}$ *is differentiable. Suppose there exist constants* $c_1, \ldots, c_d$ *such that, for all* $\boldsymbol{y} \in \mathbb{R}^d$ *and* $i = 1, \ldots, d$, *we have*

$$|\partial_i U(\boldsymbol{x})| \leq c_i \quad \textit{for all } \boldsymbol{x},\ i.$$

Then

$$\lambda_i(\boldsymbol{x}_t, \boldsymbol{v}_t) \leq c_i\sqrt{|x_0^i|^2 + |v_0^i|^2}.$$

Lemma 2.5. *Suppose for all* i *we have that*

$$\sqrt{\sum_j \partial_i\partial_j U(\boldsymbol{x})^2} \leq M_i,$$

and

$$|\partial_i U(\mathbf{0})| \leq m_i.$$

Then

$$\lambda_i(\boldsymbol{x}_t, \boldsymbol{v}_t) \leq (a_i(\boldsymbol{x}_0, \boldsymbol{v}_0) + b_i(\boldsymbol{x}_0, \boldsymbol{v}_0)t)^+$$

where

$$a_i(\boldsymbol{x}, \boldsymbol{v}) = (v^i\partial_i U(\boldsymbol{x}))^+$$

$$\begin{aligned}&b_i(\boldsymbol{x}, \boldsymbol{v}) \\ &= \sqrt{(x^i)^2 + (v^i)^2}\left(m_i + M_i\sqrt{|\boldsymbol{x}|^2 + |\boldsymbol{v}|^2}\right).\end{aligned}$$

Proof. We compute

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t} v_t^i \partial_i U(\boldsymbol{x}_t) \\ &= -x_t^i\partial_i U(\boldsymbol{x}_t) + v_t^i \sum_{j=1}^d \partial_i\partial_j U(\boldsymbol{x}_t) v_t^j \\ &= -x_t^i \partial_i U(\mathbf{0}) - \int_0^1 x_t^i \sum_{j=1}^d \partial_i\partial_j U(s\boldsymbol{x}_t) x_t^j \mathrm{d}s \\ &\qquad + v_t^i \sum_{j=1}^d \partial_i\partial_j U(\boldsymbol{x}_t) v_t^j \\ &\leq \sqrt{(x_t^i)^2 + (v_t^i)^2}|\partial_i U(\mathbf{0})| + M_i|x_t^i||\boldsymbol{x}_t| + M_i|v_t^i||\boldsymbol{v}_t| \\ &\leq \sqrt{(x_t^i)^2 + (v_t^i)^2}|\partial_i U(\mathbf{0})| \\ &\quad + M_i/2\left(\alpha(|x_t^i|^2 + |v_t^i|^2) + (1/\alpha)(|\boldsymbol{x}_t|^2 + |\boldsymbol{v}_t|^2)\right).\end{aligned}$$

Optimising over α, and using that $|x_t^i|^2 + |v_t^i|^2$ is constant along Factorised Boomerang Trajectories, yields the stated result. □

2.1. Computational bounds for subsampling

In the case of subsampling we use the unbiased estimator of Eq. (9) of the paper.

Lemma 2.6. *Suppose that for some positive definite matrix $\boldsymbol{Q}$ we have that, for all i, and $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathbb{R}^d$,*

$$\nabla^2 E^i(\boldsymbol{y}_1) - \nabla^2 E^i(\boldsymbol{y}_2) \preceq \boldsymbol{Q}, \tag{2}$$

where $A \preceq B$ means $B - A$ is positive semidefinite. Suppose $\widehat{\nabla U(\boldsymbol{x})}$ is given by Eq. (9) of the paper, and $\nabla E(\boldsymbol{0}) = \boldsymbol{0}$. Along a trajectory $(\boldsymbol{x}_t, \boldsymbol{v}_t)$ satisfying the Hamiltonian dynamics of Eq. (1) of the paper, we have, for all $t \geq 0$, that

$$\langle \boldsymbol{v}_t, \widehat{\nabla U(\boldsymbol{x}_t)} \rangle \leq \tfrac{1}{2}(|\boldsymbol{Q}^{1/2}\boldsymbol{x}_0|^2 + |\boldsymbol{Q}^{1/2}\boldsymbol{v}_0|^2), \quad \textit{a.s.}$$

where the almost sure statement is with respect to all random (subsampling) realisations of the switching intensity.

Remark 2.7. Lemma (2.6) is easily extended to the case in which $\nabla E(\boldsymbol{0}) \neq \boldsymbol{0}$. In this case we have

$$\begin{aligned}\langle \boldsymbol{v}_t, \widehat{U(\boldsymbol{x}_t)} \rangle \leq \tfrac{1}{2}(|\boldsymbol{Q}^{1/2}\boldsymbol{x}_0|^2 + |\boldsymbol{Q}^{1/2}\boldsymbol{v}_0|^2) \\ + (|\boldsymbol{v}_0|^2 + |\boldsymbol{x}_0|^2)^{1/2} |\nabla E(\boldsymbol{0})|, \quad \text{a.s.}\end{aligned}$$

Remark 2.8. In practice one may wish to take $\boldsymbol{Q}$ to be a diagonal matrix, which reduces the computation of the computational bound to a $\mathcal{O}(d)$ computation instead of $\mathcal{O}(d^2)$. For example one could take $\boldsymbol{Q} = c\boldsymbol{I}$ for a suitable constant $c > 0$ such that (2) is satisfied.

Remark 2.9 (Affine bound for subsampling is strictly worse). When we try to obtain an affine bound, of the form

$$\lambda\widehat{(\boldsymbol{x}_t, \boldsymbol{v}_t)} \leq a(\boldsymbol{x}_0, \boldsymbol{v}_0) + b(\boldsymbol{x}_0, \boldsymbol{v}_0),$$

then it seems we cannot avoid an expression for a of the form of the bound in Lemma 2.6. As a consequence, the affine bound is strictly worse than the constant bound.

Proof (of Lemma 2.6). Suppose we have $I = i$ for the random index I in Eq. (9) of the paper. We compute

$$\begin{aligned}&\langle \boldsymbol{v}_t, \widehat{\nabla U(\boldsymbol{x}_t)} \rangle \\ &= \langle \boldsymbol{v}_t, \nabla E^i(\boldsymbol{x}_t) - \nabla^2 E^i(\boldsymbol{0})\boldsymbol{x}_t - \nabla E^i(\boldsymbol{0}) \rangle \\ &= \langle \boldsymbol{v}_t, \int_0^1 \nabla^2 E^i(s\boldsymbol{x}_t)\boldsymbol{x}_t \, \mathrm{d}s - \nabla^2 E^i(\boldsymbol{0})\boldsymbol{x}_t \rangle.\end{aligned}$$

Then we may continue the above computation to find, using Lemma 2.10 below, that

$$\begin{aligned}\langle \boldsymbol{v}_t, \widehat{\nabla U(\boldsymbol{x}_t)} \rangle &= \int_0^1 \langle \boldsymbol{v}_t, [\nabla^2 E^i(s\boldsymbol{x}_t) - \nabla^2 E^i(\boldsymbol{0})]\boldsymbol{x}_t \rangle \, \mathrm{d}s \\ &\leq \int_0^1 |\boldsymbol{Q}^{1/2}\boldsymbol{v}_t| \, |\boldsymbol{Q}^{1/2}\boldsymbol{x}_t| \, \mathrm{d}s \\ &\leq \tfrac{1}{2}(|\boldsymbol{Q}^{1/2}\boldsymbol{v}_t|^2 + |\boldsymbol{Q}^{1/2}\boldsymbol{x}_t|^2).\end{aligned}$$

Since $\frac{1}{2}(|\boldsymbol{Q}^{1/2}\boldsymbol{v}_t|^2 + |\boldsymbol{Q}^{1/2}\boldsymbol{x}_t|^2)$ is invariant under the dynamics, the stated conclusion follows. □

Lemma 2.10. *Suppose $\boldsymbol{M}, \boldsymbol{P} \in \mathbb{R}^{d \times d}$ are symmetric matrices with $\boldsymbol{P}$ positive definite and such that $-\boldsymbol{P} \preceq \boldsymbol{M} \preceq \boldsymbol{P}$. Then $\langle \boldsymbol{M}\boldsymbol{y}, \boldsymbol{z} \rangle \leq |\boldsymbol{P}^{1/2}\boldsymbol{y}| \, |\boldsymbol{P}^{1/2}\boldsymbol{z}|$ for all $\boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^{d \times d}$.*

Proof. Taking $\boldsymbol{y} = \boldsymbol{P}^{-1/2}\boldsymbol{x}$, we find

$$|\langle \boldsymbol{P}^{-1/2}\boldsymbol{M}\boldsymbol{P}^{-1/2}\boldsymbol{x}, \boldsymbol{x} \rangle| = |\langle \boldsymbol{M}\boldsymbol{y}, \boldsymbol{y} \rangle| \leq \langle \boldsymbol{P}\boldsymbol{y}, \boldsymbol{y} \rangle = |\boldsymbol{x}|^2,$$

which establishes that $\|\boldsymbol{P}^{-1/2}\boldsymbol{M}\boldsymbol{P}^{-1/2}\| \leq 1$. Using this observation we arrive at

$$\langle \boldsymbol{M}\boldsymbol{y}, \boldsymbol{z} \rangle \leq \underbrace{\|\boldsymbol{P}^{-1/2}\boldsymbol{M}\boldsymbol{P}^{-1/2}\|}_{\leq 1} \, |\boldsymbol{P}^{1/2}\boldsymbol{y}| \, |\boldsymbol{P}^{1/2}\boldsymbol{z}|.$$

□

3. Scaling with dimension

In Section 3.2 of the paper, we discuss the scaling of the Boomerang Sampler with dimension. The argument in that section is self contained, but relies on the observation that the change of $E_d(\boldsymbol{x}_t)$ over a time interval of order 1 is at least of order $d^{1/2}$. Here we motivate this observation.

In the following arguments, we assume stationarity of the process for simplicity. Let U_d, $\boldsymbol{\Sigma}_d$, E_d, Π_d, $\mathbb{E}_d$ be as described in Section 3.2 of the manuscript. For simplicity and without loss of generality we assume that $E_d(\boldsymbol{x})$ is normalised as $\mathbb{E}_d[E_d(\boldsymbol{x})] = 0$. Furthermore, for simplicity we assume that $\mathbb{E}_d[\boldsymbol{x}] = \boldsymbol{0}$ although this condition can be relaxed.

As discussed we suppose that the sequence (U_d) satisfies

$$\sup_{d \in \mathbb{N}} \mathbb{E}_d[|\boldsymbol{\Sigma}_d^{1/2} \nabla U_d(\boldsymbol{x})|^2] \leq \kappa \tag{3}$$

for some $\kappa > 0$. Furthermore, we assume that the following form of the Poincaré inequality is satisfied for $\Pi_d(\mathrm{d}\boldsymbol{x}) \propto \exp(-E_d(\boldsymbol{x}))\mathrm{d}\boldsymbol{x}$:

$$C \, \mathbb{E}_d \left[f_d(\boldsymbol{x})^2 \right]^{1/2} \leq \mathbb{E}_d \left[|\boldsymbol{\Sigma}_d^{1/2} \nabla f_d(\boldsymbol{x})|^2 \right]^{1/2} \tag{4}$$

for some constant $C > 0$ not depending on d, and any differentiable function $f_d : \mathbb{R}^d \to \mathbb{R}$ with mean 0 and finite variance.

By (3) the expected number of reflections per unit time $\mathbb{E}_d[\langle \boldsymbol{v}, \nabla U_d(\boldsymbol{x}) \rangle_+]$ is bounded with respect to dimension. However the process mixes well in a single time unit under suitable regularity conditions as we will discuss now.

By applying (4) to $f_d(\boldsymbol{x}) = (\boldsymbol{\Sigma}_d^{-1/2}\boldsymbol{x})_i$, where $\boldsymbol{v}_i$ denotes the i-th coordinate of $\boldsymbol{v}$, we have $C^2 \mathbb{E}_d[|\boldsymbol{\Sigma}_d^{-1/2}\boldsymbol{x}|^2] \leq$

$\mathbb{E}_d[\text{trace}(\mathbf{\Sigma}_d^{-1/2}\mathbf{\Sigma}_d\mathbf{\Sigma}_d^{-1/2})] = d$, using the stated assumption $\mathbb{E}_d[\boldsymbol{x}] = \mathbf{0}$.

Also by (4) and by Minkowski's inequality,

$$\begin{aligned}\mathbb{E}_d[E_d(\boldsymbol{x})^2]^{1/2} &\leq C^{-1}\mathbb{E}_d[|\mathbf{\Sigma}_d^{1/2}\nabla E_d(\boldsymbol{x})|^2]^{1/2} \\ &= C^{-1}\mathbb{E}_d[|\mathbf{\Sigma}_d^{1/2}\nabla U_d(\boldsymbol{x}) + \mathbf{\Sigma}_d^{-1/2}\boldsymbol{x}|^2]^{1/2} \\ &= C^{-1}(\kappa^{1/2} + C^{-1}d^{1/2}) = \mathcal{O}(d^{1/2}).\end{aligned}$$

If $(\boldsymbol{x}_t, \boldsymbol{v}_t)$ satisfies the ODE Eq. (1) of the paper, the unit time difference $E_d(\boldsymbol{x}_t) - E_d(\boldsymbol{x}_0)$ is

$$\int_0^t \langle \nabla E_d(\boldsymbol{x}_s), \boldsymbol{v}_s\rangle \mathrm{d}s \approx \int_0^t \langle \mathbf{\Sigma}_d^{-1}\boldsymbol{x}_s, \boldsymbol{v}_s\rangle \mathrm{d}s.$$

Here, the difference between the left- and the right-hand sides is $\int_0^t \langle \mathbf{\Sigma}_d^{1/2}\nabla U(\boldsymbol{x}_s), \mathbf{\Sigma}_d^{-1/2}\boldsymbol{v}_s\rangle \mathrm{d}s$ which is of order $d^{1/2}$ under the assumption of stationarity by (3) and the Cauchy-Schwarz inequality, using that $\mathbb{E}_d[|\mathbf{\Sigma}_d^{-1/2}\boldsymbol{v}_s|^2] = d$. The right-hand may be simplified to

$$\begin{aligned}&\int_0^t \langle \mathbf{\Sigma}_d^{-1}(\boldsymbol{x}_0\cos s + \boldsymbol{v}_0 \sin s), -\boldsymbol{x}_0 \sin s + \boldsymbol{v}_0\cos s\rangle \mathrm{d}s \\ &= A_0\int_0^t 2\sin s\ \cos s\ \mathrm{d}s + B_0\int_0^t (\cos^2 s - \sin^2 s)\mathrm{d}s \\ &= A_0(1-\cos 2t)/2 + B_0(\sin 2t)/2\end{aligned}$$

where $A_0 = (\langle \boldsymbol{v}_0, \mathbf{\Sigma}_d^{-1}\boldsymbol{v}_0\rangle - \langle \boldsymbol{x}_0, \mathbf{\Sigma}_d^{-1}\boldsymbol{x}_0\rangle)/2$ and $B_0 = \langle \boldsymbol{x}_0, \mathbf{\Sigma}_d^{-1}\boldsymbol{v}_0\rangle$. Then A_0 and B_0 are uncorrelated since $\mathbf{\Sigma}_d^{-1/2}\boldsymbol{v}_0$ follows the standard normal distribution. Also, $\mathbb{E}_d[A_0^2] \geq \text{Var}(A_0) \geq \text{Var}(\langle \boldsymbol{v}_0, \mathbf{\Sigma}_d^{-1}\boldsymbol{v}_0\rangle) = 2d$. Therefore,

$$\begin{aligned}\mathbb{E}_d[|E_d(\boldsymbol{x}_t) - E_d(\boldsymbol{x}_0)|^2] &\gtrsim \mathbb{E}_d[A_0^2]\left(\frac{1-\cos 2t}{2}\right)^2 \\ &\geq 2d\left(\frac{1-\cos 2t}{2}\right)^2.\end{aligned}$$

Thus the change of $E_d(\boldsymbol{x}_t)$ over a term interval of $\mathcal{O}(1)$ is of order $d^{1/2}$ whereas $E_d(\boldsymbol{x}_t)$ itself has the same order. These informal arguments suggest that dynamics of the Boomerang sampler in a finite time interval sufficiently changes the log density even in high dimension. However, further study should be made in this direction.

4. Logistic regression

We assume a prior distribution $\pi_0(\boldsymbol{x}) \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I})$ on $\mathbb{R}^d$. Given predictors $\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(n)}$ in $\mathbb{R}^d$, and outcomes $z^{(1)}, \ldots, z^{(n)}$ in $\{0, 1\}$, we obtain the negative log posterior distribution as

$$E(\boldsymbol{x}) = \sum_{i=1}^n \left\{\log(1 + e^{\boldsymbol{x}^\top \boldsymbol{y}^{(i)}}) - z^{(i)}\boldsymbol{x}^\top\boldsymbol{y}^{(i)}\right\} + |\boldsymbol{x}|^2/2\sigma^2.$$

We then have

$$\nabla E(\boldsymbol{x}) = \boldsymbol{x}/\sigma^2 + \sum_{i=1}^n \boldsymbol{y}^{(i)}\left[\frac{e^{\boldsymbol{x}^\top\boldsymbol{y}^{(i)}}}{1+e^{\boldsymbol{x}^\top\boldsymbol{y}^{(i)}}} - z^{(i)}\right],$$

$$\nabla^2 E(\boldsymbol{x}) = \boldsymbol{I}/\sigma^2 + \sum_{i=1}^n \frac{\boldsymbol{y}^{(i)}(\boldsymbol{y}^{(i)})^\top e^{\boldsymbol{x}^\top\boldsymbol{y}^{(i)}}}{\left(1+e^{\boldsymbol{x}^\top\boldsymbol{y}^{(i)}}\right)^2}.$$

In the experiments in this paper we take a flat prior, i.e. $\sigma^2 = \infty$.

Let

$$\boldsymbol{x}_\star = \arg\min_{\boldsymbol{x}\in\mathbb{R}^d} E(\boldsymbol{x}).$$

We take $\mathbf{\Sigma}^{-1} = \nabla^2 E(\boldsymbol{x}_\star)$. We have $U(\boldsymbol{x}) = E(\boldsymbol{x}) - (\boldsymbol{x}-\boldsymbol{x}_\star)^\top\nabla^2E(\boldsymbol{x}_\star)(\boldsymbol{x}-\boldsymbol{x}_\star)/2$, which is a difference of two positive definite matrices. Using the general inequality $a \mapsto |a|/(1+a)^2 \leq 1/4$, we find

$$-\tfrac{1}{4}\sum_{i=1}^n \boldsymbol{y}^{(i)}(\boldsymbol{y}^{(i)})^\top \preceq \nabla^2 U(\boldsymbol{x}) \preceq \tfrac{1}{4}\sum_{i=1}^n \boldsymbol{y}^{(i)}(\boldsymbol{y}^{(i)})^\top.$$

We then simply have

$$\|\nabla^2 U(\boldsymbol{y})\| \leq M := \tfrac{1}{4}\|\sum_{i=1}^n \boldsymbol{y}^{(i)}(\boldsymbol{y}^{(i)})^\top\|.$$

These observations may be applied in conjunction with the lemmas of Section 2 in this supplement to obtain useful constant and affine computational bounds for the switching intensities.

5. Diffusion bridge simulation

We consider diffusion bridges of the form

$$\mathrm{d}X_t = \alpha\sin(X_t)\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = u, X_T = v, t\in[0,T] \tag{5}$$

where W is a scalar Brownian motion and $\alpha \geq 0$. The diffusion path is expanded with a truncated Faber-Schauder basis such that

$$X_t^N = \bar{\bar{\phi}}(t)u + \bar{\phi}(t)v + \sum_{i=0}^N\sum_{j=0}^{2^i-1}\phi_{i,j}(t)x_{i,j},$$

where N is the truncation of the expansion and

$$\begin{aligned}&\bar{\phi}(t) = t/T, \qquad \bar{\bar{\phi}}(t) = 1 - t/T, \\ &\phi_{0,0}(t) = \sqrt{T}\big((t/T)\mathbf{1}_{[0,T/2]}(t) + (1-t/T)\mathbf{1}_{(T/2,T]}(t)\big), \\ &\phi_{i,j}(t) = 2^{-i/2}\phi_{0,0}(2^it - jT) \quad i\geq 0, \quad 0\leq j\leq 2^i-1,\end{aligned}$$

are the Faber-Schauder functions. As shown in (Bierkens et al., 2020), the measure of the coefficients corresponding to (5) is derived from the Girsanov formula and given by

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(\boldsymbol{x},\boldsymbol{v}) \propto \exp\left\{\frac{-\alpha}{2}\int_0^T \left(\alpha\sin^2(X_s^N) + \cos(X_s^N)\right)\mathrm{d}s\right\}$$

where $\mu_0 = \mathcal{N}(\mathbf{0}, \boldsymbol{I}) \otimes \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ with $\boldsymbol{I}$ the $2^{N+1} - 1$ dimensional identity matrix. By standard trigonometric identities we have that

$$\partial_{x_{i,j}} U(\boldsymbol{x}) = \frac{\alpha}{2} \int_{S_{i,j}} \phi_{i,j}(t) \left(\alpha \sin\left(2X_t^N\right) - \sin\left(X_t^N\right)\right) \mathrm{d}t$$

where $S_{i,j}$ is the support of the basis function $\phi_{i,j}$. Similarly to (Bierkens et al., 2020), for each i, j, we use subsampling and consider the unbiased estimator for $\partial_{x_{i,j}} U(\boldsymbol{x})$ given by

$$\widehat{\partial_{x_{i,j}} U(\boldsymbol{x})} = S_{i,j} \phi_{i,j}(\tau_{i,j}) \left(\alpha^2 \sin\left(2X^N_{\tau_{i,j}}\right) - \alpha \sin\left(X^N_{\tau_{i,j}}\right)\right)$$

where $\tau_{i,j}$ is a uniform random variable on $S_{i,j}$. This gives Poisson rates $\widehat{\lambda_{i,j}(\boldsymbol{x}, \boldsymbol{v})} = \langle \boldsymbol{v}, \widehat{\partial_{x_{i,j}} U(\boldsymbol{x})} \rangle_+$. In this case, for all i, j, $|\widehat{\partial_{x_{i,j}} U(\boldsymbol{x})}|$ is globally bounded, say by $m_{i,j}$. We use the constant Poisson bounding rates given, in similar spirit as in Section 2.3 of the paper, by

$$\overline{\lambda}_{i,j}(\boldsymbol{x}_t, \boldsymbol{v}_t) = m_{i,j} \sqrt{|x_0^{i,j}|^2 + |v_0^{i,j}|^2},$$

where we used that $t \to |x_t^{i,j}|^2 + |v_t^{i,j}|^2$ is constant under the Factorised Boomerang trajectories. Similarly to (Bierkens et al., 2020), the FBS gains computational efficiency by a local implementation which exploits the fact that each $\overline{\lambda}_{i,j}(\boldsymbol{x}, \boldsymbol{v})$ is a function of just the coefficient $x_{i,j}$ (see (Bierkens et al., 2020), Algorithm 3, for an algorithmic description of the local implementation of a factorised PDMP).

References

Bierkens, J., Grazzi, S., van der Meulen, F., and Schauer, M. A piecewise deterministic Monte Carlo method for diffusion bridges. *preprint arXiv:2001.05889*, 2020. URL http://arxiv.org/abs/2001.05889.

Holderrieth, P. Cores for Piecewise-Deterministic Markov Processes used in Markov Chain Monte Carlo. oct 2019. URL http://arxiv.org/abs/1910.11429.